



CURSO 2020-2021

MÁSTER EN BUSINESS INTELLIGENCE Y DATA SCIENCE

MODULO:

**Introducción al Big Data: Nivel de madurez de una
organización**

Nombre estudiante: Marc Faravelli Rodríguez

Nº Expediente: 2047104

E-mail: marcfaravelli@gmail.com

Índice

Fase 1 – Análisis organizativo: funcional y tecnológico

**Fase 2 – Levantamiento, análisis y priorización de
necesidades de negocio**

Fase 3 – Definición detallada del caso de uso

Fase 4 – Plan de proyecto y business case

Fase 1 – Análisis organizativo: funcional y tecnológico

Para resolver el caso práctico correspondiente al módulo “Introducción al Big Data: Nivel de madurez de una organización” se ha escogido como referencia a Gastrofira, el departamento interno de comida, bebidas y eventos de Fira Barcelona.

Gastrofira fue fundada en 2001 por decisión del plan estratégico de Fira de Barcelona con el objetivo de ofrecer un valor extra a sus servicios y lograr la mejor satisfacción de los invitados a los eventos.

Todos los departamentos de Fira están conectados con la misma VPN. Esto permite una extensión segura de la red y facilita el acceso a todos los departamentos en las diferentes herramientas tecnológicas y a toda la documentación que se quiera compartir. A través de esta red privada, se crea una conexión interna con todos los trabajadores de Fira.

Gastrofira tiene una unidad, dentro del servidor, que permite compartir información (equipo, clientes, programas informáticos,...) entre sus departamentos internos. Hay que matizar que cada uno de sus departamentos comparten sus informaciones entre los conformantes de cada equipo y solo alguna información es compartida a nivel general con todos los demás.

En las tareas habituales se suele recopilar información de diferentes formas y fuentes y es compartida entre el departamento de ventas y operaciones:

- 1) SAP: todos los pedidos (ordenes de servicio) que se realizan quedan registrados en este programa informático donde se puede identificar que se ha vendido a cada cliente, el volumen, los días, ...
- 2) Tablas Excel: tanto el seguimiento de los clientes previos al evento como el debrief que se hace al acabar, se registran en unas tablas para que se analicen los resultados, especialmente las incidencias, para la toma de decisiones en las siguientes ediciones y diferentes eventos.

Cabe destacar, que el uso de técnicas de Big Data ya se está usando a nivel de Fira y ahora se está intentando implementar también a Gastrofira. Analizando una gran empresa es más complicado encontrar necesidades para poder solventar, ya que difícilmente se encuentran impreparados, pero seguramente puede convertirse en una

buena herramienta para ayudar a predecir la demanda de los clientes (tanto volumen como tipo de producto) y reducir costes.

Ahora que ya se ha seleccionado el departamento de manera justificada, se empezará con su análisis que se ejecutará en dos ejes:

1. Negocio o funcional: Gastrofira está compuesta por diferentes departamentos que están estrictamente relacionados y conectados como si de una cadena se tratara. Estos departamentos son los siguientes:

- Ventas / comercial
- Operaciones / logística
- Producción
- Compras
- Recursos humanos

Ventas / comercial

Este es el departamento más crucial ya que es el que está en estrecho contacto con los clientes y determina, juntamente con operaciones, cual serán los servicios vendidos, como se harán y los ingresos (precio de venta).

Operaciones / logística

Su labor principal es la de organizar, preparar y ejecutar (logística) los servicios vendidos y acordados previamente con el departamento de ventas. Además, acuerdan presupuestos de algunos servicios en concreto (baristas, bartenders,...)

Producción

Son los que controlan la producción de comida de todos los servicios, hacen propuestas de ofertas gastronómicas al departamento comercial y buscan acuerdos con proveedores.

Compras

Son los encargados de buscar acuerdos con las marcas para F&B de grande stock (Ej.: Coca-cola, Estrella Damm,...), establecen contratos con proveedores y se ocupan del I+D buscando nuevos productos que se acorden a las necesidades del momento.

Recursos Humanos

Un punto clave de Gastrofira y el nivel de servicio que se brinda es el hecho de que todas las personas adicionales (camareros y cocineros) han sido contratadas por el propio departamento de recursos humanos. Los requisitos buscados y que deberá tener el personal seleccionado son los siguientes:

- Orientación al cliente
- Habilidades de comunicación
- Resistencia y eficiencia
- Autonomía
- Jugador de equipo

Gastrofira se caracteriza por tener una ética y seguir unos valores bien claros:

- Orientación al cliente: escuchar y entender sus necesidades, priorizar su satisfacción
- Ser un valor diferencial de Fira Barcelona
- Flexibilidad: el cliente y sus necesidades cambian y las soluciones de Gastrofira deben de adaptarse a cualquier contexto
- Anticipación: tomar la iniciativa. Estudiar las oportunidades y los riesgos y planificar la mejor respuesta
- Eficiencia: optimizar el tiempo y los recursos. Conseguir los mejores resultados
- Medición, evaluación, mejora: medir y analizar los resultados para mejorarlos y tomar decisiones
- Visión global: entender el mercado. Conocer sus productos y servicios
- Excelencia: máxima calidad, en todo momento y en todos los servicios.

En cuanto a las líneas de negocio, se rigen en 3 ejes:

- 1) Locales de restauración y concesiones:
 - Gestión de locales fijos
 - Gestión de locales efímeros
 - Coordinación de puntos de venta móviles
 - Concesiones y maquinas vending
- 2) Acuerdos con marcas: ingresos por exclusividad de marcas
- 3) Servicios de catering:
 - Catering a stand
 - AEX: área del expositor (E-commerce)

2. Tecnológico: el departamento de TIC es el que se encarga de administrar y gestionar todo el ecosistema tecnológico de Fira. Este equipo se dedica a diseñar soluciones personalizadas para garantizar la continuidad del negocio, facilitar la comunicación interna entre recursos y proteger la seguridad informática de toda la estructura.

Los principales recursos utilizados en Fira y Gastrofira son los siguientes:

- **Outlook, Microsoft Teams y Skype:** son ambientes de trabajo que facilitan la comunicación interna entre los miembros de Fira, incluido Gastrofira. De hecho, es posible poder encontrar cualquier persona de Fira en el mismo buscador ya que están todos registrados. Son una ayuda para implicar todos los trabajadores a construir un sentimiento de comunidad, a favorecer mejores relaciones, a intercambiar libremente ideas, acelerar la ejecución de algunas tareas y a crear transparencia entre los repartos.
- **CRM:** es la base de datos propiamente de Fira. Todos los departamentos tienen acceso a la misma y se encuentra información puramente de clientes, nifs, direcciones,...

A las empresas con las que se interactúa, se llaman clientes. Los clientes se identifican siempre por un código llamado IDSAP. Este código tiene siempre 7 cifras. Hay dos tipos de clientes, se distinguen por el número con el que empieza su IDSAP. Si es un 1, es un cliente nacional, mientras que, si empieza por 2, se trata de un cliente extranjero.

Un cliente, además, puede estar en estado POTENCIAL o ADMINISTRATIVO.

Potencial: Empresa que se acaba de crear en el sistema o que ya existía pero que nunca ha tenido una relación comercial/financiera con Fira.

Administrativo: empresa con la que ya se ha tenido algún tipo de relación comercial o financiera.

Una vez se accede a la ficha de un cliente, se puede ver su principal información y con la que se hace uso habitualmente:

- Datos Empresa:
Aquí se encuentran los datos relativos a la empresa, direcciones y telefono de contacto.
- Cliente/contacto/Salón:
Aquí se encuentran los datos relativos a los contactos de la empresa.
- Datos auxiliares:
Para consultar los distintos nombres comerciales que pueda tener la empresa.
- Gestion comercial y pedidos y facturas:
Para consultar el historico de la actividad comercial que pueda tener el cliente

- **SAP:** es el programa por excelencia usado en todos los departamentos tanto de Fira como Gastrofira en cuanto lleva el registro de todas las operaciones de negocio realizadas. Es la herramienta que se utiliza para hacer pedidos además de poder extraer informes y ver todas las ordenes de servicios. Está sincronizado con el CRM para poder registrar los clientes en las ordenes de servicios creadas por el mismo.
- **FootPrints:** es una herramienta de servicio al cliente y soporte técnico fácil de usar y administrar. Permite registrar informaciones importantes (pedidos de catering onsite para el mismo día, observaciones de entrega de productos, ...). En definitiva, ofrece una experiencia excepcional al cliente al tiempo que crean valor comercial.
- **Sharepoint:** es una plataforma que permite la gestión documental y el trabajo en equipo. En especial para Gastrofira, es el referente para registrar todo el catálogo fotográfico de los productos presentes en los brochures que se envían a los clientes. Están incluidas también fotos de ofertas de años anteriores y algunas fotos de algún servicio exclusivo y de gama alta que se ha hecho en algún evento en concreto.
- **E-commerce:** es un sistema que permite el proceso de ventas en una base omnicanal (gestión de producto y catálogos, precios y promociones, pagos y seguridad, pedidos, inventario y logística,...). Es una plataforma muy útil para los comerciales porque les permite enfocarse en la gestión de servicios personalizados que requieren más labor, más dedicación y que también implican personas de otros departamentos para poder

delegar las peticiones de simples entregas de F&B en los stands a este sistema informático.

- **Redes sociales - Instagram:** es la última incorporación. En ello, se publican los servicios y productos más novedosos de Gastrofira. Es una forma de poder tener más visibilidad en las redes sociales y poder interactuar también con clientes.

Haciendo hincapié a Gastrofira, es importante matizar que no todas estas herramientas tecnológicas se aplican a todos sus departamentos. El único departamento que sí las utiliza todas es el de ventas / comercial. Por otro lado, el SAP es la única herramienta que todos los departamentos tienen en común.

Fase 2 – Levantamiento, análisis y priorización de necesidades de negocio

El proceso de análisis y priorización de las necesidades empresariales de Gastrofira, empieza después de haber especificado con claridad cuáles son los valores, la ética y las líneas de negocio del mismo departamento.

Para empezar a identificar algunas carencias o necesidades que se puedan encontrar en el negocio, es importante conocer los objetivos marcados por Fira y en específico por Gastrofira, derivados del plan estratégico.

El plan de gestión tiene la voluntad de explorar el entorno de Fira, apoderándose y desarrollando el talento de la casa, articulando una mayor transversalidad con la ambición de hacer efectivo el trabajo en red.

Los aspectos cualitativos, extrapolables a Gastrofira, que se proponen trabajar para avanzar en la consecución de los ejes fijados por el plan estratégico son:

1. Crecimiento rentable
2. Impacto económico y social
3. Organización de alto rendimiento

Los 3 ejes, están agrupados en los siguientes leitmotiv:

- Orientación al cliente
- Transformación digital
- Eficiencia y productividad
- Comunicación excelente
- Datos / Información / Conocimiento
- Focus comercial
- Compromiso social y medioambiental
- Organización de alto rendimiento

Dentro este paradigma y a partir de un análisis general de los objetivos decretados por el departamento de Gastrofira, siguiendo también los leitmotiv establecidos para dichos, se han podido detectar las siguientes necesidades:

- 1) Reducción del desperdicio de alimentos
- 2) Ayuda y participación con acciones sociales
- 3) Reducción y reciclaje de materiales
- 4) Predicción de la cantidad de visitantes que acudirán a los locales de restauración durante los eventos
- 5) Mejora del E-commerce (detalle del producto con fotos, características,...)
- 6) Pop ups para avisos de contratación de maquinaria y productos anticipables (SAP y E-commerce)
- 7) Control de stock (SAP y E-commerce)
- 8) Reducción del impacto ambiental (compromiso con el planeta)
- 9) Predicción del volumen de clientes que pedirán servicio de catering a Gastrofira para el Mobile Word Congress 2021
- 10) Entregas puntuales

Una vez listadas todas las necesidades de mejora detectadas, se analizarán más en específico para averiguar a cuáles de ellas sería apropiado usar técnicas del Big Data y cuáles no. Algunas de ellas se agruparán por objetivos comunes:

Sin necesidad de recurrir a técnicas de Big Data:

- Reducción del desperdicio de alimentos
- Ayuda y participación con acciones sociales
- Reducción y reciclaje de materiales
- Reducción del impacto ambiental

Se alinean con el crecimiento rentable y con el compromiso social. La información requerida para encontrar soluciones de mejora se encuentra disponible en el departamento de Gastrofira (comercial, operaciones y producción). No requiere técnicas de Big Data ya que independientemente es posible poder encontrar soluciones al problema todas con el fin de aumentar su valor social (muy valorable por parte de

muchos expositores y organizadores de eventos) y económico (ayudas por parte de gobiernos para la contratación de personas vulnerables, mayor demanda de materiales compostables = mayor venta de productos, ...):

- Buscar ONGs o programas de recogida de alimentos para que lo puedan aprovechar;
 - Colaborar con comunidades locales para facilitar la inserción laboral a personas con riesgo de exclusión social;
 - Cambiar los envases tanto de las bebidas como de los alimentos en vidrio y material compostable de alta calidad;
 - Reducir emisiones de CO2 con la importación de productos locales y disminuyendo el trayecto de transporte de alimentos.
-
- Mejora del E-commerce
 - Pop ups para avisos de contratación de maquinaria y productos anticipables
 - Control de stock

Estas necesidades se relacionan con el ecosistema tecnológico. Se deberían mejorar por tal de aumentar la calidad del servicio al cliente y facilitar tanto su labor de contratación de los productos como también del equipo comercial. Toda la información requerida es posible obtenerla dentro de Gastrofira con la ayuda y soporte del departamento de TIC de Fira.

Necesidad de recurrir a técnicas de Big Data:

- 1) Predicción de la cantidad de visitantes que acudirán a los locales de restauración durante los eventos

Gastrofira quiere analizar y poder predecir cuantos locales debería de abrir durante los eventos para intentar reducir costes y ser más eficiente. La decisión que tienen que tomar es bastante frecuente ya que hay una media de 2 o 3 eventos por mes. Obviamente, esta dificultad radica sobre todo en las primeras ediciones de los eventos y en eventos de tamaño mediano-grandes.

La información requerida para solventar esta necesidad se podría encontrar mediante diferentes fuentes de datos tanto internos como externos: histórico de facturación (en

caso de que no sea un evento nuevo), listado de clientes y sus respectivas características, históricos de participación de visitantes del mismo evento en ediciones anteriores en otras instalaciones feriales, condiciones meteorológicas, interacciones con el organizador de evento. Se necesitará de bastante información procedentes de diversas fuentes de datos y utilizar diferentes patrones para poder hacer una predicción lo más posible verídica.

2) Predicción del volumen de clientes que pedirán servicio de càtering para el Mobile Word Congress 2021

El Mobile World Congress, es un evento muy relevante para Fira Barcelona. Por esta razón hay la intención, debido a la crisis económica del Covid 19, de maximizar los beneficios económicos derivados del evento. Para esta predicción es necesario tener en consideración datos internos, pero sobre todo externos: políticas adoptadas por los gobiernos, evolución económica-sanitaria, restricciones de vuelos por parte de las compañías aéreas, medidas de seguridad adoptadas tanto por Fira como por la OMS.

La información será variada: páginas webs de diferentes fuentes fiables, hojas de cálculos, documentos estatales,...

La recopilación de esta información será constantemente actualizada a lo largo de los días previos al evento ya que, al ser una situación inestable, es más susceptible a cambios.

3) Evitar largas colas en los locales de restauración más demandados durante grandes eventos

Uno de los problemas individualizados es la formación de grandes colas en determinados locales de restauración durante la celebración de grandes eventos como por ejemplo Mobile World Congress, Itma, ISE, etc...

Si se consiguiera evitar estas colas, sería beneficioso para Gastrofira en cuanto no perdería clientela ya que algunos visitantes preferirían ir a bares o restaurantes a las afueras de las instalaciones feriales.

Para hacer estimaciones tendremos que encontrar un algoritmo que tenga en cuenta diferentes variables como por ejemplo el tiempo promedio de cola de los visitantes, las

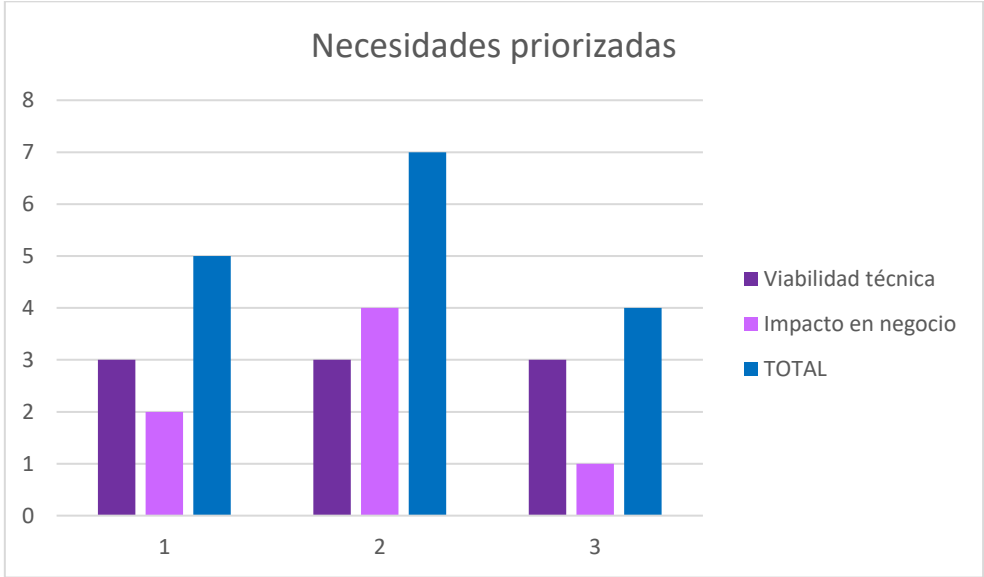
variables del día basadas en varios tipos de información, incluido el clima, la cantidad de accesos y otros parámetros monitoreados y comparados con los datos históricos en tiempo real, lo que permitirá maximizar la capacidad de los locales respetando sus limitaciones estructurales.

Para resumir las tres necesidades en las cuales es aplicable el Big Data, se ha creado una tabla donde se recoge de manera estructurada la información:

ID	Descripción	Impacto en negocio	Fuentes de datos	Enfoque analítico	Volumen de información	Frecuencia de análisis
Predicción de la cantidad de clientes que acudirán a los locales de restauración durante los eventos	Quiero identificar la cantidad de clientes que irán a los locales de restauración durante los eventos	Aumento del 10% de los ingresos	<p>Listado de clientes y características</p> <p>Histórico de facturación (no primeras ediciones)</p> <p>Histórico participación ediciones anteriores en otras instalaciones feriales</p> <p>Condiciones meteorológicas</p> <p>Interacciones con el organizador</p>	Identificar patrones internos como externos para predecir el volumen de visitantes que acudirán a los locales	Alto	Semanal
Predicción del volumen de clientes que pedirán servicios de catering para el MWC21	Quiero predecir la cantidad de clientes que pedirán servicios de catering para el MWC21	Aumentar las ventas y reducir costes	<p>Políticas adoptadas por los gobiernos</p> <p>Evolución económica-sanitaria</p> <p>Medidas de seguridad (Fira y OMS)</p> <p>Restricciones de vuelo</p>	Identificar patrones para poder predecir el volumen que cogerá la próxima edición del MWC21	Alto	Diaria o semanal

Evitar colas en los locales de restauración más demandados en los grandes eventos	Quiero identificar un nuevo sistema de gestión de las colas	Aumentar ventas	Listado de clientes y características Datos (previsiones) meteorológicas Histórico de participación	Identificar patrones que nos permita crear un algoritmo para minimizar las colas	Medio	Diaria o semanal
---	---	-----------------	---	--	-------	------------------

A continuación, se ha querido valorar cada una de las necesidades mediante un grafico para poder resaltar cual ha sido la que se ha elegido como prioritaria para el caso de uso:



Fuente: Elaboración propia

Para esta valoración se ha tenido en cuenta dos variables: la viabilidad técnica o analítica (color morado) y el impacto en negocio (color fucsia). El total, representa la suma de las dos variables para cada necesidad y, como se puede observar, la segunda ha sido la que ha tenido más puntuación y por lo tanto la elegida (prioritaria). El número que se muestra debajo de cada grafico es el correspondiente a cada necesidad mientras que la escala en la que se basa va de 0 a 5.

Fase 3 – Definición detallada del caso de uso

La epidemia de Covid-19 representa una amenaza significativa para todo el sector turístico, en particular al de negocio, tanto para quienes trabajan allí como para quienes desean continuar viajando. Ante esta situación imprevisible en todo el planeta, ha llevado a muchas empresas a replantearse su gestión económica y rectificar algunas directrices del plan estratégico previamente establecido. En el caso de Fira, se ha querido focalizar la atención en su evento estrella: el Mobile World Congress. Este es el evento que más beneficios aporta a la empresa y no es un caso que haya una cierta preocupación al respecto.

Como se ha podido observar en el gráfico anterior, después de haber comparado cada necesidad respecto las variables viabilidad técnica e impacto en negocio, se ha decidido priorizar la siguiente:

- Predicción del volumen de clientes que pedirán servicio de catering para el Mobile Word Congress 2021

Para la resolución del caso de uso, se implementarán técnicas de Machine Learning Supervisado. Una de las características extraordinarias del aprendizaje automático es su capacidad predictiva. En el pasado, las decisiones comerciales a menudo se tomaban en función de resultados históricos. Hoy, el aprendizaje automático utiliza análisis de datos avanzados para hacer predicciones. En este caso, Gastrofira puede tomar decisiones proactivas y prospectivas en lugar de depender de datos pasados; con mayor razón viviendo una pandemia, es ideal implementar estas técnicas para predecir y prevenir ya que el panorama económico mundial ha cambiado sin aun conocer sus reales consecuencias.

Lo que realmente se quiere conseguir es, mediante algunos patrones analizados, poder identificar y predecir una cantidad aproximativa de clientes que pedirán servicio de catering durante el MWC21.

El esperado impacto en el negocio de Gastrofira se podría desglosar en 3 puntos:

- Poder “atacar” al cliente con antelación (especialmente los considerados “VIP”) y poderles ofrecer un servicio lo más personalizado posible. Este servicio normalmente cuesta más debido al alto volumen de peticiones y de visitantes (m2 disponibles en el congreso);
- Poder concretar mejores negociaciones con los clientes y proveedores (como por ejemplo descuentos o “congelación” de una parte de dinero pagado para usarlo en otro evento sin la necesidad de devolvérselo al momento);
- Hacer un balance y entender el volumen de trabajo que podría haber para la preparación de la próxima edición. De esta manera, poder reducir los costes: contratación de menos personal (comerciales de refuerzo de 6 meses, camareros, cocineros, transportistas, contratación con proveedores...).

Las fuentes de datos necesarias para el caso de uso son principalmente externas, ya que, aunque se trate de un gran evento consolidado en Barcelona, las circunstancias adversas que se presentan no dependen exclusivamente de Fira y Gastrofira.

Fuentes externas

- Listado de clientes y características: edad, sexo, países de origen, cultura, ...
- Políticas adoptadas por los países: oberturas fronteras, cuarentena, ...
- Listado sobre las medidas de seguridad declaradas por la OMS: uso de mascarillas, guantes, gel de mano, ...
- Informes sobre la evolución económica tanto de España como del resto de países: daño económico causado por el Covid 19 a las empresas, menos presupuesto para apuntarse en ferias o congresos, cambio de moneda (p.ej.: € - \$)
- Informes sobre la evolución sanitaria tanto de España como del resto de países: condiciones hospitales: nº de personas enfermas, nº de personas inmunes, nº de personas consideradas en alto riesgo,...
- Informe sobre las restricciones de vuelos por parte de las compañías aéreas y los precios (coste del vuelo)

Fuentes internas

- Base de datos de Fira/Gastrofira (CRM y SAP) con información de los clientes
- Normativas sobre las medidas de seguridad para la prevención de la propagación del Covid 19

En cuanto al enfoque analítico, los pasos para poder dar respuesta a la necesidad de negocio tratada se basarán en las siguientes fases:

1. IDENTIFICAR LAS TECNOLOGÍAS ADECUADAS PARA RECOPILAR Y ALMACENAR DATOS

Para empezar este proyecto es necesario tener un sistema integrado y poderoso, es decir, integrar las infraestructuras informáticas y las tecnologías de inteligencia artificial para poder trabajar en equipo formando una sinergia entre ellas. En resumen, almacenamiento en nube, potencia computacional, software de análisis (por ejemplo, Python y R) y networking constituirán un sistema que pueda almacenar a los grandes volúmenes de datos heterogéneos que viajan en la red y realizar análisis complejos en tiempo real.

2. ANALIZAR LOS DATOS

El tipo de análisis que más se ajusta a la necesidad de negocio es el análisis predictivo con herramientas avanzadas de Machine Learning (aprendizaje automático). Esta metodología nos ayudará en el análisis de datos históricos, pero sobre todo datos actuales de la evolución del COVID 19, con sus respectivos efectos, para desarrollar modelos predictivos sobre eventos y tendencias futuras (¿qué podría pasar?). El análisis predictivo lo utilizaremos principalmente para reducir el riesgo de transacciones financieras, pero también para optimizar procesos internos con información obtenida sobre el comportamiento/tendencia de compra que podrán tener los clientes.

Para este análisis se seguirán los siguientes pasos de forma cronológica:

- **Preselección de los datos útiles** para lograr el objetivo (necesidad de negocio);
- **Limpieza y preprocesamiento de datos:** separación adicional entre datos válidos e inútiles, elección de cómo tratar campos incompletos o vacíos, selección definitiva de la información fundamental para el modelo de referencia ideal (patrones elegidos);
- **Transformación:** ¿el formato en el que se representan los datos es válido para alimentar el software de análisis? Si la respuesta es no, los datos deben convertirse;
- **Minería de datos:** es, por supuesto, el paso más importante. Se elige el mejor software para el caso individual, que escanea el almacén de datos de forma selectiva para proporcionar la respuesta buscada. La minería de datos generalmente consta de varios pasos, incluso repetidos varias veces, para refinar el procedimiento y verificar gradualmente los resultados logrados.

Una herramienta que nos podría servir para la minería de datos es la red neuronal. El programa está equipado con un conjunto de instrucciones y un algoritmo de aprendizaje que le permite evolucionar con la experiencia, ampliando su capacidad para resolver ciertos tipos de problemas.

Esta red neuronal de aprendizaje supervisado se entrena proporcionando una cierta cantidad de entradas (problemas) y salidas (soluciones), de modo que al detectar las asociaciones aprende a producir salidas correctas de forma independiente. Para la elaboración de dicho algoritmo, tendremos que analizar, por ejemplo, mediante líneas de regresión lineal, un conjunto de subvariables que constituyen la variable que usaremos para nuestro análisis. Un ejemplo sería evaluar las personas dispuestas a volar (variable) teniendo en cuenta dos subvariables: relación entre el número de vuelos y los precios del billete.

- **Interpretación de los resultados:** se evalúa si se alcanza el objetivo, y si la respuesta es no, se repite el paso anterior (y posiblemente la modificación) del paso anterior;

- **Visualización de los resultados en un formato comprensible:** estamos avanzando en la era de los grandes datos y la visualización es una herramienta cada vez más necesaria para dar sentido a la miríada de datos que se generan todos los días. Ver los datos ayuda a explicarnos los contenidos, organizar los datos de una manera más comprensible y resaltar las tendencias y los valores atípicos. La visualización efectiva nos permite exponer contenido al eliminar datos innecesarios y traer información útil al primer plano. La disponibilidad de información y la rapidez en la toma de decisiones son factores clave para el éxito de nuestro objetivo del caso de uso. Por estas razones, la opción que encajaría mejor en nuestro planteamiento sería usar un dashboard. Es un sistema de agregación que nos permite mostrar toda esta información de una manera óptima e inmediata.

Las principales ventajas son:

- a. inmediatez visual
- b. personalización de la interfaz gráfica
- c. los datos se pueden agrupar y filtrar según sea necesario
- d. informes dinámicos, en tiempo real
- e. visualización simultánea de información de diferentes fuentes

Fase 4 – Plan del proyecto

El desarrollo del caso se estructura en 3 fases de las cuales, algunas de ellas, tienen diferentes tareas. Este proyecto está planeado para llevarlo a cabo en 4 meses como se muestra en el calendario Gantt a continuación:

CALENDARIO GANTT																															
Tareas	Inicio	Fin	Días	Julio							Agosto							Setiembre							Octubre						
Fase 1: Identificar las tecnologías adecuadas para recopilar y almacenar datos																															
Definir el almacenamiento en nube			2																												
Verificar la potencia computacional			3																												
Elegir e instalar software de analisis			2																												
Fase 2: Analizar datos																															
Preselección de los datos utiles			20																												
Limpieza y preprocesamiento de datos			25																												
Transformación			10																												
Mineria de datos			15																												
Interpretación de los resultados			10																												
Fase 3: Activación de resultados																															
Visualización/Presentación			4																												

Fuente: Elaboración propia

Cada fase requiere de un tiempo determinado y de algunos perfiles bien concretos para poder desarrollar correctamente todas las diferentes tareas diseñadas en el plan:

Fase 1: Identificar las tecnologías adecuadas para recopilar y almacenar datos

La primera fase está pensada para poderla llevar a cabo en 7 días repartidos en 2 semanas.

El perfil más indicado para estas tareas es el **arquitecto de datos**. A estos profesionales se les suele relacionar con la parte de infraestructura y administración de sistemas. Montan y mantienen la infraestructura (clúster Big Data, servidores).

Fase 2: Analizar datos

Esta fase es el “corazón” del proyecto. Consta de 80 días repartidos en 14 semanas. Aquí se encuentran todas las tareas más destacadas a nivel analítico.

Podríamos dividir esta fase en 2 partes:

- La primera incluye la preselección de los datos útiles, la limpieza y preprocesamiento de datos. Para estas tareas es adecuado que se encargue un **analista de datos**. Su función será la de recopilar todo tipo de datos, para realizar una consultoría previa al proyecto. Con los datos recopilados, realizará informes para definir el proyecto o para corroborar que lo que se está haciendo va por buen camino.
- La segunda consta de las tareas de transformación, minería de datos e interpretación de los resultados. Los datos hablan, pero para entenderlos se tiene que conocer su “idioma”. La Data Science requiere habilidades analíticas para interpretar datos heterogéneos, garantizar la coherencia y veracidad. El profesional capaz de realizar esta actividad es el **científico de datos**, capaz de interpretar los datos gracias a una formación académica específica: informática, economía, estadística, física o matemáticas.

Fase 3: Activación de resultados

Esta fase es la más corta ya que tiene una duración de 4 días. Durante este breve periodo se configurará el dashboard con todos los resultados y conclusiones extraídas y se hará la presentación.

Como se podrá observar la fase que necesitará mucho más tiempo de dedicación es la fase 2 de análisis de datos. En particular, cabe destacar que la recopilación y criba de toda la información es la tarea que más atención conlleva, especialmente si la mayoría de las fuentes son externas.