

<b>EVALUACIÓN</b>	Obligatorio 1	<b>GRUPO</b>	Todos	<b>FECHA</b>	Abril 2025
<b>MATERIA</b>	Analítica para Data Warehousing				
<b>CARRERA</b>	Ingeniería en Sistemas y Licenciatura en Sistemas				
<b>CONDICIONES</b>	<p><b>Uso de material de apoyo y/o consulta</b>  <b>Inteligencia Artificial Generativa</b></p> <ul style="list-style-type: none"> <li>- Seguir las pautas de los docentes: Se deben seguir las instrucciones específicas de los docentes sobre cómo utilizar la IA en cada curso.</li> <li>- Citar correctamente las fuentes y usos de IA: Siempre que se utilice una herramienta de IA para generar contenido, se debe citar adecuadamente la fuente y la forma en que se utilizó.</li> <li>- Verificar el contenido generado por la IA: No todo el contenido generado por la IA es correcto o preciso. Es esencial que los estudiantes verifiquen la información antes de usarla.</li> <li>- Ser responsables con el uso de la IA: Conocer los riesgos y desafíos, como la creación de "alucinaciones", los peligros para la privacidad, las cuestiones de propiedad intelectual, los sesgos inherentes y la producción de contenido falso.</li> <li>- En caso de existir dudas sobre la autoría, plagio o uso no atribuido de IAG, el equipo docente tendrá la opción de convocar al equipo de obligatorio a una defensa específica e individual sobre el tema.</li> </ul> <p><b>IMPORTANTE:</b></p> <ol style="list-style-type: none"> <li>1) Inscribirse.</li> <li>2) Formar grupos de hasta 3 estudiantes.</li> <li>3) Se responderán consultas en el foro de Aulas hasta 24 h antes de la entrega.</li> <li>4) El trabajo es de cada grupo. Les recordamos que copiar, cometer plagio o recibir ayuda no autorizada de terceros en la realización de trabajos académicos es considerado una falta grave según el Art. 51 del Reglamento Estudiantil (<a href="http://www.ort.edu.uy/variantes/pdf/documento001.pdf">http://www.ort.edu.uy/variantes/pdf/documento001.pdf</a>).</li> <li>5) Subir el trabajo a Gestión antes de la hora indicada (ver al final del documento: "RECORDATORIO").</li> </ol> <p>Si se presenta alguna <u>dificultad con la inscripción o algún inconveniente técnico</u>, por favor contactarse con la Coordinadora de cursos o Coordinación adjunta <b>antes de las 20 h del día de la entrega</b>, a través de los mails <a href="mailto:crossa@ort.edu.uy">crossa@ort.edu.uy</a> / <a href="mailto:posada_l@ort.edu.uy">posada_l@ort.edu.uy</a> (matutino) / <a href="mailto:larrosa@ort.edu.uy">larrosa@ort.edu.uy</a> (nocturno), o vía MS Teams.</p>				

## Origen y disponibilidad de los datos

Descargar los datos de precios de útiles escolares, recabados mediante relevamientos de precios en diferentes cadenas de comercios. Este conjunto contiene información consolidada sobre el costo de diversos productos escolares en Uruguay. Los datos se encuentran disponibles en los siguientes enlaces:

- 2025: <https://www.precios.uy/utiles-escolares-2025/>
- 2024: <https://www.precios.uy/utiles-escolares-2024/>
- 2023: <https://www.precios.uy/utiles-escolares-2023/>
- 2022: <https://www.precios.uy/utiles-escolares-2022/>
- 2021: <https://www.precios.uy/utiles-escolares-2021/>

Se deberán descargar los siguientes archivos:

- a) **Precios medios de muestras:**
- **Descargar precios medios muestra 2022-2025**
  - **Descargar precios medios muestra 2020-2024**

De estos archivos se debe consolidar la información en un único conjunto de datos, agregando una columna que indique el año correspondiente.

- b) **Cuadro Comparativo de Cadenas:** Descargar los cuadros comparativos de precios de útiles escolares para los años desde el 2021 al 2025.

Estos cuadros contienen tablas de precios medianos por producto y por cadena comercial en formato PDF. Procesamiento requerido:

- Cada archivo PDF debe ser procesado para estructurar su contenido de manera que pueda ser cargado a una base de datos o utilizado en Power BI. Para ello, pueden optar por:
  - **Opción A:** Procesar los PDFs utilizando Python (por ejemplo, librerías como pdfplumber o camelot) o alguna otra herramienta, para extraer las tablas y guardarlas en un archivo Excel o CSV.
  - **Opción B:** Cargar los datos en una base de datos MySQL, construyendo una tabla donde cada fila represente un producto en una determinada cadena.
- Es obligatorio consolidar toda la información en un único dataset final, indicando de qué año y fuente proviene cada registro.

## Ejercicio 1

Entregables:

- Los archivos pbix generados.
  - El dump de la base de datos, los datos y scripts (sql y/o Python) requeridos que permitan a los correctores reproducir lo realizado en el desarrollo del obligatorio.
  - Un único archivo en formato pdf (con índice y numeración de páginas). Presentar un tutorial con capturas de pantalla de cada paso realizado en que se muestre el desarrollo de este ejercicio. Tener en cuenta que el tutorial se utilizará en un curso de capacitación de Analítica para Datawarehousing.
  - NO se admiten links a archivos que formen parte de la entrega. Se corregirá solamente lo cargado en [gestion.ort.edu.uy](http://gestion.ort.edu.uy)
1. Cargar los datos en una base de datos en MySQL
    - 1.1. Especificar los pasos para extraer la información de los PDF usando Python u otra herramienta o IA.
    - 1.2. Especificar los pasos para realizar la carga de la información a MySQL
  2. Analizar los metadatos y diseñar una tabla de hechos (diseño star) apropiada para analizar los datos
    - 2.1. Explicar el procedimiento realizado en SQL
    - 2.2. Adjuntar el script (DDL) para la creación de las tablas de dimensiones y la tabla de hechos
    - 2.3. Adjuntar el script (DML) para la carga de datos en cada una de las tablas anteriores
  3. Implementar las tablas de hechos con las relaciones apropiadas en Power BI Desktop
  4. En caso de realizar alguna transformación en los datos en el Power BI, se deberá de explicarlos
  5. Realizar un dashboard que responda las siguientes preguntas:
    - 5.1. ¿Cuál es el producto escolar con mayor precio promedio en 2025?
    - 5.2. ¿Qué cadena ofrece los precios más bajos en promedio en 2025?
    - 5.3. ¿Cómo varió el precio promedio de los útiles escolares entre 2022 y 2025?
    - 5.4. ¿Qué productos tuvieron el mayor aumento porcentual de precio entre 2024 y 2025?
    - 5.5. ¿Qué productos bajaron de precio entre 2024 y 2025?
    - 5.6. ¿Cuál es la cadena con mayor dispersión de precios (diferencia entre máximo y mínimo)?
    - 5.7. ¿Qué útil escolar es el más barato considerando todas las cadenas?
    - 5.8. ¿Cómo varía el precio del mismo producto entre cadenas distintas en un mismo año?
    - 5.9. ¿Qué cadena presenta precios más estables (menor desviación estándar) entre los años?
    - 5.10. ¿Qué categorías de productos (ej: cuadernos, lápices, adhesivos) presentan mayor inflación?
    - 5.11. ¿Cuál es el top 5 de productos más caros en 2025?
    - 5.12. ¿Cuál es el top 5 de productos más baratos en 2025?
    - 5.13. ¿Qué cadena incrementó más sus precios entre 2024 y 2025?
    - 5.14. ¿Cómo evolucionaron los precios promedios año a año? (de 2021 a 2025)
    - 5.15. ¿Qué porcentaje del total de productos aumentó su precio en cada año?

5.16. ¿En qué año hubo el mayor incremento general de precios?

5.17. ¿Qué cadena ofrece los mejores precios promedio para una lista básica de útiles?

Considerar lo visto en clase:

- En caso de que se requiera, crea medidas (como sumas, promedios, máximos, etc.) utilizando DAX para calcular valores que responderán a tus preguntas específicas.
  - Escoger los tipos de gráficos que mejor representen los datos para cada pregunta. Por ejemplo, un gráfico de barras para comparar precios, un gráfico de líneas para tendencias de tiempo, o mapas si tienes datos geográficos.
  - Ajustar los colores, etiquetas y herramientas de filtrado para mejorar la legibilidad y la interacción.
  - Organizar las visualizaciones en el lienzo de Power BI de manera que el flujo sea lógico y fácil de entender. Agrupa visualizaciones relacionadas y utiliza títulos y etiquetas para clarificar las secciones.
  - Agregar segmentaciones de datos, filtros y otras herramientas interactivas para permitir a los usuarios explorar los datos de manera más dinámica.
6. Formule 3 preguntas a su elección, diferentes a las planteadas en el punto anterior, y represente las respuestas en una pestaña nueva (entre 4 y 6 objetos visuales)
7. Crear un nuevo archivo que cuente una historia con los datos en Power BI, utilizando técnicas de storytelling (narrativa de datos)
- 7.1. Definir claramente qué historia quieres contar con tus datos.
  - 7.2. Decidir el flujo de la historia utilizando la estructura típica de una introducción, desarrollo y conclusión.
  - 7.3. Seleccionar y crear las visualizaciones que considere.
  - 7.4. Configurar la interactividad utilizando la segmentación de datos para ver más detalle de los datos.
  - 7.5. Utiliza cuadros de texto para añadir explicaciones, comentarios o preguntas que guíen a los usuarios a través de la historia.
8. Presente un resumen no más de 2 página, respondiendo las siguientes preguntas del artículo "[Visualization Requirements for Business Intelligence Analytics: A Goal-Based, Iterative Framework](#)":
- 8.1. Explique cómo el establecimiento del objetivo estratégico “Reducir las facturas impagas” influye en la definición del tipo de análisis que debe realizarse.
  - 8.2. Analice las ventajas y posibles limitaciones de utilizar este enfoque en entornos empresariales reales.

## Ejercicio 2

En este ejercicio se continuará trabajando con la información consolidada en el ejercicio anterior, correspondiente a los precios de útiles escolares relevados en los años 2019 a 2025.

Con el objetivo de realizar un análisis más completo y permitir la comparación de precios a valores constantes, se deberán incorporar además los datos del Índice de Precios al Consumo (IPC) publicados por el Instituto Nacional de Estadística (INE) de Uruguay.

La serie histórica del IPC necesaria puede ser consultada y descargada desde el sitio oficial del INE: <https://www.ine.gub.uy/web/guest/ipc>

Se deberá ajustar el valor de los precios nominales de útiles escolares a precios constantes tomando como año base 2024. Para ello, se utilizarán los índices de IPC correspondientes a cada año de la serie.

Entregables:

- Los archivos pbix generados.
  - Los datos y scripts requeridos que permitan a los correctores reproducir lo realizado en el desarrollo del obligatorio.
  - Un único archivo en formato pdf (con índice y numeración de páginas). Presentar un tutorial con capturas de pantalla de cada paso realizado en que se muestre el desarrollo de este ejercicio y los resultados obtenidos. Tener en cuenta que el tutorial se utilizará en un curso de capacitación de Analítica para Datawarehousing.
  - NO se admiten links a archivos que formen parte de la entrega. Se corregirá solamente lo cargado en [gestion.ort.edu.uy](http://gestion.ort.edu.uy)
1. Diseñar una tabla de hechos (diseño star) apropiada para analizar los datos e implementarla en PowerBIDesktop.
  2. Realizar 2 tablas y 2 gráficos teniendo en cuenta los datos del IPC. Comentar los resultados para un usuario de negocio.
  3. Explorar los datos con el objetivo de identificar si existen (o no) problemas de calidad que requieran transformaciones de los datos.
  4. Presentar un informe de dos páginas en que evalúe la calidad de los datos dirigido a responsables de negocio. .
  5. Datos duplicados:
    - 5.1. Modificar los datos de origen para que presenten dos casos de datos duplicados. Explicar los datos que se modificaron y el motivo por el que pueden haberse generado los datos duplicados.
    - 5.2. Explicar la forma de resolver el problema de duplicación en este caso.
    - 5.3. Implementar lo propuesto en la parte anterior
  6. Datos inconsistentes:

- 5.1. Modificar los datos de origen para que presenten dos casos de datos inconsistentes.
- 5.2. Explicar los datos que se modificaron indicando un motivo por el que pueden haberse generado. Presente (si es posible) una forma de visualizar los datos inconsistentes.
- 5.3. Explicar la forma de resolver el problema de datos inconsistentes en este caso.
- 5.4. Implementar lo propuesto en la parte anterior
7. Datos incorrectos:
  - 6.1. Modificar los datos de origen para que presenten dos casos de datos incorrectos. Explicar los datos que se modificaron y el motivo por el que pueden haberse generado los datos incorrectos.
  - 6.2. Explicar la forma de resolver el problema de datos incorrectos en este caso.
  - 6.3. Implementar lo propuesto en la parte anterior
7. Datos nulos:
  - 7.1. Modificar los datos de origen para que presenten dos casos de datos nulos. Explicar los datos que se modificaron y el motivo por el que pueden haberse generado los datos nulos.
  - 7.2. Explicar la forma de resolver el problema de datos nulos en este caso.
  - 7.3. Implementar lo propuesto en la parte anterior
8. Datos atípicos:
  - 8.1. Modificar los datos de origen para que presenten dos casos de datos atípicos. Explicar los datos que se modificaron. Presente una forma de visualizar los datos atípicos. Indique un motivo por el que pueden haberse generado los datos atípicos.
  - 8.2. Explicar la forma de resolver el problema de datos atípicos en este caso.
  - 8.3. Implementar lo propuesto en la parte anterior
9. Ajuste por inflación
  - 9.1. Ajustar los precios nominales de los útiles escolares utilizando los índices de IPC obtenidos del Instituto Nacional de Estadística (INE Uruguay)
  - 9.2. Presente mínimo una tabla y un gráfico que permitan comparar los valores corrientes y los valores constantes de gastos.
- 10 En base a los datos de 10 productos a su elección, analizar si los precios anteriores son útiles para estimar el precio más reciente, y mostrar los resultados obtenidos en base a un gráfico y una tabla.
- 11 Para el caso considerado en el Ejercicio 2
  - 11.1. Explique el motivo por el cual puede considerarse necesario contar con un Data Lake.
  - 11.2. Indique en este caso dos propiedades del Data Lake, dos ventajas y dos inconvenientes, explicando el motivo por el cual dichas propiedades son relevantes en el caso del Ejercicio 2.
  - 11.3. Indique en este caso dos diferencias entre Data Lake y Data Warehouse, adaptándolas al caso del Ejercicio 2, explicando los motivos con detalle para un usuario de negocio.

**Porcentaje de puntaje de los ejercicios:**

Ejercicio 1: 50%

Ejercicio 2: 50%

Importante: Se tomará en cuenta la prolijidad de la documentación entregada.

Tener en cuenta:

- i Para realizar este obligatorio es necesario acceder a datos a través de Internet, puede llegar a ocurrir que los sitios no estén disponibles, por lo que se recomienda no dejarlo para último momento.
- ii Puede llegar a ocurrir que la dirección indicada para descargar los datos haya cambiado. Ante cualquier consulta recurra al foro de la asignatura en [aulas.ort.edu.uy](http://aulas.ort.edu.uy)
- iii Si el obligatorio se entrega por varios estudiantes, todos los integrantes del grupo deben poder responder preguntas relativas a todo el obligatorio y la forma en que se desarrolló el obligatorio en su totalidad.

**NOTA: Es requisito entregar por separado los ejercicios 1 y 2 en dos carpetas comprimidas, nombradas Ejercicio 1 y Ejercicio 2.**