**Tipología y ciclo de vida de los datos**

**PRA1: ¿Cómo podemos capturar los datos de la web?**

**Miguel Ángel Quesada Fernández**

**Criterios de evaluación generales de la PEC.**

* Demostrar que se han asimilado correctamente los contenidos teóricos y que se aplican con precisión.
* Razonar todas las respuestas de forma analítica y sintética.
* Estructurar y maquetar correctamente el documento.
* Esquematizar las respuestas mediante el uso de gráficos, infografías y diagramas explicativos.
* Utilizar con precisión las citas y las referencias bibliográficas.
* Ajustar la extensión del documento a un máximo de 20 páginas.

**Contexto: *La revolución de Twitter***

1. **Pregunta 1 (10% puntuación)**
   1. **En la reestructuración de la plantilla que plantea Elon Musk.**

* **¿Crees necesaria la presencia de Científicos de Datos?**
* **¿Qué tareas realizarían?**

En primer lugar, antes de determinar la posible necesidad de un perfil de científico de datos, se debe definir qué es. El científico de datos hace uso de sus conocimientos en matemáticas, ciencias de la computación y del negocio en cuestión, con el fin de extraer conocimiento y comprender patrones de comportamiento a partir de conjuntos de datos. A diferencia del analista de datos, el científico construye nuevos modelos de aprendizaje y busca conocimiento sobre indicadores que la organización aún no sabe que son importantes para su negocio.

Explicado el rol del científico de datos, y su función, podemos enlazarlo con las necesidades que requiere Twitter tras la adquisición de Elon Musk, y de qué manera un científico de datos puede ayudar a cumplir esas metas, siendo un papel clave en la reestructuración de la plantilla que plantea Elon Musk. Atendiendo a Ramón Armero en su artículo *“Los 5 grupos de interés a los que Twitter y Elon Musk deben hacer frente en 2023” (*Bussiness Insider, 2023), describe los 5 retos principales a los que tiene que hacer frente Twitter tras su compra por Elon Musk:

1. **Los usuarios: aumentar su cifra y conseguir que se queden**

Los usuarios son la razón de ser de una red social:

* Son los generados de contenido, una red social se alimenta de los contenidos que publican sus usuarios. Además, los usuarios no sólo enriquecen la plataforma con su contenido propio, sino que a través de las distintas reacciones con otros usuarios (likes, retwettes) crean interacciones enriquecedoras.
* Son los principales anunciantes de la plataforma, la compartición de tweets en otros canales hace que otros usuarios accedan / interactúen con la red social, creando tráfico.
* Son los principales consumidores de los anunciantes de la plataforma, finalmente, como en todo producto digital, la existencia de anunciantes se ve justificada con la existencia de usuarios que consumen una publicidad con la que puede financiar los gastos de la red social.

Por todos estos motivos, aumentar la cifra de usuarios y mantener a dichos usuarios, especialmente tras las últimas controversias en Twitter tras la llegada de Elon Musk, es uno de los principales retos a los que se debe centrar la compañía.

Uno de los usos de la ciencia de datos definidos en *“Los roles, ámbitos y nombres de la ciencia de datos”* (Mora, 2019) es la retención de clientes. En el caso que nos ocupa las tareas que podrían realizar los científicos de datos podrían ser:

* Comprender a los usuarios existentes: analizar los usuarios de la plataforma, comprender sus patrones de uso y cuáles son los temas de interes que les motivan a utilizar la plataforma.
* Conocer las demandas no satisfechas por los usuarios para comprender las áreas en las que la red social debe mejorar.
* Recopilar feedback de la competencia, qué motiva que los usuarios abandonen Twitter y vayan a la competencia.

1. **Los trabajadores: contratar nuevos perfiles y cuidar a los que hayan decidido quedarse**

Atendiendo al artículo *(*Bussiness Insider, 2023), se estima que a principios de 2023 quedaban unos 1800 de los 7500 trabajadores que tenía la organización. Una disminución de este nivel debe venir acompañado por un plan de mejora de productividad, en este caso el científico de datos puede jugar un papel clave en la automatización de tareas y consiguiente ahorro de trabajo humano.

Las tareas que se podrían realizar en esta área podrían ayudar a sufragar la pérdida de capital humano, usando la tecnología para realizar aquellas tareas que conlleven esfuerzo y que puedan ser realizadas por máquinas:

* La utilización de análisis textual o de sentimiento, para identificar posibles infracciones de los términos de servicio de la plataforma.
* La identificación de las necesidades más inmediatas de desarrollo que se deben acometer en base al feedback de los usuarios y obtener el mayor rendimiento posible.

1. **Los inversores y los anunciantes: recuperar la estabilidad y demostrar confianza**

El ámbito del marketing es uno de los principales puntos fuertes dónde los científicos de datos pueden aportar una ventaja competitiva. En esta área, los científicos de datos podrían realizar las siguientes tareas:

* Segmentación: aprovechar el estudio de la información puede ayudar a segmentar demográficamente y los targets de audiencia de los que se obtiene el máximo retorno posible.
* Análisis de campañas: recoger información sobre el rendimiento y la eficacia de las estrategias, estas tareas pueden mejorar la estabilidad y confianza con los anunciantes si se realizan con transparencia, garantizando un acceso a información clara y precisa sobre el rendimiento de las campañas en las que han invertido dinero.

1. **La competencia: ofrecer innovación y dibujar una hoja de ruta con propósito**

Como se ha trasladado en los retos anteriores, el análisis del feedback de los usuarios, para conocer las carencias que observan en la red social es una tarea que puede realizar en científico de datos. El primer punto para hacer frente a la competencia es conocer que ven tus usuarios en ellos.

Los científicos de datos pueden aportar un alto valor mediante la clasificación de tweets y el análisis de sentimiento.

1. **Los organismos reguladores: cumplir con la legislación vigente sin caer en la censura**

Finalmente, el último reto plantea cuestiones éticas en relación a la libertad de expresión. La existencia de analizadores automáticos de contenido, que permitan cumplir con la legislación y eviten cualquier posible sesgo es uno de los puntos que promtía Elon Musk.

1. **Pregunta 2 (25% puntuación)**

**Para concretar la configuración de los equipos:**

* 1. **Define los perfiles involucrados**

En la configuración de los equipos que se deberían formar o reforzar tras la reestructuración de Twitter estarían los siguientes perfiles:

* **Ingeniero de datos (data engineering):** Es un perfil encargado de preparar la infraestructura necesaria para la gestión, almacenamiento y procesamiento de los datos en Twitter garantizando la disponibilidad, integridad y confidencialidad de los datos. Teniendo en cuenta el volumen de datos que mueve Twitter, uno de los requisitos que se espera es que el tratamiento de los datos se realice de forma eficiente y escalable.

Los ingenieros de datos también pueden encargarse de la homogenización de los datos procedentes de distintas fuentes. No obstante, en una empresa tan grande como Twitter, se espera que este perfil de limpieza de datos esté altamente especializado en esta tarea, siendo independiente del ingeniero de datos.

Un ingeniero de datos en Twitter trabajaría con tecnologías orientadas al big data como Hadoop o Spark, así como bases de datos no tradicionales, en el Twitter se dispone de tecnologías propias como Manhattan.

Es un perfil que debe seguir estando en el nuevo Twitter tras la compra de Elon Musk, ya que es los datos son la pieza clave del negocio de Twitter, y es imprescindible una infraestructura adecuada que los sustente.

* **Científico de datos**: Es un perfil altamente capacitado en matemáticas, estadística y tecnología de la información. Su trabajo se enfoca en extraer conocimiento de grandes conjuntos de datos mediante técnicas de minería de datos, aprendizaje automático, inteligencia artificial y análisis estadístico.

El científico de datos debe ser capaz de identificar qué preguntas deben ser formuladas para lograr los objetivos de la empresa y encontrar soluciones a través del análisis de los datos disponibles. El científico de datos debe trabajar en conjunto con los ingenieros de datos, quienes son los responsables de recolectar, procesar y almacenar los datos, para poder obtener resultados precisos.

Hay que resaltar que el trabajo de un científico de datos no termina con el análisis de datos, sino que debe ser capaz de validar los resultados obtenidos y garantizar que las conclusiones obtenidas sean precias y útiles.

Un científico de datos en Twitter tendría que, Para realizar su trabajo, el científico de datos utiliza herramientas como Python o R, que son lenguajes de programación especializados en análisis de datos.

* **Analista de datos:** Es un perfil especializado en recopilar, procesar y analizar grandes cantidades de datos, con el objetivo de obtener información útil y relevante que pueda ser utilizada para la toma de decisiones en una empresa u organización. Una de las principales tareas de un analista de datos es generar informes y visualizaciones que presenten los resultados obtenidos de una manera clara y fácilmente comprensible para el equipo de trabajo y los líderes de la empresa. En equipos grandes, como puede ser el caso de Twitter se puede plantear la necesidad de disponer del perfil de experto en visualización.
* **Expertos de visualización:** En organizaciones grandes, es conveniente la presencia de expertos en visualización. La visualización se ha erigido como una disciplina en si misma, dada la complejidad de comunicar de forma eficaz las conclusiones obtenidas por los científicos y analistas de datos.

Estos expertos hacen uso de herramientas que permitan generar infografías, así como visualizaciones dinámicas. Algunos ejemplos de herramientas serían Tableau o Qlik

* 1. **¿Qué niveles de competencia analítica deben tener cada uno de ellos?**

En “*Organizaciones orientadas al dato*” (Curto, 2019) se describe los distintos niveles de competencias analíticas que pueden darse. En el caso de Twitter pueden verse ejemplos de uso de los distintos niveles de análisis:

* **Análisis descriptivo:** la organización es capaz de entender qué pasó en las interacciones de los usuarios. En el caso de estudio de Twitter podríamos querer responder a la siguiente pregunta:
  + ¿Cómo una cuenta ha llegado a ser tan influente?
* **Análisis de diagnóstico:** la organización es capaz de entender las razones por las que las interacciones con los usuarios suceden. Por ejemplo:
  + ¿Por qué una determinada etiqueta es tendencia?
  + ¿Por qué se ha producido una disminución en la cantidad de Tweets en una determinada región?
* **Análisis predictivo:** la organización es capaz de predecir ciertas interacciones de los usuarios. Por ejemplo:
  + Predecir el nivel de engagement que tendrá una futura cuenta o hashtag en particular.
  + Predicción de tendencias.
  + Predicción de influencia de una cuenta sobre otros usuarios de la plataforma.
* **Análisis prescriptivo:** la organización es capaz de tomar decisiones vinculadas con las interacciones de los usuarios basadas en escenarios. Por ejemplo:
  + Recomendaciones de cuentas, tweets, y en general contenido que pueda ser relevante en función del comportamiento y preferencias de un usuario.
* **Análisis preventivo:** la organización es capaz de actuar con antelación a las necesidades de los usuarios. Por ejemplo:
  + Detección temprana de crisis de imagen en una marca

El científico de dato debe ser capaz de crear modelos de análisis que puedan dar respuesta a las preguntas planteadas anteriormente.

* 1. **¿Qué buenas prácticas deberían realizar como trabajadores analíticos?**

Los trabajadores analíticos deben observar una serie de buenas prácticas, con carácter transversal, en un proyecto se debe:

* Desarrollar un buen conocimiento del negocio y del proceso que se quiere analizar.
* Comprender el modo de pensar de los usuarios de negocio y cómo los resultados del análisis pueden influir en su toma de decisiones.
* Establecer una relación de trabajo colaborativa con las personas clave del proyecto.
* Usar el lenguaje de negocio para comunicar los beneficios y mejoras del proyecto.

De forma más específica, atendiendo a la fase del proyecto en la que nos encontremos se podrían considerar otras buenas prácticas que deberían ser realizadas por los trabajadores analíticos. En la fase inicial de definición del problema se deben considerar las siguientes buenas prácticas:

* Al recopilar las necesidades para un análisis, es importante identificar las decisiones que los stakeholders quieren tomar. Por ejemplo, si una empresa quiere lanzar un nuevo producto en el mercado o el propio Twitter quiere lanzar una nueva funcionalidad o promoción, los stakeholders podrían querer saber qué opinan los usuarios de Twitter sobre el producto, cómo reaccionan a las promociones de la empresa y si están dispuestos a comprarlo.
* También es importante acotar el alcance del problema a analizar mediante la identificación de hipótesis que puedan ser validadas. En el ejemplo que nos ocupa, el alcance del problema se podría acotar a los tweets que mencionan el nuevo producto de la empresa y su reacción a las promociones. Esto ayudaría a evitar la definición de objetivos no medibles y a enfocar el análisis en una pregunta específica.
* También es crucial identificar objetivos específicos en lugar de objetivos amplios para evitar diferentes análisis potenciales con resultados incompatibles. Por ejemplo, el objetivo específico podría ser medir el sentimiento de los usuarios de Twitter hacia el nuevo producto de la empresa y determinar si es positivo o negativo. Con este objetivo específico, se podrían identificar las palabras clave que se usan con mayor frecuencia en los tweets relacionados con el producto y analizar cómo se usan en el contexto de los tweets. Esto ayudaría a obtener una visión más clara de la opinión de los usuarios de Twitter sobre el producto.

Finalmente, en el momento de comunicación de los resultados, es buena práctica contar la historia a partir de los datos, y teniendo siempre en cuenta el lenguaje de negocio.

1. **Pregunta 3 (20% puntuación)**
   1. **Describe los distintos datos que utiliza Twitter y clasifícalos según sean Simples, Estructurados, Semiestructurados o No Estructurados**

En el ecosistema de Twitter hay una serie de conceptos fundamentales para su funcionamiento, estos conceptos están erigidos en base a unos datos que se describen a continuación:

* **Tweet:** Un tweet es la unidad mínima de comunicación en el ecosistema de Twitter, cuando nos referimos a un tweet no sólo estamos hablando del mensaje textual que se transmite, sino del conjunto de metadatos que envuelven al tweet. El tweet constituye un dato de tipo estructurado, compuesto de los siguientes datos que se enumeran (sin pretender ser exhaustivo), y que de forma individualiza podrían ser considerados como datos simples en su gran mayoría:
  + **Texto**, de forma individual, el texto de un tweet es un dato simple de tipo cadena de caracteres. Los tweets pueden contener otro tipo de datos adicionales al propio texto como imágenes o videos, que constituirán en sí mismo un dato compuesto.
  + **Fecha y hora de publicación,** considerado como un timestamp, es un dato simple.
  + **Lugar de publicación,** las coordenadas de publicación (longitud, latitud) o la ubicación como tal (cadena de caracteres), sería un dato simple.
  + **Nombre del usuario autor del tweet**, hace referencia a un dato simple, del tipo cadena de caracteres o un identificador interno numérico.
  + **Número de retweets, me gustas o de comentarios,** valor numérico simple.
  + **Listado de usuarios que han retweeteado, gustado,** dato estructurado compuesto de identificadores de usuarios.
  + **Listado de respuestas a un tweet,** dato estructurado compuesto por los identificadores de los tweets respuesta a un tweet.
  + **Menciones**, menciones asociadas al tweet, cada mención sería considerado como un dato simple, ya que se asocia a un identificador de un usuario de la red social.
  + **Hashtag,** sería considera como una cadena de caracteres.
* **Perfil de usuario:** Otro concepto fundamental en Twitter, y en el resto de redes sociales, sería el de perfil de un usuario. Al igual que un tweet, un perfil constituye un dato estructurado, que está conformado por una serie de datos que en si mismos de forma individualizada podrían ser considerados en su mayoría datos simples. Sin ser exhaustivos, un perfil de usuario en Twitter estaría conformado por los siguientes elementos:
  + **Nombre de la cuenta**, dato simple con formato de cadena de caracteres.
  + **Nombre del usuario completo**, dato simple con formato de cadena de caracteres.
  + **Foto del perfil**, dato compuesto.
  + **Biografía** **del perfil**, cadena de caracteres.
  + **Ubicación**, cadena de caracteres, ejemplo “Madrid”.
  + **Listado de cuentas seguidas y de seguidores**, dato compuesto constituido por la relación de identificadores asociados a las respectivas cuentas seguidas o de seguidores.
  + **Listado** **de tweets publicados, retweeteados, y reaccionados con me gustas**, dato compuesto constituido por la relación de identificadores de tweets referidos.
  1. **Describe el Ciclo de Vida de los datos detectados en el punto anterior**

Atendiendo a *“Fundamentos de data science”* (Minguillón), en la literatura tradicional de la gestión de datos, los datos son generados, almacenados, preprocesados, analizados, visualizados y publicados. Cada una de estas fases tiene un objetivo, generando valor a partir de los datos disponibles en ellas, pero no todas las fases son necesarias, e incluso algunas pueden llegar a solaparse.

Podemos identificar las siguientes fases en el ciclo de vida de un Tweet o la creación de un perfil en Twitter:

* **Captura:** El dato comienza con la creación de un tweet o la creación de una cuenta en Twitter por parte de usuario. En este sentido, un usuario puede crear un tweet a través de distintas vías: haciendo uso de la web de Twitter, la app de Twitter, o de plataformas de terceros que se comunican con Twitter a través de APIs para la creación de contenido. La creación de una cuenta se limita a las plataformas oficiales de Twitter.

La información de un dato de tweet es obtenida directamente a partir del propio usuario que introduce la información (texto del tweet, menciones…), y también de forma indirecta, el sistema captura metadatos como la ubicación, fecha y hora…

En esta fase de captura se aplicarán una serie de reglas de calidad para garantizar que la información entre en el sistema respetando una serie de condiciones.

Algunos ejemplos de validaciones serían:

* + - Validar el tamaño máximo de un tweet (ampliado a 280 caracteres en 2017)
    - En el caso de creación de una cuenta, que el nombre de la cuenta no exista previamente
    - E incluso, se podría plantear una primera fase pequeña de análisis del contenido del tweet para identificar posibles incumplimientos, se podría realizar un análisis del texto para descartar tweets que contengan determinadas palabras vetadas, etc.
* **Almacenamiento:** Los datos capturados en el punto anterior, son almacenados en un formato que permita su posterior explotación. En este sentido, a pesar de no tener el detalle exacto sobre las tecnologías internas que se utilizan en Twitter, podemos presuponer el uso de varios sistemas de almacenamiento:
  + Por un lado, **un sistema de almacenamiento orientado al ámbito operativo,** Twitter utiliza el sistema de almacenamiento Manhattan, una base de datos NoSQL de clave valor que ha sido desarrollada internamente por Twitter, y que es utilizada para almacenar Tweets y datos de la plataforma, con una alta velocidad y disponibilidad (Twitter, 2014). También se hace uso de la tecnología Blobstore [(Twitter,](https://blog.twitter.com/engineering/en_us/a/2012/blobstore-twitter-s-in-house-photo-storage-system) 2012), una tecnología de almacenamiento de objetos en la nube proporcionado por Google Cloud Platform, y que proporciona una solución de almacenamiento escalable para el almacenamiento de archivos y videos multimedia, como imágenes y videos.
  + Por otra parte, **un sistema de análisis de información basado en warehouse,** que son un tipo de bases de datos orientadas a almacenar datos no normalizados y que están orientadas a optimizar consultas y generar informes y visualizaciones de datos que puedan responder a las preguntas de los analistas.

En esta fase de almacenamiento intervendrían de forma más directa los Ingenieros de datos.

* **Preprocesado:** El objeto de esta etapa es el de preparar los datos para su posterior tratamiento y que puedan ser utilizado por los científicos y analistas de datos. Esta fase estaría ligadamente con la fase justamente anterior, ya que se extrae información del ámbito operativo, y se realizan una serie de transformaciones para finalmente almacenar los datos en sistemas de almacenamiento que permitan un análisis de la información.

Algunos de los elementos que se podrían realizar en Twitter podrían ser los siguientes:

* + - Identificación de las palabras claves de un tweet: Uno de los elementos clave de la plataforma Twitter es la identificación de tendencias, estas tendencias se obtienen en base a los elementos más discutidos en un corto lapso de tiempo. Para la identificación de tendencias el primer paso es la identificación de las palabras constituyentes de un tweet. Nótese que este preprocesamiento está muy orientado a ser utilizado también en el ámbito operacional y tiene una fuerte restricción temporal, es decir se requiere que se realice de forma continua online, y no puede esperar a un procesamiento en batch.

En esta fase intervendría el ingeniero de datos, aunque también pueden participar el científico de datos.

* **Análisis:** El objetivo de esta etapa es el crear modelos que puedan explicar como son los datos, y sean capaces de extraer conocimiento, respondiendo a las preguntas planteadas por los analistas. En el primer punto de este trabajo se han identificado distintos retos que presenta Twitter con la compra de Elon Musk, estos retos pueden ser alcanzados a través del análisis de la información generada por los usuarios: tweets publicados, reacciones, perfiles de usuarios y sus interconexiones.

Algunos de los análisis que se podrían plantear sobre los Tweets serían los siguientes:

* + - Identificación de temas, Twitter puede hacer uso de técnicas de procesamiento de lenguaje natural que permita capturar el tema del tweet, mejorando la calidad y precisión de los análisis posteriores. Uno de los retos que se ponían de manifiesto era la captura de feedback de los usuarios de Twitter.
    - Análisis de sentimiento, complementando el punto anterior, el análisis de sentimiento puede clasificar los tweets como positivos, negativos o neutrales. Esto ayuda a Twitter a comprender como percibe el usuario la plataforma, o ciertos temas y productos.

En esta fase trabaja principalmente el científico y el analista de datos.

* **Visualización:** El objetivo de la fase final es el de transmitir de forma visual y eficiente los datos obtenidos en los diferentes análisis realizados y que puedan servir de ayuda a la toma de decisiones.
* **Publicación:** Finalmente, y cerrando el circulo, el resultado de las fases anteriores puede ser publicado con distintos fines.

Por ejemplo:

* + - El análisis de los resultados de las campañas de marketing contratadas por los anunciantes para dar transparencia.
    - Por otro lado, destaca la publicación de los Tweets y el contenido de Twitter a través de APIS, que permita el desarrollo de herramientas de terceros que puedan acceder a los datos deseados.

1. **Pregunta 4 (20% puntuación)**
   1. **¿Twitter cumple con los criterios para ser considerada una Organización Orientada al Dato?**

La primera cuestión sería definir que es una Organización al Dato, esta definición no es simple, pero podría resumirse en considerar los datos como un activo estratégico que fundamente la toma de decisiones. Así, atendiendo a *“Organizaciones orientadas al dato”* (Curto), una organización puede considerar orientada al dato si cumple tres elementos fundamentales:

1. La organización debe ser capaz de identificar, combinar y gestionar múltiples fuentes de datos.
2. La organización debe ser capaz de construir modelos de análisis avanzado para predecir y optimizar sus resultados.
3. La organización debe poseer la capacidad de transformar la organización para que los datos y los modelos realmente deriven en mejores decisiones.

Si tenemos en cuenta la definición anterior, y consideramos el funcionamiento de Twitter podemos afirmar que es una organización orientada al dato.

* En primer lugar, la razón de ser de Twitter son los datos (principalmente Tweets) generados por los usuarios. A partir de estas unidades de información, y teniendo en cuenta los metadatos, la plataforma gestiona esta información para dar cumplimiento a los dos puntos siguientes. Además, Twitter puede combinar datos de múltiples fuentes de datos para crear un perfil de usuario más completo, por ejemplo, incorporando información demográfica.
* En segundo lugar, Twitter es una organización que construye modelos de datos avanzados para identificar patrones de uso y preferencias de uso. Algunos ejemplos ya han sido tratados en este trabajo, como el análisis de tendencias o análisis de sentimiento. Sin embargo, el ámbito de actuación de Twitter va mucho más allá, y uno de los más importantes, por constituir una de sus principales fuentes de ingresos es la publicidad, siendo uno de los principales retos a los que tiene que enfrentarse Twitter recuperar la estabilidad y demostrar confianza a los inversores y los anunciantes.

En este sentido, Twitter puede combinar distintas fuentes de datos para ofrecer los anuncios pertinentes al mejor target de usuarios, combinando la información de la que ya disponen junto con información demográfica, creando segmentos de audiencia específicos según que producto.

* En tercer lugar, Twitter tiene la capacidad de usar los datos obtenidos mediante los análisis del punto anterior, para mejorar no solamente la toma de decisiones de la propia plataforma sino también de sus anunciantes.

Un ejemplo de mejora en la toma de decisiones lo podemos encontrar con los datos de uso y rendimiento que son capturados por la plataforma, datos de gran importancia a nivel operativo para la identificación de problemas técnicos. Como se tratado anteriormente, la captura del feedback de los usuarios sobre las distintas funciona-lidades de la plataforma constituye un punto de partida a la hora de determinar posibles mejoras en la plataforma.

Finalmente, en relación con los anunciantes, Twitter es uno de los mayores sitios de opinión que podemos encontrar en internet y, por tanto, un sitio del que poder extraer la opinión de los clientes en base a un determinado producto o servicio, ofreciendo información útil a los anunciantes sobre qué funciona y que no funciona en sus productos.

* 1. **Analiza el modelo de madurez DELTA de la red social**

DELTA es un framework para la evaluación de la madurez de una determinada organización en base a su capacidad para implementar y utilizar tecnologías de gestión y análisis de datos. *“Organizaciones orientadas al dato”* (Curto). Este modelo se erige en base a cinco factores críticos:

* Datos, como punto de partida
* Organización, ámbito de aplicación de los procesos analíticos
* Liderazgo, para fomentar el uso de los procesos analíticos
* Objetivos, con impacto a nivel tanto estratégico, táctico como operativo
* Analistas, con habilidades necesarias para alcanzar los objetivos.

Cada uno de estos factores críticos puede estar en un nivel de madurez distinto.

Teniendo en cuenta que no se puede tener un conocimiento completo y exacto del funcionamiento de la organización Twitter, se va a analizar cada uno de los puntos, intentando justificar la fase en la que se encontraría para cada uno de los factores de éxito.

* Datos: La plataforma está fundamenta en la recopilación y análisis de grandes cantidades de datos generados por los usuarios. En este sentido, Twitter ha demostrado una alta capacidad para recopilar y almacenar los datos, desarrollando las tecnologías necesarias para ello (Cassandra, Manhattan). Es una compañía que ha superado el simple uso de un Data Warehouse, y que tiene marcada una estrategia de mejora de la tecnología que de respuesta a su necesidades de datos.
* Organización: Twitter tiene una cultura de datos y de toma de decisiones basada en dichos datos, al menos, hasta la compra por parte de Elon Musk. El uso de procesos analíticos se extiende a lo largo de toda la organización, tanto para la mejora de la experiencia de los usuarios (tendencias, baneo automático) o la mejora de la propia plataforma (captura de feedback o la monitorización de incidencias).
* Liderazgo: El fomento de los procesos analíticos es contemplado desde sus lideres, que ven en los procesos analíticos una forma de ahorrar recursos humanos,
* Objetivos: Twitter ha establecido objetivos claros y medibles para su estrategia de analítica de datos. La compañía utiliza KPIs (Indicadores Clave de Desempeño) para medir el éxito en áreas como la interacción del usuario, la tasa de retención y la monetización. Además, Twitter utiliza una amplia variedad de herramientas de análisis de datos para monitorear y medir su rendimiento.
* Analistas: Twitter ha contratado a algunos de los mejores talentos en el área de analítica de datos, incluyendo científicos de datos y analistas. La compañía ha establecido un fuerte enfoque en el desarrollo de habilidades en análisis de datos y ha proporcionado oportunidades de capacitación y desarrollo para sus empleados.

En base a la anterior argumentación, podría considerarse que Twitter está en la fase 5 de competición basada en analítica, que considera que la organización ha embebido la cultura analística en la organización y tiene articulada una estrategia completamente coordinada.

1. **Pregunta 5 (15% puntuación)**

Ponte en la piel de un miembro del equipo de visualización de datos

* 1. **¿Qué visualización propondrías sobre la información en Twitter para incrementar el “engagement” de los usuarios?**

El engagment de los usuarios en Twitter hace referencia a la interacción y participación que los usuarios tienen con la plataforma, teniendo en cuenta las funcionalidades de Twitter, esto incluiría acciones como publicar un Tweet, retwittear, dar un like, o seguir una cuenta.

Entonces, ¿cómo podría incrementarse la interacción y participación de los usuarios en la plataforma? Para que exista interacción y participación debe existir contenido que permita incentivar esa interacción.

Actualmente existen visualizaciones muy sencillas en Twitter que se basan en esta idea, la más conocida es el TOP10 de tendencias, que no deja de ser un listado con el TOP de palabras clave que da lugar a que los usuarios puedan participar en temas de actualidad.

Se propondría desarrollar una visualización que fuese la puerta de entrada de un usuario cuando accede a Twitter. Actualmente, al acceder a Twitter lo primero que se muestra es un listado de Tweets en tu home, que requiere que el usuario tenga que ir leyendo y analizando de uno en uno. Hoy en día, los usuarios buscan ingerir una gran cantidad de datos en un corto lapso de tiempo (TikTok), abandonando cada vez más el consumo de información ingerida lentamente (Youtube). La presencia de una visualización que haga las veces de resumen de lo que ha pasado en la cuenta del usuario desde su última visita ayudaría a que el usuario supiera de una forma más rápida de qué están hablando sus seguidores, sus seguidos, o incluso le recomendase según la hora y seguidores, de que tema podría hablar para generar impacto.

La visualización podría mostrar información como la siguiente:

* Visualización animada de nubes de palabras, que permitiría mostrar los temas más populares de los que están hablando tus seguidores y seguidos, filtrando por seguidores / seguidos / ciudad. Al presionar sobre los temas se mostraría el listado de tweets relacionados.
* Mapa de conversaciones, generar un mapa de red que permita visualizar quién esta hablando con quién y de qué. Se podría incorporar además información de tipo geográfica, para conocer desde que puntos geográficos están participando en las conversaciones. Este tipo de visualización podrían fomentar.
* Nube de palabras de temáticas qué según la información de tus seguidores, ubicación y gustos, podría tener un alto impacto.

1. **Pregunta 6 (10% puntuación)**

Destaca dos de las intervenciones del primer debate.

* 1. **¿Qué resaltarías de cada una de ellas? ¿Hay algún aspecto en el que no coincidas?**

Me gustaría destacar dos intervenciones:

* Intervención de Francisco Alamilla Martínez, en su intervención del día 12 de Marzo titulada **“¿Qué conocimientos y habilidades crees que deben tener?”.** 
  + **Resumen:** En esta intervención se hace un repaso del diagrama de Venn con las habilidades que son esperables de un científico de datos, estableciéndose la figura del científico de datos como un perfil que debe tener un conocimiento del dominio, matemáticas y ciencias de la computación. Y la dificultad que entra que un perfil tenga un conocimiento difícil de estas tres áreas.
  + **A destacar:** Sobre esta intervención me gusta la observación realizada sobre como una disciplina como la ciencia de datos, que nace inicialmente con un dominio acotado y abarcable, ha ido evolucionando hasta convertirse en una disciplina difícilmente abarcable por una persona, y con un alto nivel de especialización.
  + **¿En que no coincido?** Comparto la opinión, en el sentido de que es necesario disponer siempre de personas que entiendan y ayuden a interpretar el negocio en cuestión con el que se está trabajando. Sin embargo, en mi opinión, otro elemento fundamental en el futuro más inmediato va a ser el de contar con perfiles muy especializados en determinadas tecnologías (GPT, generación de imágenes…), ya que la evolución en la ciencia de datos, e inteligencia artificial en general, va a ser tan grande que encontrar perfiles que sepan un poco de todo va a ser cada vez más complicado.
* Intervención de Antonio Santos Parrado, en su intervención del día 5 de Marzo titulada “**Efectos negativos que podrían derivarse del uso de redes sociales por parte de las empresas”** 
  + **Resumen:** La intervención destaca los riesgos asociados a la presencia de las empresas en las redes sociales, comentando tres situaciones que pueden tener consecuencias negativas, como son: la presencia de comentarios que pueden afectar a la reputación, los comunicados que pueden volverse en contra de la compañía y la dependencia de las redes sociales como media publicitario, lo que puede generar riesgos en caso de cambios en las normas o problemas técnicos.
  + **A destacar:** El autor pone de manifiesto que el uso de las redes sociales puede ser una herramienta muy valiosa para las empresas, pero pone de manifiesto la existencia de riesgos y la necesidad de tomar medidas preventivas para mitigar su impacto.
  + **¿En que no coincido?** Comparto la opinión y relación de riesgos que se ha comentado, y me gustaría que se hubiese tratado otros riesgos como:
    - **Riesgos de seguridad:** La presencia de la empresa en otra plataforma, es otro punto posible de hackeo y filtración de información sensible de la plataforma.
    - **Problemas de gestión de crisis:** La gestión de crisis se complica con la presencia de una marca en redes sociales, ya que amplia el radio de acción sobre el que tienen que actuar las empresas. Cabe destacar la campaña de descrito que ha habido los primeros meses en la empresa Mercadona a través de la redes sociales por su aumento de precios, esto ha provocado situaciones comprometidas a los gestores de las redes sociales ante la abalanza de comentarios negativos. The Objective (2023)

1. **Bibliografía**

**Curto Díaz, Josep**. *Organizaciones orientadas al dato*

**Minguillón, Julià.** *Fundamentos de data science*

**Mora Cantallops, Marçal.** *Los roles, ámbitos y nombres de la ciencia de datos*

**Bussiness Insider (2023)** Disponible a: <https://www.businessinsider.es/5-retos-twitter-elon-musk-deben-afrontar-2023-1168600> (Consultado 10 de marzo)

**Twitter (2014)** Disponible a: <https://blog.twitter.com/engineering/en_us/a/2014/manhattan-our-real-time-multi-tenant-distributed-database-for-twitter-scale> (Consultado 15 de marzo)

**Twitter (2012)** Disponible a: <https://blog.twitter.com/engineering/en_us/a/2012/blobstore-twitter-s-in-house-photo-storage-system> (Consultado 15 de marzo)

**The Objective (2023)** Disponible a: <https://theobjective.com/economia/2023-01-22/mercadona-subida-precios/> (Consultado 1 de abril)