

Universidad ORT Uruguay

Facultad de Ingeniería

Proyectos de big data en organizaciones actuales

Implementación CRM de tarjetas de crédito

Entregado como requisito para la obtención del título de

Master en Big Data

Juan Ignacio Saldías - 153699

Tutor: Alejandro Bianchi

2021

DECLARACIÓN DE AUTORÍA

Yo, Juan Ignacio Saldías, declaro que el trabajo que se presenta en esa obra es de mi propia mano.

Puedo asegurar que:

- La obra fue producida en su totalidad mientras realizaba el trabajo final de la carrera Master en Big Data;
- Cuando he consultado el trabajo publicado por otros, lo he atribuido con claridad;
- Cuando he citado obras de otros, he indicado las fuentes. Con excepción de estas citas, la obra es enteramente mía;
- En la obra, he acusado recibo de las ayudas recibidas;
- Cuando la obra se basa en trabajo realizado conjuntamente con otros, he explicado claramente qué fue contribuido por otros, y qué fue contribuido por mí;
- Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.



Juan Ignacio Saldías

24 de septiembre de 2021

RESUMEN EJECUTIVO

El objetivo de este documento es a partir del análisis de un proyecto de implementación de un CRM de tarjetas de crédito, y de la estructura y procesos de la organización en la que se llevó a cabo, identificar aspectos relativos a la implementación de proyectos de *big data* en esta organización. Que, adicionalmente, entendemos que se pueden extrapolar a otras organizaciones de características similares. Esto con el objetivo de lograr al final del documento, un panorama claro sobre los temas más comunes que afrontan hoy las organizaciones en sus procesos de transformación hacia organizaciones más orientadas a los datos y las nuevas tecnologías. Los acelerados y grandes cambios que se han dado en esta materia en los últimos años, hacen que la búsqueda de la mejor estrategia para convertirse en una organización que obtenga el máximo rédito posible de la información que genera y maneja, y de las nuevas tecnologías disponibles, hacen que este sea un tema recurrente en las agendas de discusión en la alta dirección de las grandes empresas. En los últimos años, las empresas que lo han logrado hacer exitosamente, han logrado una ventaja competitiva que se refleja en sus resultados y crecimientos en el volumen de negocio. Y las que no, poco a poco se están debilitado y ven surgir y crecer a nuevos competidores.

Comenzaremos el documento por el análisis de la gobernanza de TI y datos en la organización, y la evaluaremos respecto al marco teórico de referencia. Entendemos que este es un paso necesario para entender con mayor claridad el proceso de implementación del proyecto en cuestión y poder evaluarlo adecuadamente. No es posible comentar respecto al proceso realizado durante el proyecto, sin antes haber repasado el contexto del negocio y las capacidades actuales de tecnología de la empresa.

A los efectos de enriquecer este análisis y para mostrar el potencial que presentan los proyectos de *big data* en las organizaciones de la actualidad, se hará un ejemplo simplificado de cómo podría apalancarse la organización en los activos de información logrados con la implementación del CRM de tarjetas para seguir agregando valor al negocio. Para ello se mostrará el proceso de implementación de un algoritmo basado en la información del CRM de tarjetas, pero quedará por fuera del alcance del documento el perfeccionamiento y ajuste de este algoritmo. También se adjuntan con la entrega de este documento, los archivos correspondientes a la implementación de este algoritmo en R Studio.

La organización seleccionada para el caso de estudio y que ha prestado su información ha sido uno de los bancos de plaza, pero ha solicitado que se mantenga el anonimato durante todo el documento y que se evite cualquier información que lo pueda identificar. Con lo cual se han tomado estos recaudos necesarios en el desarrollo del documento. Cabe aclarar que, dado que la organización analizada es una organización internacional, de gran porte, gran estructura y volumen de negocio, puede que mucha de las observaciones, recomendaciones y conclusiones a las que se llegue en este documento no apliquen ni sean extrapolables para organizaciones de menor tamaño y escala. Con lo cual, es importante tomar las observaciones y conclusiones que aquí se mencionan con ese criterio.

Al finalizar este documento, nos debería quedar una idea clara del caso real, la empresa analizada y el proyecto implementado. Así como el contraste entre cada uno de los aspectos referidos a su gobernanza de TI y datos, y el marco teórico de referencia. También se resumirán las principales recomendaciones para empresas que se encuentren en una situación similar, basadas también en el marco teórico y mejores prácticas de referencia.

ÍNDICE

DECLARACIÓN DE AUTORÍA	2
RESUMEN EJECUTIVO.....	3
1 CONTEXTO	6
1.1 Presentación de la empresa	6
1.2 Estructura de TI	7
1.2.1 Contexto banco	7
1.2.2 Análisis según marco teórico	9
1.3 Estructura gestión datos	15
1.3.1 Contexto banco	15
1.3.2 Análisis según marco teórico	18
2 PRESENTACION DEL PROYECTO	25
2.1 Contexto	25
2.2 Problemática	28
2.3 Equipo.....	32
2.4 Etapas de la implementación	33
2.4.1 Marco teórico.....	33
2.4.2 Aplicación al caso de estudio	34
3 APLICACIÓN	59
3.1 Consideraciones preliminares	59
3.2 Análisis del <i>dataset</i>	60
3.2.1 Visualización de la estructura.....	61
3.2.2 Depuración de ceros, nas, infinitos	63
3.2.3 Análisis de las variables numéricas	63
3.2.4 Análisis de las variables categóricas.....	65
3.3 Selección del algoritmo	66
3.4 Aplicación y resultados.....	70
4 CONCLUSIONES	76
REFERENCIAS BIBLIOGRÁFICAS	82

1 CONTEXTO

1.1 Presentación de la empresa

El banco analizado en este documento, es una empresa multinacional con presencia en varios países de América Latina, Europa, EEUU y Asia. Aunque su fuerte es la operación en las dos primeras regiones, y su casa matriz se encuentra ubicada en Europa. La empresa tiene una estrategia comercial agresiva que plantea llegar a todos los segmentos de clientes y productos de la actividad financiera, con lo cual mantiene una posición de liderazgo en cuanto a su share de mercado en todas las principales geografías en las que opera. Puntualmente en Uruguay, mantiene un share de mercado que se encuentra aproximadamente en el 28% del volumen de negocios del mercado financiero.

Respecto a su trayectoria y antigüedad de la operación, estas datan aproximadamente desde el año 1857 con sus comienzos en Europa. Aunque no fue hasta el año 1874 que comenzó a focalizarse en el negocio de banca comercial, y los años 1923 cuando comienza su expansión a nuevas geografías y 1946 cuando comienza su expansión a otras regionales del mundo.

Como todas las cosas, estas características que presenta la organización, tienen tanto ventajas como desventajas. Entre las principales ventajas que vale la pena destacar para el análisis en cuestión, podemos destacar el alto poder económico y capital con el que cuenta la organización, para invertir en aspectos en los que la alta dirección considere estratégicos. Así como también, la capacidad de recoger experiencias y mejores prácticas en las distintas geografías en las que opera, que luego puede replicar en otros mercados. Esto último gracias a un *governance* centralizado y coordinado desde la casa matriz, que ha permitido mantener una identidad única a pesar de la diversidad de geografías en las que opera. Por otro lado, el hecho de ser una empresa tan grande y con tanta antigüedad, hace que tenga ciertos aspectos culturales y de estructura con los que debe convivir y muchas veces jueguen en su contra. La realidad actual en épocas donde la tecnología evoluciona de formas muy disruptivas y aceleradas, requieren de cambios drásticos en la forma de llevar adelante varios modelos de negocio. Y el negocio de la intervención financiera, es uno de los que más está sufriendo este tipo de cambios. En los últimos años han surgido menos barreras a la entrada en el negocio, nuevos competidores, nuevas reglas, nuevas exigencias de los clientes y presiones a la baja de los precios. Esto obliga a las empresas que hace años están en el mercado, a modificar sus estructuras y su forma de gestionar en el negocio, para mantenerse competitivas. Y empresas con

este banco, con tantos años de trayectoria, culturas heredadas, estructuras heredadas y demás, tienden a ser muy lentas e ineficientes para transitar esos cambios. Típicamente, los intereses personales de la alta dirección, los gremios, la presión para mantener retornos a los accionistas, y la falta de comprensión del nuevo contexto y visión de largo plazo del personal superior, dificultan enormemente este tipo de cambios. Recordemos que, en la mayoría de los casos de este tipo de empresas, hablamos de personal superior que llevan en el mercado laboral más de 20 años y se formaron incluso antes. Para entonces, la revolución digital que sufren hoy varios modelos de negocio era aún muy incipiente, incluso si pensamos en 10 años atrás lo seguía siendo. Con lo cual, es esperable y normal que las organizaciones de este tipo estén atravesando este tipo de problemas. Y seguro son muy distintos sus problemas, a problemas como los que pueden estar atravesando empresas como Uber, PedidosYa, Mercadolibre, D Local o cualquier otra de las conocidas como *fintech* o unicornios nativos digitales. Ninguna de estas empresas tiene que lidiar con una estructura heredada que se formó en un contexto totalmente distinto al actual.

Para finalizar esta introducción a la organización, vale la pena repasar también brevemente su estructura y organigrama a alto nivel. La empresa cuenta con un gerente general (CEO) del cual dependen 10 directores encargados de liderar las principales áreas de la empresa: comercial, tecnología y operaciones, comunicación, auditoría interna, cumplimiento normativo, tesorería, control financiero, recursos humanos y administración, riesgos y legales. Cada uno de estos directores cuenta a su vez con gerencias encargadas de liderar equipos de 2 a 30 personas. Los equipos de menor tamaño no tienen mandos medios, mientras que en los de mayor tamaño existen jefes y subjefes encargados de coordinar las tareas de equipos más pequeños. Como excepción, cabe mencionar a la gerencia de distribución (ventas y atención de clientes) que depende de la dirección comercial. Esta tiene a su cargo casi 400 personas que se encargan entre otras cosas, de atender una red de 38 sucursales en todo el país. Con lo cual, existen gerentes zonales y por sucursal que diferencian a la gerencia de distribución del resto. En total al día de hoy, la empresa tiene en su conjunto aproximadamente 700 funcionarios activos en su plantilla.

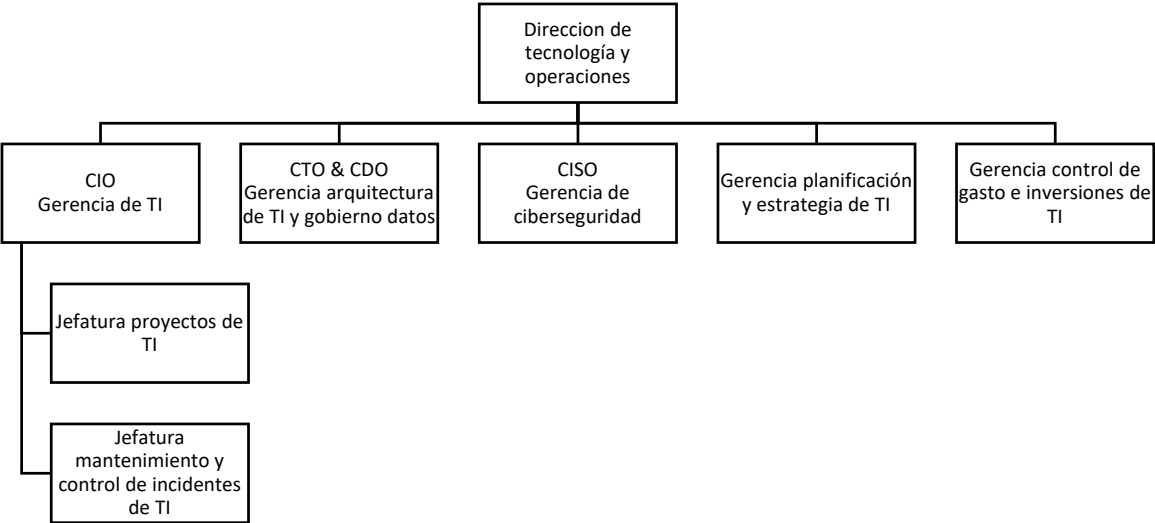
1.2 Estructura de TI

1.2.1 Contexto banco

El área de tecnología del banco cuenta con aproximadamente 30 personas que se distribuyen entre 5 gerencias. La mayoría del personal se concentra en la gerencia del CIO, que concentra 18 personas

de esta estructura: 10 personas dependen de la jefatura de proyectos de TI y otras 5 dependen de la jefatura de mantenimiento y control de incidentes de TI. En el resto de las gerencias, ninguna cuenta con más de 4 integrantes en el equipo. Como otros aspectos a mencionar de este equipo, cabe destacar que el promedio de edad del personal que ocupa cualquiera de estos puestos mencionados, es de no menos de 45 años de edad. Y adicionalmente, el nivel de formación en TI es bajo. El porcentaje de funcionarios en este equipo con formación terciaria en TI y/o experiencia en otras empresas del rubro es muy bajo (por debajo del 50%).

Se adjunta aquí debajo un esquema de la estructura de TI dentro de la dirección de tecnología y operaciones. No incluye estructura de la gerencia de operaciones que depende de la misma dirección.



Los funcionarios que forman parte de los equipos de proyectos de TI (10 personas), son responsables de la implementación del *software* que da soporte a toda la actividad del banco. Pero en ningún caso son personas con formación o habilidades en desarrollo de ningún tipo de lenguaje. Sus actividades se limitan a relevar requerimientos, tercerizar el desarrollo con proveedores, hacer el seguimiento con estos terceros externos y gestionar internamente la instalación de los nuevos *softwares* en la infraestructura del banco. Todos ellos, se caracterizan por tener varios años de trayectoria en la empresa y conocer en profundidad los sistemas actuales y procesos del banco.

En el relevamiento realizado para el resto de las áreas (jefatura de mantenimiento, CTO & CDO y CISO) se detectan situaciones similares. Poco personal con conocimientos técnicos profundos y un elevado grado de tercerización y dependencia de proveedores externos que cuenten con los conocimientos que se requieren.

1.2.2 Análisis según marco teórico

Para evaluar aspectos referidos al gobierno corporativo de TI, siempre es de utilidad referirnos a los estándares planteados por la ISO 38500. Esta norma brinda un marco de mejores prácticas para que las empresas que buscan la mejora continua en este aspecto, puedan tomarla como referencia y lograrla. Esta ISO define al gobierno corporativo de TI como *“el sistema por el cual el uso actual y futuro de TI es dirigido y controlado. La gobernanza de TI implica evaluar y dirigir el uso de las TI para dar soporte a la organización y monitorear el uso para lograr los planes. Incluye la estrategia y las políticas para la utilización de las TI en la organización.”*¹ La ISO habla además de la importancia que tiene en la actualidad, la alineación entre la gobernanza de TI y las áreas de negocio tanto a nivel de estrategia como funcional. Asimismo, plantea una serie de prácticas y estructuras que deberían ser compartidas por todas las organizaciones que deseen tener un adecuado gobierno de TI. En la tabla a continuación, se enumeran las principales y se agregan comentarios relevados para el caso del banco en cuestión.

PRÁCTICAS TI ESTRUCTURAS TI	OBSERVACIONES RELEVAMIENTO BANCO	COMENTARIOS SOBRE DEBILIDADES BANCO
Comité de estrategia de TI a nivel de junta directiva. TI debe ser un elemento regular de agenda e informe continuo al comité directivo.	No existe un espacio específico dedicado a tratar temas de estrategia de TI a nivel de dirección de la empresa. Existe un comité de dirección en el cual se presentan temas variados según la agenda y prioridades del momento, y regularmente se	TI debería tener un espacio exclusivo en el cual participe todo el comité de dirección. Esto ayuda a entender relevancia y extender una visión de TI a lo largo de toda la empresa.

¹ Organización Internacional de Estandarización & Comisión Electrotécnica Internacional (2015). *ISO/IEC 38500:2015 Corporate Governance for Information Technology*.

	<p>manejan allí temas referidos a TI. También hay espacios exclusivos entre dirección de TI y CEO, pero en los que no se participa al resto.</p>	
CIO en comité ejecutivo	<p>El CIO no forma parte del comité ejecutivo directamente, quien forma parte es el director de TI y operaciones.</p>	
CIO como reporte directo al CEO o COO en su defecto	<p>En este caso el CIO depende de un director, quien tiene a su cargo además al COO.</p>	
<p>Planificación estratégica de sistemas de información: proceso formal para definir y actualizar la estrategia de TI.</p>	<p>En lo que respecta a planificación de TI, el área de TI se encarga de anualmente hacer un relevamiento de las demandas de proyectos del resto de las áreas. Y en función de este relevamiento se planifica el año internamente desde TI.</p> <p>En lo que respecta a mejoras y mantenimiento de la TI actual, la planificación se hace exclusivamente desde TI.</p> <p>La dirección comercial es quien fija el presupuesto de clientes, volúmenes de negocio e ingresos. Y el área de administración es quien fija el presupuesto de gastos. En ambos procesos de</p>	<p>Se observa una separación entre la planificación estratégica y la gestión de TI. Esta es una visión antigua en la que TI era un área más de soporte a la cual el negocio u otras áreas realizaban demandas para cumplir en mediano plazo sin mayores inconvenientes.</p> <p>Hoy TI debe tener un rol mucho más activo en la planificación estratégica, lo cual incide por ejemplo en la forma de evaluación de inversiones de TI.</p>

	presupuestación anual la participación de TI es mínima.	
Comité de dirección de TI: evaluación de inversiones de TI a nivel de alta dirección.	<p>Las inversiones en TI no se discuten a nivel de dirección. Estas inversiones tienen un sponsor que puede ser de cualquier área de la organización (en su mayoría del área comercial). Para poder presentarlas cada sponsor debe tener el visto bueno de su dirección y de la jefatura de proyectos de TI. Con ambos vistos buenos, debe finalmente conseguir la aprobación de la gerencia de control de gasto en TI. No hay procesos únicos formales definidos para obtener los 3 vistos buenos requeridos.</p> <p>Al momento de fijación de los presupuestos anuales de volumen de negocio y gastos, rara vez se conoce con exactitud el número de proyectos de TI que se precisan llevar a cabo para que sean posibles y mucho menos el alcance funcional de cada uno.</p>	<p>En este punto también queda en evidencia la división entre la estrategia general de la empresa y la estrategia de definición de los proyectos de TI.</p> <p>Hoy en día, en un contexto en que las posibilidades de mejorar los servicios de intermediación financiera dependen en gran parte de las capacidades de digitalizar procesos, y por ende de las capacidades de TI. ¿Cómo es posible que se fije un presupuesto de ingresos sin haber pasado por una evaluación de TI?</p> <p>Esta forma de planificar, evidentemente distorsiona también los procesos de evaluación y gestión del gasto en TI.</p>
Comité de dirección de proyectos de TI: priorización y	No existe un comité de dirección de proyectos de TI. Existe un comité de TI donde se hace	En línea con los dos puntos anteriores. Queda en evidencia una separación errónea y

<p>seguimiento de proyectos de TI en forma conjunta entre negocio y TI.</p>	<p>seguimiento a temas varios del área, y al que se invita a algunos roles clave de otras direcciones.</p> <p>Algunos proyectos que lo ameriten por su escala, pueden tener sus comité de seguimiento específicos.</p>	<p>antigua entre los proyectos de TI y la estrategia de la organización, que impacta también en los procesos de seguimiento de los proyectos de TI.</p>
<p>Marco de control y gobernanza de TI basado en procesos formales y estandarizados.</p>	<p>No hay procesos estandarizados escritos formalmente. Se siguen prácticas habituales y cada gestor impone su impronta personal.</p>	<p>El hecho de que los procesos dependan de la impronta personal de cada gestor, dificulta la eficiencia y el logro de una visión única de TI a lo largo de toda la organización.</p>
<p>Metodología en gestión de proyectos de TI.</p>	<p>Hay poco personal con formación en metodología de gestión de proyectos. La organización tampoco promueve la formación en ese campo. La metodología depende de quién sea el gestor o sponsor de cada tema.</p>	<p>En la actualidad existe amplia y diversa bibliografía sobre la gestión de proyectos. Algunas más estrictas como el PMBOOK y otras más incipientes como las metodologías ágiles. Más allá de la discusión de cuál es la mejor, es importante que la empresa tenga claro cual se adapta mejor a su realidad, capacite y evalúe a su personal al respecto.</p>
<p>Informes y control presupuestario de TI.</p>	<p>Existe un área específica dentro de la estructura de TI encargada de este tema.</p>	

<p>Rotación de trabajo entre personal de negocio y personal de TI.</p>	<p>No existe ni sería viable con la dotación y estructura actual de la empresa.</p>	
<p>Liderazgo y rol del CIO: capacidad y autoridad del CIO para articular la visión de TI a lo largo de todos los niveles de la organización</p>	<p>CIO no tiene formación en liderazgo, ni tiene autoridad o espacios en la organización que le puedan permitir trabajar por una visión compartida de TI en toda la organización.</p>	<p>Se deberían generar los espacios necesarios para que la preocupación sobre los temas de TI sea compartida por toda la organización. Adicionalmente, si se tiene la visión de que el negocio de intermediación financiera tendrá que ver cada vez más con tecnología, es importante que cada vez más personal debería estar capacitado al respecto.</p>
<p>Evaluaciones periódicas o actividades que aseguren una adecuada gobernanza y control de TI.</p>	<p>No existen auditorías o controles internos que evalúen los procesos de gobernanza de TI. Hoy el foco está en la realización de controles o auditorías sobre procesos operativos que aseguren la continuidad del negocio. Pero no en la calidad de la gobernanza.</p>	<p>Debe abandonarse la visión de TI como un área más de soporte. La salud e idoneidad de los procesos de gobernanza de TI deben ser una preocupación también de las áreas de control y auditoría. Ya no son meramente un tema de gestión interna de la dirección de TI.</p>
<p>Campañas corporativas de concientización sobre la importancia y rol de TI.</p>	<p>No existen campañas de comunicación internas que ayuden a concientizar a toda la organización sobre el nuevo rol</p>	<p>Si se tiene la visión de que en el mediano plazo el negocio de intermediación financiera tendrá que ver cada vez más</p>

	de TI y su importancia. Si las hay dentro del propio sector de TI pero no se expande al resto de la organización.	con tecnología, es importante que todo el personal vaya tomando conciencia de ello, y sea incentivado a facilitar el proceso desde su rol.
--	---	--

A modo de resumen de este repaso de la estructura de TI del banco, podemos decir que llama la atención como en un contexto actual en el cual todo indica que el futuro de la intermediación financiera va hacia la digitalización y simplificación de los procesos, con clientes que están cada vez más acostumbrados a auto atenderse a través de canales digitales soportados en nuevas tecnologías, al día de hoy, solo 30 de 700 funcionarios estén trabajando en el área de TI (4%). Más aún, cuando a esto agregamos el hecho del bajo grado de capacitación en temas técnicos y de gestión de proyectos, que genera como consecuencia ineficiencias y elevado grado de dependencia de proveedores que sí poseen estos conocimientos.

No obstante, estas debilidades son consistentes con el bajo grado de participación de TI en la definición de la estrategia de la empresa, que se percibe al hacer un análisis de su realidad actual versus las prácticas de TI sugeridas por la ISO 38500. A su vez, esto último es consistente con las clásicas dificultades de lentitud para adaptarse a los cambios, mencionada también anteriormente para empresas de este tamaño y antigüedad. Con lo cual seguramente, la problemática de esta empresa no sea muy distinto a los que están afrontando otras del rubro que no sean “nativas digitales”.

Es de esperarse que muchas de estas características mencionadas respecto a la gobernanza de TI de la empresa, tengan su impacto en el proyecto de implementación del sistema de información de tarjetas de crédito, que se repasara en este documento.

Finalmente, la ISO 38500 propone también una clasificación de las organizaciones según la madurez de su modelo de gobierno corporativo de TI. Pudiendo ser un modelo inexistente en el caso de las empresas peor ubicadas en esta clasificación, hasta un modelo optimizado para el caso de las organizaciones más desarrolladas en este sentido. A la vista de los elementos comentados en la tabla anterior, el modelo de gobierno de TI del banco podría encasillarse según esta clasificación como repetible, que es el tercer escalón evolutivo entre los 6 que propone. Un modelo de gobierno

repetible se define como un modelo en que *“los procesos se han desarrollado al nivel en que procedimientos similares son desarrollados por diferentes personas que realizan la misma tarea. No hay capacitación formal o comunicación de procedimientos estándar, la responsabilidad se deja al individuo. Existe un alto grado de confianza en el conocimiento de las personas y, por lo tanto, es probable que se produzcan errores.”*² En función de la información relevada, el modelo de la empresa seguro no alcanza a clasificar como definido (siguiente nivel) porque no hay procedimientos 100% estandarizados y documentados, y puede ser discutible si no clasifica como un nivel anterior (nivel inicial), que se caracteriza por enfoque más ad hoc de cada uno de los temas y una menor coordinación en general. Esto es consistente con lo concluido anteriormente, respecto a que la empresa aún tiene mucho por hacer en lo que refiere a gobernanza de TI.

1.3 Estructura gestión datos

1.3.1 Contexto banco

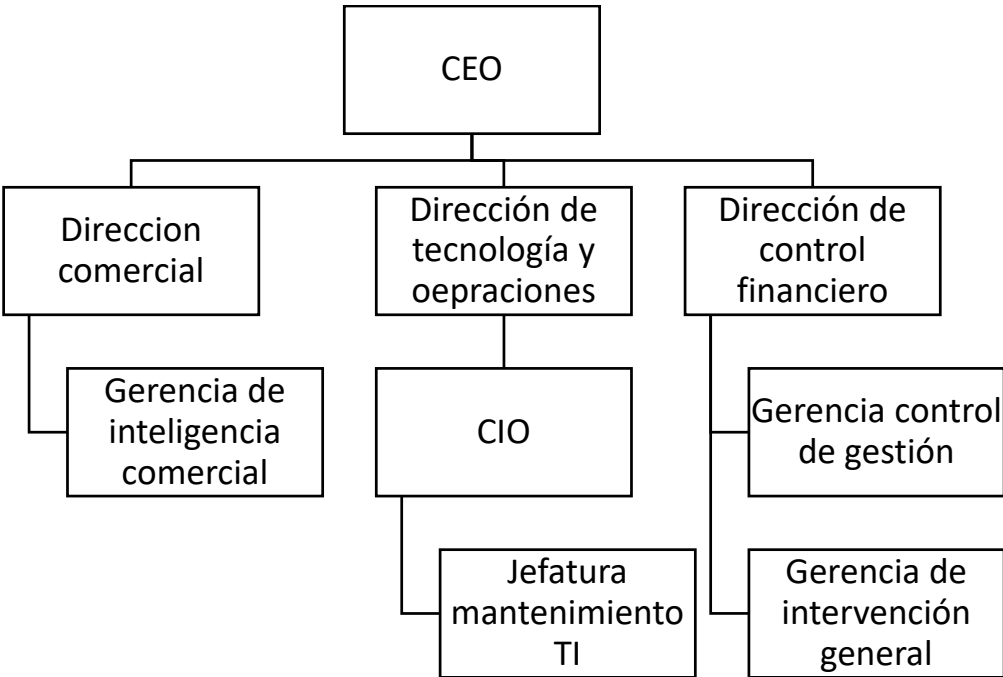
Según información relevada en la empresa, la estructura dedicada a la gestión de los datos ha sufrido algunos cambios en los últimos 3 años y aún continúa ajustándose y mejorando. Hasta el momento en que comenzó este proceso de mejora, la función de gestión de la información se encontraba dividida fundamentalmente entre la dirección comercial y la dirección de control financiero. Dentro del área comercial existe una gerencia de inteligencia comercial, que es quien se encarga de almacenar, depurar y generar informes de soporte para toda la gestión comercial de todos los productos y segmentos de clientes del banco. Por otro lado, dentro de la dirección de control financiero existen dos gerencias denominadas control de gestión e intervención general. Ambas áreas son responsables por el procesamiento, depuración y generación de otros informes, que, en muchos casos, se aplican sobre los mismos datos usados por la gerencia de inteligencia comercial. Estas gerencias se encargan en grandes líneas, de controles contables, *reporting* a reguladores, *reporting* a casa matriz y *reporting* al comité de dirección local.

La recopilación de los datos, está a cargo de TI, quien se encarga en este modelo de recibir las demandas de información de cualquiera de las gerencias mencionadas anteriormente, y asegurar la adecuada recopilación y almacenamiento de los datos necesarios en las bases de datos, así como su

² Organización Internacional de Estandarización & Comisión Electrotécnica Internacional (2015). *ISO/IEC 38500:2015 Corporate Governance for Information Technology*.

disponibilización de forma oportuna a las áreas que la procesarán. Específicamente, personal dentro de la jefatura de mantenimiento de TI es quien tiene esta responsabilidad. Es decir que, en este modelo, el rol de TI es un rol pasivo y limitado a asegurarse de que no se pierda información y disponibilizarla en tiempo y forma a quienes se encargarán de procesarla. No existe ningún tipo de coordinación formal que asegure que cuando procesen los mismos datos, las gerencias en cuestión lo hagan de la misma forma.

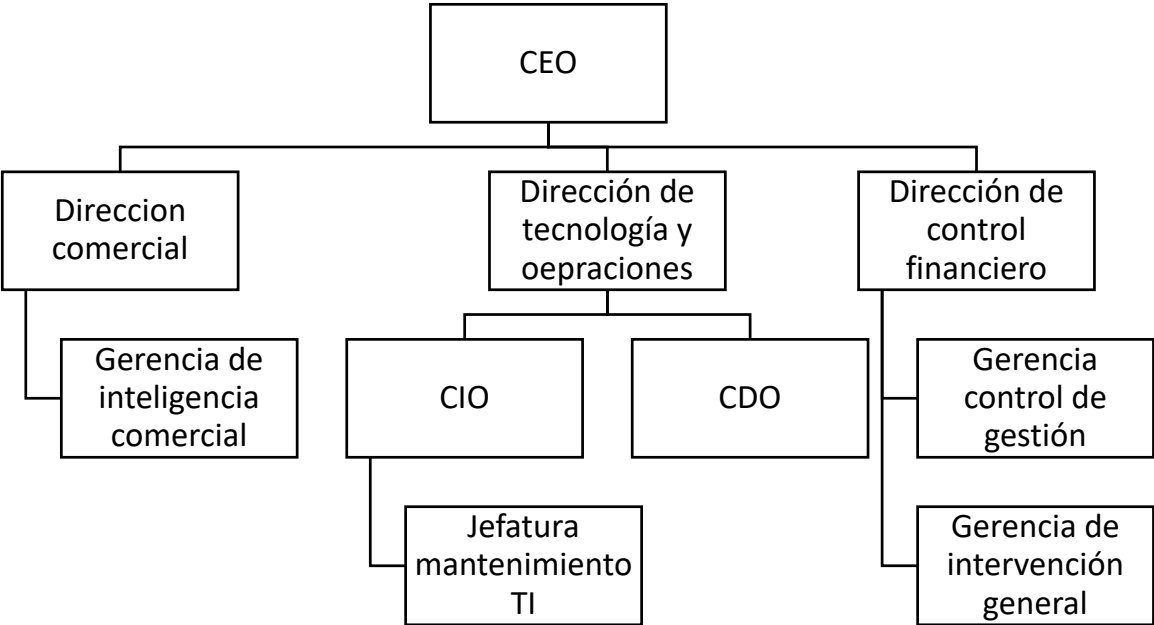
Gráficamente, vale la pena entonces representar a todos estos actores mencionados para el gobierno de datos en un organigrama como el que se presenta a continuación. Cabe aclarar, que no se están incluyendo en el organigrama a las áreas que no están vinculadas directamente con este proceso de procesamiento de información. Si bien luego, muchas otras áreas realizan demandas de información a una o más de estas gerencias, ya que son las únicas que tienen acceso directo a los datos y por lo tanto, tienen capacidad de procesarlos y compartirlos.



El cambio que se ha producido hace 3 años en la organización, es la incorporación de un CDO dentro del área de tecnología, con el objetivo de tener una visión transversal respecto a la gestión de los datos. Esta oficina del CDO está compuesta por 4 personas encargadas de apoyar en todos los proyectos de gestión de la información que se están intentando llevar adelante desde la creación de esta gerencia. Como todo desafío para una empresa que se encuentra en marcha, con un negocio funcionando y que no puede detenerse, el rol de esta nueva gerencia dentro de la organización es

aún muy insipiente. Aún siguen focalizados mayormente en gestionar proyectos, que le permitan tener herramientas y procesos necesarios para hacerse del control y administración de todos los flujos de información que suceden dentro de la empresa. Que son muchos y cada uno con sus sistemas, bases de datos y particularidades asociadas. Esto quiere decir, que la mayoría de los flujos de recopilación, almacenamiento, depuración, procesamiento y análisis de la información dentro de la organización, siguen aún sucediendo de forma similar a como lo hacían previo a este cambio. Según comentarios recibidos de personal de la empresa, también se percibe una cierta resistencia de las áreas que hoy manejan y controlan la información en ceder ese lugar de “autoridad” sobre el acceso a la información a esta nueva gerencia. Al parecer, en muchos casos se percibe como una amenaza a perder un lugar que hoy tienen de relevancia, y a su vez un riesgo, de que a que muchas de las tareas que hoy realizan y están acostumbrados a realizar de una cierta manera, puedan complicarse. Esto genera, en resumen, una resistencia al cambio por parte de ciertos actores clave y que tienen conocimiento de las bases de datos y la información de la empresa, que no es buena para el progreso hacia una organización mucho más orientada al manejo de datos.

Luego de este cambio, el esquema que veíamos sobre las áreas que intervienen en el proceso de gestión de la información se podría resumir de la siguiente manera.



A partir de este esquema, y sin haber entrado aún en detalle en el marco teórico relacionado con el gobierno de datos, podemos decir que pareciera que la responsabilidad sobre la gestión de la

información se encuentra distribuida en varias partes de la empresa. Lo cual seguramente, pueda estar generando ineficiencias en los procesos, duplicación de trabajo, inconsistencias y problemas adicionales asociados a los desafíos de coordinación entre varias partes en una empresa tan grande.

1.3.2 Análisis según marco teórico

1.3.2.1 Teoría de los silos

“Según un estudio de Price Waterhouse Coopers, el 50% de las compañías trabaja con silos organizacionales. Se le llaman silos organizacionales a la incapacidad para trabajar eficientemente entre las áreas o unidades de negocio que integran la empresa, es decir, que cada departamento dentro de la empresa se aísla para trabajar de su propia manera. Adicionalmente, está demostrado que las empresas que no cuentan con un sistema integrado de gestión son más propensas a tener este tipo de problemas.

Los silos complican los flujos de información, porque las organizaciones se vuelven lentas y poco ágiles y puede impedir que la organización tome decisiones basadas en datos fiables.

Aún en la actualidad, existen numerosas empresas donde cada departamento dispone de un sistema de gestión de información independiente y aislado. Debido a la gran cantidad de aplicaciones con las que generan y manejan información (sobre todo en las de gran porte), suele ocurrir que estas den origen a silos de información.”³

Un ejemplo claro sobre este problema, es el mencionado en el apartado anterior respecto al sentimiento de las áreas que manejan la información de la empresa respecto a la incorporación de una gerencia de CDO. Si bien la dirección de la empresa, seguramente este haciendo este cambio porque entiende la necesidad de mejorar aspectos de su gobernanza de los datos, las áreas que hoy manejan la información y están acostumbradas a hacerlo a su manera, lejos de percibirlo como un beneficio lo perciben como una amenaza. Anteponen sus intereses personales y no se comportan como facilitadores de un cambio que es necesario para la competitividad de la empresa en el mediano y largo plazo.

³ Blog Ibermática Industria. Accedido el 5 de agosto, 2021, desde <https://ibermaticaindustria.com/blog/trucos-para-romper-los-silos-organizacionales-con-el-sistema-de-gestion-de-la-informacion/>

También se puede decir que en *“tecnología de la información, un silo describe cualquier sistema de gestión que no puede funcionar con ningún otro sistema, lo que significa que está aislado de los demás sistemas. Los silos crean un entorno de sistemas individuales y dispares dentro de una organización.*

En otras palabras, un silo de información es un sistema de gestión de información que no puede comunicarse libremente con otros sistemas. La comunicación dentro de un silo de información siempre es vertical, lo que dificulta o imposibilita que el sistema funcione con sistemas no relacionados.

Un silo de información puede ocasionar problemas como la duplicación de esfuerzos y roles de trabajo redundantes. Conduciendo al desarrollo de sistemas contrastantes que pueden conducir a un aumento de los costos y la falta de sinergia.

Un cuello de botella en la información resulta en ineficiencia ya que diferentes departamentos pueden estar trabajando con una serie de entendimientos alternativos para completar un proyecto. Esto puede conducir fácilmente a una serie de oportunidades perdidas para el negocio, o en el peor de los casos, contribuir al fracaso general de una empresa.”⁴

Según algunas preguntas realizadas al personal del banco objetivo de análisis, entendemos que esta problemática que plantea la teoría de los silos no es ajena a la empresa. La ausencia por tanto tiempo de un rol que unifique las políticas para el manejo de la información y el tamaño de la organización hacen incluso natural que esto sucediera. Las áreas de inteligencia comercial, control de gestión e intervención general fundamentalmente, se han transformado en silos independientes que procesan y gestionan su propia información y la que otras áreas usuarias les demandan. Esto se suma a un rol muy pasivo de TI, que se limita a dar soporte para la recopilación y almacenamiento de la información a cada uno de estos silos. Con una alta probabilidad, la estructura resumida en el apartado anterior, sin duda tuvo una elevada incidencia en que, al día de hoy, la situación sea de esta manera. Otra posibilidad es que la estructura presentada sea una consecuencia de trabajar de esta manera, que, si bien tal vez se adaptaba a otros tiempos, no es acorde con la importancia y el rol que tiene la gobernanza de datos en la actualidad.

⁴ Blog Wearedrew. Accedido el 08 de agosto, 2021, <https://blog.wearedrew.co/departamentos-a-ciegas-silos-de-informaci%C3%B3n-y-su-efecto-en-tu-empresa>

Adicionalmente, por lo relevado con personal involucrado en estas tareas de gestión de la información, también se confirma la existencia de silos a nivel de sistemas de gestión de la información. Se pudieron identificar al menos 7 sistemas de gestión de información distintos, solo para los principales negocios de *retail* del banco. Muchos de ellos incluyen refieren a la misma información.

Según manifiestan los propios funcionarios del banco con los que se comentó sobre este tema, si bien la incorporación de la gerencia de CDO es un paso hacia lograr una mejora necesaria de la gobernanza de datos, la empresa aún se encuentra lejos de contar con un único sistema de información. Que permita agilizar y hacer más eficiente la gestión de la información, así como asegurar información confiable, consistente y compartida entre todos los niveles de la organización.

1.3.2.2 DAMA (DAMA Guide to the data management body of knowledge)

Este marco teórico, define a la *“gestión de los datos como a una función asociada a un proceso empresarial de alto nivel, que consiste en la planificación y ejecución de políticas, prácticas y proyectos que capturan, controlan, protegen, entregan y avalan el valor de los activos de datos e información de la organización. Su objetivo, es satisfacer las necesidades de información de todas las partes interesadas de la empresa en términos de disponibilidad, seguridad y calidad. Entre las funciones más importantes de la gestión de datos se pueden enumerar:*

- *Control o gobierno de datos: desarrollo e implementación de un conjunto de reglas, políticas, procesos y estándares para manejar los datos corporativos considerados como activos estratégicos de la organización.*
- *Gestión de arquitectura de datos: definir las necesidades de información de la empresa y diseñar los planos maestros requeridos para satisfacer estas necesidades.*
- *Desarrollo de datos: implementar y mantener soluciones que permitan satisfacer necesidades de datos.*
- *Gestión de operaciones de base de datos: adquisición, creación, almacenamiento y purga de las BBDD.*
- *Gestión de seguridad de datos.*
- *Gestión de datos maestros y de referencia: control y consistencia de la información.*
- *Almacenamiento de datos y gestión de inteligencia de negocios: disponibilizar información procesada para toma de decisiones.*

- *Gestión de contenido y documentación: control y protección del almacenamiento de datos.*
- *Gestión de metadatos.*
- *Gestión de calidad de los datos.*

Este mismo marco teórico, también destacan algunas actividades propias del gobierno de datos (primera de las enumeradas en el listado anterior), que vale la pena destacar para tener en cuenta en el análisis de esta organización en particular. Estas son

- *Estrategia de datos: programa para mantener y mejorar la calidad, la integridad, la seguridad y el acceso a los datos. Es recomendable que provenga de un entendimiento de las necesidades del negocio.*
- *Políticas de datos: pueden ser políticas sobre modelos, arquitectura, calidad, confidencialidad, seguridad, recuperación, intercambio o almacenamiento de los datos.*
- *Arquitectura de datos: debe ser desarrollado por los arquitectos de datos y administradores de datos como un equipo de custodia de los datos.*
- *Estandarización y definición de procedimientos para la recopilación, uso y mantenimiento de los datos.*
- *Cumplimiento normativo: en nuestro país las principales son Ley 18331 (datos personales y acción de habeas data), Ley 17616 (protección de derechos de autor) y Ley 17250 (Defensa del consumidor).*
- *Gestión de problemas asociados a la recopilación, uso o mantenimiento de datos.*
- *Gestión de proyectos asociados a la gestión de datos.*
- *Coordinación de la gestión de datos en la organización.*
- *Valoración de activos intangibles que representan los datos.*
- *Comunicación y difusión interna de asuntos relacionados a la información.”⁵*

Según lo comentado por personal contactado de la empresa en cuestión, pocas de estas funciones de la gobernanza de datos se estaban llevando adelante hasta la incorporación de la gerencia del CDO. De hecho, manifiestan que solo algunas de estas se llevaban adelante en la organización y no unificadas en un rol específico ni formalizadas en ningún tipo de documentación. Por ejemplo, TI se encargaba de definir algunos procesos para recopilación y almacenamiento de los datos. Y en la

⁵ Deborah Henderson. (2010) *DAMA Guide to the Data Management Body of Knowledge*. Boston.

dirección legal se velaba por el cumplimiento normativo en el manejo de información. Luego cada uno de los silos mencionados en el apartado anterior que manejan y gestionan información, tenían su propia estrategia de datos.

Este mismo personal contactado, afirma que si bien han habido avances con la incorporación de un CDO. Como por ejemplo un proyecto aún en curso, de migración de datos a la nube y creación de un sistema de información que consuma y disponibilice esta información a las diferentes áreas de la empresa procesadas según sus necesidades. Aún queda mucho camino por recorrer, y que los grandes proveedores de información interna siguen siendo las áreas de inteligencia comercial, control de gestión e intervención general. Y son también quienes aún imponen su impronta en temas referidos a gestión de la información.

1.3.2.3 Roles fundamentales para la gestión de big data

Según la literatura referida a cuáles son los perfiles con los que una organización que aspire lograr una adecuada gobernanza de *big data* debe contar, se reconocen 3 fundamentales que se detallan a continuación.

“ 1) Data Analyst: entendiendo el negocio

Con el propósito de quitarle al Data Scientist responsabilidades que lo pueden desenfocar de su principal tarea (que es el proyecto científico de datos en sí), la figura del Data Analyst es fundamental para brindar al equipo un entendimiento completo del negocio.

En líneas generales, se ocupa de analizar los datos que existen en la empresa y se utilizan para la gestión, se comunica e intercambia con las áreas consumidoras de la información, define alcances y es quien entiende si el modelo responde o no a las necesidades de los usuarios.

2) Data Engineer: que todo funcione y los datos lleguen

El Data Engineer arma el entorno de trabajo y el repositorio de datos para que el Data Scientist y el Data Analyst trabajen. Debe garantizar que, desde el punto de vista técnico, todo funcione. También que la infraestructura que soporte las bases de datos y que los procesos sean confiables. Asimismo, se encarga de testear nuevas propuestas y tecnologías que surgen para, eventualmente, continuar mejorando la infraestructura disponible, los procesos y las capacidades de recolección, almacenamiento y procesamiento.

3) Data Scientist

El Data Scientist se enfocará exclusivamente en todo lo relacionado con el problema científico.

- *En función del análisis exploratorio de datos define los experimentos a realizar, variables o características que son parte de los modelos.*
- *Tuneo de hiperparámetros y otras configuraciones necesarias para el desarrollo de los experimentos y la selección del mejor modelo posible.*
- *Colabora en la selección e ingeniería de features.*
- *Hace research de papers para poder aplicar nuevas ideas a los proyectos*
- *Estudia los sesgos en el modelado para proponer mejoras.*

Es decir, que es quien se encarga de construir los modelos de machine learning: toma el resultado del proceso de feature selection, diseña el modelo inicial, lo construye, parametriza y exporta para su entrenamiento. También se encarga de la mejora continua de estos modelos.”⁶

En función de esta información, y el comentario de la reciente incorporación de la gerencia del CDO en la empresa, cualquiera tendería a pensar que esta nueva área se ha armado contemplando estas mejores prácticas en lo que refiere a sus integrantes. En ocasiones esto puede ser complicado para empresas con estructuras heredadas en áreas de manejo de datos, pero dado que en este caso se está formando un área específica para esto y nueva, sería razonable y esperable comenzar con los perfiles adecuados. Sin embargo, según información relevada se detectó que:

- No hay perfiles como el del *Data Analyst* en este equipo. En general por cómo se encuentra estructurada la empresa, lo más similar que podemos encontrar a un *Data Analyst* son los propios responsables de los distintos segmentos de clientes o productos. Que son personal del área comercial encargados de administrar de forma centralizada cada uno de los negocios del banco. Ejemplo: cuentas, tarjetas de crédito, créditos, banca corporativa, banca *premium*, entre otros. Este es personal que históricamente ha demandado información a los sectores que la procesan, la ha utilizado para la gestión, se encarga de presupuestación, entre otras tareas similares. Con lo cual, tienen un elevado conocimiento del negocio y su información asociada.
- No hay perfiles como el del *Data Engineer* en este equipo. El propio CDO, que además es quien tiene también el rol del CTO es quien tienen el mayor conocimiento técnico para

⁶ Blog 7 puentes. Accedido el 10 de julio, 2021, desde <http://www.7puentes.com/blog/2021/06/02/data-analyst-engineer-scientist-jugadores-clave-en-un-proyecto-de-datos/>

poder cubrir este rol. Pero no hay dentro de su equipo personas con conocimiento de la infraestructura actual del banco ni tampoco de las mejores prácticas en cuanto a infraestructura hacia la que debería migrar el banco si quisiera lograr mejorar la gestión de su información (como por ejemplo podría ser: experiencia en manejo de infraestructura en *cloud*).

- No se contrataron *Data Scientists* con experiencia previa en esta tarea. Para este rol se tomaron a 2 personas con formación en estadística y negocio, que desempeñaban tareas en otras áreas del banco anteriormente, y se las está apoyando en su proceso de formación para desempeñar este nuevo rol. Si bien en este caso al menos está pensado cubrir la necesidad, se optó por una opción con una curva de aprendizaje mucho más lenta y que tal vez no parezca lo más adecuado para una empresa en marcha del porte del banco. Según lo indagado, tampoco hay mucho otro personal en el banco que tenga los conocimientos y experiencia que requeriría este puesto. Con el cual, en caso de precisarlos, deberían continuar este proceso de formación o salir a contratarlos fuera.

A la vista de esta información respecto al equipo de la nueva gerencia del CDO, podemos afirmar que aún es débil para afrontar proyectos de *big data* de gran porte. Ya que, desde el punto de vista técnico, aún tiene varias carencias que le impedirían llevarlo a cabo de la manera correcta y/o completa. Esto sin duda, no favorece la transformación del modelo de gestión de la información de la empresa y la mejora de su gobernanza de datos. Mucho menos, hacia un modelo de gobernanza de *big data*.

2 PRESENTACION DEL PROYECTO

2.1 Contexto

Como se comentara anteriormente, el banco atiende a variedad de segmentos de clientes, cada uno de los cuales tiene sus propias necesidades y por lo tanto requiere sus propios productos y canales de atención específicos, entre otras cosas. A modo de resumen, el banco abarca segmentos de clientes personas físicas, empresas unipersonales, pequeñas o grandes empresas. A su vez, dentro de cada uno de estos segmentos podemos identificar otras divisiones como personas físicas de altos ingresos, ingresos medios y bajos ingresos. Consecuentemente, el banco tiene una amplia variedad de productos, servicios y canales de atención, que en muchos casos son específicos para uno de estos grupos, y en otros casos, pueden ser compartidos por más de uno. Para que se entienda a lo que nos referimos como productos, tomemos como ejemplo el caso de personas físicas, cuyos productos principales son cuentas bancarias, tarjetas de débito, tarjetas de crédito, créditos al consumo y créditos hipotecarios. Lo importante de todo esto, es notar que el banco es una empresa en la que conviven varios modelos de negocio que, si bien comparten una marca, estructura, procesos y áreas de soporte, poco tienen que ver el uno con el otro.

Este contexto es importante, ya que como menciona una de las metodologías más populares para la gestión de proyectos de *big data* como CRISP-DM de IBM, todo proyecto de este tipo debe comenzar por un entendimiento del negocio sobre el que se está aplicando. Dejando de lado los proyectos que se llevan adelante con fines académicos o estadísticos, todos los proyectos de *big data* que se llevan adelante en empresas comerciales deben estar aplicados a un caso de uso y para un modelo de negocio particular. Con lo cual, el proyecto objeto de este apartado no es ajeno a esta realidad y es importante conocer algunas generalidades del negocio al cual se aplicó.

El proyecto en cuestión se trata de la implementación de un sistema de información para la gestión de los negocios de tarjetas de crédito del banco. Entonces, a los efectos de entender de forma resumida de que se trata el negocio de tarjetas de crédito del banco, podemos mencionar sus principales palancas que son:

1. Captación de nuevos clientes: venta de tarjetas a no clientes.
2. Uso y transaccionalidad de actuales clientes: en el sentido amplio, esto se trata de lograr que aquellos clientes que tienen una tarjeta la utilicen recurrentemente. E idealmente, lograr que el cliente centralice la mayor parte del gasto que realiza en su vida cotidiana a través

de estos medios de pago. El objetivo es que el cliente prefiera gastar con esta tarjeta antes que usar efectivo u otras tarjetas.

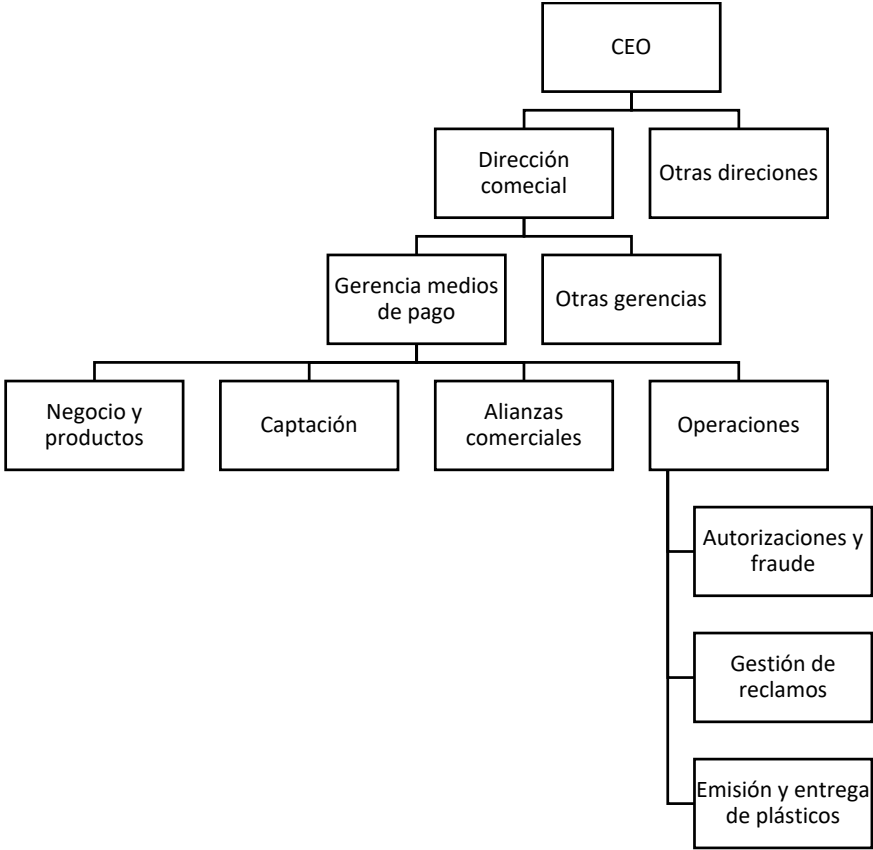
3. Transaccionalidad y gestión del fraude: vinculado al punto anterior, para lograr buenos ratios de uso, es importante que la tarjeta de crédito tenga buenos niveles de aceptación en la mayor cantidad de comercios posibles. Tanto físicos como digitales, y tanto locales como del exterior. Esto tiene dos caras. Una que refiere a las integraciones y tecnologías que son necesarias para lograr esa capilaridad en la aceptación del producto. Pero hay otra que refiere a la seguridad del cliente, las tecnologías y procesos, que son necesarios para que el cliente no sufra ningún tipo de fraude en el uso de su producto. El desafío de encuentra en lograr un equilibrio entre ambos.
4. Acceso al crédito: la tarjeta de crédito es una herramienta para facilitar el acceso al crédito al cliente. Esto tiene una doble cara en términos de gestión, una que refiere a incentivar a que los clientes utilicen este instrumento para financiar aquellas compras para las cuales no cuentan con suficiente liquidez y, por otro lado, evitar ofrecer crédito a clientes que luego no podrán pagarlo y por lo tanto generarán pérdidas al negocio.
5. Retención de clientes: lograr que el cliente mantenga sus productos a lo largo del tiempo y no los cambie por los de un competidor o deje de usarlos. Esto es un desafío principalmente porque: a) las necesidades del cliente son cambiantes a lo largo del tiempo, b) el contexto y la oferta de la competencia es cambiante y c) hay que procurar que el cliente tenga una buena experiencia en todas las interacciones que vaya manteniendo con el banco por un período prolongado de tiempo (incluso por otros negocios).
6. Fidelización de clientes: típicamente, los productos de tarjetas de crédito brindan acceso a sus tenedores a programas de beneficios. Estos son generalmente programas de puntos, programas de descuentos en comercios o programas de beneficios para viajeros. Estos beneficios, son importantes herramientas de vinculación y fidelización de clientes, que potencian no solamente todos los objetivos comentados en puntos anteriores propios del negocio de tarjetas de crédito, sino que también generan un efecto positivo en otros negocios del banco.

En definitiva, el área encargada de administrar el negocio de tarjetas de crédito, se debe encargar de diseñar productos competitivos que permitan captar clientes, y de realizar análisis que le permitan mantener buenos niveles de transaccionalidad sobre la cartera, buena participación sobre

el gasto total del cliente y un acompañamiento durante todo su ciclo de vida. También se encarga de implementar las acciones necesarias para concretar todos estos objetivos, y asegurar un adecuado nivel de servicio al cliente.

En lo que refiere a la monetización de este negocio. Fundamentalmente, el banco recibe ingresos en comisiones e intereses por la actividad de los clientes con sus tarjetas. Las principales líneas de ingreso refieren al cobro de anualidades de tarjetas de crédito, a los aranceles que pagan los comercios por el pago de clientes con tarjetas emitidas por el banco, y los intereses a los clientes que hacen uso de las opciones de financiación. De estos ingresos se netean varios gastos, asociados fundamentalmente a los programas de beneficios y comisiones cobradas por las redes (Visa y Mastercard).

Los negocios de tarjetas de crédito del banco, se gestionan desde una gerencia específica de Medios de Pago que depende directamente de la dirección comercial. Bajo esta gerencia trabajan aproximadamente 20 personas que se dividen tareas que refieren a la administración de este negocio. Aquí debajo se adjunta un organigrama resumen de referencia, con la ubicación de esta gerencia en el organigrama general y un resumen de los roles que existen dentro del área.



Cada una de estas áreas, tiene una jefatura y un equipo dedicado a resolver las tareas distribuidas por esta jefatura. La mitad del equipo total está concentrado en el área de operaciones. A los efectos de asociar los roles que aparecen en este organigrama con el negocio descrito en los párrafos anteriores, agregamos algunos comentarios sobre las tareas que realizan quienes se desempeñan en cada uno de estos roles.

1. Negocio y productos: se encargan del diseño de los productos de tarjetas y su análisis constante para mantenerlos competitivos y acordes a las necesidades de los clientes. También se encargan del análisis recurrente de las carteras para asegurar el cumplimiento de los objetivos y detectar oportunidades de mejora en cualquiera de las palancas del negocio.
2. Captación: se encargan de asegurar el cumplimiento de los objetivos de captación, fundamentalmente como nexo con los canales de venta (sucursales, call center, otros) y áreas de apoyo que son necesarias para que se puede llevar adelante la entrega y activación de las tarjetas.
3. Alianzas comerciales: se encargan del relacionamiento con los socios comerciales con los que el banco debe mantener una relación para poder brindar a sus clientes programas de beneficios. También se encargan del análisis de cada uno de los programas de beneficio en términos de inversión, rentabilidad, uso, etc.
4. Operaciones: se encargan de tareas operativas asociadas a que todos los procesos específicos del negocio de tarjetas de crédito funcionen correctamente. Así como también para que las tarjetas tengan un adecuado nivel de aceptación en los puntos de venta, y de analizar y dar respuesta a todos los reclamos realizados por clientes internos y externos sobre su experiencia con la tarjeta.

Esta introducción en líneas generales al negocio de tarjetas y la estructura del equipo que lo gestiona en el banco, es importante a los efectos de entender el caso de uso que da origen al proyecto en cuestión.

2.2 Problemática

Habiendo entendido la presentación anterior sobre el negocio y la estructura con la cual se trabaja, podemos profundizar ahora en la problemática que dio origen al proyecto en cuestión.

Como es habitual, la adecuada gestión de un negocio tan grande y variado como este requiere tener buena información que permita monitorearlo y actuar en consecuencia. Pero no nos olvidemos sobre el contexto de la gobernanza de datos en el banco. Según habíamos comentado, no existe en el banco una figura única consolidada que centralice temas de estrategia, políticas o arquitectura de datos. O incluso que centralice la gestión de problemas o proyectos de datos. La inexistencia de esta figura y una visión holística de la gestión de la información en la organización, generó que negocios como el de tarjetas de crédito, que han tenido un crecimiento acelerado y sostenido en los últimos 10 años fueran generando demandas de información aisladas a los departamentos de TI, inteligencia comercial, control de gestión o incluso proveedores externos. Y en función de estos reportes, armando y modificando el sistema de información que hoy utilizan para la gestión. Muy probablemente, la historia hubiera sido distinta si durante el proceso de crecimiento del negocio en el cual la visibilidad y el manejo de cierta información se vuelve crucial, los referentes del negocio hubieran tenido una figura de referencia con un adecuado rol sobre la gobernanza de datos. Pero, por otro lado, también es esperable que así fuera tratándose de un negocio que tiene varios años de funcionamiento, y que comenzó este proceso de crecimiento en un momento en el cual no se le daba a los datos y la información el rol relevante que hoy tiene para la gestión empresarial. Seguramente, nuevas empresas que están comenzando hoy su proceso de crecimiento, no enfrentarán este tipo de problemas. Consecuentemente, se puede decir que el proceso de generación y armado del sistema de información sobre el que hoy se basa todo el negocio es el resultado de un proceso liderado desde el área comercial del banco, específicamente, las jefaturas dentro de la gerencia de medios de pago. Proceso que se ha basado en demandas a otros sectores del banco que tienen acceso a la información, y en la implementación de mejores prácticas sugeridas desde proveedores o incluso desde la casa matriz. El resultado de este proceso, es un conjunto de planillas de Excel en las que mensualmente se agregan, de forma más o menos automatizada según el caso, el conjunto de datos que se quiere analizar.

Como comentamos también en el apartado de gobernanza de datos, si bien en el negocio se cuentan con personas con un rol que se asemeja mucho al rol de un *data analyst*, no hay en la organización personal con formación en el resto de los perfiles que se requieren para una adecuada gestión de *big data*. Con lo cual, sin un *data engineer*, parece razonable que exista todo un sistema de información tan crítico basado en planillas de Excel y no en cualquiera de las tecnologías hoy existen disponibles para almacenamiento, manejo y visualización de la información. También es coherente

que no existan *data scientists* en la empresa, ya que no hay información procesada disponible sobre las tecnologías necesarias como para generar modelos que agreguen valor. La única forma de que la empresa hubiera avanzado de otra manera en este sentido, aún sin contar con estos perfiles dentro de su planilla, podría haber sido tecerizándolos. Pero para eso, hubiera sido necesaria una visión de mediano y largo plazo centralizada, con formación sobre este tipo de temas. Para que lo priorizara y permitiera este tipo de subcontratación de recursos especializados y costosos.

Por otro lado, según comentarios relevados con empleados del área en cuestión el banco. Se pueden ver en el área de medios de pago y en la gestión de este negocio, varias evidencias del problema de silos comentado en el marco teórico. En primer lugar, en la medida que el negocio ha ido creciendo se han ido incorporando nuevos sistemas, necesarios por diferentes temas operativos o de gestión del negocio. Cada uno de estos sistemas con sus propias bases de datos, de las que surgen diferentes reportes con las mismas u otras métricas. En segundo lugar, vimos que la gerencia de medios de pago se divide en 4 jefaturas que se concentran cada una en determinadas funciones específicas que tienen que ver con el negocio. Esta estructura en el contexto del banco, ha generado que cada una de las jefaturas haga sus propios requerimientos de información a las áreas de TI, inteligencia comercial o control de gestión. Multiplicando la cantidad de reportes que se utilizan e incluso en muchos casos, repitiendo la información entre ellos. Esto es otro ejemplo, de cómo cada una de estas jefaturas dentro del área también han ido generando sus propios silos, con sus propios flujos de información y reportes.

Lo que agrava aún más la situación, es la combinación de todas las cosas: la falta de una visión unificada de la información, la demanda de la misma información desde diferentes jefaturas hacia diferentes áreas del banco y la existencia de diferentes bases de datos desde donde se puede extraer esta información. El resultado de esto, no es solo la variedad de reportes para mostrar la misma información, que en definitiva representan ineficiencia para la organización. Sino que también, existe un elevado riesgo en que exista inconsistencias entre los diferentes reportes. Ya que son diferentes equipos quienes los elaboran, sin una guía (criterios) sobre de qué sistema debería extraerse esa información en particular y cómo. Nada asegura que cuando existan coincidencias, los datos que reciban las diferentes jefaturas del área sean los mismos.

Según comentarios recibidos por parte del personal del área en cuestión, su sentir es que esta situación descripta llevó a que

- Se duplicarán y solaparán tareas entre las diferentes jefaturas en lo que refiere a armado de informes.
- Se generará desconfianza en la información cuando se detectan diferencias.
- Se invierte tiempo y esfuerzo en encontrar, identificar, explicar y corregir las diferencias cuando corresponde.
- Se generen malos entendidos entre los equipos producto de situaciones generadas por las diferencias en la información.
- No se cuente con un sistema de información con capacidad de procesar grandes cantidades de datos. Según lo requiere un entorno cada vez más competitivo y exigente.
- No se cuente con un sistema de información que permita evolucionar hacia la aplicación de modelos y algoritmos. Que requieren de la existencia de datos procesados y ordenados que viabilicen este tipo de inteligencia. Necesaria también en el entorno competitivo actual.
- No se respeten mejores prácticas en lo que refiere a almacenamiento y mantenimiento de datos.
- No existan definiciones o metadata sobre indicadores básicos del negocio. Muchas veces, aunque ciertos indicadores parezcan triviales, no podemos estar seguros sobre si estamos hablando de lo mismo salvo que exista una definición consensuada o estemos tomando el mismo dato del mismo lugar. Por ejemplo, se mencionan como casos típicos para este caso de estudio: ¿qué es una tarjeta activa?, ¿cuándo decimos que un cliente ha dado de baja el producto?, ¿en qué etapa del flujo de venta se considera a una tarjeta como vendida?

Esta problemática, es la que dio origen al proyecto de implementación de un nuevo sistema de información para la administración del negocio de tarjetas de crédito del banco objeto de este trabajo. Queda en evidencia entonces, que el objetivo principal de esta implementación fue la unificación de todas las fuentes de información que se usan dentro de la gerencia de medios de pago, de forma tal de eliminar las ineficiencias y problemas asociados a la duplicidad de información y tareas. Al mismo tiempo que, el proceso serviría para contribuir con la organización en su proceso hacia convertirse en una organización más orientada hacia los datos, con una visión única de la gestión de la información, que le permita procesar mayores volúmenes de información y aplicar nuevas y mejores técnicas de explotación de la información.

2.3 Equipo

El proyecto de implementación de este sistema de información de tarjetas, fue llevando adelante entre la jefatura de negocio y productos de medios de pago, y una persona designada dentro de la gerencia de inteligencia comercial. No hubo otras personas directamente involucradas en el equipo designado para el proyecto, sino que estas personas ocasionalmente realizaron demandas a otros en la organización cuando así lo consideraron y necesitaron.

El solo hecho de que el equipo haya sido conformado de esta manera, deja en evidencia varias de las debilidades comentadas en los apartados de gobernanza de TI y de datos. Si lo pensamos, este es un proyecto que tiene una componente mucho más fuerte de TI que de negocio. Ya que se trata de analizar los procesos de generación, recopilación, almacenamiento y disponibilización de la información. Además de evaluar y seleccionar la infraestructura y softwares más adecuados para ello. El rol del negocio podría pasar simplemente por definir los requerimientos funcionales y enumerar los KPIs o indicadores que precisa monitorear durante su gestión. Sin embargo, la dirección comercial se encargó de la ejecución punta a punta del proyecto. Esto es deja en evidencia varias carencias a nivel de las estructuras de TI sugeridas por la ISO 38500 como podrían ser:

- Falta de coordinación de estrategia de TI a nivel de la junta directiva.
- Falta de planificación estratégica de los sistemas de información.
- Carencias en el control y gestión de proyectos de TI en la organización.

Adicionalmente, siendo que este proyecto comenzó y sucedió en el período durante el cual la organización ya contaba con un CDO, resulta difícil de explicar cómo este proyecto no fue liderado y gestionado a través de esa gerencia. Si esta gerencia será la que defina la estrategia, política y arquitectura de datos en la organización, resulta raro que sea otra dirección la que lidere el proceso de cambio del sistema de información de uno de los principales negocios del banco. A primera vista, esto parecería ser una clara evidencia de lo mucho que falta posicionar a la gerencia del CDO dentro de la organización, así como también lo mucho que le hace falta evolucionar a la cultura de la organización hacia una mucho más orientada a los datos. Esto solo es viable lograr, dándole al CIO y CDO la autoridad, relevancia y ámbitos, que propone la ISO 38500.

Otro aspecto que genera dudas, es el no haber incluido dentro del proyecto a todas las áreas que hoy intervienen en la generación de información para este negocio. En el análisis de la gobernanza de datos comentamos que tanto las áreas de inteligencia comercial, control de gestión, tecnología

e intervención general, participan de una forma u otra en estos procesos. Con lo cual, dada la relevancia para la estrategia del banco del negocio en cuestión, pareciera importante contar con representantes de todas las áreas en el equipo que llevará adelante el proyecto. Fundamentalmente, para asegurarse de entender todas las fuentes de información existentes al día de hoy, y que no se interrumpan ninguno de los procesos que hoy se estén llevando adelante en la organización. También puede ser discutible, si no deberían haberse incluido en el equipo del proyecto a un *stakeholder* de cada una de las jefaturas de medios de pago, ya que son todos usuarios finales de la información con visiones e intereses diferentes. Toda esta participación, ya que en ausencia de un liderazgo de la gerencia de CDO, y siendo una organización que muestra varias evidencias de silos, lo mejor sobre todo en las fases iniciales del proyecto es contar con representantes de todas las partes que estar interviniendo en los procesos de información actuales. Luego, en la medida que el proyecto se encauce y avance, probablemente no se requiera de todos estos actores con la misma intensidad que al inicio, o incluso se pueda prescindir de alguno de ellos.

2.4 Etapas de la implementación

2.4.1 Marco teórico

Al buscar en internet, podemos encontrar varias teorías y procesos sugeridos para llevar adelante un proyecto de *big data*. Si bien cada uno tiene sus particularidades, es válido decir que la mayoría de ellos cuentan en grandes rasgos, con las etapas que se definen a continuación. Si bien en varios casos las podemos encontrar aún más desagregadas, en mucha bibliografía podemos identificar estas etapas a alto nivel.

“Etapa 1: Definición de los elementos básicos como objetivo y equipo.

Como cualquier proyecto, el primer paso básico es definir el objetivo que queremos conseguir. A partir de ello, será más fácil determinar qué recursos necesitamos y desglosar el proyecto en tareas. Posteriormente, cada una de las tareas puede tener objetivos propios.

Etapa 2: Determinar los datos a integrar.

En esta fase habrá que decidir qué fuentes de datos se van a utilizar. Es decir, de dónde se van a sacar los datos. Para ello, lo mejor es trabajar en dos áreas. Por un lado, desde el área de negocio de la empresa se debe determinar qué datos realmente necesitan y para qué. Por otro lado, hay que trabajar el área de las propias fuentes de datos. Es decir, analizar los sistemas de la empresa para

entender qué fuentes de datos existen. Esta fase debe acabar con un documento que incluya las diferentes dimensiones y métricas imprescindibles para cubrir el objetivo del proyecto.

Etapa 3: Integrar los datos.

Una vez se tienen los datos y la utilidad que queremos darle, el siguiente paso es integrarlos. Esto se conoce como procesos ETL (Extract, Transform and Load). Herramientas habituales en proyectos de Big Data como Hadoop o Bases de Datos NoSQL (MongoDB, Cassandra o HBase) tienen su propio enfoque a la hora de extraer y subir datos.

Se deben desarrollar servicios que permitan cualificar el dato como consistente y real. Por último, se debe integrar en el timing adecuado para que la operativa o los análisis sean relevantes para la organización.

Etapa 4: Análisis de datos dentro del proyecto.

La última fase es el análisis de los propios datos. Esto permitirá que el proyecto sea operativo y se pueda controlar su correcto funcionamiento.”⁷

2.4.2 Aplicación al caso de estudio

Si bien según información relevada en la empresa, el desarrollo de este proyecto en los hechos no se llevó adelante siguiendo conscientemente una metodología o un proceso formal, podemos identificar varias tareas realizadas que hacen sentido con el marco teórico presentado anteriormente. Con lo cual, para presentar las diferentes tareas que fueron realizadas por la empresa según el relevamiento realizado, las agruparemos según las etapas vistas en el marco teórico anterior.

2.4.2.1 Etapa 1: Definición de los elementos básicos

1) Reunión de *kick off*:

Se realizó una reunión de *kick off* con presencia de la dirección comercial, gerencia de medios de pago, gerencia de inteligencia comercial y las dos personas involucradas directamente en el proyecto. En esta reunión se sentaron las bases respecto a lo que se esperaba del proyecto y las necesidades y situación afrontadas por el negocio, ya comentadas en el apartado de descripción de

⁷ Web EALDE. Accedido el 30 de julio, 2021, desde: <https://www.ealde.es/fases-proyecto-bog-data-empresarial/>

la problemática. Al igual que se mencionó en el apartado de definición del equipo, esto deja evidencias debilidades respecto a la gobernanza de TI y datos. Un proyecto como este debería surgir de reuniones de planificación estratégica de TI y proyectos de TI. Y discutido previamente en ese ámbito. Adicionalmente, debería haber sido liderado por la gerencia de CDO y haber participado a todos los *stakeholders* e involucrados en el proyecto. Como se describe aquí, el proyecto fue gestado como un tema interno de la dirección comercial. Esta reunión de *kick off* fue el visto bueno para comenzar a trabajar en el proyecto. Sin por ejemplo hasta el momento, haber evaluado impacto en tecnologías necesarias, costos en los que se podría incurrir, sin comunicarlo al departamento de TI y sin consultar disponibilidad de otros equipos que pudieran estar involucrados. Todo esto, por no seguir prácticas sugeridas para una adecuada gobernanza de TI, lo cual representa un riesgo elevado para el éxito del proyecto.

2) Reuniones de planificación:

Existieron reuniones posteriores de trabajo y planificación entre las personas designadas al proyecto del equipo de inteligencia comercial y medios de pago del banco. En estas sesiones se buscaron acuerdos sobre las tareas que se debían llevar adelante, dinámica de trabajo y plazos estimados para el desarrollo del proyecto. Un resumen del resultado de estas sesiones, fue presentado luego a las gerencias de medio de pago e inteligencia comercial para la aprobación del plan de trabajo. Una vez obtenido este visto bueno, comenzó a ejecutarse el plan.

2.4.2.2 Etapa 2: Determinar los datos a integrar

3) Reuniones con equipo de medios de pago:

Como punto de partida, se realizaron sesiones de trabajo con los usuarios finales de la información para que estos ayudaran a especificar exactamente cuáles eran sus necesidades de información. A modo de resumen, se despejaron con estos usuarios dudas respecto de:

- ¿Qué información y datos utilizan actualmente para sus gestiones?
- ¿De dónde y cómo obtiene esa información actualmente?
- ¿Qué información precisan actualmente pero no cuentan?
- ¿Qué problemas o dudas tiene sobre la información que maneja actualmente?
- ¿Con qué frecuencia requiere cada tipo de información?
- ¿De qué formas prefiere visualizar y acceder a la información?

- ¿Qué usos y aplicaciones le da a la información? ¿Es solo para visualización o realiza algún procesamiento posterior?
- ¿Tiene conocimientos de alguna herramienta de manejo de datos? ¿Cuál es su nivel de conocimientos y habilidades técnicas para el manejo de datos?

Toda la información recabada al respecto durante estas primeras sesiones de trabajo, es clave para entender el alcance funcional de nuevo sistema de gestión de la información que se debía crear. Así como también para identificar similitudes y diferencias entre la información que requieren los distintos usuarios finales. La última pregunta específicamente, tiene que ver con información útil para la selección de la tecnología y software en la que se basará el programa.

4) Definición de indicadores de negocio incluidos en el sistema

A partir de inputs obtenidos en reuniones con el equipo de medios de pago. Se realizó un listado taxativo de los indicadores que se debían incluir en la primera versión del nuevo modelo de información. Este listado fue armado por la persona del equipo de medios de pago que integró el proyecto.

A modo de resumen, se identificaron los siguientes indicadores que debían ser incluidos de forma indispensable en el nuevo sistema de información.

a) Información sobre cantidades de tarjetas

- a. Cantidad de tarjetas titulares activas. (habilitadas para consumir)
- b. Cantidad de tarjetas adicionales activas.
- c. Altas de nuevas tarjetas. (ventas)
- d. Bajas de tarjetas. (cancelaciones)
- e. Cantidad de tarjetas transaccionales. (con consumos)

b) Información específica por tarjeta

- a. Número de contrato (único por titular).
- b. Número de tarjeta: uno por cada tarjeta titular o adicional del contrato.
- c. Límite de crédito.
- d. Fecha de alta.
- e. Condición económica: indica si el cliente tiene alguna bonificación o no en el costo anual de la tarjeta.

- f. Fecha de cobro de cuota: fecha en que cliente paga la próxima primera cuota de anualidad de la tarjeta.
 - g. Fecha de vencimiento.
- c) Información sobre consumos con tarjetas
 - a. Cantidad de transacciones.
 - b. Cantidad de facturación. (consumos en puntos de venta)
 - c. Apertura indicadores de transacciones y facturación por varios indicadores de interés: por moneda, contado/cuotas, local/externo, punto de venta físico/internet, otras.
 - d. Transacciones y volumen de intentos de consumo no confirmado. Es decir, cantidad de intentos de autorización de transacciones rechazados. Para este punto interesa particularmente entender también el motivo del rechazo.
- d) Información de ingresos e inversión asociada a negocios de tarjetas
 - a. Ingresos por cobros de anualidades.
 - b. Ingresos por aranceles generados por consumos de cliente.
 - c. Ingresos por financiación de estado de cuenta.
 - d. Otros ingresos asociados a la operativa con tarjetas.
 - e. Inversión en programas de fidelidad.
 - f. Inversión en programas de descuento en puntos de venta.
 - g. Inversión en pagos a sellos Visa y Mastercard.
 - h. Otras inversiones asociadas al negocio de tarjetas.
- e) Formas de pago por cliente
 - a. Canales de pago de estado de cuenta de tarjetas utilizados: digital, redes de cobranza, cajas, débito automático en cuenta, otros.
 - b. Cantidad y volumen de pagos realizados por cliente.
 - c. Cantidad y volumen de saldos atrasados. (tarjetas con estados de cuenta vencidos)
 - d. Cantidad y volumen de saldos financiados. (clientes que optan por no pagar el total de su estado de cuenta y financiar el remanente)
- f) Información del titular de la tarjeta
 - a. País, tipo y número de documento.
 - b. Nombre completo.

- c. Fecha de nacimiento.
 - d. Sexo.
 - e. Datos de contacto: celular, mail, dirección.
 - f. Fecha de alta cliente.
 - g. Segmento: según clasificación de clientes interna del banco.
 - h. Subsegmento: según clasificación de clientes interna del banco.
- g) Información de otra vinculación con el banco
- a. Tenencia de otros productos: cuentas, créditos, seguros, etc.
 - b. Volumen asociado a otros productos.
 - c. Ingresos a inversión asociado a otros productos.

Sobre estos indicadores, se definieron en este punto también algunos lineamientos que serían restricciones importantes para considerar en los próximos pasos de armado del modelo:

- a) Toda la información debería estar disponible para los usuarios con una frecuencia mínima mensual. Idealmente, los usuarios solicitaron poder contar con esta información en tiempo real, pero por su conocimiento de la situación actual de los sistemas, entienden que esto no es viable.
- b) Toda la información debería estar desagregada de tal forma de poder entenderla por cada tarjeta de forma individual. Es decir, que la información agregada debería ser el resultado de la sumatoria de ese indicador para el total de tarjetas activas. A modo de ejemplo y para que se entienda al punto, si el total de transacciones con tarjetas en un mes específico es de 10, se quiere poder llegar a la información de cuáles son las tarjetas que generaron esas transacciones y cuantas realizó cada una. Esto para todos los indicadores de transacciones, volumen de facturación, inversión, ingresos y demás.
- c) Se quiere conservar información histórica mensualizada. Es decir, que se requiere poder consultar todos estos indicadores para cualquier mes en particular. De forma tal de poder realizar análisis versus información histórica de la cartera, y cumplir con varios reportes exigidos desde reguladores, dirección local y dirección corporativa que son exigidos con esta recurrencia.

Respecto a esta etapa, vale la pena resaltar como claramente se va cumpliendo lo comentado en el apartado de gobernanza de datos en el sentido que es la persona del equipo de medios de pago quien está cumpliendo el rol de *data analyst* en este proyecto de *big data*. Ya que es quien, a partir

de las charlas realizadas con los usuarios finales ha definido cuales son los indicadores de negocio que se quieren medir e incorporarán al nuevo modelo.

Adicionalmente, vuelven a quedar en evidencia en este punto debilidades respecto a la gobernanza de TI y datos dentro de la organización:

- Sin haber incorporado aún al equipo de TI al proyecto, se están realizando definiciones que tienen un elevado grado de dependencia de su gestión. A modo de ejemplo, se están realizando definiciones respecto a recurrencia en la disponibilidad de la información y almacenamiento de información histórica. Ambas cosas, son aspectos que deberían ser definidos por el equipo de TI de la organización o por un *data engineer*. Esto podría ocasionar re trabajos o trabas a futuro en el proyecto, que podrían haberse anticipado con una metodología más adecuada.
- No se está sistematizando ni guardando en un lugar centralizado información que es relevante para una adecuada gobernanza de datos. Cada uno de los indicadores de negocio identificados en este apartado implican definiciones, como por ejemplo: ¿qué se entiende por una tarjeta activa?, ¿cómo fecha de alta de una tarjeta se considera la fecha de emisión o la de entrega al cliente? o ¿cómo se asignan ciertos ingresos o inversiones a cada tarjeta?. Esta información generalmente se almacena como “metadata” que acompaña a los datos propiamente, y asegura que cualquier usuario en un momento posterior del tiempo tendrá información adicional suficiente como para interpretar de forma adecuada los datos que se le están presentando. Con esta metodología de trabajo y la falta de involucramiento de la gerencia del CDO, esta es información que en el proyecto se está perdiendo.
- No se generaron espacios de discusión de los indicadores seleccionados con todos los sectores que hoy intervienen en el manejo de la información. Según lo visto en el repaso de los procesos actuales de la organización las áreas de control de gestión, intervención general y tecnología también forman parte activa de los procesos de procesamiento y disponibilización actuales de la información. Con lo cual, si bien no tienen por qué tener incidencia directa sobre el análisis y requerimientos del negocio, si hubiera sido una buena práctica incorporarlos al trabajo conjunto para estas definiciones y análisis por su conocimiento de los flujos de información actuales.

5) Reuniones con equipo de tecnología

En este punto de grado de avance del proyecto, se generaron las primeras instancias de puesta a punto y requerimientos a las gerencias del CIO y CDO. El objetivo de esta instancia, fue poner al tanto a estas gerencias del proyecto en curso, los avances hasta el momento y requerirle su apoyo para definiciones vinculadas a la recopilación, almacenamiento y disponibilización de toda la información definida en el punto anterior.

Como era de esperarse, el resultado de esta instancia no fue bueno. Ambas gerencias explicaron fundamentalmente que:

- Se estaba minimizando y simplificando las implicancias de los requerimientos realizados. Con esto, se referían a que la mayoría de los requerimientos que se estaban realizando requerían de bastante tiempo de análisis para ser respondidos de forma adecuada y que adicionalmente, el proyecto no estaba en la planificación del equipo de tecnología y por lo tanto no había personal disponible para realizar los análisis solicitados en el corto o mediano plazo.
- La organización se encuentra en un proceso de implementación de un sistema de información en la nube, en el cual se buscará nuclear toda la información para la gestión de todos los negocios del banco. Con lo cual, estuvieron en desacuerdo en trabajar y destinar recursos en este proyecto paralelo.

La propuesta del equipo de tecnología fue entonces, poner en pausa este proyecto hasta que se pudiera seguir adelante en el marco de la implementación del sistema de información para la gestión de información en la nube que se está implementando desde la gerencia del CDO. Adicionalmente, sugieren que este proyecto debe sea tomado por la gerencia del CDO cuando sea posible.

Dado que la dirección comercial entendía que precisaba una mejora urgente en lo que refiere a la información del negocio de tarjetas, y que no podía esperar a que se logran avances en el proyecto comentado por el equipo de TI, se decidió seguir avanzando de todas formas con el proyecto. Ahora con la premisa, de que no contarían con el apoyo del área de TI para hacerlo, con lo cual, habría que implementar el sistema aprovechando la infraestructura, softwares y procesos ya disponibles. O como máximo incorporar solamente aquellas cosas que pudieran ser fácilmente tercerizables con un proveedor y de bajo costo. Ya que sería muy difícil conseguir aprobación de inversiones para un proyecto que no está avalado por el área de TI.

En esta etapa del proyecto, quedan en evidencia varios de los puntos que veíamos anticipando respecto a las debilidades en la gobernanza de datos y TI de la organización. Queda demostrada la falta de coordinación entre las direcciones de TI y comercial. Si existieran los ámbitos adecuados para discusión de proyectos de este tipo, y procesos conocidos y compartidos para su aprobación, esto no podría haber sucedido. También queda en evidencia, la falta de autoridad del CIO y CDO en temas relativos al manejo de la información, ya que a pesar de que el proyecto no contara con su visto bueno, se decidió avanzar. Es clave el empoderamiento de estas figuras en temas de este tipo si se quiere lograr una visión única y suficientemente profesional en temas referidos al manejo de datos.

También quedan en evidencia problemas de comunicación interna. Ya que si la organización se encuentra en procesos de implementación de un sistema de información que beneficiará a todas las áreas y que cumple con todos los estándares técnicos requeridos, no tiene sentido que una dirección esté impulsando otro proyecto en forma paralela, buscando satisfacer la misma necesidad.

6) Relevamiento de las fuentes de información disponibles

Una vez entendida la información que se quiere disponibilizar con este sistema de información, y conociendo la premisa de que el sistema debería minimizar los requerimientos al área de tecnología del banco, se realizaron algunas reuniones involucrando a personas de los equipos de inteligencia comercial, control de gestión e intervención general para conocer las fuentes de información que se utilizan actualmente.

En estas reuniones de trabajo, se detectó que los sistemas de los cuales se podría obtener la información requerida eran principalmente los que se detallan a continuación (no se incluye el nombre de cada software a pedido de la empresa):

- Sistema propio del banco encargado de la autorización de transacciones de crédito.
- Sistema propio del banco encargado de administración de productos de tarjetas de crédito, estados de cuenta y transaccionalidad por cliente.
- Sistema de seguimiento de ventas de productos.
- Sistema de administración de programas de fidelidad y programas de descuentos.
- Sistema contable del banco.
- Sistema de administración de clientes del banco.

Algo que quedó en evidencia en este punto, es el hecho de que algunos indicadores relevantes se pueden tomar de más de uno de estos sistemas. Y pueden existir diferencias entre ellos, incluso tratándose de información básica. Por ejemplo, la información de transaccionalidad se encuentra en más de uno de estos sistemas y en todos hay diferencias. Esto se debe a cómo sucede el flujo de una transacción de tarjeta de crédito, y desde dónde y cuándo recibe la información de transacciones cada uno de los sistemas. Cada sistema tiene un universo de total de transacciones diferente en cada momento del tiempo, lo cual habitualmente genera diferencias. Como este, se identificaron varios detalles que hicieron necesario definir para cada indicador cuál es el dato más confiable y que se quiere que todos los usuarios manejen para su gestión. Esto tiene que ver y es una evidencia del problema de silos en los sistemas y procesos de información que se comentara en el apartado de gobernanza de datos.

También se identificó en este punto, que hay mucha información que no necesariamente se encuentra registrada en las bases de datos de un sistema. Sino que solo existe en planillas o documentos auxiliares de los gestores del negocio. El ejemplo más claro, podría ser lo que sucede con la inversión realizada en pagos a los sellos Visa y Mastercard. A nivel de los sistemas del banco, lo que se registra es un pago mensual a estos proveedores, pero no la composición de ese pago a nivel de cada concepto o cada tarjeta de cliente que generó el gasto. En la composición de ese pago, es que se puede entender la inversión que se está realizando por cada tarjeta activa con la que cuenta el banco, lo cual era parte del alcance solicitado para el sistema en construcción. Pero esta información solo existe a nivel de planillas auxiliares que maneja el equipo de medios de pago que controla estos pagos a los sellos. No se generan registros a nivel de ningún sistema y por lo tanto no es consultable posteriormente. Como este, se detectaron varios otros ejemplos de información requerida para las que no existía al momento una fuente única y confiable. Sobre todo, en lo que refiere a asignación de algunas inversiones a cada tarjeta de crédito en específico. Si bien lo ideal hubiera sido en este punto realizar alguna reingeniería de procesos que permitiera no perder esta información a nivel de los sistemas en el proceso de pago mensual, para minimizar los requerimientos al área de TI y simplificar el proyecto, lo que se decidió para estos casos fue definir criterios que permitieran asignar esta información a nivel desagregado por tarjeta de crédito. Esto último, para poder cumplir con el requerimiento inicial de contar con toda la información desagregada a nivel de cada tarjeta activa.

El resultado de esta etapa del proyecto, fue entonces una matriz en donde se vinculó cada uno de los indicadores relevados dentro del alcance del sistema con la fuente desde donde se debería obtener y la frecuencia con que podría hacerse. También se definieron en este punto, los criterios de asignación de aquella información que no se encuentra registrada en las bases de datos, sino que solo se puede hallar por el momento, en planillas auxiliares (como el ejemplo comentado más arriba de los pagos a los sellos).

DATO	FUENTE SUGERIDA	FRECUENCIA MÁXIMA DISPONIBLE
Cantidad de tarjetas titulares	Sistema administración de tarjetas	Mensual
Cantidad tarjetas adicionales	Sistema de administración de tarjetas	Mensual
Altas nuevas tarjetas	Sistema de seguimiento de ventas de productos	Diario
Bajas tarjetas	Sistema de administración de tarjetas	Mensual
Cantidad tarjetas transaccionales	Sistema de administración de tarjetas	Mensual
Número de contrato	Sistema de administración de tarjetas	Mensual
Número de tarjeta	Sistema de administración de tarjetas	Mensual
Límite de crédito	Sistema de administración de tarjetas	Mensual
Fecha de alta tarjeta	Sistema de administración de tarjetas	Mensual
Condición económica	Sistema de administración de tarjetas	Mensual

Fecha de cobro cuota	Sistema de administración de tarjetas	Mensual
Fecha de vencimiento	Sistema de administración de tarjetas	Mensual
Transacciones	Sistema de administración de tarjetas	Mensual
Facturación	Sistema de administración de tarjetas	Mensual
Transacciones no confirmadas	Sistema autorización de transacciones.	Mensual
Facturación no confirmada	Sistema de autorización de transacciones	Mensual
Ingresos cobro anualidades	Sistema contable del banco	Mensual
Ingresos aranceles	Sistema contable del banco	Mensual
Ingresos financiación	Sistema contable del banco	Mensual
Otros ingresos	Sistema contable del banco	Mensual
Inversión programas fidelidad	Sistema programas de fidelidad y descuentos	Diario
Inversión programas de descuento	Sistema programas de fidelidad y descuentos	Diario
Inversión pagos a sellos	Sistema contable del banco	Mensual
Otras inversiones	Sistema contable del banco	Mensual
Pagos de tarjetas	Sistema de administración de tarjetas	Mensual
Saldos atrasados	Sistema de administración de tarjetas	Mensual

Saldos financiados	Sistema de administración de tarjetas	Mensual
Documento cliente	Sistema administración de clientes	Diario
Nombre cliente	Sistema administración de clientes	Diario
Fecha nacimiento	Sistema administración de clientes	Diario
Sexo	Sistema administración de clientes	Diario
Datos de contacto	Sistema administración de clientes	Diario
Fecha alta cliente	Sistema administración de clientes	Diario
Segmento	Sistema administración de clientes	Diario
Subsegmento	Sistema administración de clientes	Diario
Tenencia otros productos	Sistema contable del banco	Mensual
Volumen otros productos	Sistema contable del banco	Mensual
Resultados otros productos	Sistema contable del banco	Mensual

De este relevamiento, surge que no sería viable la implementación de un sistema que disponibilice toda esta información con una frecuencia superior a mensual, ya que la mayor parte de los datos requeridos no están disponibles con una recurrencia menor. Cabe aclarar, que el principal motivo por lo cual esto sucede de esta manera es que el departamento de inteligencia comercial, no tiene acceso directo a las bases de datos de los sistemas, sino que se nutre de descargas de información de estas bases que son proporcionadas por el departamento de TI. En su mayoría con frecuencia

mensual. Dado que el departamento de TI no aceptó nuevos requerimientos de procesos o accesos para este proyecto, no se pudieron mejorar estos plazos para la versión inicial del sistema. Este es otro ejemplo, de cómo las carencias mencionadas a nivel de gobernanza de TI y datos, están teniendo un impacto negativo en el proyecto y la calidad del entregable. Lo esperable en este punto, hubiera sido que se trabajara junto con el departamento de TI para encontrar oportunidades de mejora de estos plazos y evaluar, por ejemplo, la viabilidad con un sistema de información lo más cercano a tiempo real posible.

Lo que es importante entender en este punto, es que todas estas deficiencias del sistema (en este caso la falta de recurrencia y oportunidad de la información) tienen efectos negativos en la gestión del negocio. Ya que no se pueden hacer mediciones de acciones y tomar medidas correctivas en tiempo real, sino que se contará con información a mes vencido lo cual muchas veces puede ser inoportuno o tarde para tomar acciones correctivas. Es decir que, está quedando en evidencia como defectos en la gobernanza de TI y los datos, tienen un impacto negativo directo en el negocio. Esto ayuda a entender que, una adecuada gobernanza de TI y datos en el contexto actual, no son un fin en sí mismos, sino que son una necesidad para que la empresa tenga una ventaja competitiva y pueda desempeñar mejor su negocio.

7) Evaluación y selección de tecnología

Como se comentara anteriormente, dado que el proyecto no se realizó de forma coordinada entre los equipos de dirección comercial y dirección de TI de la empresa, no hubo espacio durante su implementación para la evaluación e incorporación de nuevas tecnologías. Idealmente en esta instancia del proyecto, se requeriría de bastante apoyo de un *data engineer* que identificara los requerimientos de infraestructura y *software* más apropiados para la implementación de este proyecto. Que además deberían ser consistentes con la arquitectura de la empresa y la visión que exista para esta en un mediano plazo. Sobre este punto podemos decir que, aunque se hubiera podido solicitar el apoyo de un *data engineer*, ya comentamos que la empresa no cuenta al día de hoy aún con personal de esas características. No obstante, se podría haber tercerizado puntualmente para este proyecto si se hubiera planificado de manera adecuada.

En este contexto, lo que se sugirió desde el área de inteligencia comercial del banco y se utilizó para avanzar, fue seleccionar de las herramientas disponibles al día de hoy en la empresa las que mejor se adaptaban para el caso. Las herramientas seleccionadas entonces para la implementación fueron SQL Server, Alterix y Qlickview. A continuación, comentaré brevemente sobre cada una.

SQL Server:

“Microsoft SQL Server es un sistema de gestión de bases de datos relacionales (RDBMS) que admite una amplia variedad de aplicaciones de procesamiento de transacciones, inteligencia empresarial y análisis en entornos informáticos corporativos. Microsoft SQL Server es una de las tres tecnologías de bases de datos líderes del mercado, junto con Oracle Database y DB2 de IBM.

Al igual que otros programas RDBMS, Microsoft SQL Server se basa en SQL, un lenguaje de programación estandarizado que los administradores de bases de datos (DBA) y otros profesionales de TI utilizan para gestionar las bases de datos y consultar los datos que contienen. SQL Server está vinculado a Transact-SQL (T-SQL), una implementación de SQL de Microsoft que añade un conjunto de extensiones de programación propias al lenguaje estándar.

Al igual que otras tecnologías RDBMS, SQL Server se construye principalmente en torno a una estructura de tablas basada en filas que conecta los elementos de datos relacionados en diferentes tablas entre sí, evitando la necesidad de almacenar datos de forma redundante en varios lugares dentro de una base de datos. El modelo relacional también proporciona integridad referencial y otras restricciones de integridad para mantener la exactitud de los datos. Estas comprobaciones forman parte de una adhesión más amplia a los principios de atomicidad, consistencia, aislamiento y durabilidad, conocidos colectivamente como propiedades ACID, y están diseñadas para garantizar que las transacciones de la base de datos se procesen de forma fiable.

El componente principal de Microsoft SQL Server es el motor de base de datos de SQL Server, que controla el almacenamiento, el procesamiento y la seguridad de los datos. Incluye un motor relacional que procesa los comandos y las consultas y un motor de almacenamiento que gestiona los archivos de la base de datos, las tablas, las páginas, los índices, los búferes de datos y las transacciones. El motor de base de datos también crea y ejecuta procedimientos almacenados, desencadenantes, vistas y otros objetos de la base de datos.

Debajo del motor de base de datos se encuentra el sistema operativo de SQL Server o SQLOS. SQLOS se encarga de las funciones de nivel inferior, como la gestión de la memoria, la programación de trabajos y el bloqueo de datos para evitar actualizaciones conflictivas. Por encima del motor de base de datos se encuentra una capa de interfaz de red que utiliza el protocolo Tabular Data Stream de Microsoft para facilitar las interacciones de solicitud y respuesta con los servidores de base de datos. Y a nivel de usuario, los administradores de bases de datos y desarrolladores de SQL Server escriben

sentencias T-SQL para construir y modificar estructuras de bases de datos, manipular datos, implementar protecciones de seguridad y realizar copias de seguridad de las bases de datos, entre otras tareas.”⁸

El departamento de inteligencia comercial del banco tiene toda su información en una base de datos relacional SQL Server, y por lo tanto toda la información requerida para el armado de este nuevo de sistema de información debió volcarse previamente en tablas dentro de esta base de datos. Gran parte de este trabajo ya se hacía, y poco más se tuvo que incorporar a los flujos mensuales de aprovisionamiento de datos.

Cabe aclarar que, para proyectos de analítica de datos, muchas veces se evalúa la utilización de bases de datos indexadas u orientadas a columnas, ya que presentan grandes ventajas en lo que respecta a la velocidad de lectura y procesamiento para análisis posterior. Aquí identificamos otra desventaja en la realización de este tipo de proyectos sin la intervención de un *data engineer* u otro personal técnico idóneo para este tipo de decisiones. La elección de la mejor base de datos tiene impacto en la performance del sistema y, por lo tanto, en las demoras, tiempos de trabajo y eficiencia de los *data analyst* y *data scientists*. Es decir que, en este caso, podríamos estar incurriendo en ineficiencias por mayores demoras, que en definitiva es más tiempo improductivo y mayor inversión económica para realizar el mismo trabajo para la empresa. Todo esto de nuevo, como consecuencia de una mala gobernanza de TI y datos.

Alterix:

“Es una herramienta multidimensional que nos permite desde preparar nuestro conjunto de datos hasta obtener modelos predictivos pasando por la limpieza de éstos, la aplicación de análisis de diferentes orígenes y la capacidad de unir o mapearlos con una enorme facilidad.

Alteryx permite operar a través de las diferentes herramientas con una enorme variedad de opciones. Podemos pasar desde hacer modificaciones estándar a través de fórmulas a nivel de fila así como operar vertical y horizontalmente hasta análisis predictivo o análisis espacial junto con la creación de aplicaciones analíticas.

⁸ Web Searchdatacenter. Accedido el 20 de agosto, 2021, desde <https://searchdatacenter.techtarget.com/es/definicion/SQL-Server>

Alteryx dispone de hasta 260 posibles herramientas que se pueden utilizar como parte de un flujo de trabajo. Comprende los diferentes procesos desde la obtención de datos pudiendo realizarse desde archivos locales así como conexiones a múltiples bases de datos.

Alteryx no es una herramienta que simplemente permite realizar operaciones de ETL puesto que dispone, también, herramientas de reporting e incluso podría sustituir o, por lo menos, complementar el uso de otras herramientas de business intelligence.”⁹

“En muchos proyectos de big data, 80% del tiempo se pasa preparando datos para analítica. Alteryx cambia las reglas del juego al permitir que los analistas obtengan y combinen datos, los limpien y los enriquezcan con insight predictivo o espacial, de manera rápida e intuitiva. Potencia Tableau, QlikView, Microsoft Power BI u otras herramientas de visualización con datos confiables y ágiles.

A modo de resumen, sus principales ventajas son:

- Combinación de datos más rápida: accede y combina datos de todas las fuentes, incluso datos locales y en la nube, todo en un solo flujo de trabajo intuitivo.*
- Crea visualizaciones más inteligentes: agrega fácilmente análisis predictivo, prescriptivo, espacial y estadístico a los datos.*
- Preparación visual: un entorno de preparación de datos completamente con clics y sin códigos permite a los analistas crear un conjunto de datos confiable y perfecto con facilidad.*
- Plantillas analíticas y el soporte nativo para cargar Tableau, Power BI y Qlik con datos combinados y enriquecidos directamente.*
- Preparación de datos sin supervisión: las funcionalidades de programación y automatización integradas permiten programar y ejecutar flujos de trabajo analíticos según sea necesario.*
- Comparte flujos de trabajo de preparación de datos: un repositorio centralizado permite la colaboración, el control de versiones y el uso compartido de funcionalidades, asegurando sólida gobernanza en torno a la preparación de datos.*
- Escala: pasa de cargar directamente herramientas de analítica visual a cargar un datamart o almacén de datos.”¹⁰*

⁹ Web The Information Lab. Accedido el 1 de septiembre, 2021, desde <https://www.theinformationlab.es/que-es-alteryx-y-para-que-sirve/>

¹⁰ Web Alteryx. Accedido el 1 de septiembre, 2021, desde <https://www.alteryx.com/es-419/solutions/analytics-need/bi-visualization/visual-analytics>

En el caso del proyecto en cuestión, Alteryx permitió implementar flujos de trabajo que permitan de forma recurrente y automatizada:

- Unificar los datos que provienen de las diferentes fuentes mencionadas: sistema de autorización de transacciones, sistema de administración de tarjetas, sistema de seguimiento de ventas, sistema de programas de beneficios, sistema contable y sistema de clientes.
- Trabajar y realizar los cálculos y transformaciones necesarias sobre los datos para que todos queden asignados correctamente a nivel de cada tarjeta activa.
- Disponibilizar los datos unificados y transformados para que puedan ser utilizados de forma correcta por el sistema que se disponibilizará para que los usuarios accedan y visualicen la información. Que en este caso fue Qlickview.

Qlickview:

“Qlikview es una herramienta business intelligence y analytics poderosa y de fácil utilización para la construcción de cuadros de mandos y la navegación de los datos.

Como definición, QlikView proporciona un análisis empresarial rápido, potente, altamente interactivo y residente en memoria, sin las limitaciones, el coste o la complejidad de las herramientas business intelligence tradicionales. QlikView se despliega en días y los usuarios se forman en minutos.

Principales características:

- *Es una herramienta con la que se puede de forma fácil acercarse a los datos de una base de datos relacional y navegar por ellos de forma natural.*
- *Puede conectarse con diversos orígenes de datos: desde bases de datos Access a conexiones ODBC hacia cualquier servidor de bases de datos u otro origen de datos, como ficheros Excel, o simplemente ficheros de texto.*
- *La carga de datos pasa a través de un script de carga que utiliza un lenguaje muy parecido al SQL y, por tanto, de fácil comprensión. Una vez que los datos han sido cargados se pueden, o no, almacenar en un fichero de formato Qlickview (fichero QVD). También pueden utilizarse directamente en el mismo cuadro de mandos.*

- *El cuadro de mando se puede construir fácilmente utilizando distintos objetos: tablas pivot, tablas planas, gráficos, texto, menús con filtros, entre otros. Todo en entorno Windows y sin necesidad de escribir ni una línea de código.*
- *Si queremos personalizar el cuadro de mando o disponer de botones con código asociado también se pueden desarrollar subrutinas y conectarlas a eventos.*
- *Todos los objetos se pueden colocar en reportes, imprimirlos individualmente, exportarlos a excel o transformarlos en tablas o gráficos.*

El sistema utiliza toda la RAM disponible con la ventaja de tener siempre a disposición, de forma rápida, todos los datos que queremos navegar en Qlikview.

QlikView ofrece una nueva perspectiva para las aplicaciones basadas en bases de datos. De forma distinta respecto a las tradicionales herramientas de BI sobre bases de datos, QlikView presenta al usuario distintas posibilidades de desarrollo grafico solucionando de esta manera las más diversas necesidades.

En lugar de trabajar con tablas, forms o queries complejas, el usuario trabaja con una interface sencilla basada en point and click. De esta manera, QlikView ayuda a comprender e interpretar fácilmente los datos, algo que los usuarios de negocio aprecian especialmente.”¹¹

En el proyecto en cuestión, se elaboraron tableros y cuadros de mando en esta herramienta para que todos los usuarios que lo deseen del banco puedan acceder a la información procesada por Alteryx a través de una herramienta intuitiva y sencilla. Se elaboraron cuadros de mando que apuntaron inicialmente, a satisfacer las necesidades y simplificar la gestión del equipo de medios de pago como principales usuarios finales de la organización. El objetivo de estos cuadros de mando es que los usuarios puedan acceder a las vistas que requieren para su trabajo diario sin tener que hacer ningún tipo de transformación o trabajo sobre la información. Y que, además, siempre que lo requieran para un mayor análisis tengan la posibilidad de descargar la información a Excel u otras herramientas útiles para este tipo de tarea. Incluso a nivel de desagregación de cada tarjeta que compone el dato que están visualizando el cuadro de mando de forma agregada. Todo esto, a través de una interface sencilla que no requiere que el usuario tenga conocimientos de programación de ningún tipo.

¹¹ Web Tech-BI. Accedido el 5 de septiembre, 2021, desde <https://www.tech-bi.com/es/tecnologias/qlikview/>

Es decir que, a modo de resumen, se puede decir que se aprovecharon 3 herramientas que ya existían en el banco para la ejecución de este proyecto:

- SQL Server como motor de base de datos. Se elaboraron procesos que permitieran volcar de forma mensual toda la información requerida de los diferentes sistemas listados. En algunos casos estos procesos ya existían.
- Alteryx como herramienta que permitió la unificación y transformación de forma automatizada y recurrente de la información provenientes de diferentes tablas de la base de datos.
- Qlickview como herramienta para facilitar y centralizar el acceso de los usuarios finales a la información.

2.4.2.3 Etapa 3: Integrar los datos

8) Implementación de procesos de extracción de información

Una vez definidos los datos que se desean obtener con el nuevo sistema, sus respectivas fuentes de información y la tecnología que se utilizaría, se estuvo en condiciones de avanzar con los procesos de aprovisionamiento de la base de datos en SQL Server y los procesos de transformación de la información en Alteryx. Este proceso fue llevado adelante por parte del personal de inteligencia comercial que integró el proyecto, apoyándose de manera ad hoc de referentes en el área de TI para lo que tuvo que ver con bases de datos de los sistemas mencionados del banco, y con un proveedor externo para asesoramiento en la implementación de procesos de transformación de datos en Alteryx.

El primer punto no presentó mayores dificultades, ya que son procedimientos a los que ambas áreas (TI e Inteligencia Comercial) ya están acostumbrados y forman parte de su modelo de gestión actual.

El segundo punto tomó un poco más de tiempo ya que la herramienta Alteryx era de reciente implementación en la organización, y no existían precedentes en la organización de un trabajo de esta magnitud realizado sobre la herramienta. El resultado del procesamiento de toda la información en Alteryx es una tabla cuya clave primaria es la combinación de número de contrato (número único por tarjeta), mes y año. Y en cada campo de esta tabla se asignan los datos del cliente, transaccionalidad, facturación y resultados, para cada contrato en un mes en particular. De esa forma, se lograban satisfacer exitosamente los requerimientos de contar con información de todos

los indicadores solicitados de forma desagregada por tarjeta y mes del año. La apertura por mes año es lo que permite el registro de información histórica que viabiliza los comparativos con los que se quiere contar para el seguimiento mensual del negocio.

La asignación de la información asociada a cada contrato, se asignó de dos formas:

- Información que ya estaba asociada a un contrato de tarjeta específico en otro sistema: se vinculó usando el dato del número de contrato para vincular las tablas y tomar la información deseada.
- Información que no estaba asociada a un contrato de tarjetas específico en otro sistema: en estos casos se toma el dato disponible en otro sistema y define un criterio para asignarlo por contrato. Este tipo de transformaciones y asignaciones pueden ser realizadas utilizando Alteryx. El ejemplo más claro, es el ya comentado respecto a los pagos a los sellos Visa y Mastercard. Si bien estos pagos en su mayoría dependen de la cantidad de transacciones realizado por cada tarjeta, el único dato del que se dispone en el sistema contable es el total de pagos realizados a cada sello en el mes. Entonces, lo que se hizo en este caso es que Alteryx tome el dato del total de pagos y realice un prorrateo por contrato según las transacciones realizadas ese mismo mes por cada uno. Ya que el dato de las transacciones si se encuentra disponible por número de contrato en otras de las tablas. Este mismo tipo de lógica se aplicó para varios de los indicadores incluidos en la tabla final.

A pesar de ser reiterativo, vale la pena aclarar que en este punto vuelven a quedar en evidencia debilidades respecto a la gestión de datos en la organización. En primer lugar, el departamento de inteligencia comercial sigue actuando como un silo con sus propios criterios y metodología para el tratamiento de la información. Sin supervisión ni bajo los lineamientos del CIO o CDO. Esto no ayuda al objetivo de lograr una visión única de organización orientada hacia los datos en el mediano plazo. Por otro lado, se mencionó una inversión económica y de tiempo en formar a parte del personal en esta nueva herramienta Alteryx, lo cual si bien es positivo para el equipo no está alineado con las herramientas que están promoviendo el CIO y CDO para este tipo de tareas. Con lo cual, aquí hay potenciales problemas de mantenimiento, soporte, renovación de licencias y gestión del conocimiento dentro de la organización. Se están apalancando procesos clave para la gestión en herramientas que estarán por fuera de los planes de TI, formación de recursos humanos y captación de recursos humanos. Lo cual no lo hace sostenible en el largo plazo. Esto es otra clara evidencia,

de fallas en los procesos de gobernanza de TI a nivel de alta dirección. Y son riesgos que la organización está asumiendo en este proyecto como consecuencia de esas fallas.

Otra debilidad sobre la gobernanza de datos que queda en evidencia en este punto tiene que ver con la gestión de la metadata. Todas las decisiones respecto a las fuentes de origen de cada dato o las transformaciones que se aplican a ellos previo a ser disponibilizados al usuario final, son información que es crucial para la interpretación de los datos y para el traspaso de conocimiento dentro de la organización. En una organización con una adecuada gobernanza de datos, toda esta información se almacena de forma estandarizada y permanece accesible para cualquier usuario que la precise en el futuro. En el caso de este proyecto, es información que queda únicamente en poder de quienes participaron del proyecto y hay un alto riesgo que se pierda en el futuro.

9) Iteraciones, validación y ajustes sobre el dataset

Una vez implementados procesos de extracción y transformación de la información, fue necesaria la realización de validaciones sobre la información resultante de estos procesos. Como se comentara en el apartado de presentación de la problemática, el objetivo final es que el nuevo sistema se vuelva la única fuente de información del negocio de tarjetas de crédito para todos sus *stakeholders*. Para que ello suceda, es fundamental que la información incluida en el sistema sea debidamente validada y por lo tanto exista un alto grado de seguridad y exactitud sobre estos datos.

Para cumplir con este requisito se realizaron varias iteraciones en las cuales:

- Personal de inteligencia comercial corría los procesos para algunos meses históricos, y enviaba los resultados al equipo de medios de pago (usuarios finales).
- El equipo de medios de pago contrastaba estos resultados con la información que actualmente utilizan para la gestión o directamente contra los sistemas sobre los cuales se definió que se extraería la información según el dato.
- El equipo de medios de pago enviaba compilado de las diferencias u observaciones que pudieran surgir de esos comparativos.
- El equipo de inteligencia comercial evaluaba comentarios recibidos y en caso que correspondiera, realizaba ajustes sobre los procesos de extracción y/o transformación de datos.
- El equipo de inteligencia comercial vuelve a correr los procesos con ajustes, y enviar los resultados al equipo de medios de pago para su análisis nuevamente.

Esta secuencia se repitió hasta que el equipo de medios de pago no detectó ningún comentario o diferencia entre la información recibida desde inteligencia comercial y las otras fuentes. Para finalizar esta etapa, se corrieron los procesos para los últimos 24 meses, de manera tal de cumplir con el requerimiento de contar con información histórica que es importante para los usuarios que analizan tendencias sobre algunos de los indicadores.

En este proceso quedan en evidencia nuevas debilidades en la gobernanza de datos de la empresa, en este caso respecto a los estándares y exigencias de calidad de la información que se disponibiliza para la gestión. Si bien el proceso realizado y descrito parece ser bueno y coherente, no es un proceso que se encuentre estandarizado en la organización y haya sido validado por la gerencia del CDO. Fueron los integrantes del proyecto quienes definieron el estándar de calidad que exigirían para el sistema y también el proceso que llevarían adelante para validarlo. Si existiera un gobierno centralizado de los datos, estos estándares y procedimientos serían definidos por esta figura y ofrecerían mayores garantías de éxito. Adicionalmente, se podrían capitalizar los aprendizajes de la aplicación de estos procedimientos en cada proyecto replicándolos en otros. Nuevamente, como el proyecto está gestionado en este caso por personal que no tiene una visión global de la organización, los aprendizajes difícilmente se capitalicen y repliquen en nuevos y futuros proyectos.

2.4.2.4 Etapa 4: Análisis de datos dentro del proyecto

10) Implementación de modelo para visualización de datos

Una vez que se logró contar con un dataset con una estructura definida y estable, se pudo comenzar a trabajar en la implementación de tableros y cuadros de mando que permitieran a los usuarios finales visualizar y acceder a la información de forma sencilla, y sin necesidad de continuar aplicando transformaciones o procesamientos posteriores. Estos tableros son los que se implementarían en la herramienta Qlickview. Quien se encargó nuevamente de la implementación de esta implementación, fue la persona de inteligencia comercial involucrada en el proyecto con el apoyo de un proveedor externo.

Previo a comenzar a trabajar con el proveedor externo, la persona responsable de esta tarea se encargó de coordinar algunas reuniones con los usuarios finales del equipo de medios de pago para acordar los tableros que se estarían disponibilizando en esta primera versión. También algunas funcionalidades que los usuarios consideraban importantes y fueran viables de implementar en la

herramienta. De estas reuniones de trabajo surgió que en la primera versión del sistema se incluirían:

- Tabla con datos mensuales de cantidad de tarjetas: altas, bajas y stock.
- Tabla con datos mensuales de transacciones confirmadas y rechazadas: totales y abiertas por cada tipo.
- Tabla con datos mensuales de volumen de compras confirmadas y rechazadas: totales y abiertas por cada tipo.
- Tabla con estado de resultados mensual: vista de todos los ingresos y egresos.
- Filtros que permitieran obtener todas estas vistas por: sello, tipo de tarjeta, programa de fidelidad, clientes específicos u otros filtros.
- Posibilidad de descargar las tablas.
- Posibilidad de descargar el detalle de clientes en cuestión para un mes en particular, únicamente con los indicadores deseados y seleccionados por el usuario.

De todo esto, se lograron implementar todos los puntos con el proveedor externo con excepción de la posibilidad de descargar el detalle de clientes en cuestión para un mes en particular. Esto significa que, los usuarios finales podrían visualizar datos agregados en las tablas solicitadas, aplicar filtros sobre esas tablas y descargarlas. Pero no saber el grupo de clientes que componen ese dato agregado. (ejemplo: el usuario sabrá que hubo 10.000 transacciones rechazadas en marzo, pero no podrá definir de cuales tarjetas) Esto estaba comprendido dentro del alcance original, y es uno de los grandes valores agregados del sistema para aquellos usuarios que además de analizar datos macro y tendencias, deben realizar análisis sobre bases de clientes específicas y/o ejecutar acciones concretas sobre colectivos de clientes. Con lo cual, es una funcionalidad que condiciona negativamente de gran forma los resultados obtenidos con esta implementación. Consultado con personal involucrado en el proyecto respecto al porqué de esta situación, explicaron que se debió a una limitante de la infraestructura. A modo de resumen, el servidor donde se encuentra instalado Qlickview no tiene la capacidad de procesamiento requerida para armar el detalle de clientes con la información de los indicadores deseados por el usuario y descargarlo en un archivo de Excel. Y dado que como ya comentamos, el proyecto no fue avalado por el departamento de TI, no fue viable plantear una mejora en este servidor que permitiera aprovechar mejor el trabajo realizado y agregar esta funcionalidad. Consecuentemente, en esta primer etapa los usuarios que quisieran acceder a este tipo de detalle por cliente, pueden hacerlo, pero solo haciendo el requerimiento al equipo de

inteligencia comercial. Que son quienes tienen acceso a la base y pueden descargarlo directamente desde allí y no a través de Qlickview. Esto si bien es una alternativa, genera un cuello de botella en el equipo de inteligencia comercial, dependencias e ineficiencias innecesarias y evitables.

Esta situación última, es otra consecuencia de varias de las debilidades en la gobernanza de TI y datos ya comentadas. Más específicamente, en este caso queda en evidencia la falta de un *data engineer* que podría haber detectado este problema al momento de la selección de la tecnología y buscar una alternativa. Y ni que hablar que, de haberse realizado el proyecto de manera coordinada con la dirección de TI, lo razonable hubiera sido gestionar este inconveniente del proyecto y evaluar alternativas de mejoras en la infraestructura que permitieran habilitar esta funcionalidad para los usuarios finales. Nuevamente aquí, una muestra de cómo las deficiencias en la gobernanza de TI y datos, tienen un impacto negativo en los usuarios finales de la información y por transitiva, en el negocio propiamente dicho.

11) Presentación de la nueva herramienta a *stakeholders*

Una vez finalizada la implementación el Qlickview, el nuevo sistema estuvo pronto para presentar a los usuarios. Como se comentó en el apartado de selección de las tecnologías, una de las principales ventajas de la herramienta de Qlickview es que es muy intuitiva y fácil de usar, y permite que cualquier usuario con conocimientos muy básicos pueda utilizar los tableros armados allí. Con lo cual, esto hizo prescindible dedicar mucha cantidad de horas en capacitación a los usuarios finales.

Previo a la organización de la instancia de presentación, se gestionaron licencias y accesos para todos los usuarios finales del equipo de medios de pago a la herramienta. Para que pudieran comenzarla a utilizar de forma inmediata posterior a su presentación formal.

La presentación formal de la herramienta consistió en una reunión a la que asistieron los equipos de medios de pago, inteligencia comercial, control de gestión, intervención general y tecnología. La reunión fue liderada por los referentes de medios de pago e inteligencia comercial para el proyecto. Quienes mostraron brevemente el proceso realizado y varios casos de uso sobre Qlickview. Además, se avisó a los usuarios de medios de pago que ya podrían comenzar a utilizarlo y qué hacer en caso de detectar cualquier inconsistencia en la información. Según comenta el personal consultado respecto a la instancia, el nuevo sistema tuvo muy buena recepción por parte del equipo de medios de pago, pero muchos cuestionamientos de varios tipos desde las áreas que no participaron o participaron muy poco de todo el proceso (tecnología, control de gestión, intervención general).

Incluso comentaron que este hito tuvo algunas repercusiones negativas como, por ejemplo, que otros sectores comenzaran a demandar herramientas similares, a lo cual como era de esperarse, el departamento de TI se opone por los motivos detallados anteriormente. Otros también comenzaron a solicitar que se les diera acceso a la herramienta, lo cual no era viable, por temas de licencias que tienen un costo asociado y quien gestiona ese presupuesto es el área de TI.

Es decir que, en algún punto la implementación de este sistema, en vez de percibirse como un avance hacia una nueva forma de trabajar y gestionar los datos, fue percibido como un desvío respecto a la ruta que quiere seguir la empresa en este sentido. Esto era esperable luego de lo sucedido al comentarle sobre el proyecto al equipo de TI. Adicionalmente vemos como aquí, el aplicar una buena práctica que es difundir y compartir un caso de éxito vinculado a la gestión de datos con otras áreas, puede tener un efecto contrario si no se trata de algo avalado por las figuras que se supone que debe centralizar y liderar este tipo de proyectos. E incluso indirectamente está alentando a otras áreas a intentar replicar actitudes similares, desautorizando totalmente al CIO y CDO y complicando la materialización de la visión y estrategia de datos que estén intentando implementar. Nuevamente, una muestra de las consecuencias de la falta de coordinación a nivel de dirección en la coordinación de proyectos de TI y datos.

12) Etapa de *testing* y ajustes con *stakeholders*

Actualmente el proyecto se encuentra en los primeros meses desde su implementación. En esta fase, cuando los usuarios finales detectan cualquier inconveniente con la información o el sistema, lo reportan al equipo de inteligencia comercial para su análisis y ajuste. En paralelo, no se han discontinuado ninguno de los reportes que se utilizaban hasta el momento para que sigan funcionando como una contingencia.

Sobre este último punto, podemos ver un potencial problema que se ha generado producto de este proyecto. El equipo de inteligencia comercial está brindando soporte sobre una solución de datos y TI al equipo de medios de pago. Esto lo pueden asumir durante los primeros meses luego de finalizado el proyecto, pero probablemente no en el mediano y largo plazo. Naturalmente, en una adecuada gobernanza de datos y de TI a la que apunta la organización esto debería ser gestionado desde el área del CIO y CDO. Pero ahora, difícilmente quieran asumir la administración y el mantenimiento de un sistema sobre el que no estuvieron de acuerdo en su implementación. Eso es un claro posible conflicto que surgirá entre ambas áreas, producto de cómo surgió el proyecto, que no existiría si el proyecto hubiera sido producto de adecuados procesos de gobernanza de TI.

3 APLICACIÓN

3.1 Consideraciones preliminares

El objetivo de esta parte del documento, es dejar en evidencia el potencial que podría tener dentro de una organización como la del caso de estudio, el logro de una adecuada gobernanza de *big data* que le permita y facilite la realización de proyectos de *big data*. Concretamente lo que se hará en esta sección, es buscar un ejemplo de aplicación de una herramienta que se podría haber construido apalancándose en el trabajo realizado para la implementación del CRM de tarjetas y que sin duda traería beneficios positivos para el negocio.

El ejemplo que se llevará a cabo consiste en la implementación de un algoritmo de clusterización que permita, a partir de la información agrupada por tarjeta (número de contrato), generar clusters de clientes y compararlos con la segmentación que el banco utiliza actualmente. Esto podrá tener potencialmente dos conclusiones. En primer lugar, podría ser que el algoritmo nos devuelva clusters que sean muy similares a los segmentos de clientes que hoy utiliza el banco para trabajar. Lo cual sería un indicio de que esta aplicación no habría agregado mucho valor a la empresa y que el banco ya está usando modelos muy buenos para segmentar a sus clientes. O, en segundo lugar, podríamos llegar a clusters de clientes muy distintos a los que hoy utiliza el banco lo cual nos llevaría a pensar que el banco está utilizando criterios muy distintos para segmentar a sus clientes o bien no está lo suficientemente avanzado o desarrollado en este sentido.

Una mala segmentación de clientes, puede llevar a análisis incorrectos del negocio que se manifiestan en: desarrollo de productos que no son lo que buscan los clientes, publicidad ineficiente, canales de comunicación ineficientes, canales de atención ineficientes, ofertas que no se adecuan a las necesidades del cliente, entre otros inconvenientes. Lo que es lo mismo, un mal análisis lleva a clientes insatisfechos (o menos satisfechos de lo que podrían estar) y consecuentemente un negocio menos sustentable y rentable. Con lo cual, este debería ser un ejemplo de problema relevante para cualquier organización y en el que merece la pena trabajar siempre que se mantenga una adecuada relación costo – beneficio.

Cabe aclarar que lo que se hará en este apartado es un repaso del proceso que debiera llevarse adelante para la implementación de un algoritmo de este tipo, mostrando resultados concretos aplicándolos sobre un dataset concreto brindado por la empresa en cuestión. Pero no se intentará

de modo alguno, llevar adelante un proceso de mejora continua del algoritmo que permita optimizar sus resultados y garantizar que es el mejor que podría obtener la empresa. Ya que esto queda fuera del alcance de este trabajo de tesis y requeriría incluir un trabajo mucho más detallado de aplicación de diferentes algoritmos, parámetros y *datasets*, y comparación de resultados. Adicionalmente, es importante dejar en claro que este apartado no cambia ninguna de las conclusiones sobre los inconvenientes mencionados para esta empresa respecto a su gobernanza de TI y datos. Y que antes de embarcarse en cualquier proyecto de este tipo, la recomendación sigue siendo que logre y realice esas mejoras en su gobernanza, para que los proyectos de *big data* se puedan llevar delante de forma eficiente y beneficiosa para el negocio.

El ejemplo en cuestión fue implementado en RStudio y se describe detalladamente en el archivo formato .rmd llamado “Tesis Juan Saldías” entregado junto a este documento de tesis. Este archivo adjunto, se basa en el archivo formato .rmd llamado Multivariado I compartido por la profesora Joselyna Davyt en el curso de Taller de aprendizaje automático de esta maestría.

3.2 Análisis del *dataset*

El *dataset* utilizado para la aplicación de este algoritmo, es un extracto de un *dataset* total de información por contrato de tarjeta de crédito de aproximadamente cuatrocientos mil registros y 100 variables. Es decir, que es un extracto del *dataset* que luego de ser procesado por Alteryx, se disponibiliza para alimentar los tableros implementados en Qlickview, según se comentó en el desarrollo del proyecto del apartado anterior.

Para este ejemplo puntual se utilizó un *dataset* de solamente 4.081 registros y 24 variables, por dos motivos principales:

1. La empresa no estuvo de acuerdo en brindar información completa para realizar el trabajo, para mantener la confidencialidad y seguridad de varios aspectos de su negocio.
2. La capacidad de procesamiento de la computadora utilizada para el trabajo, no permitía trabajar en la implementación de este tipo de algoritmos con *datasets* de gran tamaño. Y el trabajo tampoco ameritaba una gestión de recursos de este tipo ya que como se comentó anteriormente, está fuera del alcance del trabajo encontrar el algoritmo óptimo para resolver esta necesidad.

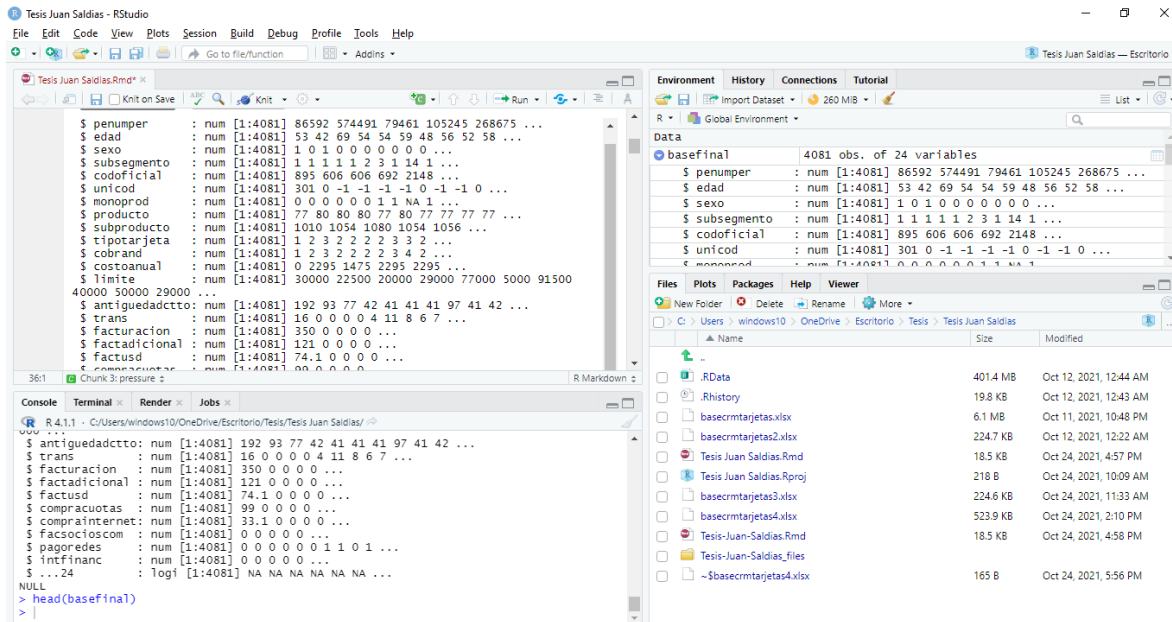
De todas formas, es importante aclarar que, en caso de querer llevar adelante un proyecto de implementación de este tipo de algoritmos, siempre se lograrán mejores resultados cuanto mayor sea la cantidad de datos reales disponibles que se le pueda brindar al algoritmo. Es importante antes de embarcarse en proyectos de este tipo siempre contar con ambos: un *dataset* con una cantidad de registros suficientes y confiables, y una adecuada capacidad de procesamiento para el problema que se quiere resolver. Ambos deberían ser el paso inicial de todo proyecto, y son parte de las tareas que se deben adjudicar a quienes lideren este tipo de proyectos.

Aclarado esto, el primer paso en cualquier proyecto de implementación de algoritmos de este tipo siempre consistirá en el análisis exploratorio del *dataset*. Este trabajo consiste básicamente en un trabajo que puede ser realizado tanto por el *data analyst* como por el *data scientist*, y tiene dos objetivos principales. El primero es el entendimiento de la información con la que se va a trabajar: tipos de datos, cantidad de datos, distribución de las observaciones, medias, medianas, etc. Ya que este entendimiento es clave para decisiones que tienen que ver con qué parte de las variables y observaciones se va a trabajar (pueden ser todas o solo una parte de ellas), y que algoritmos se va a intentar aplicar. No todos los algoritmos son aplicables ni performan igual con todo tipo de variables y cantidad de observaciones, con lo cual este puede ser un primer filtro. El segundo es depurar el *dataset* de todo aquello que no sirva, como pueden ser variables duplicadas, variables que no estén relacionadas al caso de estudio, registros en blanco, registros con valores en N/A o defectos similares. Cuanto más depurado esté el *dataset* sobre el que se trabaje de todas estas posibles variables u observaciones con imperfecciones, más acertados serán los resultados obtenidos.

En este sentido, en la primera parte del archivo `.rmd` entregado junto con este documento, se puede ver el análisis exploratorio del *dataset* que se decidió hacer en este caso. A modo de resumen, se realizan algunos comentarios a continuación.

3.2.1 Visualización de la estructura

De la visualización rápida de la estructura (ver pantalla adjunta debajo), surge que el *dataset* se compone de 24 variables y 4.081 observaciones. También se puede ver rápidamente que las 24 variables que lo componen son numéricas y tienen variedad de valores.



De esta visualización de la estructura y diálogo con personal de la empresa que brindó el *dataset*, surge que, si bien todas las variables tienen un formato numérico, no todas lo son conceptualmente. En varios casos estos números representan un universo finito de categorías que se asignan a cada tarjeta para cada variable. Algunos ejemplos de esto pueden ser:

- Sexo: indican si el cliente es hombre o mujer.
- Subsegmento: clasifica a los clientes entre rentas masivas, *select*, rentas medias y *select premium*.
- Codoficial: representa la sucursal del banco que tiene la tutela del cliente.
- Tipotarjeta: indica si la tarjeta es regional, internacional, oro, *platinum* o *select*.
- Cobrand: indica el programa de fidelidad de la tarjeta.
- Producto: indica si la tarjeta es Visa o Mastercard.

El tratamiento de estas variables debería ser bien distinto al de las variables numéricas continuas como ser:

- Limite: indica el límite de crédito asignado a cada tarjeta.
- Antiguedadcto: indica la antigüedad en años del contrato de tarjeta de crédito en cuestión.
- Facturación: indica el gasto realizado con la tarjeta por el cliente en el período de análisis.
- Trans: indica la cantidad de transacciones realizadas con la tarjeta en el período de análisis.

Esto deja en evidencia como este análisis preliminar de la estructura y entendimiento de cada una de las variables que se realiza en un primer momento, es vital y condiciona el abordaje del trabajo que se va a llevar adelante. El entendimiento de las variables a utilizar es clave tanto para priorizarlas (no todas son igualmente importantes desde lo conceptual para el problema que se quiere explicar) como para definir como haremos su análisis. Siempre en este primer encuentro, es fundamental la participación del *data analyst*, que es quien debería conocer con exactitud lo que representa cada una de las variables. En el caso de ejemplo de aplicación simplificado, quien cubrió ese rol fue la persona de la empresa quien brindó la información.

3.2.2 Depuración de ceros, nas, infinitos

The screenshot shows the RStudio interface with a data frame summary table. The table has the following data:

variable	q.zeros	p.zeros	q.na	p.na	q.inf	p.inf
22 pagoredes	1432	35.09	0	0.00	0	0
23 intfinanc	3182	77.97	0	0.00	0	0
7 monoprod	1494	36.61	52	1.27	0	0
24 ...24	0	0.00	4081	100.00	0	0

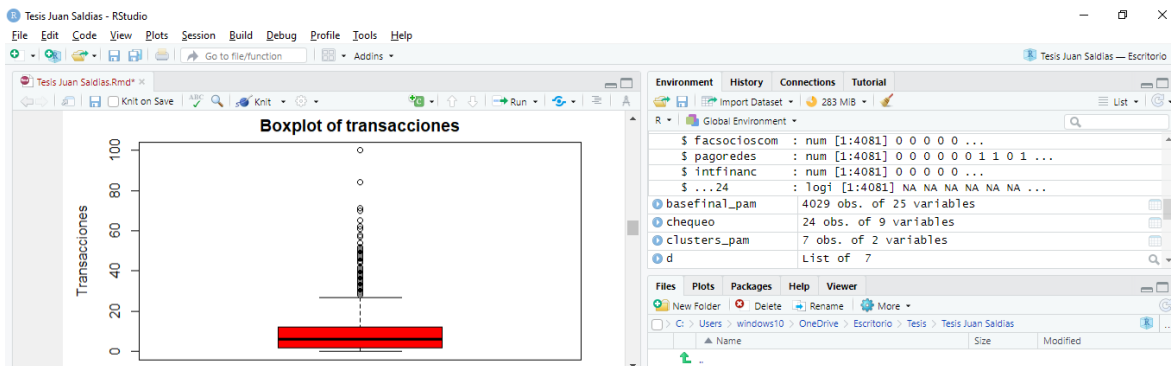
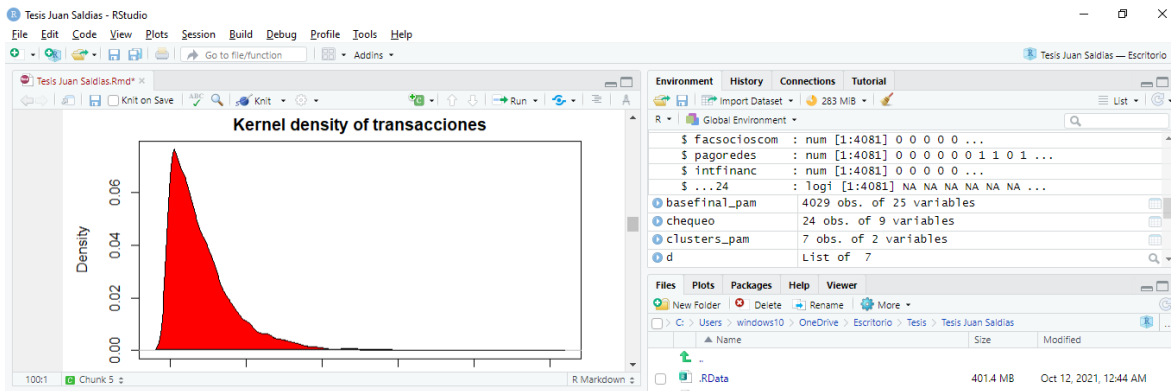
Se realizó el análisis que se evidencia en la imagen adjunta arriba para detectar si entre las observaciones del *dataset* existían valores con ceros, n/as y/o infinitos. De este análisis surgió que para la variable *monoprod* (que indica si el titular de la tarjeta tiene o no otros productos contratados con el banco), tenía algunas observaciones con n/a que se procedieron a quitar.

Por otro lado, se detectó también que varias variables tenían una cantidad significativa de ceros, pero dado que en todos los casos se trata de variables binarias y con una alta relevancia para el caso de estudio se decidió mantenerlas en su totalidad.

3.2.3 Análisis de las variables numéricas

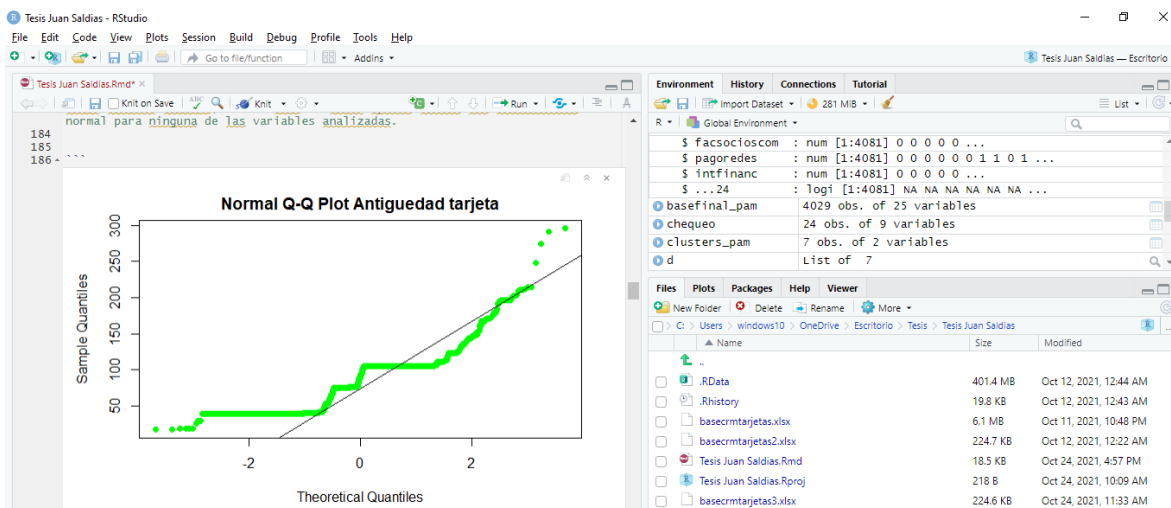
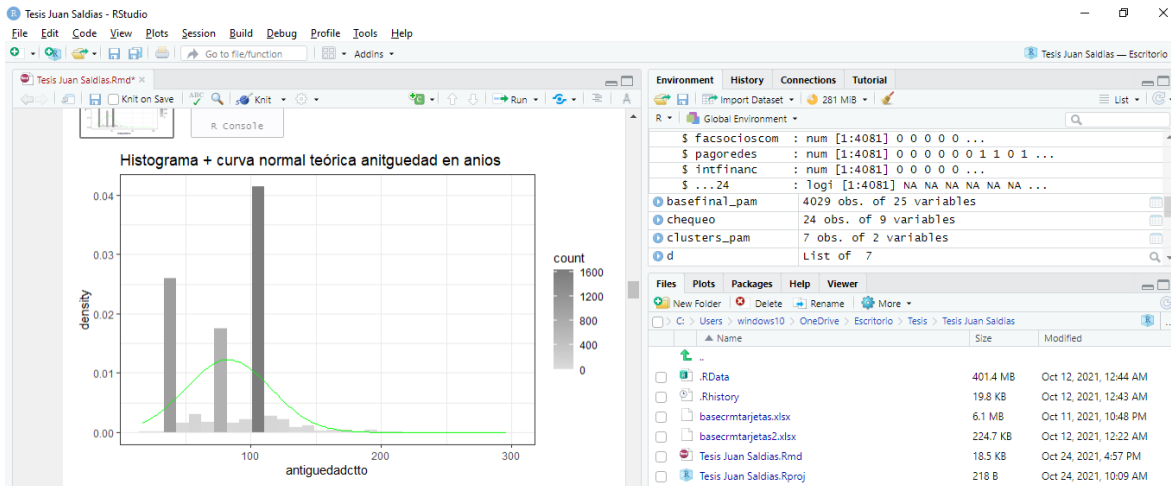
Como se mencionara durante la visualización de la estructura del *dataset*, una parte de las variables incluidas podemos definir las como numéricas. Para esta parte del análisis se realizaron varios análisis gráficos de las variables seleccionadas como más representativas para el caso de estudio que fueron: transacciones, facturación, antigüedadctto y limite. Idealmente en un caso real y completo, se debería repetir este análisis con todas las variables numéricas del *dataset*.

En primer lugar, se realizó un análisis con gráficos de distribución. Estos gráficos se utilizan para visualizar la distribución de las observaciones de cada una de las variables. Para aplicaciones como la de un algoritmo de clusterización, se deberían descartar aquellas que con este análisis se detectara que tiene las observaciones tienen una concentración muy elevada sobre algún valor o intervalo específico. También se podría aprovechar para detectar y eliminar observaciones con valores que sean *outliers* y pudieran desvirtuar las conclusiones. Debajo se ilustran ejemplos para una de las variables.



No se descartaron variables ni observaciones específicas con este análisis por entenderse luego del análisis de todos los gráficos, que no correspondía.

Posteriormente, se realizó una evaluación de la normalidad de cada uno de estas variables. Este análisis se utiliza para comparar la distribución de las observaciones de cada variable con una distribución normal teórica. Esto se hace fundamentalmente, para entender si alguna de las propiedades conocidas para una distribución normal (simetría alrededor de la media que coincide con la mediana, entre otras), podría aplicarse en el análisis y utilización de las variables en cuestión. Debajo se ilustran ejemplos para una de las variables.



El último gráfico muestra los cuantiles teóricos de una distribución normal con la misma media y desviación estándar que las observaciones. Esta herramienta es muy útil para evaluar la normalidad de una variable. Cuanto más se aproximen los datos a una distribución normal, más alineados estarán entorno a la recta teórica que se proyecta.

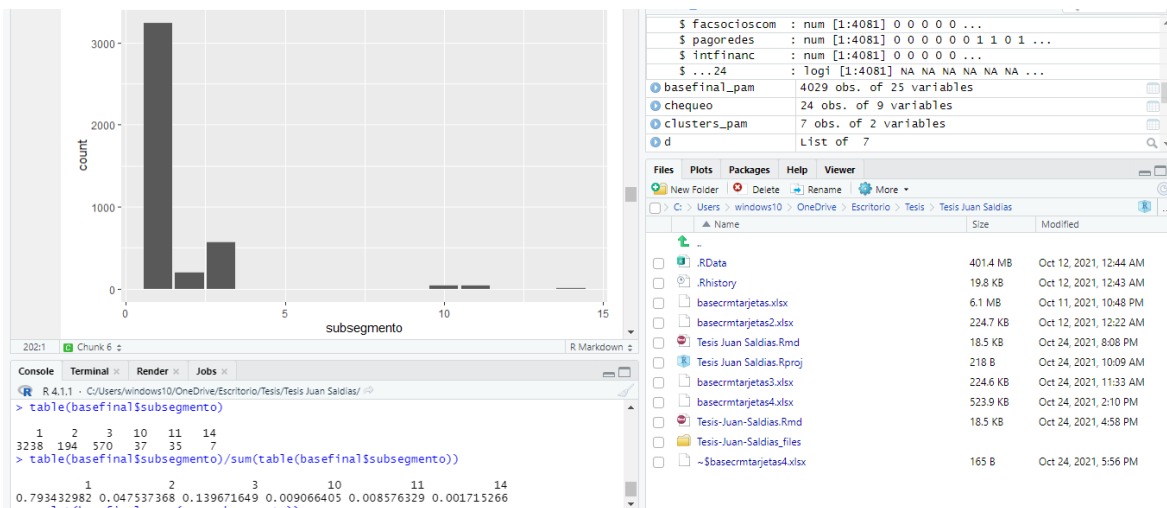
No surgieron conclusiones relevantes de este análisis de normalidad realizado para las variables numéricas del *dataset*.

3.2.4 Análisis de las variables categóricas

Respecto a las variables categóricas, se hicieron algunos análisis gráficos y con tablas, con el objetivo de entender la proporción de observaciones en cada una de las categorías de las variables

seleccionadas. La razón para realizar este tipo de análisis, es que, en caso de detectar una elevada concentración de observaciones para una categoría de una variable en particular, podríamos asumir que esta variable no agrega valor para la división de las observaciones entre diferentes clusters y quitarlas.

Al igual que para las variables numéricas, para este ejercicio simplificado solo se eligieron las variables relevadas como más importantes para el caso de estudio. Que en este caso fueron subsegmento, tipotarjeta y cobrand. Debajo se adjuntan algunas imágenes de cómo se hizo este trabajo en RStudio.



De este análisis, surgió que hay una elevada concentración de las observaciones en las clasificaciones de rentas masivas, tarjetas internacionales, tarjetas regionales y los programas de fidelidad que corresponden a públicos más masivos (acuerdo con cadena de farmacias, cadena de supermercados y programa propio del banco). Consultado con persona que brindó el *dataset*, esta confirmó que ello es coherente con la realidad del banco con lo cual no se quitaron ninguna de estas variables.

3.3 Selección del algoritmo

En lo que refiere a algoritmos de clusterización, es importante diferenciar en primer lugar para su elección, si se trata de un *dataset* que tiene variables numéricas, categóricas o ambas. Lo más habitual es que muchos de los algoritmos no tengan la capacidad de resolver el problema para un tipo de dato u otro, sino que se especifican en resolver el problema de clusterización para un tipo de datos en particular. Cuando se trata de *datasets* de tipo mixto (que tienen ambos tipos de

variables), lo más común es optar por algoritmos que utilicen una métrica de distancia compatible con ambos tipos de datos, y luego usar alguna forma de agrupamiento que dependa solo de las distancias entre los puntos de datos.

En lo que respecta a estas métricas de distancia, una de las opciones más comunes para resolver problemas de datos mixtos, es la de usar la llamada distancia de Gower. *“La distancia de Gower o disimilitud de Gower es una métrica que mide la diferencia de dos elementos con datos mixtos numéricos y no numéricos. Un posible uso de la distancia de Gower es la agrupación de k-medias con datos mixtos porque k-medias necesita la distancia numérica entre elementos de datos. A modo de resumen, para calcular la distancia de Gower entre dos elementos, se debe comparar cada elemento y calcular un término. Si el elemento es numérico, el término es el valor absoluto de la diferencia dividido por el rango. Si el elemento no es numérico, el término es 1 si los elementos son diferentes o el término es 0 si los elementos son iguales. La distancia de Gower es el promedio de los términos.”*¹²

Una vez definida la métrica de distancia, se debe definir el algoritmo con el que se utilizará. A continuación, hago una breve reseña de los algoritmos de clusterización más comunes y vistos durante el curso.

Algoritmo k-means:

“K-means es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster. Se suele usar la distancia cuadrática.

El algoritmo consta de tres pasos:

- 1. Inicialización: una vez escogido el número de grupos, k, se establecen k centroides en el espacio de los datos, por ejemplo, escogiéndolos aleatoriamente.*
- 2. Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroide más cercano.*
- 3. Actualización centroides: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.*

¹² Web James McCaffrey. Accedido el 3 de septiembre, 2021, desde <https://jamesmccaffrey.wordpress.com/2020/04/21/example-of-calculating-the-gower-distance/>

Se repiten los pasos 2 y 3 hasta que los centroides no se mueven, o se mueven por debajo de una distancia umbral en cada paso.”¹³

Este algoritmo se puede utilizar únicamente si todas las variables son numéricas.

Algoritmo k-medoids:

Es una alternativa robusta a la agrupación de k-means, menos sensible a los valores atípicos. Un medoide se podría definir como el objeto perteneciente a un cluster o conglomerado, cuyo promedio de disimilaridad a todos los objetos en el conglomerado es mínima. Es decir, que se puede considerar como el punto más céntrico de la agrupación considerada. K-medoids es un método de *clustering* muy similar a k-means en cuanto a que ambos agrupan las observaciones en k *clusters*, donde k es un valor pre establecido por el analista. También es exclusivo para variables numéricas.

Algoritmo de *clustering* jerárquico:

“El agrupamiento jerárquico es un método de análisis de grupos puntuales, el cual busca construir una jerarquía de grupos. Estrategias para agrupamiento jerárquico generalmente caen en dos tipos:

- *Aglomerativas: Este es un acercamiento ascendente: cada observación comienza en su propio grupo, y los pares de grupos son mezclados mientras uno sube en la jerarquía.*
- *Divisivas: Este es un acercamiento descendente: todas las observaciones comienzan en un grupo, y se realizan divisiones mientras uno baja en la jerarquía.*

En general, las mezclas y divisiones son determinadas con un algoritmo voraz. Los resultados del agrupamiento jerárquico son usualmente presentados en un dendrograma.

En orden de decidir qué grupos deberían ser combinados (para aglomerativo), o cuando un grupo debería ser dividido (para divisivo), una medida de disimilitud entre conjuntos de observaciones es requerida. En la mayoría de los métodos de agrupamiento jerárquico, esto es logrado mediante uso de una métrica apropiada (una medida de distancia entre pares de observaciones), y un criterio de

¹³ Web Unioviedo. Accedido el 3 de septiembre, 2021, desde https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html

enlace el cual especifica la disimilitud de conjuntos como una función de las distancias dos a dos entre observaciones en los conjuntos.”¹⁴

En el caso del problema de nuestro caso de estudio en cuestión, correspondería la utilización de un algoritmo jerárquico divisivo.

Algoritmo Modha-Spangler:

El agrupamiento Modha-Spangler calcula la ponderación óptima para las variables continuas y categóricas mediante una estrategia de búsqueda de fuerza bruta.

Como se puede ver en este resumen de las opciones disponibles para resolver este tipo de problemas, la elección del algoritmo no es trivial. Solo leyendo una breve descripción de lo que hace cada algoritmo, queda en evidencia que su elección no es una tarea que pueda ser abordada por ninguno de los perfiles que, por ejemplo, participaron en el proyecto de implementación del CRM de tarjetas. Estos temas vinculados a la implementación de algoritmos requieren de mucha experiencia en el tema y conocimientos técnicos vinculados a estadística, programación y manejo de las herramientas sobre las que se hará su implementación. Por todo lo comentado sobre la empresa del caso de estudio, es evidente que actualmente no cuentan con este tipo de conocimiento, y mucho menos lo están generando exponiendo a personas con potencial a este tipo de desafíos. Si se quiere lograr implementar este tipo de proyectos que pueden tener un valor agregado significativo para el negocio, la empresa tendrá que definir dentro de su estrategia de gobernanza de *big data* como querrá manejar estos temas más técnicos. Podrá ser contratando personal que cuente con el conocimiento y lo expanda, ayudando a formar al personal actual o quizá tercerizándolo. Estas son tareas de las que se debería ocupar un *data scientist*.

Para el caso del ejemplo simplificado de este trabajo, no profundizaré en la comparación y evaluación de los diferentes tipos de algoritmos. Utilizaré para el caso en cuestión la distancia de Gower y el algoritmo k-medoids. Cabe aclarar que, en un caso real, correspondería evaluar ventajas y desventajas de más de una métrica de distancia y algoritmo, descartar algunos de forma preliminar y aplicar más de uno de los no descartados sobre el *dataset* para comparar resultados obtenidos. En este caso se implementará una única opción por estar el resto fuera del alcance del objetivo de este documento.

¹⁴ Web Wikipedia. Accedido el 3 de septiembre, 2021, desde https://es.wikipedia.org/wiki/Agrupamiento_jer%C3%A1rquico

3.4 Aplicación y resultados

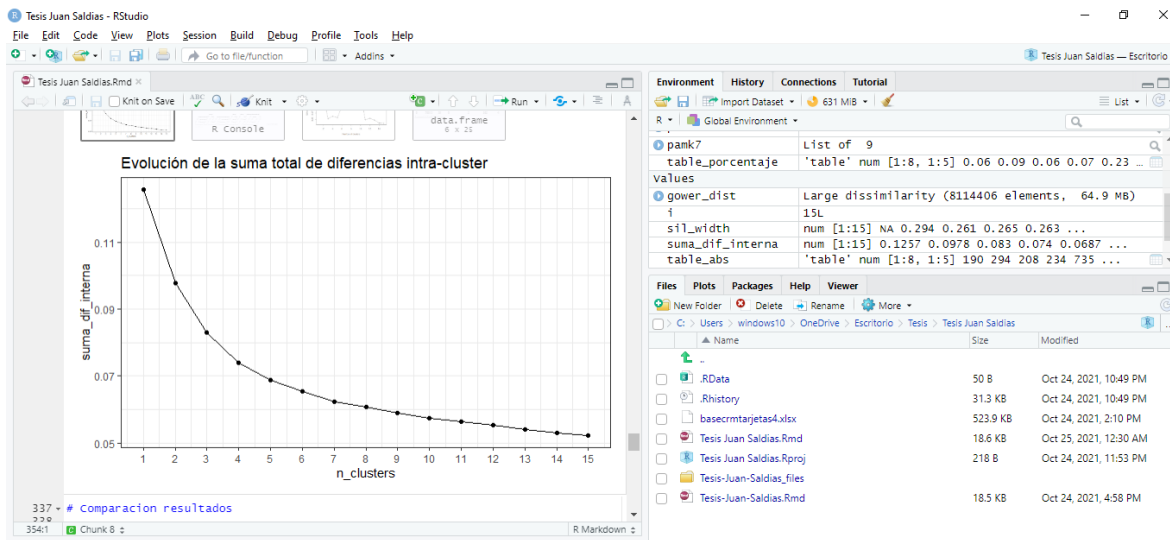
Para la aplicación del algoritmo k – medoids es necesario utilizar una técnica que permita definir la cantidad de k óptima, o lo que es lo mismo, la cantidad de clusters que permita tener un resultado óptimo en términos de la similitud y homogeneidad de las observaciones que se contemplan dentro de cada uno de los clusters. En el caso de la aplicación en cuestión, determiné el k óptimo a partir del análisis gráfico del método *elbow* y el análisis de siluetas de cada cluster. Previamente a estos análisis, es necesario el cálculo de la distancia de Gower para el *dataset*.

A los efectos de trabajar en el algoritmo de clasificación, quité dos de las variables del dataset original:

- La variable que corresponde a un identificador único por cliente, ya que para este análisis no aporta valor.
- La variable que corresponde al subsegmento, que el valor con el que clasifica a los clientes el banco. El objetivo de esto es que la clasificación del banco no sesgue el ejercicio de clusterización, y que después pueda comparar ambos resultados: los grupos de clientes definidos por el banco y por el algoritmo implementado.

“En el método elbow se tiene que calcular la distorsión promedio de los clusters, que es la distancia promedio del centroide a todos los puntos del cluster. Así, cuando se va de una situación en la que el número de clusters es inferior al correcto a una situación en la que el número es el adecuado, el valor de la dispersión disminuye bruscamente, mientras que, si aumenta el número de clusters al adecuado, el valor de la dispersión se reducirá más lentamente, formando un codo en la gráfica.”¹⁵

¹⁵ Web Slashmobility. Accedido el 3 de septiembre, 2021, desde <https://slashmobility.com/blog/2019/07/clustering-como-obtener-agrupaciones-inherentes-en-los-datos/>

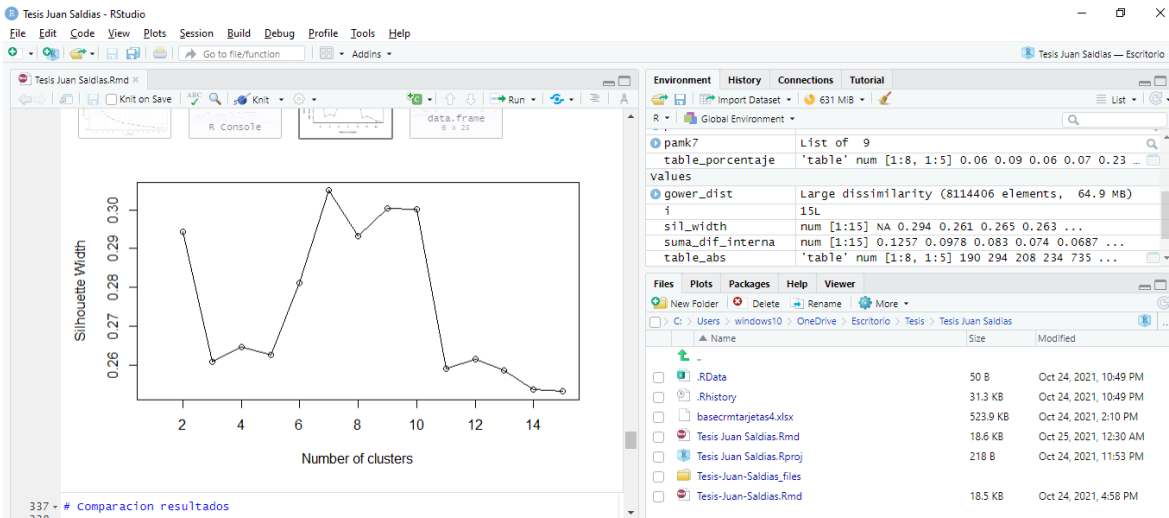


Aplicando este método para el caso de estudio en cuestión, podemos ver que el k óptimo buscado está entre 5 y 7. Ya que como se ve en el gráfico hasta k igual a 7 las divisiones del *dataset* se justifican por una ganancia de información significativa. A partir de allí, la ganancia de información con cada nueva división es poco significativa y cada vez menor. Con lo cual parecería que no justifica dividir el *dataset* en más allá de 7 grupos.

“El método de la silueta se denomina así porque utiliza el coeficiente de la Silhouette. Éste se define como la diferencia entre la distancia media a los elementos del clúster más cercano y a distancia intra-clúster media de los elementos de un clúster dividido por el máximo de los dos. En el momento que se alcance el número de clústeres óptimos para un conjunto de datos, la Silhouette, en esta situación, se maximiza.”¹⁶

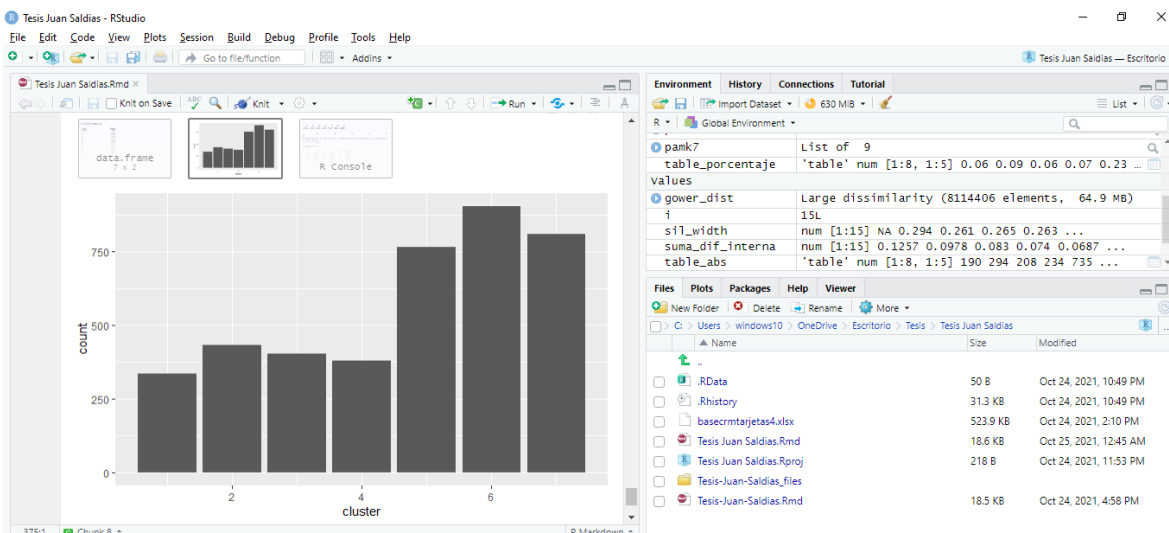
La silueta mide la similitud de un objeto con los otros objetos de su propio cluster, frente a los del cluster vecino. Sus valores pueden ir de 1 a -1. Un valor de silueta cercano a 1, indica que el objeto está bien agrupado. O, en otras palabras, que el objeto es similar a los otros objetos de su grupo. Un valor de silueta cercano a -1 indica que el objeto está mal agrupado y que la asignación a algún otro cluster probablemente mejoraría los resultados generales. Por lo tanto, para este ejercicio debemos tratar de maximizar el ancho de la silueta.

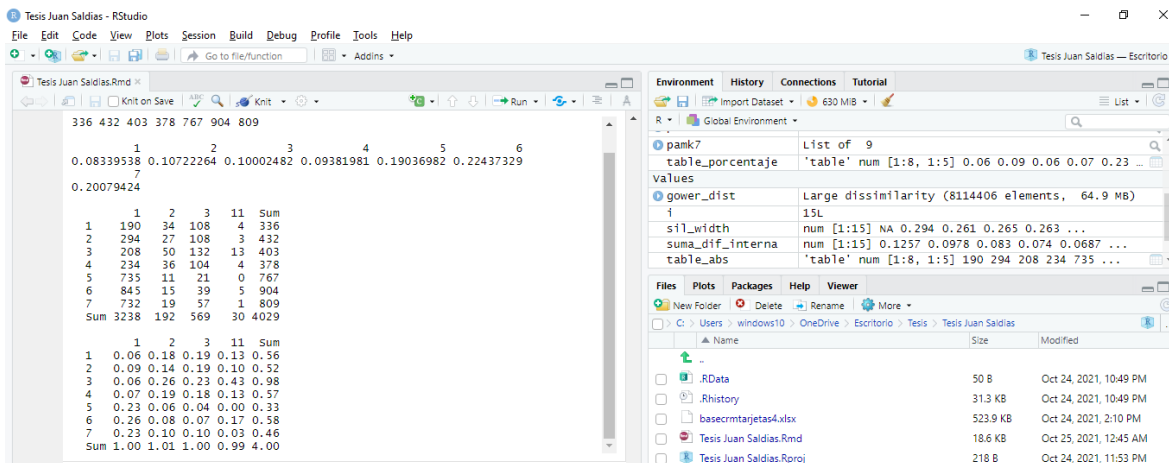
¹⁶ Web Slashmobility. Accedido el 3 de septiembre, 2021, desde <https://slashmobility.com/blog/2019/07/clustering-como-obtener-agrupaciones-inherentes-en-los-datos/>



En la gráfica presentada anteriormente, se puede apreciar como el valor óptimo de k según el ancho de la silueta parecería ser 7. Seguido a continuación de las opciones de 9 y 10. Si mezclamos esto con los resultados obtenidos con el método *elbow* podemos concluir entonces que el k óptimo para nuestro algoritmo es 7, ya que es la opción que maximiza el resultado en ambos análisis. Con lo cual, implementamos el algoritmo k -medoids utilizando esa opción.

Las tablas y gráficos adjuntos a continuación resumen los resultados obtenidos con la aplicación del algoritmo.





Se logró agrupar a las observaciones en 7 clusters, cada uno con una cantidad razonable de observaciones. Las tablas adjuntas en la segunda imagen, permiten visualizar un cruzamiento entre la cantidad de observaciones en cada uno de los subsegmentos propuestos por el banco (4 columnas) con las observaciones en cada uno de los nuevos clusters definidos (7 filas).

Esta tabla permite hacer algunas reflexiones interesantes al respecto que vale la pena mencionar:

- El 72% de los clientes de rentas masivas (subsegmento 1) se reparten entre los grupos 5, 6 y 7. Con lo cual como anticipábamos, tal vez el criterio que se está utilizando actualmente para agruparlos es demasiado amplio y generalista. Y se podría lograr mayor exactitud buscando dividir este subsegmento en más de uno.
- El 63% de los clientes *select* (subsegmento 2), se dividen entre los clusters 1, 3 y 4. Con lo cual podríamos sugerir generar nuevos subsegmentos en este caso también.
- El 79% de los clientes rentas medias (subsegmento 3) se concentran en los clusters 1, 2, 3 y 4. Lo cual nos lleva a pensar que, en primer lugar, tienen bastante en común con los clientes *select* que comentamos en el punto anterior. Ya que se da una distribución similar del colectivo entre los nuevos grupos generados por el algoritmo.
- El 43% de los clientes *select premium* se concentran en el cluster 3. Lo cual nos lleva a inferir que este cluster debe ser el que concentre a los clientes de mayores consumos y transaccionalidad. No obstante, hay que considerar que la muestra de *select premium* es solo de 30 contratos, con lo cual, deberían requerirse nuevos datos antes de sacar conclusiones ya que podría haber algún sesgo en la muestra.
- Ninguno de los subsegmentos propuestos por el banco presenta una concentración muy elevada (por ejemplo, superior al 80%) de las observaciones con uno de los clusters que

arrojo el algoritmo. Es decir que, en todos los casos, el algoritmo entendió como necesario subdividir a estos subsegmentos originales. Con lo cual, es muy probable es que el modelo del banco este tomando menos o distintas variables para hacer su agrupación.

Indagando con personal del banco, nos comentan que los subsegmentos que se utilizan actualmente se definen según un criterio corporativo que tiene ya varios años y que considera para dividir a los clientes: la cantidad de productos contratados, su volumen total de operativa de crédito y depósitos con el banco y la cantidad de transacciones voluntarias realizadas por el cliente. Esto, confirma las sospechas que se detectan al comparar los resultados obtenidos por el algoritmo con la clasificación de clientes utilizada por el banco. Y refuerza también el potencial que existe si el banco comenzara a explotar más la información de formas como esta.

Si bien el análisis realizado con esta implementación no permite sacar conclusiones contundentes por la profundidad del trabajo realizado, quedan claros varios puntos que se querían ilustrar con este ejemplo que son:

- El banco actualmente cuenta con mucha información que podría explotar para maximizar su beneficio y el de sus clientes de varias formas.
- El banco hoy utiliza información como la del ejemplo del subsegmento de sus clientes, que ni siquiera cuestiona pero podría no ser la mejor o incluso incorrecta. Esto podría ser aceptable hace unos años cuando era lo mejor que se podría lograr, pero los avances en el *big data* hacen que hoy existan más y mejores formas de obtener y analizar información. Y las organizaciones que las logran utilizar, tienen una ventaja competitiva.
- Poder implementar proyectos de *big data* requiere de conocimientos de estadística, programación, herramientas de análisis de datos y mucha experiencia. Actualmente no es fácil conseguir este tipo de recursos (por su costo y disponibilidad), y tampoco formarlos. Lo cual es un desafío para todas las organizaciones. Sobre todos las que no han nacido con una cultura orientada hacia los datos.
- Cualquier intento en implementar proyectos complejos de este tipo, en una organización que no tiene una adecuada gobernanza de *big data* con todo lo que ello implica, no tendrá oportunidades de éxito. En el mejor de los casos, podrá con mucho esfuerzo lograr concretar algún proyecto aislado. Pero como queda en evidencia en este breve repaso de implementación, las oportunidades de éxito con impactos significativos en el negocio son nulas si no se lleva adelante los proyectos de este tipo en un entorno controlado que

permita capturar y mejorar continuamente la estructura, procesos y conocimiento necesario.

4 CONCLUSIONES

Luego de realizado este análisis, podemos concluir que, tanto de la comparación contra el marco teórico sobre gobernanza de tecnología y datos, como del repaso del proyecto de implementación del CRM de tarjetas de crédito, llegamos a conclusiones homogéneas sobre la situación de la empresa. Que seguramente no sean conclusiones ajenas lo que sucede en otras empresas del rubro, o que compartan alguna de sus características de tamaño, cultura, estructura, trayectoria u otras.

Encontramos varias debilidades en su gobernanza de TI y datos, las cuales asociamos fuertemente a una cultura y estructura heredadas de otro contexto, y muy difíciles de cambiar de forma ágil. Hablamos de una empresa que opera desde el año 1.857 en varias geografías, y por lo tanto tiene una estructura, procesos, modelo de negocio, *governance* y capital humano, pensados para otro contexto y forma de trabajar. Si bien se ha ido renovando a lo largo del tiempo, las tecnologías y la gestión de los datos tenían un papel secundario hasta hace muy pocos años atrás. O lo que es lo mismo, han cambiado a un ritmo que la organización no ha podido acompañar. Y sin duda este ha sido uno de los mayores y más abruptos cambios del último tiempo. Esto, sumado a un gran tamaño, compromisos con el personal actual, un negocio en marcha que no puede detenerse y compromisos de retorno a los accionistas, hacen que el proceso de cambio hacia una cultura y modelo de negocio más orientado a las nuevas tecnologías y los datos, sea lento. Y esto queda en evidencia en varios de los síntomas que hemos repasado en este documento. Ninguna las características mencionadas en este párrafo son exclusivas de este banco en particular ni del rubro financiero, lo cual nos invita a pensar que seguramente también esté sucediendo en mayor o menor medida en otras empresas que comparten estas características.

Los síntomas negativos más relevantes que identificamos en lo que respecta a su gobernanza de TI fueron:

- En una empresa donde su misión es convertirse en una plataforma abierta de servicios financieros, solo el 4% del personal forma parte del departamento de TI. Los cuales a su vez tienen en promedio una edad superior a 45 años y baja formación técnica, tanto en temas de TI como en temas de gestión de proyectos.
- Varios problemas de sinergia y coordinación entre el área comercial y área de tecnología. Lo cual por la elevada incidencia que tiene hoy la gestión de TI en las capacidades de llevar adelante el negocio, comprometen cada vez más los resultados. Existe una incorrecta

separación entre la planificación estratégica y la gestión de TI. Esta es una visión antigua en la que TI era un área más de soporte a la cual se realizaban demandas para cumplir en mediano plazo sin mayores inconvenientes.

Toda empresa de este porte en la actualidad debería tener instancias de discusión estratégicas donde se entienda cual es el rol que se le quiere dar a TI en la organización. En caso que las TI tengan un rol relevante para el cumplimiento de su misión, deberían asegurarse de en primer lugar, contar con la cantidad y calidad de personal requerido para hacer frente a una adecuada gestión de TI. Habrá otras empresas en las que el rol de la TI no sea tan significativo para su modelo de negocio, y puedan optar por ejemplo en tercerizar esa gestión y apalancarse en otras empresas expertas en esa materia y contratarle sus servicios. Este no pareciera ser el caso del banco, que debiera si quiere asegurar el cumplimiento de su misión y mantenerse competitivo frente a las nuevas empresas llamadas *fintech*, aumentar su plantilla de TI y comenzar planes de capacitación para el personal actual. Solo el 4% de su personal dedicado a TI es inadmisibles para una empresa de este tipo en la actualidad, y peor aún, si no cumplen con los requisitos básicos de formación técnica. El principal desafío seguramente esté, en cómo hacer frente a estos planes de captación de nuevo personal y capacitación, sin generar sobrecostos insostenibles y manteniendo los niveles de rentabilidad necesarios. Con lo cual este análisis de la estrategia y proceso de recambio del personal, es una primera recomendación o punto de revisión que sin duda haría a cualquier empresa de características similares en esta situación.

Por otro lado, es fundamental que las empresas trabajen en nuevas y mejores formas de coordinación entre el negocio y las áreas de TI. Antiguamente el negocio era quien llevaba la estrategia y TI era un área más de soporte. A la que se hacían requerimientos ad hoc que debía atender. Al día de hoy, cada vez en más negocios y rubros, TI pasó a formar parte del corazón de la estrategia, de los factores diferenciales, de las barreras a la entrada, de la competitividad, de la eficiencia y de la calidad del servicio. Con lo cual, la estrategia debe ser planificada, ejecutada y controlada de forma conjunta entre TI y el negocio. Algunas empresas han llevado esto a un extremo tal, que están comenzando a concentrar en una única autoridad la dirección comercial y de TI de la empresa. De hecho, existen varios ejemplos de esto por ejemplo en la industria financiera. Otro cambio menos drástico pero necesario, es re adecuar todos los procesos de planificación anual, gestión de proyectos, seguimiento de resultados y *governance* de alta dirección, de forma tal que comiencen a tener otro grado de participación mucho más elevado de los sectores de TI. TI debe

dejar de ser un área más de soporte y pasar a ser parte de la gestión del negocio y la estrategia. De hecho, este es un paso que el que el banco analizado precisa, y no puede demorar en dar si quiere mantenerse competitivo. El área comercial no puede ponerse objetivos de ventas, calidad de servicio o lanzamiento de productos nuevos por sí solo. Ya que en la mayoría de las industrias esto depende de la infraestructura, software y capacidades del área de TI. La capacidad de hacer negocios hoy ya no está ligada al músculo comercial y de gestión financiera, sino que depende cada vez más de la capacidad de digitalizar procesos y adaptación a las nuevas tecnologías. En los sectores financiero y de *retail*, esto resulta bastante evidente, mientras que en otros puede no serlo tanto. Pero en todos está afectando en mayor o menor medida y deberían al menos cuestionarse si sus procesos de coordinación entre estrategia y gobernanza de TI no deberían ser actualizados.

También como parte de una forma de trabajar incorrecta y heredada, vemos en el caso de estudio una organización que, si bien está intentando avanzar hacia una mejor gobernanza de datos, aún tiene mucho por hacer en este sentido. Por mucho tiempo no existió una función única y centralizada encargada de llevar adelante la gestión de la información dentro de la organización. Esto llevo a que estas funciones estén divididas entre varios otros sectores. Lo cual desencadenó a su vez que se generaran silos en la organización. Tanto desde el punto de vista de los sistemas, como de los propios analistas y consumidores de la información. Esto sumado a un rol de TI muy pasivo en la gestión de datos, limitado a la recopilación y almacenamiento de la información en bases de datos. Esto se ha intentado comenzar a cambiar con la incorporación de un CDO en la empresa, pero su figura aún no se encuentra debidamente posicionada y empoderada en la organización como para ejercer el liderazgo necesario para un cambio de este tipo. Esto en parte, por la resistencia al cambio de quienes gestionan la información actualmente en la organización y por la falta de autoridad detectada para el CIO y CDO. Nuevamente, parecería lógico que ninguno de estos inconvenientes fuera exclusivo del banco, sino que con seguridad aplique para varias empresas con características similares. Es clave que, ante esta situación, la empresa trabaje en un plan de empoderamiento de CIO y CDO, generando los espacios necesarios para hacerlo a nivel de alta dirección, avalando únicamente proyectos de gestión de la información patrocinados por estas figuras y apoyando con una comunicación interna que ayude a concientizar sobre la importancia de la tecnología y los datos a aquellos que por su rol tengan visibilidad. Adicionalmente, es importante que para ayudar en el proceso de empoderamiento del CDO, la empresa se asegure que su equipo cuente con todos los perfiles necesarios para dar un adecuado nivel de soporte a toda la

organización. Como mínimo: un *data analyst*, un *data engineer* y un *data scientist*. Más allá de las particularidades que pueda tener una u otra organización, y su grado de avance en este proceso, todas las empresas desde un determinado porte deberían estar pensando en comenzar este proceso de ajuste en su forma de gestionar los datos. El proceso sin duda será distinto y menos prioritario en empresas de menor tamaño.

Varios de estos temas quedaron además en evidencia en el repaso del proyecto de implementación de CRM de tarjetas en la organización del caso de estudio:

- Un área de negocio que ha tenido un alto crecimiento y estaba experimentando problemas e ineficiencias para el manejo de su información. Duplicidad, inconsistencias e ineficiencias, entre otros.
- Un proyecto que no surgió de reuniones de planificación estratégica de alto nivel entre negocio y tecnología. Sino que surgió exclusivamente del negocio.
- Un proyecto de gestión de la información que se implementó, pero no fue liderado por la gerencia del CDO. Y fue llevado adelante a pesar de incluso no contar con la aprobación de CIO ni CDO.
- Un proyecto que no se llevó adelante siguiendo una metodología convalidada y promovida por la propia organización. Si se quiere una visión única de la gestión de la información, todos los proyectos de gestión de datos deberían compartir una metodología de implementación. Que debería además ser propuesta y promovida por CDO.
- No se capitalizarán los conocimientos sobre las bases de datos, sistemas y negocio trabajados. Ya que el personal que lo llevó adelante no es el que desarrollará proyectos similares para otros de los negocios del banco. Tampoco se capitalizarán aprendizajes en lo que respecta a gestión de proyectos.
- Decisiones tomadas por personal fuera del área de TI que tendrán impacto en temas de mantenimiento de software, soporte, renovación de licencias y gestión de la información dentro de la organización en el futuro.
- No existió gestión de la metadata para el caso puntual del CRM de tarjetas. El cual dará soporte a uno de los negocios más importantes del banco.

Todo esto sin duda, refuerza la necesidad de empoderamiento de CIO y CDO. Y la urgencia de mejorar sinergias entre áreas de negocio y de TI de forma tal de evitar cualquiera de las situaciones

mencionadas para el proyecto ejemplo. Cuando se solucionen estos defectos en la gestión, sin duda los resultados obtenidos con proyectos de este tipo serán aún mejores. En el repaso paso a paso de este proyecto se mencionaron varios aspectos que podrían haberse realizado mejor, de forma más eficiente y con mejores resultados para la empresa, si se hubieran llevado adelante de la forma correcta. Con lo cual queda claro que el no atacar estas situaciones, tiene un impacto negativo en el negocio y los resultados obtenidos en la última línea.

En lo que respecta al ejemplo de implementación del algoritmo, interesa destacar dos aspectos principales:

- Al día de hoy existe un enorme potencial y nuevas herramientas para que las empresas puedan explotar la información que poseen. Y se pueden encontrar ejemplos concretos de aplicación que pueden tener un impacto directo en la forma de clasificar sus clientes, las ofertas que se realizan, los productos que se diseñan y/o los canales que se utilizan para atenderlos o captarlos, entre otros. Es decir, existen muchas formas de explotación de la información que tienen impacto directo y tangible en la última línea de resultados.
- Del repaso del ejemplo simplificado de aplicación, queda claro también que estas nuevas herramientas requieren mucho conocimiento técnico de las herramientas necesarias, conocimientos estadísticos y de programación, y mucha práctica. Con lo cual una empresa no puede pensar en embarcarse en proyectos de este tipo sin invertir en recursos especializados.

Para empresas del porte como la del banco, sin duda entiendo que deberían hacer una apuesta a invertir en recursos de este tipo, que le permitan posicionarse y lograr una ventaja competitiva a través de la explotación de la información que genera y maneja. Máxime en empresas del rubro financiero, que son quienes por excelencia manejan mucha información confidencial sobre los hábitos de consumo, ingresos, patrimonio, ahorro y más sobre sus clientes. En el caso del banco analizado debería dar un paso más, reforzando el equipo del CDO y empoderándolo como ya mencionamos. Para empresas de menor porte, tal vez esta sea una inversión innecesaria o que podrían tercerizar apoyándose en otras que ya cuentan con el conocimiento.

Una conclusión importante y en la que podemos resumir mucho de lo comentado, es que es un error común en la actualidad buscar los motivos del fracaso o dificultades en la implementación de un proyecto de *big data*, en una causa puntual o la falta de capacidad de un líder o un área

específica dentro de la organización. Cuando en realidad los motivos para el fracaso de un proyecto de este tipo seguramente sean varios y mucho más profundos. Pedirle a una organización que lleve adelante proyectos de *big data* sin antes haberse cuestionado si su gobernanza de TI y datos se encuentran preparadas y adaptadas para el contexto actual, es un error. Es como intentar correr sin antes haber logrado caminar. Entre las dificultades más comunes que hacen al fracaso de un proyecto de este tipo, podemos encontrar la subestimación o falta del capital humano necesario, la descoordinación y falta de sinergia entre áreas de negocio y TI, la falta de empoderamiento de CIO y CDO y la subestimación de los problemas de silos en la gestión de la información y los sistemas actuales. Encarar proyectos de este tipo sin la revisión de estos problemas clásicos, implica un alto riesgo para el éxito del proyecto y los resultados que se obtendrán. Es importante que en la alta dirección exista conciencia de este tipo de problemas habituales, y se aseguren de ayudar al equipo a superarlos antes de exigir resultados en la gestión de proyectos de *big data*.

REFERENCIAS BIBLIOGRÁFICAS

Blog Ibermática Industria. Accedido el 5 de agosto, 2021, desde <https://ibermaticaindustria.com/blog/trucos-para-romper-los-silos-organizacionales-con-el-sistema-de-gestion-de-la-informacion/>

Blog Wearedrew. Accedido el 8 de agosto, 2021, <https://blog.wearedrew.co/departamentos-ciegas-silos-de-informaci%C3%B3n-y-su-efecto-en-tu-empresa>

Blog 7 puentes. Accedido el 10 de julio, 2021, desde <http://www.7puentes.com/blog/2021/06/02/data-analyst-engineer-scientist-jugadores-clave-en-un-proyecto-de-datos/>

Deborah Henderson. (2010) *DAMA Guide to the Data Management Body of Knowledge*. Boston.

Organización Internacional de Estandarización & Comisión Electrotécnica Internacional (2015). *ISO/IEC 38500:2015 Corporate Governance for Information Technology*.

Web EALDE. Accedido el 30 de julio, 2021, desde: <https://www.ealde.es/fases-proyecto-bog-data-empresarial/>

Web Alteryx. Accedido el 1 de septiembre, 2021, desde <https://www.alteryx.com/es-419/solutions/analytics-need/bi-visualization/visual-analytics>

Web James McCaffrey. Accedido el 3 de septiembre, 2021, desde <https://jamesmccaffrey.wordpress.com/2020/04/21/example-of-calculating-the-gower-distance/>

Web Searchdatacenter. Accedido el 20 de agosto, 2021, desde <https://searchdatacenter.techtarget.com/es/definicion/SQL-Server>

Web Slashmobility. Accedido el 3 de septiembre, 2021, desde <https://slashmobility.com/blog/2019/07/clustering-como-obtener-agrupaciones-inherentes-en-los-datos/>

Web Tech-BI. Accedido el 5 de septiembre, 2021, desde <https://www.tech-bi.com/es/tecnologias/qlikview/>

Web The Information Lab. Accedido el 1 de septiembre, 2021, desde <https://www.theinformationlab.es/que-es-alteryx-y-para-que-sirve/>

Web Unioviedo. Accedido el 3 de septiembre, 2021, desde https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html

Web Wikipedia. Accedido el 3 de septiembre, 2021, desde https://es.wikipedia.org/wiki/Agrupamiento_jer%C3%A1rquico