

# Apuntes

Basado en las transcripciones de las clases, especialmente la **Clase 11** donde el profesor detalla la estructura final, y la **Clase 01** donde dio las primeras pinceladas, el examen tendrá las siguientes características, formato y tipos de ejercicios:

## 1. Formato General y Peso

- **Modalidad:** Es un examen online realizado en la plataforma **Hexam**,.
- **Peso:** Supone el **60%** de la calificación final de la asignatura. Es obligatorio aprobarlo (mínimo un 5) para pasar la materia.,
- **Estructura:** Se divide en una **parte teórica (aprox. 40%)** y una **parte práctica (aprox. 60%)**,.

## 2. La Parte Práctica (El Caso de Negocio)

El profesor ha sido muy explícito sobre cómo será esta sección. Consistirá en un **caso de negocio** simulado.

- **El Escenario:** Se presentará una situación de negocio (ej. una empresa de bicicletas, ropa o software con un nombre ficticio) y se pedirá generar un dashboard para analizar datos concretos (ej. ventas por región o evolución del negocio),.
- **Herramientas:** Deberás usar **Power BI** o **Tableau** (tú eliges). También podrás usar Excel para visualizar los datos brutos.
- **El Dataset:** Usarás un archivo específico llamado “**base de datos examen.zip**” (que estará protegido con contraseña hasta el momento del examen). Es un dataset similar a los vistos en clase (tipo *retail*), pero “guarreado” (con errores) a propósito para obligarte a limpiarlo.,
- **Entregable:** Deberás subir el archivo del dashboard (.pbix o .twb) a la plataforma. Las respuestas escritas se redactan directamente en Hexam.

**Pasos o Tareas Específicas del Ejercicio Práctico:** El enunciado te guiará por 4 partes claramente diferenciadas:

1. **Revisión y Pre-análisis:** Deberás describir el dataset, identificar variables principales y comentar sobre la calidad de los datos (ej. celdas vacías, formatos inconsistentes),.
2. **Limpieza y Transformación (ETL):** Se te pedirá corregir errores, valores anómalos o crear variables nuevas (ej. un ratio de facturación/unidades).  
**Importante:** Debes usar **Power Query** y documentar (o que se vean) los pasos aplicados.,
3. **Generación del Dashboard:** Se te pedirán visualizaciones concretas (ej. 3 o 4 gráficos específicos, como un mapa geográfico o un gráfico de barras

comparativo). No tendrás que ser excesivamente creativo, te dirán qué quieren ver.

4. **Interpretación y Conclusiones:** Habrá unas 5 o 6 preguntas concretas que deberás responder mirando tu dashboard (ej. “¿Cuál es el producto que más se vendió en verano?”). Aquí demuestras que tu dashboard funciona y sabes leerlo.

**Sobre Fórmulas DAX:** \* No se pedirán fórmulas DAX complejas. Si se necesita alguna (ej. multiplicar precio por unidades), será sencilla o se darán instrucciones claras sobre cómo hacerla.,

### 3. La Parte Teórica

- **Tipo de preguntas:** **NO es tipo test.** Son preguntas de desarrollo o respuesta corta, pero objetivas (no de opinión subjetiva),,
- **Contenido:**
  - Preguntas sobre conceptos vistos en los temas (el profesor mencionó un índice de repaso en la Clase 11),,
  - **Ejemplo de pregunta:** “Defineme qué es un grafo de red, cuáles son sus elementos principales y pon un ejemplo”.
  - **Interpretación:** Te pueden poner un gráfico y preguntarte qué interpretas o qué conclusiones sacas de él.
  - **Power Automate:** Aunque no habrá ejercicio práctico de automatización (porque requiere licencias PRO), **sí puede caer teoría.** Por ejemplo, te pueden poner un pantallazo de un flujo y preguntarte qué está haciendo o para qué sirve,,

### 4. Consejos del Profesor

- **Limpieza de datos:** Es fundamental. Una parte importante de la nota dependerá de que limpies bien el dataset antes de graficar,,
- **No memorizar definiciones:** Prefiere que entiendas el concepto y sepas explicar qué hace una técnica (ej. para qué sirve un árbol de decisión) a que memorices el nombre exacto.
- **Hoja de trucos (Checklist):** El profesor recomendó revisar las diapositivas de resumen al final de los temas para saber qué conceptos son los más preguntables.

Basado en las transcripciones de las clases (especialmente la **Clase 11** donde se hizo un repaso general) y los documentos de los temas, aquí tienes la teoría clave que debes dominar para la parte teórica del examen (que supone aproximadamente el 40% de la nota).

El profesor ha insistido en que **no busques memorizar definiciones exactas**, sino **entender los conceptos** para poder explicarlos o interpretarlos ante un caso o gráfico dado.

Aquí tienes el desglose por bloques temáticos:

## 1. Visualización Avanzada y Diseño (Temas 1 y 2)

Esta es la base para interpretar gráficos y diseñar dashboards.

- **Visualización Multivariante:** Debes saber qué técnicas usar cuando tienes muchas variables.
  - **Gráfico de Burbujas:** Permite ver hasta 4 dimensiones (Eje X, Eje Y, Tamaño de la burbuja, Color).
  - **Matriz de Dispersión (Scatter Matrix):** Útil para ver correlaciones entre pares de variables.
  - **Small Multiples:** Varios gráficos pequeños con la misma escala para comparar categorías.
- **Perfiles de Usuario:** Saber diferenciar para quién diseñas el dashboard.
  - **Ejecutivo (CEO):** Necesita KPIs agregados, tendencias macro y ver excepciones. No quiere gran detalle.
  - **Analista/Técnico:** Necesita granularidad, filtros y capacidad de explorar el dato (drill-down).
- **Principios de Diseño (Gestalt):** Entender cómo el cerebro agrupa la información.
  - **Proximidad:** Elementos cerca se perciben como grupo.
  - **Semejanza:** Mismo color o forma implica relación.
  - **Cierre:** El cerebro completa formas incompletas.,
- **Tipos de Gráficos Especiales:**
  - **Mapas Coropléticos:** Usan colores para representar valores (tasas/ratios, no absolutos) en áreas geográficas.
  - **Grafos de Red:** Entender qué son **nodos** (entidades) y **aristas** (relaciones). Ejemplo: aeropuertos y rutas,.
  - **Diagrama de Sankey:** Para visualizar flujos y cuellos de botella (anchura de la línea representa volumen).

## 2. Preprocesamiento y Calidad del Dato (Tema 3)

Aunque esto se aplicará en la práctica, pueden caer preguntas conceptuales sobre la importancia de la limpieza.

- **Tidy Data (Datos Ordenados):** La regla de oro:
  1. Cada variable es una columna.
  2. Cada observación es una fila.
  3. Cada valor es una celda.
- **Importancia del preprocesamiento:** Supone el 70% del tiempo de un proyecto. Sin datos limpios (sin nulos, duplicados o errores de formato), la visualización es errónea (“Basura entra, basura sale”),.

## 3. Modelización y Análisis Exploratorio (Tema 4)

Este bloque es muy importante para la teoría, especialmente la diferencia entre describir y predecir.

- **Diferencia Descriptiva vs. Predictiva:**
  - **Descriptiva:** Mirar por el retrovisor. Explicar qué ha pasado y por qué (patrones históricos).
  - **Predictiva:** Mirar con prismáticos. Estimar qué pasará en el futuro.

- **Técnicas de Agrupamiento (Clustering):**
  - **K-Means:** Agrupa datos basándose en la distancia a un centroide. Útil para segmentar clientes.
  - **Clustering Jerárquico:** Crea un árbol (dendrograma) de relaciones.
- **Reducción de Dimensionalidad (PCA vs. t-SNE):** (Muy probable que pregunten diferencias o cuándo usar cuál).
  - **PCA (Análisis de Componentes Principales):** Técnica **lineal**. Rápida, buena para resumir y mantener la varianza global. Reduce ruido.,,
  - **t-SNE:** Técnica **no lineal**. Excelente para visualizar clústeres locales y relaciones complejas. No conserva bien las distancias globales.,
- **Correlación vs. Causalidad:** Entender que dos variables se muevan juntas (correlación) no significa que una cause la otra.

## 4. Automatización y Gobernanza (Tema 5)

Aunque no harás prácticas de Power Automate en el examen, **sí entra la teoría** sobre cómo funciona un flujo.

- **Arquitectura de un Flujo:** Debes saber identificar los tres componentes clave de una automatización:
  1. **Trigger (Desencadenador):** El evento que inicia el proceso (ej. “Llega un correo”, “Una alerta de datos salta”).
  2. **Conectores:** Las herramientas que interactúan (Outlook, Excel, Power BI).
  3. **Acciones:** Lo que el sistema hace (Guardar archivo, enviar email).
- **Gobernanza:** Conceptos sobre seguridad (quién ve qué datos) y control de acceso (Row Level Security).

### Ejemplo de Pregunta Teórica (según el profesor)

*“Defineme qué es un grafo de red, cuáles son sus elementos principales y pon un ejemplo.”* \* **Respuesta esperada:** Es una visualización para ver relaciones. Sus elementos son Nodos (puntos) y Aristas (líneas). Ejemplo: Conexiones entre ordenadores en una red de ciberseguridad.,

**Recomendación final:** Revisa las diapositivas de resumen (“Checklist” o “Resumen Tema”) que aparecen al final de las presentaciones de los Temas 4, 5 y 7, ya que el profesor indicó que son una buena guía de estudio.,