

Fundamentos de Ciencia de Datos

PEC1: ¿Ciencia en los datos?

Criterios de evaluación generales de la PEC.

- *Claridad en la respuesta y en los razonamientos.*
- *No omitir ninguna de las preguntas de cada ejercicio.*
- *Capacidad de síntesis.*
- *Originalidad.*
- *Aportación de alguna referencia externa que complemente o sustente los razonamientos.*
- *Formato estructurado del documento que facilite la lectura.*

CASO DE ESTUDIO

Supongamos que eres el CDO o *Chief Data Officer* de una multinacional con sede en tu país. Dicha organización, se dedica a la producción y distribución de productos de telefonía, informática y multimedia.

En este escenario, tu principal responsabilidad es la de gestionar la **plataforma de gobierno de datos**, cuyo principal objetivo es el de planificar, vigilar y gestionar el uso de los datos.

Como responsable de gobernanza de los datos, mantienes reuniones periódicas de seguimiento con miembros de otros departamentos de la organización, como el de ventas y el de marketing, en donde se presenta el *reporting* correspondiente a la situación y evolución de los productos, la marca y otros datos internos de la compañía.

Actualmente, el departamento de *Marketing Online*, es el departamento que más datos hace llegar al seno de la organización a través de las herramientas y técnicas de *marketing online* de las que dispone: *SEO* (*Search Engine Optimization*), *SEM* (*Search Engine Marketing*), *E-mail marketing*, agregadores de productos y publicaciones en medios de terceros.

Además de las técnicas de *marketing online* anteriores, la organización utiliza técnicas de **Marketing Analytics**, que, en base a todos los datos disponibles, permiten evaluar diferentes estrategias y poder tomar así las decisiones más beneficiosas para el negocio.

Con este fin, se definen y se analizan las **métricas** más importantes, y en base a los datos que se disponen, se aplicarán técnicas de **análisis predictivo** para la elaboración de modelos analíticos que permitan analizar los posibles escenarios futuros.

En este sentido, el uso de herramientas de medición y de visualización geográficas pueden ayudarnos a realizar análisis precisos y captar, por ejemplo, correlaciones existentes entre las diferentes campañas de marketing y la distribución de las ventas de nuestros productos por todo el territorio, teniendo en cuenta variables de análisis como la edad, genero, ubicación de los clientes, etc.

Por otro lado, la organización utiliza las redes sociales para informar a los clientes sobre la aparición de nuevos productos y también para recibir el grado de satisfacción de los clientes con los productos de la compañía.

Sin embargo, el uso que la multinacional hace de las redes sociales no va mas allá, y en este sentido, durante la última reunión del comité de dirección de la compañía, se nos informa de la necesidad de potenciar dicho uso para obtener ventaja competitiva respecto al resto de empresas del sector.

El principal objetivo es detectar a personas, con elevada influencia e impacto en las redes sociales, cuyas opiniones a cerca de nuestros productos nos permitan captar nuevos clientes e incluso nuevos mercados. Estas personas se considerarán **embajadores de la marca** en las redes sociales.

En consecuencia, se trata de un proyecto de **social analytics** que la multinacional quiere llevar a cabo en todos los países donde tiene representación, con el fin de difundir los valores de la compañía, aumentar el numero de clientes de su público objetivo, incrementar ventas y servicios en los mercados actuales y aumentar la presencia de la marca online.

Al tratarse de un proyecto de carácter global, los datos y análisis obtenidos deberán ser compartidos por todas las delegaciones.

En este sentido la compañía es consciente de los problemas de los **sistemas de computación actuales**:

- No son lo suficientemente rápidos, para capturar y almacenar esta información.
- No pueden alojar el volumen de datos relacionados con los productos de la organización que a diario se generan en las distintas redes sociales.

- No pueden gestionar las múltiples fuentes de procedencia y la heterogeneidad de la información (mensajes y fotos en Twitter, Facebook e Instagram, videos en YouTube etc....).

Pregunta 1 (20% puntuación)

Enunciado

1. Definid los conceptos de Business Intelligence (BI), Data Science y Big Data.
2. Contextualizad estos tres conceptos en el caso de uso descrito en el enunciado anterior, argumentando en qué momento entrarían en juego los diferentes conceptos, indicando la conveniencia o necesidad de aplicación para el caso que se plantea.

Criterios de evaluación

- Comprender correctamente las diferencias entre Big Data, BI y Data Science.

Pregunta 2 (25 % puntuación)

Enunciado

Como responsable del área de gobernanza de los datos y en particular de la función de gestionar los datos maestros (MDM):

- ¿Qué **etapas o fases** debería contemplar un programa de gestión de datos maestros en una organización como la que se describe en el enunciado? ¿Cuál es el objetivo de cada una de ellas en el contexto de la organización descrita?
- ¿Qué **modelo de madurez** en la gestión de datos maestros, crees que debe tener una organización como la del caso de estudio? ¿Que capacidades crees que debería tener?
- ¿Qué impacto (ventajas/inconvenientes) crees que puede tener el proyecto de detección de embajadores en las redes sociales en el MDM de nuestra organización.

Criterios de evaluación

- Entender las fases de un programa de gestión de datos maestros y la importancia para los sistemas de inteligencia de negocio.
- Conocer los distintos modelos de madurez en relación a los datos maestros.
- Contextualizar las argumentaciones en el caso de estudio.

Pregunta 3 (30 % puntuación)

Enunciado

Como responsable de la gestión de los datos, eres conocedor de las principales métricas o **KPIs** que gestiona la organización.

Supón que has de supervisar los datos relacionados con un proyecto de **Marketing Analytics** como el descrito en el apartado anterior, y que además intervinieran los datos procedentes de las redes sociales (opiniones de producto por zona/edad/genero, tendencias por país, competidores principales etc...) obtenidos en el proyecto de **Social Analytics** descrito en el enunciado.

Expón las diferentes fases del ciclo de los datos, destacando los aspectos que tienen que cumplir cada una de las diferentes fases, de manera que se genere valor con los datos y se pueda dar respuestas al tipo de preguntas que se plantean en el escenario descrito.

Nota: Podéis establecer y argumentar las hipótesis relacionadas con el caso de estudio que creáis oportunas.

Criterios de evaluación

- Entender y calcular al valor del dato como activo y los métodos de cálculo.
- Conocer el ciclo de vida del dato, definiendo y diferenciando cada fase.
- Contextualizar las argumentaciones en el caso de estudio.

Pregunta 4 (25% puntuación)

Enunciado

Continuando con el caso de uso responded:

1. Enumerar y argumentar los diferentes perfiles y roles que tendrían que participar en el proyecto.
2. ¿Qué cualidades relacionadas con la actitud, hábitos y conocimiento han de cumplir los directivos y usuarios en una organización orientada a los datos como la del ejemplo?
3. ¿Cómo crees que deberían ser las relaciones entre los diferentes perfiles?
4. ¿Qué factores diferenciales son comunes en este tipo de organizaciones orientadas a los datos?

Nota: Podéis establecer y argumentar las hipótesis relacionadas con el caso de estudio que creáis oportunas.

Criterios de evaluación

- Comprender las principales cualidades de las organizaciones orientadas a los datos.
- Comprender la cultura analítica de las organizaciones orientadas a los datos.
- Contextualizar las argumentaciones en el caso de estudio.