

Apuntes

GUÍA MAESTRA DE ESTUDIO: ESTRATEGIA Y GESTIÓN DE DATOS

1. PILARES TEMÁTICOS Y CONCEPTOS RECURRENTES

El curso se estructura en tres grandes bloques que debes dominar: **Estrategia y Negocio**, **Arquitectura y Gobierno del Dato**, y **Analítica Aplicada (Interpretación)**.

A. Inteligencia de Negocio (Business Intelligence - BI)

El BI no es solo tecnología. Se define recurrentemente como la integración de cinco elementos clave que deben funcionar como vasos comunicantes: 1. **Procesos:** Metodologías e infraestructuras. 2. **Tecnología:** Infraestructura (Cloud, On-premise). 3. **Herramientas:** Software (Tableau, PowerBI, Python, SAP). 4. **Personas:** El elemento central y más difícil de copiar por la competencia (ventaja competitiva). 5. **Capacidades:** Músculo financiero y organizativo,.

- **Objetivo del BI:** El flujo crítico es: Transformar datos en información → información en conocimiento → conocimiento en soporte para la decisión → **ACCIONABILIDAD** (el fin último es tomar decisiones tácticas o estratégicas),.

B. Arquitectura Técnica y Calidad del Dato

Para que el análisis sea fiable, la “cocina” del dato debe estar limpia. * **Procesos ETL (Extract, Transform, Load):** Vitales para consolidar la información de múltiples fuentes (CRM, ERP, Billing, Ficheros planos). * *Extracción:* Sacar el dato del origen. * *Transformación:* Limpieza, deduplicación, gestión de nulos. * *Carga:* Guardarlo en el Data Warehouse. * **Data Warehouse (DTW):** Repositorio centralizado y estructurado para el análisis (no para la operativa diaria). * **Modelo Estrella:** Una tabla de hechos central (ej. Ventas) rodeada de tablas de dimensiones (Tiempo, Producto, Cliente) sin jerarquías. Es más rápido para consultas. * **Modelo Copo de Nieve (Snowflake):** Las dimensiones tienen jerarquías y se normalizan en

múltiples tablas (ej. Ciudad → Provincia → País). * **Golden Record:** La versión única y veraz de un cliente, obtenida tras limpiar y unificar duplicados (fundamental para no tratar al mismo cliente como varios distintos),.

C. Estrategia de Clientes y Toma de Decisiones

- **HIPPO vs. Data-Driven:**
 - *HIPPO (Highest Paid Person's Opinion):* Decisiones basadas en la intuición subjetiva del jefe (Ejemplo: Angels Brick en la Actividad 1),.
 - *Data-Driven:* Decisiones objetivas basadas en evidencia analítica.
- **Customer Centricity:** Pasar de estrategias centradas en producto a centradas en el cliente. El mercado no es homogéneo; diferentes necesidades requieren diferentes tratamientos,.
- **Ciclo de Vida:** Captación → Desarrollo (Cross/Up-selling) → Retención.

D. Gobernanza y Ética (RGPD)

- **Dato Personal:** Cualquier información que identifique o haga identificable a una persona física (nombre, mail, ID dispositivo, geolocalización),.
- **Principios Clave (Caso Cambridge Analytica):**
 - *Limitación de la finalidad:* Los datos solo se pueden usar para lo que se informó al usuario (Facebook falló aquí al permitir uso político de datos académicos),.
 - *Minimización:* Pedir solo lo necesario.
 - *Exactitud:* Los datos deben estar actualizados (ej. no guardar CVs de hace 10 años).
 - *Limitación del plazo de conservación:* No guardar datos “por si acaso” eternamente.

E. Visualización

- **Mantra:** “Olvido lo que leo, recuerdo lo que veo”.
 - **Tableau:** Se estructura en tres niveles:
 1. **Hoja de trabajo:** Gráfico individual.
 2. **Dashboard:** Conjunto de hojas combinadas.
 3. **Historia:** Narrativa secuencial para presentaciones (diferencial respecto a Power BI),.
-

2. METODOLOGÍA ANALÍTICA E INTERPRETACIÓN

*Aunque el examen final es teórico/desarrollo y no requiere programar, debes saber **i** nterpretar estos conceptos vistos en la parte técnica y en las actividades.,*

- **Clustering (K-Means):**
 - Aprendizaje **No Supervisado** (no hay variable objetivo, el ordenador busca patrones).
 - **Interpretación:** Se analizan las medias de cada grupo para ponerles “nombre” (ej. “Clientes VIP”, “Clientes Low Cost”). A veces se usa un

árbol de decisión posterior para explicar las reglas del clúster (“engañando al ordenador”),.

- **Árbol de Decisión:**

- Aprendizaje **Supervisado** (tengo una variable respuesta).

- **Interpretación:** El nodo superior es la raíz. Cada “rama” es una regla.

Debes saber leer la probabilidad de cada clase en las hojas finales (ej. “Si Edad > 30 y Ingresos < 20k -> 90% probabilidad de No compra”),.

- **Regresión Logística:** Para predecir variables binarias (Sí/No, 0/1). Se mira el P-valor (significatividad) para saber qué variables influyen realmente en el resultado.,.

- **Series Temporales (ARIMA/TS):** Requieren una variable temporal y una frecuencia (mensual/diaria). Se usan para predicciones a futuro (forecast). *Nota: En R se definen con el objeto ts.,*

3. PREGUNTAS DE EXAMEN PROBABLES (Interpretación y Desarrollo)

Basado en lo enfatizado en las sesiones de repaso y en la estructura de las Actividades:

Bloque 1: Estrategia y Negocio (Basado en Actividad 1 y 3)

1. **Caso Práctico de Decisión (HIPPO vs Data):** “Una empresa quiere lanzar una promoción basándose en la intuición del Director Comercial. Explique por qué esto es un riesgo y describa el proceso analítico que seguiría para validar o refutar esa intuición con datos.” *Respuesta clave: Mencionar análisis exploratorio, definición de KPIs, validación de hipótesis y evitar sesgos subjetivos.,*
2. **KPIs SMART:** “Defina qué es un KPI SMART y proponga dos ejemplos para un e-commerce.” *Respuesta: Específico, Medible, Alcanzable, Relevante, Temporal. Ej: Tasa de conversión mensual, Ticket medio por cliente.,*

Bloque 2: Gobierno del Dato y Ética (Basado en Actividad 2)

1. **Análisis de Caso Legal:** “Una empresa utiliza los CVs de candidatos de hace 5 años para enviarles publicidad de productos. ¿Qué principios del RGPD se están vulnerando?” *Respuesta clave: Principio de limitación de la finalidad (se recogió para empleo, no para publicidad) y Principio de exactitud/conservación (dato antiguo/desactualizado),.*
2. **Dato Personal:** “¿Es una dirección IP o una geolocalización un dato personal?” *Sí, si permite identificar indirectamente a una persona física o singularizarla.*

Bloque 3: Infraestructura y BI

1. **Data Warehouse:** “Explique la diferencia entre un modelo Estrella y un modelo Copo de Nieve. ¿Cuál elegiría para un sistema donde prima la velocidad de consulta?” *Respuesta: Estrella (desnormalizado, menos tablas, más rápido) vs. Copo de Nieve (normalizado, más tablas, ahorra espacio pero más lento en joins). Para velocidad: Estrella.,*

2. **ETL y Calidad:** “¿Por qué es peligroso cargar datos directamente de los operacionales al sistema de análisis sin pasar por una ETL?” *Respuesta: Riesgo de duplicados, datos sucios, formatos inconsistentes y falta de una visión única del cliente (Golden Record).*

Bloque 4: Interpretación Analítica (Conceptos)

1. **Interpretación de Clusters:** “Si en un análisis de clúster obtiene 3 grupos, ¿cómo describiría el perfil de cada uno?” *Debes ser capaz de decir: ‘Miro las medias de las variables o uso un árbol de decisión para ver las reglas: El Grupo 1 son jóvenes con poco gasto, el Grupo 2 son mayores con mucho gasto...’.*
 2. **Correlación:** “¿Qué significa una correlación fuerte negativa entre Precio y Demanda?” *Significa que a mayor precio, menor demanda. Es vital saber que correlación no implica causalidad.*
-

4. CONSEJOS FINALES DEL PROFESOR PARA EL EXAMEN

- **No memorices, relaciona:** No te pedirán “Define Data Warehouse”, sino “¿Cómo ayuda un Data Warehouse a evitar silos de información en una empresa?” o preguntas basadas en un contexto empresarial,.
- **Extensión y Formato:** Tendrás un espacio limitado. Usa esquemas, gráficos simples si es presencial, o estructura de puntos. Sintetiza. No escribas “el Quijote en un grano de arroz”,.
- **Interpretación:** Si te ponen un caso, asume el rol de analista. No digas solo “el dato es 5”, di “el dato es 5, lo que implica que la estrategia no funciona y deberíamos cambiar a...”.
- **Material:** Las preguntas salen de las “Ideas Clave” de los temas,.

La clave está en entender el **flujo del dato**: desde que nace (ETL), se guarda (DTW), se analiza (Modelos/Clustering), se visualiza (Dashboards) y se protege (RGPD) para tomar decisiones.