

MAESTRIA EN ANALISIS Y VISUALIZACION DE DATOS MASIVOS

VISUALIZACION INTERACTIVA DE LA INFORMACION

Mtra. Elizabeth Martínez Ramírez

GUIA DE TRABAJO

Índice de la asignatura

- 2. Qué necesitas saber antes de empezar
- 3. Las fases
- 4. Elaboración
- 5. Tu profesor de la asignatura
- 6. Actividades y entrega
- 7. El vídeo de presentación
- 8. Calificación
- 8. Entrega extraordinaria
- 9. Referencias bibliográficas

3. Las fases

Para la realización de tu trabajo seguirás un flujo de trabajo organizado en varias fases. Estas fases o etapas están pensadas para **agrupar** diferentes actividades relacionadas entre sí, que te ayudarán a comprender el proceso general y a distribuir tu trabajo y esfuerzos:

- ▶ **Fase 1: cread vuestro equipo de trabajo y elegid la temática del proyecto de análisis de datos** que vayáis a abordar en el trabajo final, cada equipo tiene libertad para escoger el problema sobre el que hace el proyecto. El proyecto puede ser una invención del propio equipo, pero este debe estar bien fundamentada en una necesidad real. También es posible que el equipo realice el proyecto de la

mano con una empresa, pero en este caso será responsabilidad de los estudiantes dicha colaboración. La universidad no gestionará ni acordará ningún tipo de relación para tal fin.

- ▶ **Fase 2: realización del trabajo.** Se debe redactar cada una de sus partes e ir entregando en plazo cada una de sus evoluciones. **Hay cinco entregas:** la última es la presentación del trabajo final, y ya están programadas desde el inicio de la asignatura. Es la fase más prolongada y deberéis trabajar en equipo, enviando a vuestro profesor los avances que vayáis realizando.
- ▶ **Fase 3:** comprende la **evaluación final** de vuestra última entrega. Es el momento en el que el profesor, junto con todo el trabajo escrito, evaluará como un todo las entregas que definen el proyecto de análisis de datos llevado a cabo.

FASE 0 Y FASE 1

Secuencia de entregas

En las cinco entregas se deben ir entregando los siguientes aspectos de manera aproximada:

- ▶ **Entrega 1:**
 - Portada debidamente cumplimentada.
 - Borrador de resumen ejecutivo y palabras clave (en español y en inglés, máximo doscientas palabras y cinco palabras clave en cada idioma).
 - Introducción y objetivos.
 - Descripción de los datos.
 - Identificación empresa (si aplica).
 - Metodología.
- ▶ **Entrega 2.** Lo anterior más:
 - Análisis exploratorio de los datos EDA.
 - Modelo de datos.
- ▶ **Entrega 3.** Lo anterior más:
 - Evaluación de resultados.

FASE 0 Y FASE 1

- Discusión.
 - Conclusiones.
- **Entrega 4.** Lo anterior más:
- Prototipo (trabajo o desarrollo que reúne en grado máximo las características principales del proyecto de análisis de datos).
 - Recomendaciones y próximos pasos.
- **Entrega 5:**
- Trabajo completo ya finalizado para ser evaluado.

Recuerda que en cada actividad tienes el detalle de los subapartados que se recomiendan incluir en cada entrega.

4. Elaboración

Esta es la parte central del proceso y la más dilatada en el tiempo. En ella te ocuparás de trabajar sobre el tema propuesto con tus compañeros, realizando una contribución acorde con los objetivos de la titulación y la carga de créditos asociada a esta asignatura. Durante todo el período tendrás al profesor de la asignatura que revisará las entregas haciendo un seguimiento de vuestro progreso. El resultado de vuestro trabajo quedará plasmado en un documento escrito que será valorado por dicho profesor.

A continuación, tienes algunos ejemplos de trabajos que podrán servirte de guía para proponer tu proyecto.

Cómo afecta el clima al consumo eléctrico en Baleares

Este trabajo capturó datos de orígenes diferentes para utilizar técnicas predictivas y así realizar una estimación a corto plazo del consumo eléctrico en una zona geográfica concreta.

Procesado masivo y automático del formulario 13F de la SEC

Este trabajo desarrolló *software* para la automatización del proceso de Extracción, Transformación y Carga (ETL) de la información pública contenida en los formularios 13F de la SEC. El resultado final es un proceso automático que engloba todas las fases de construcción de una base de datos con la información contenida en el formulario.

Metodología para captura y análisis de datos en laboratorio de metrología

La idea central de esta metodología es dar la pauta para que se pueda implementar un modelo con el cual se puedan almacenar los datos leídos por los instrumentos de medida y, a partir de los datos almacenados, analizar y estimar los valores de interés tanto para una calibración como para estudios adicionales.

FASE 0 Y FASE 1

Líneas de trabajo

Recuerda que puedes elegir **una o varias** de las siguientes líneas de trabajo que estén relacionadas con los contenidos tratados en las diferentes asignaturas de la titulación. Piensa cuáles de los contenidos estudiados te interesa trabajar con mayor profundidad y son más apropiados para el tema de trabajo que te ocupa:

- ▶ **Línea 1.** Técnicas de captura y almacenamiento de datos.
- ▶ **Línea 2.** Técnicas estadísticas e inteligencia artificial y su aplicación para el análisis masivo de datos.
- ▶ **Línea 3.** Entornos de computación distribuida para el análisis masivo de datos o Big Data.
- ▶ **Línea 4.** Visualización de datos como técnica de análisis.
- ▶ **Línea 5.** Herramientas de apoyo a la toma de decisiones en el contexto empresarial o en otros escenarios.

En cuanto a **herramientas tecnológicas**, apóyate en productos Open Source o en plataformas como AWS, Google o Azure siempre que tengáis acceso a ellas. La gestión de estos recursos y su uso será responsabilidad del equipo de trabajo.

5. Tu profesor de la asignatura

Tu profesor, además de dar una serie de clases para todos los alumnos, será el encargado de corregir los envíos que vayáis haciendo. Entre sus **funciones** está:

- ▶ **Dirigirte y orientarte** durante la realización del trabajo. Apóyate en él para resolver las dudas que surjan sobre el contenido del trabajo. Ten en cuenta sus comentarios a los borradores, pues estarán siempre orientados a mejorar la calidad del trabajo y a preparar una evaluación positiva.
- ▶ Valorar el grado de **avance y supervisar** el proceso, velando por el cumplimiento de los objetivos. Será el encargado de corregir las distintas entregas que realices durante esta fase. Ten en cuenta que en esta asignatura las valoraciones de las entregas no tienen repercusión directa en la calificación final del trabajo, que dependerá de lo que se presente como trabajo final. Sin embargo, servirán para corregir y orientar tu progreso. Lee detenidamente lo que se dice sobre las entregas y su valoración en esta guía.
- ▶ Deberá valorar el **documento completo** que recoge los resultados de vuestro trabajo. Según su criterio y experiencia, dará la nota final de la asignatura entre trabajo (100 %: 30 % formatos y 70 % contenido).

[Sin título]

FASE 0 Y FASE 1

La comunicación con el profesor se realiza:

- ▶ **En el momento del envío de las entregas.** Tanto en el documento como en el propio envío puedes incluir consultas y otra información que el profesor podrá leer. También te ofrecerá retroalimentación en sus comentarios en la corrección de las entregas.
- ▶ **El correo interno** del aula virtual. Utiliza siempre el correo interno de la plataforma para ponerte en contacto con tu director sobre asuntos relacionados con el contenido del trabajo. No dudes en consultarle y avisarle de cualquier problema que encuentres durante el proceso. Para consultas de carácter administrativo, recuerda que siempre puedes recurrir a tu tutor.
- ▶ **Las sesiones en directo.** Tendrás varias y el calendario concreto de estas sesiones lo fijará tu profesor.

6. Actividades y entrega

En el aula virtual encontrarás la programación semanal, es decir, una planificación que marca diferentes hitos en el proceso de elaboración de tu trabajo. Revisa las fechas de entrega de estas actividades y tenlas presentes en tu planificación personal para asegurarte de que llegas a tiempo de enviar cada una de ellas. Es importante que tengas en cuenta los siguientes aspectos:

- ▶ **Temática:** el trabajo debe ser la realización de un proyecto de análisis de datos que siga las líneas de trabajo antes mencionadas. En la parte final se deberá hacer un prototipo que muestre los resultados del análisis y unas conclusiones que expliquen la efectividad de este.
- ▶ **Respetar los formatos y demás exigencias** puestas en el trabajo: el trabajo deberá comprender **entre 50 y 120 páginas**, sin contar portada, resumen e índices iniciales. El no cumplimiento de los formatos o demás exigencias, aunque el contenido sea correcto, **puede ser motivo de no superación**.

FASE 0 Y FASE 1

- ▶ **Número de entregas:** debes realizar cinco entregas que marcan distintos grados de avance incremental. Finalmente, la quinta entrega ya no admitirá modificaciones: contiene el documento tal cual quedará para su evaluación. No se admiten entregas adicionales, intermedias o extraordinarias y todas las entregas deben realizarse a través de las actividades correspondientes definidas en el campus virtual y deberán realizarse dentro del plazo temporal establecido.
- ▶ **Importancia de las entregas:** en caso de no realizar alguna entrega o que sea calificada como no apta por el director, es posible que tu trabajo no pueda ser aprobado en convocatoria ordinaria y debas trabajar más de cara a la convocatoria extraordinaria. Las entregas están pensadas para valorar tu grado de avance y compromiso con el trabajo, reflejando una dedicación acorde a la carga de créditos de la asignatura y la adquisición de las competencias requeridas. Si el grado de avance es insuficiente en el plazo ordinario, puede que el visto bueno para depositar no sea viable ni siquiera en el plazo extraordinario.

FASE 0 Y FASE 1

- ▶ **Originalidad de las entregas:** tu trabajo debe ser inédito y original. No se admiten contenidos previamente publicados o cuyo origen no esté adecuadamente mencionado, citado y referenciado. Tampoco pueden ser contenidos que no desarrolle el propio equipo. Cada entrega se somete a revisión mediante una herramienta que comprueba problemas de originalidad y sus resultados son verificados por tu profesor, que en caso de detectar algún problema te avisará de ello en la corrección de la entrega. El profesor únicamente informará del problema; en estos casos, es el estudiante el que debe aportar las evidencias necesarias para demostrar la autoría del trabajo.
- ▶ **Revisión de las entregas:** el profesor es el encargado de revisar cada entrega y corregirla, ofreciendo retroalimentación para afianzar o reorientar tu trabajo. Ten en cuenta sus comentarios y aplícalos siempre en la siguiente entrega.

FASE 0 Y FASE 1

Los **estudiantes de segunda matrícula** que tengan su trabajo comenzado en un período anterior deberán ajustarse a estas mismas entregas en cuanto a plazos y contenidos esperados. En ningún caso se modificarán las fechas ni el número de las entregas establecidas. Además, en determinados casos, el profesor podrá recomendar a estos estudiantes el cambio total del lugar sobre el que se realiza el sistema.

7. El vídeo de presentación

La última entrega se debe acompañar de un vídeo de entre quince a veinte minutos en total (no deberá ni bajar ni superar dicho tiempo) en la que todos los alumnos del grupo participen. Este se deberá acompañar de diapositivas o imágenes junto a la imagen del alumno que habla en ese momento. Todos los alumnos del grupo deberán aparecer en el vídeo, en tiempos semejantes entre cada uno.

FASE 0 Y FASE 1

El vídeo deberá contener los aspectos fundamentales del trabajo y exponer sucesivamente los puntos principales siguiendo el mismo hilo conductor. De esta manera se comenzará con la presentación, identificación del proyecto, de la problemática abordada, las fases del análisis, el prototipo construido y el resto de los puntos hasta el apartado conclusiones.

No se recomienda que sea un discurso leído de lo que aparezca en pantalla sin aportar nada más. No conviene llenar la pantalla con muchos datos o texto, es mejor que haya pocas cosas y que sean preferentemente visuales para poder ser bien entendido.

- ▶ Es importante que en todo momento muestres dominio del contenido dejando claro que la autoría del trabajo es tuya, y que lo conoces a la perfección.
- ▶ Demuestra capacidad de síntesis y no intentes resumir todo el contenido. Céntrate en los aspectos fundamentales y aquellos que más quieras reforzar, teniendo en cuenta el tiempo disponible.
- ▶ Recuerda que una parte de la rúbrica de evaluación se centra precisamente en esta parte del proceso. Ten presentes los aspectos que valorará el profesor.

FASE 0 Y FASE 1

Al tratarse de un vídeo de un trabajo académico se deberá atender:

- ▶ **Atuendo:** acude con una presencia e indumentaria acordes con la importancia del acto. Vístete de manera formal, aunque sencilla (tampoco es una boda) y evita en todo caso la ropa de sport o demasiado informal.
- ▶ **Lenguaje verbal:** emplea un lenguaje adecuado, evitan lo coloquial, pero sin caer en la pedantería.
- ▶ **Lenguaje corporal:** muéstrate relajado, pero seguro. Apóyate en la presentación para reforzar las ideas clave, pero intenta no mostrar que estás leyendo o que llevas un discurso aprendido de memoria. La presentación es útil como soporte para recordar la estructura de la información, pero se espera que tú elabores el discurso de manera natural, demostrando el dominio sobre el tema.
- ▶ **Contacto visual:** procura mantener el contacto visual con la cámara.
El vídeo debe llegar al menos al 50 % de puntuación (sacar al menos un cinco de diez) para que haga media ponderada con el trabajo escrito.

8. Calificación

El profesor **se apoyará en la rúbrica** de evaluación para unificar criterios con otros profesores y decidir la calificación que merece el trabajo. En dicha valoración tiene en cuenta **estructura (30 %) y contenidos (70 %)**.

Esta calificación se otorgará según una **escala numérica de 0 a 10**, con expresión de un decimal, a la que deberá añadirse su correspondiente calificación cualitativa:

- ▶ 0 – 4,9 Suspenso (SS).
- ▶ 5,0 – 6,9 Aprobado (AP).
- ▶ 7,0 – 8,9 Notable (NT).
- ▶ 9,0 – 10 Sobresaliente (SB).

En caso de que el trabajo sea propuesto para la mención de Matrícula de Honor por el profesor, la Comisión Académica del Título decidirá la asignación de dicha mención.

8. Entrega extraordinaria

El plazo extraordinario supone una **nueva oportunidad** para completar o terminar de mejorar algunos aspectos de tu trabajo. Es importante que tengas en cuenta que en este tiempo no se puede realizar un trabajo completo (esto no tendría sentido, según la carga de créditos de la asignatura), sino que es una **extensión de plazo** para finalizar aquellos aspectos que quedaron pendientes en el período ordinario y que impidieron a tu director dar el visto bueno al depósito.

Si el motivo de la no autorización en convocatoria ordinaria tiene que ver con la no originalidad del trabajo, ten en cuenta que deberás realizar un esfuerzo considerable para la modificación de este.

En convocatoria extraordinaria únicamente habrá **una entrega más** del trabajo, ya que no se trata de hacer el trabajo en dicho tiempo, sino de dar un poco más de tiempo a aquellos que no lo finalizaran en plazo o de que el director considerara que no estuviera correcto en ordinaria.

9. Referencias bibliográficas

Al final del trabajo incluye un listado con las referencias completas de las citas que has ido incorporando en el texto. Es importante que apoyes tus explicaciones con unas referencias sólidas para dar fuerza a tu exposición.

Te facilitarán las pautas bibliográficas de la normativa correspondiente a tu titulación **(APA)** para que las referencias estén completas y se ajusten a lo que necesitamos. Ten en cuenta que al final la bibliografía debe exponerse por **orden alfabético** (no en orden de aparición) y es recomendable exponerlo con sangría francesa para una buena visualización.

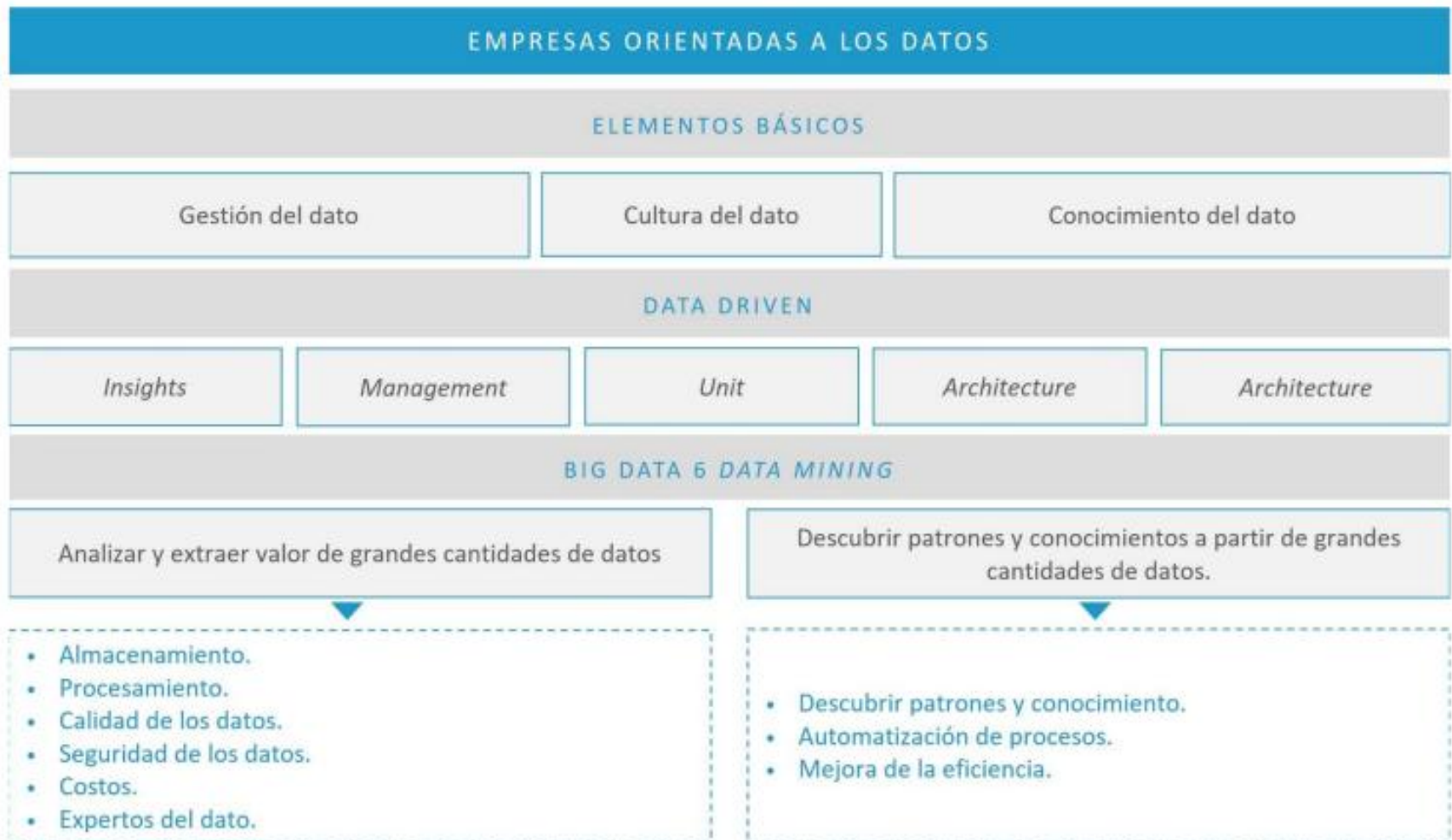
FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

Ideas clave	4
1.1. Introducción y objetivos	4
1.2. La importancia de los datos en el actual contexto económico y laboral	4
1.3. Propuesta diferencial de las empresas orientadas al dato (<i>data driven</i>)	6
1.4. Orientar una empresa al dato	10
1.5. Proyectos de <i>big data</i> y <i>data mining</i>	12
1.6. CRISP-DM como metodología de proyecto y modelo de procesos	16
1.7. La gestión de datos y la figura del CDO	19
1.8. Referencias bibliográficas	23

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato



1.1. Introducción y objetivos

En este tema se hará un repaso sobre aquellos conceptos relevantes que existen alrededor de las organizaciones orientadas a los datos. Al tratarse de un título enfocado al análisis de datos, el estudiante debe ser capaz de identificar aquellos aspectos diferenciadores de las empresas que basan sus **objetivos de negocio** en los datos y se aprovechan de estos para **mejorar** la atención al cliente y **potenciar** sus beneficios económicos.

Los objetivos que persigue este tema son:

- ▶ Comprender qué significa que una empresa está orientada al dato y cómo esto le ayuda a mejorar y personalizar la atención al cliente.
- ▶ Conocer algunos de los conceptos que son relevantes en las estrategias basadas en los datos.
- ▶ Entender la dinámica de las empresas que utilizan el dato no solo como su principal activo, sino también como fuente de información para mejorar sus procesos.
- ▶ Comprender el papel del analista de datos en la dinámica de las empresas orientadas a los datos.
- ▶ Dimensionar el alcance de un proyecto de análisis de datos teniendo en cuenta la estrategia de datos en las organizaciones.

1.2. La importancia de los datos en el actual contexto económico y laboral

Las empresas y todos sus servicios giran alrededor de los datos que estas generan o recogen de su actividad diaria. Para estas organizaciones es muy importante utilizar estos datos para **aprender y evolucionar** dentro de su propio contexto económico, y el fin de todo esto no es solo aumentar la eficiencia y eficacia de sus procesos, sino que también busca brindar los servicios que realmente **esperan** sus clientes.

No es un secreto que gracias a esta era tecnológica la recopilación de información para las empresas ha permitido que puedan ofrecer a sus clientes servicios mucho **más personalizados y cercanos** a lo que realmente quieren. Cuanta más disposición de los datos exista en una compañía, mayor libertad tendrá para **adaptar sus servicios y su negocio** a lo que verdaderamente importa: sus **clientes**.

El reto que tienen las organizaciones es aprender a gestionar **proactivamente** los datos y a utilizar la información como parte de su estrategia para beneficiarse de ellos. Con una buena estrategia ese beneficio puede ir desde **conseguir** ventajas competitivas en el mercado hasta contar con una estructura y unos procesos muy **eficientes**. Esto es posible porque la organización aprende que debe basar su cultura empresarial en el pensamiento **analítico** y en la explotación de este recurso desde todos los puntos de vista posibles.

Analizar los datos no es una tarea reciente, es algo que se viene haciendo desde hace muchos años conforme la disposición de los datos ha ido **creciendo**. El análisis de datos ha pasado por etapas que han ido en línea con lo que se ha podido hacer con ellos. En los sistemas de soporte a la decisión, por ejemplo, se tomaba la información y se miraba su utilidad a la hora de apoyar a la decisión en áreas clave para la compañía. Los sistemas de información ejecutiva eran similares, pero su ámbito era más estratégico dentro de la organización.

Llegó un momento en que no fue suficiente definir sistemas de información analíticos, sino que además fue necesario incorporar los **cuadros de mando integral** con el fin de hacer más accesible los beneficios de **«vigilar»** el comportamiento de

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

los datos y **observar** desde diferentes puntos de vista los análisis que se llevaban a cabo.

El concepto siguió creciendo y se pasó a definir así la inteligencia de negocio y el Business Analytics como ese punto en el que, con cierta madurez, la organización ya podía hablar de estrategia de negocio **basada en datos** en diferentes áreas en la organización. Ahora con el *big data* y la ciencia de los datos, el beneficio de todos estos avances alrededor del dato se ha visto mucho más impulsado, principalmente porque el dato crece de forma **exponencial** y porque hoy es posible recogerlos de donde antes era impensable.



Figura 1. Elementos básicos de las empresas orientadas al dato. Fuente: elaboración propia.

Es una realidad tangible que las empresas que gestionan proactivamente los datos se ven forzadas, de alguna forma, a cambiar drásticamente su **modelo de negocio**. Estas organizaciones actúan diferente y son capaces de reaccionar **anticipadamente** ante situaciones imprevistas. No solo se trata de un reto tecnológico relacionado con la capacidad de almacenar y procesar datos, sino que es algo que afecta a la **estructura** de la organización, a su estrategia, a esos procesos de trabajo críticos, a cómo usa y se apoya en la tecnología, a la forma en que se promueven y seleccionan a las personas, a cómo todo ello se involucra con la **cultura de la organización** en general.

1.3. Propuesta diferencial de las empresas orientadas al dato (*data driven*)

Para que las empresas puedan competir y mantenerse en esta era digital es indispensable promover su transformación hacia una cultura **orientada a los datos**. Cuando la organización logra aprovechar sus activos de datos, puede mejorar sus procesos de negocio y entender mejor a sus clientes. La premisa principal es que los datos serán un recurso prácticamente **ilimitado** que les permitirá continuar **evolucionando y adaptarse** a los cambios que propone el mercado diariamente.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

Para conseguir ese cambio, las empresas deben tomar decisiones alrededor de ese recurso ilimitado llamado «datos» y actuar en consonancia con una estrategia que garantice éxitos en sus pasos. La estrategia implica no solo un cambio de los procesos, sino que requiere una adopción de una cultura *data driven*.

La cultura *data driven* es un enfoque que usa los datos como la base principal para la **toma de decisiones** empresariales y se fundamenta en la **confianza** que se les puede tener y los beneficios que pueden salir del **análisis** de estos cuando sea necesario **apoyar** la decisión del rumbo que se debe tomar ante determinadas situaciones (Patil y Mason, 2015).

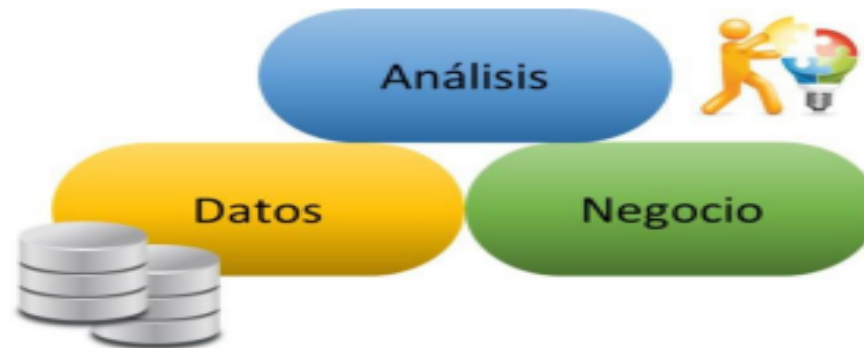


Figura 2. Elementos de una organización Data-Driven. Fuente: elaboración propia.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

Una organización será *data driven* cuando, basándose en los datos y en el análisis de la información generada, pueda tomar decisiones importantes. *Data driven* no es un concepto nuevo en el ámbito empresarial y tampoco es una forma de trabajo que solo esté en altas las áreas más estratégicas de la compañía. Su papel es **transversal** y, por ello, es muy común que se hable y aplique dicho concepto en **todos los niveles** de la organización.

El concepto *data driven*

Cuando se dice que una empresa es *data driven* se afirma, también, que dicha empresa puede y sabe **extraer conocimiento** de la gestión de sus datos para darle una **utilidad real**. Algunos de los elementos que diferencian el **potencial** de la empresa *data driven* son:

- ▶ *Data insights*: capacidad de la empresa para **identificar** los datos que necesita.
- ▶ *Data management*: la **gestión** de los datos a nivel de organización de la empresa.
- ▶ *Data unit*: promover una **unidad** dentro de la compañía que fomente y asegure la calidad de los datos.
- ▶ *Data architecture*: ante la necesidad de manejar los datos, es necesario definir arquitecturas o plataformas de datos que **soporten y almacenen** estos grandes conjuntos de datos.
- ▶ *Data skills*: el dato no es ajeno a los miembros de la organización; por ello, es necesario promover planes de **formación y capacitación** alrededor del uso, acceso, protección y calidad de los datos en todas las áreas de la compañía.

En situaciones informales es común usar indiscriminadamente los términos *dato*, *información* y *conocimiento*. En el ámbito profesional y académico, es conveniente **distinguir** estos conceptos para evitar malinterpretaciones durante las distintas fases de la analítica de datos. Existen varias aproximaciones para la distinción de estos términos; en el contexto de las organizaciones, la definición descrita por Davenport y Prusak es apropiada porque diferencia el **alcance** de cada término (Davenport y Prusak 1993).

Ventajas de las compañías *data driven*

Dentro de las ventajas del cambio cultural de las empresas *data driven*, se pueden mencionar las siguientes:

- **Aumento de la productividad:** algunos informes de mercado revelan que la productividad de las empresas *data driven*, por lo general, suele ser superior a otras empresas que no lo son. La atención al dato, desde luego, genera mayores beneficios por el enfoque centrado en ellos.

- ▶ **Ventaja competitiva:** la **visualización** constante de los datos permite conocer en tiempo real **cómo se comporta** la organización. Los expertos de negocio disponen de herramientas que usan para visualizar el estado de las decisiones, su relación directa con el comportamiento del producto o servicio en el mercado y la reacción de la competencia frente a dichas decisiones.
- ▶ **Capacidad de predicción:** la siguiente necesidad que se plantea toda empresa *data driven* es buscar la capacidad de **anticiparse y prever cambios** a tiempo. Para ello, las empresas analizan los datos y los usan como herramientas para conocer no solo su estado actual, sino su comportamiento a futuro.

El analista de datos en las compañías *data driven*

El analista de datos tiene un rol clave en una empresa *data driven*. Su papel es **recopilar, procesar y analizar** grandes cantidades de datos para obtener *insights* que ayuden a la empresa a tomar decisiones informadas y mejorar sus procesos y estrategias.

Algunas de las **responsabilidades** comunes de un analista de datos incluyen:

- ▶ Recopilación y limpieza de datos.
- ▶ Análisis de datos para identificar patrones y tendencias.
- ▶ Desarrollo de modelos y algoritmos para predecir futuros resultados.
- ▶ Visualización y presentación de resultados y hallazgos clave.
- ▶ Trabajo con equipos de toda la empresa para aplicar los hallazgos de los datos a sus áreas específicas.

1.4. Orientar una empresa al dato

Implantar una cultura *data driven* en una organización es un **esfuerzo conjunto** que impacta a todas las áreas de la compañía. Por ello, antes de abordar una misión como esta, es importante que las organizaciones valoren si están listas para dar este paso y si tienen la disposición real de **pensar, actuar y comportarse** de forma distinta, siguiendo el valor que aporten los datos.

Para hacer esta **valoración interior** en la compañía, algunas cuestiones que se recomiendan aclarar son:

- ▶ ¿Existe una estrategia de datos en la organización?
- ▶ Si no está definida la estrategia del dato o está mal definida, ¿es posible identificar problemas en los que los datos podrían contribuir, ayudar o mitigar su solución?
- ▶ ¿Alguna parte de la organización es consciente de los que datos existen?, ¿se puede confiar en la calidad de dichos datos?
- ▶ ¿Qué procesos se deben mejorar para garantizar que hay una gobernanza de datos sólida dentro de la organización?
- ▶ ¿Los empleados de la compañía conocen sus capacidades sobre los datos?

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

Es común que las organizaciones piensen que potenciar la tecnología sobre los datos por sí solo es suficiente para potenciar la estrategia de datos. Es cierto que la tecnología juega un papel muy importante en esta ecuación, pero es necesario entender que debe existir un **cambio de mentalidad** y muchos esfuerzos por parte de empleados, directores, responsables y el propio CEO de la compañía.

Algunos **pasos recomendados** para orientar una empresa hacia un enfoque *data driven* son los siguientes:

- ▶ **Establecer la importancia de los datos:** la cultura de la empresa debe cambiar para valorar la información y los datos como un recurso clave para la toma de decisiones.
- ▶ **Recopilar y centralizar los datos:** es necesario tener acceso a todos los datos relevantes en un **solo lugar** para que puedan ser analizados efectivamente.
- ▶ **Mejorar la calidad de los datos:** los datos deben ser precisos, completos y consistentes para ser útiles en la toma de decisiones.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

- ▶ **Uso de tecnología y herramientas adecuadas:** la tecnología y las herramientas de análisis de datos son esenciales para **procesar y analizar** grandes cantidades de información.
- ▶ **Capacitación y desarrollo de equipos:** el personal de la empresa, especialmente los analistas de datos y los equipos de toma de decisiones, deben estar capacitados para trabajar con datos y utilizar herramientas de análisis.
- ▶ **Integración de los datos en la toma de decisiones:** los datos deben ser un factor clave en la toma de decisiones y utilizados de manera **integral** en toda la empresa.
- ▶ **Monitoreo y mejora continua:** la empresa debe monitorear continuamente sus procesos y resultados y utilizar los datos para mejorar de manera continua.

Es importante destacar que, aunque se indiquen que son pasos, realmente estas tareas hacen parte de un **proceso continuo** que requiere tiempo y esfuerzo constante para impulsar verdaderamente un enfoque *data driven* que ofrezca un **impacto positivo** en la toma de decisiones y los resultados empresariales.

Cuando se plantea un proyecto de análisis de datos dentro de una organización *data driven* es recomendable tener en cuenta los pasos anteriores para **identificar posibles problemas** que puedan afectar el resultado de las tareas del proyecto. En este contexto, el criterio del analista es muy útil a la hora de valorar la **credibilidad** de los datos; su trabajo lo que persigue es, justamente, apoyar la estrategia del dato y hacer confiable la toma de decisiones sobre los datos analizados.

1.5. Proyectos de *big data* y *data mining*

Hasta ahora se ha mencionado la importancia de una buena estrategia orientada al dato dentro de las organizaciones, pero poco se ha dicho del reto que supone abordar proyectos de datos en empresas con un alto volumen de estos.

En las organizaciones donde el dato crece de forma exponencial, el reto de asumir un enfoque *data driven* supone, además, la necesidad de afrontar proyectos de *big data* con el fin de poder **procesar y sacar valor** de esos volúmenes de datos. Lo mismo ocurre con los proyectos de *data mining*, los cuales son necesarios cuando existe la necesidad de descubrir **patrones y conocimientos** a partir de esos grandes volúmenes de datos.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

Los proyectos de *data mining* son iniciativas que, basándose en técnicas y algoritmos de minería de datos, se utilizan para extraer **información valiosa** de los datos (Goldschmidt, Passos y Bezerra, 2015). No siempre ambos proyectos van de la mano dentro de las organizaciones. En muchos casos los proyectos de *big data* brindan a las compañías la capacidad de **almacenar, tratar y procesar** los datos, mientras que los proyectos de *data mining* luego aprovechan esta capacidad para extraer **valor, información y conocimiento**.

En un Trabajo de Final de Maestría es importante medir el alcance del trabajo a desarrollar, en especial cuando se trate de diferenciar muy bien la necesidad de desarrollar un proyecto de *big data* o llevar a cabo un proyecto de *data mining*.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

Características de un proyecto *big data*

Como se ha mencionado antes, un proyecto de *big data* es una iniciativa cuyo objetivo principal es **analizar y extraer valor** de grandes cantidades de datos, más allá de los límites que deparan las tecnologías de almacenamiento y las técnicas de análisis convencionales.

Algunas de las **características** más importantes de los proyectos de *big data* son las siguientes:

- ▶ **Volumen:** estos proyectos se enfrentan a grandes cantidades de datos (terabytes o petabytes de información).
- ▶ **Velocidad:** los datos pueden generarse y recopilarse a una velocidad exponencial y, por ello, en ocasiones, es necesario procesarlos y analizarlos en tiempo real.
- ▶ **Variedad:** estos proyectos deben ser capaces de manejar diferentes tipos de datos, estructurados, no estructurados y semiestructurados.
- ▶ **Veracidad:** un objetivo claro es garantizar la calidad y la precisión de los datos en el análisis dentro de los entornos *big data*.
- ▶ **Valor:** los proyectos buscan obtener *insights* y valor de los datos que sean útiles para la empresa.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

Estas características hacen que los proyectos de *big data* conlleven mucha complejidad, pero también ofrecen un **gran potencial** para mejorar la toma de decisiones y aumentar la eficiencia de los procesos en las empresas.

Problemas comunes en los proyectos *big data*

Los proyectos de *big data* pueden presentar una serie de desafíos y dificultades tales como:

- ▶ **Almacenamiento:** no es una tarea fácil almacenar grandes cantidades de datos, esto puede llegar a ser costoso y demandar tecnologías especializadas.
- ▶ **Procesamiento:** procesar y analizar grandes cantidades de datos puede ser complejo y necesitar de recursos informáticos adecuados.
- ▶ **Calidad de los datos:** un desafío constante en los proyectos de *big data* es garantizar la calidad y la precisión de los datos.
- ▶ **Seguridad de los datos:** quizás es una de las tareas más críticas dentro de la organización; se debe garantizar la seguridad y privacidad de los datos en todo momento.
- ▶ **Profesionales especializados:** en el mercado actual puede ser un obstáculo no contar con el personal capacitado en habilidades de ciencia de datos, tecnología y análisis de información.

- ▶ **Integración con los sistemas existentes:** no solo hay que desarrollar el proyecto, también es importante integrar los resultados con los sistemas y procesos existentes en la organización. La integración puede ser otro gran inconveniente en este tipo de proyectos, dada la necesidad de ajustar los sistemas actuales a los nuevos procesos.
- ▶ **Costos:** mantener e implementar un proyecto de *big data* no es una tarea de bajo coste, todo lo contrario. La inversión puede ser alta especialmente cuando se necesitan de plataformas robustas que se ajusten a las demandas de la organización.

Características de un proyecto de *data mining*

Recordemos que *data mining* es el proceso de descubrir **patrones y conocimientos** a partir de grandes cantidades de datos. Los proyectos de *data mining* son iniciativas que buscan utilizar **técnicas y algoritmos** de minería de datos para extraer información valiosa de ellos.

Dentro de las **características** más importantes de los proyectos de *data mining* están:

- ▶ **Descubrir patrones y conocimiento:** un objetivo claro del proyecto es identificar patrones y conocimientos en los datos que puedan ser útiles para la toma de decisiones dentro de la organización.
- ▶ **Automatización de procesos:** en muchos casos, el descubrimiento de patrones y conocimiento permite automatizar tareas o procesos dentro de las empresas. Dicha automatización irá de la mano con obtener resultados eficientes y certeros.
- ▶ **Mejora de la eficiencia:** los resultados eficientes de los proyectos de *data mining* pueden mejorar la eficiencia de los procesos empresariales y, por ende, promover la toma de decisiones en distintos ámbitos.

Problemas comunes de los proyectos de *data mining*

A continuación, se describen algunos problemas comunes en los proyectos de *data mining*:

- ▶ **Complejidad:** por lo general, los proyectos pueden llegar a ser complejos y requerir **habilidades especializadas** en ciencia y minería de datos.
- ▶ **Costos:** pueden ser costosos y requerir una **inversión significativa** en tecnología y en especialistas.
- ▶ **Calidad de los datos:** para el éxito del proyecto es imprescindible garantizar la calidad y la precisión de los datos; es crítico para el éxito del proyecto asegurar que los datos sean **confiables y precisos**.
- ▶ **Interpretación:** los resultados de *data mining* pueden ser difíciles de interpretar y requieren una comprensión profunda de las técnicas y algoritmos utilizados.

- *Bias*: no hay que olvidar los problemas de **sesgo** que genera el uso de datos que no sean representativos o que estén mal etiquetados.

Al igual que los proyectos de *big data*, un proyecto *data mining* puede brindar un gran valor y mejorar la eficiencia de los procesos dentro de las organizaciones, pero si no se aborda cuidadosamente, puede ser un desafío importante cuyo éxito no esté del todo asegurado.

En tu Trabajo de Final de Maestría es importante que identifiques posibles problemas que debas documentar y abordar de alguna forma. Recuerda que no tienes que resolver todos los problemas mencionados en este tema, pero sí puedes mostrar que los conoces y prever su aparición o no.

1.6. CRISP-DM como metodología de proyecto y modelo de procesos

Cross-Industry Standard Process for Data Mining, o por sus siglas, CRISP-DM, es un método o marco de referencia ampliamente utilizado en la industria para planificar y llevar a cabo proyectos de minería de datos (Chapman et al., 1999).

Desde el punto de vista de una **metodología**, esta **propone y describe** las fases de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre dichas tareas. Desde el punto de vista de **modelo de proceso**, CRISP-DM ofrece un resumen del **ciclo de vida** de los proyectos de minería de datos.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

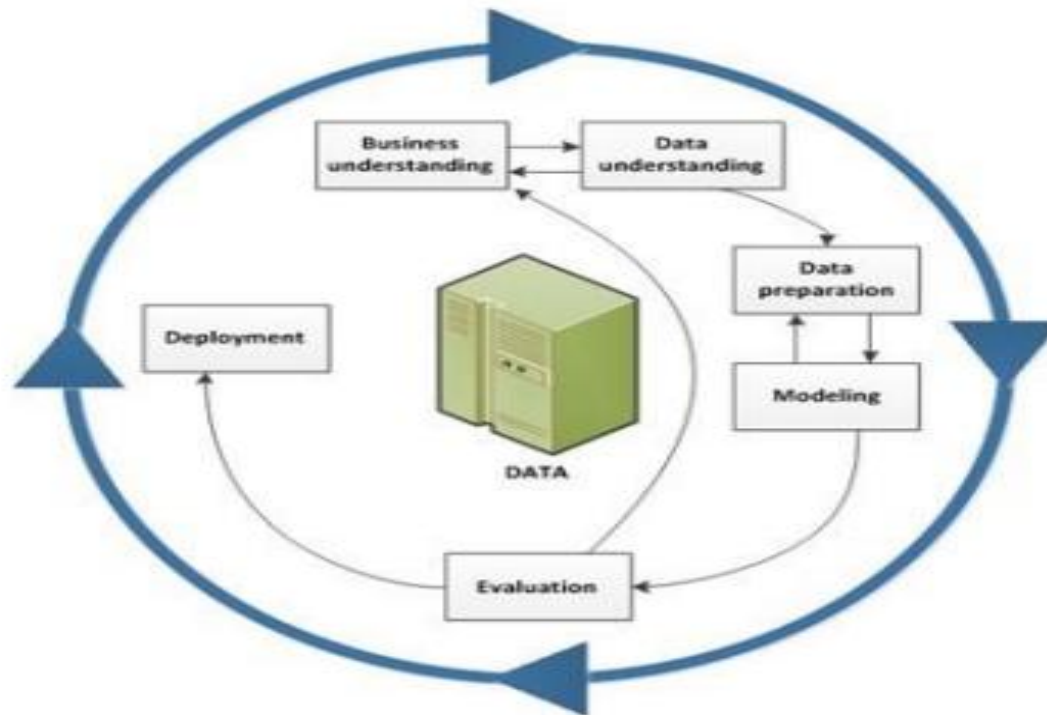


Figura 3. Ciclo de Vida de la minería de datos- Fuente: IBM, s. f.

CRISP-DM consta de seis fases que abarcan todo el **ciclo de vida** del proyecto de minería de datos, desde la identificación de oportunidades hasta la implementación de soluciones. Las seis fases son:

CRISP-DM consta de seis fases que abarcan todo el **ciclo de vida** del proyecto de minería de datos, desde la identificación de oportunidades hasta la implementación de soluciones. Las seis fases son:

- ▶ **Comprensión del negocio:** la primera fase consiste en comprender los objetivos y desafíos del negocio y determinar cómo la minería de datos puede contribuir a la solución de estos problemas.
- ▶ **Preparación de los datos:** esta segunda fase incluye la recopilación y limpieza de los datos necesarios para el proyecto.
- ▶ **Modelado:** en esta tercera fase se utilizan técnicas y algoritmos de minería de datos para identificar patrones y conocimientos en los datos.
- ▶ **Evaluación:** en esta cuarta fase se evalúan los resultados de la minería de datos para determinar si cumplen con los objetivos del proyecto y si son aceptables para su implementación.
- ▶ **Implementación:** es la quinta fase y en ella se implementan las soluciones identificadas a través de la minería de datos en el negocio.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

- **Monitoreo:** última fase en la que se monitorean los resultados de la implementación y se realiza cualquier ajuste necesario.

El ciclo de vida del modelo describe tanto las **fases** como las **dependencias** más importantes y frecuentes entre ellas. El orden de las fases es importante, pero la secuencia entre ellas no es una obligatoriedad y tampoco se debe seguir al pie de la letra en todos los casos. En muchos proyectos es completamente normal que se avance y se retroceda entre fases si las condiciones del proyecto así lo sugieren.



Figura 4. [Adaptación del enfoque a proyectos de análisis de datos]. Fuente: DECIDATA.

Las ventajas de apoyarse en este enfoque son la **flexibilidad** de uso y la posibilidad de **adaptarlo** a las necesidades y limitaciones del proyporecto. En ciertos casos, en lugar de realizar el modelado, el trabajo se puede centrar en la exploración y visualización para descubrir patrones de interés en los datos utilizados. En este contexto, las fases de modelado, evaluación y despliegue pueden ser menos relevantes que las fases de preparación y comprensión de datos, o simplemente no ser necesarias porque así el proyecto lo indique.

Sin embargo, es muy importante que en todo proyecto basado en este enfoque se tengan en cuenta las **pautas y recomendaciones** que sugiere el propio enfoque para abordar en un futuro las fases que no se hayan ejecutado.

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato

FASE 0 Y FASE 1

TEMA 1. - Introducción a las empresas orientadas al dato