

e            eg            h            he  
          T            g    eh            g    eh

## h g

Requisitos.....	3
-----------------	---

## g g

- A continuación se solicita realizar el preprocesado y procesado de acuerdo a un conjunto de requisitos dados por los usuarios.
- Debes cargar los datos en dataframes de Pandas y realizar las operaciones necesarias para dar respuesta a los requisitos en un Notebook de Jupyter.
- Utiliza los markdowns, creando una estructura que permita seguir correctamente la ejecución. Comenta también las instrucciones de código para aportar mayor claridad.
- Lee los detalles y los requisitos a implementar antes de ejecutar la tarea.
- Carga únicamente las librerías que vayas a utilizar.
- Asegúrate de ejecutar todas las celdas. Una opción para verificar que no te olvidas de ninguna es ir a *Kernel* → *Restart Kernel and Run All Cells...*
- No te olvides de incluir comentarios en el código.
- **I** **e** tanto el notebook en **e** **f** como exportado a **e** **h**.
- **MUY IMPORTANTE:** Asegúrate de que en el notebook justificas las decisiones tomadas. Por ejemplo, columnas que elimines, sustituciones de nulos que realices, transformaciones de categóricos, etc.

Dispones de dos fuentes de datos distintas:

- Un fichero csv **eeeg** con un conjunto de registros correspondientes a ventas que ha realizado el grupo de empresas. En este fichero encontrarás:
  - **ID Pedido:** identificador del pedido realizado, que puede estar compuesto por ventas a diferentes comercios de tu grupo.
  - **ID Factura:** Identificador de la factura. Cada comercio tiene una distinta.
  - **Comercio:** Nombre del negocio que factura.
  - **Ciudad:** Ciudad del negocio.
  - **Tipo cliente:** Existente o nuevo.
  - **Sexo:** Hombre o Mujer.
  - **Línea de negocio:** sector, con 5 valores posibles.
  - **Precio unitario:** Precio en euros sin IVA de cada unidad del producto.
  - **Cantidad:** Cantidad del producto adquirido.
  - **Total:** Importe total en euros sin IVA.
  - **IVA:** Impuesto del 21% correspondiente.
  - **Fecha:** fecha de la compra.
  - **Pago:** tipo de pago empleado. 3 valores posibles.
- Un registro en una **bbdd MySQL** almacenado en una tabla llamada **MI GMSRI** que tiene los siguientes campos:
  - **ID:** clave primaria.
  - **ID\_pedido:** Identificador del pedido. Se corresponde con alguno de los del fichero.
  - **dir\_entrega:** Dirección postal de entrega del pedido
  - **tipo\_entrega:** tipo de entrega. 3 valores posibles
  - email: dirección de correo electrónico de contacto

Te has reunido con tus usuarios para entender cómo necesitan disponer de la información una vez cargada.

Estos son los requisitos:

- Se desea disponer de la información en un dataframe llamado **h\_cg** que contenga la información del fichero **Dataset-Ventas.csv**. En ese dataframe se debe añadir:
  - el **g\_h** de la dirección que figura en la tabla de base de datos asociada al pedido.
  - El **h** del pedido.
  - El **f\_h** e **eh\_h eh\_gg\_h**. Es la cadena anterior a la “@”. En caso de que no exista correo asociado, pon el valor **“DESCONOCIDO”**
- Este dataframe **df\_completo** solo debe tener como índice el **ID\_Factura**.
- **R\_h\_f** e **G\_g**, ya que no le aporta información. Tampoco debe figurar otra información que sea redundante.
- **e\_g** e **he e g eh\_g** e, no será usada.
- **h\_f** g e los datos para dotar a la información de calidad. Incluye **ee h h geh**, por ejemplo.
- Además de realizar consultas sobre la información, **e e e e** **eh eg P e** con estos datos, con lo que debes preparar de forma oportuna los datos categóricos. **R g e e g**.

Son las columnas **de Ciudad, Tipo de cliente, Sexo, Línea de negocio, Pago y Tipo de entrega**.

Realiza las operaciones oportunas para dar respuesta a los requisitos.

Para obtener el **g\_h** e asociado a una dirección **h e e f e** que ya vimos. En la propiedad **address** devuelta por la consulta, el código postal es el penúltimo elemento.

## IMPORTANTE PARA LA REALIZACIÓN DE LA TAREA:

1. Debes cargar las dos fuentes indicadas. Dispones de:
  - Fichero **Dataset-Ventas.csv** con la información en formato texto.
  - Fichero **MI GMSRI .sql**
2. Conéctate a tu bbdd mysql a través de phpmyAdmin y carga en la bbdd *mibbdd* la tabla DIRECCIONES con todo el contenido. Lo puedes hacer fácilmente tal y como hicimos en clase con la tabla COMPRAS. Pincha en la bbdd y selecciona *Importar*. En esa pestaña puedes subir el fichero e importar. Se creará la tabla con el contenido.
3. Con el fichero **Dataset-Ventas.csv** y la tabla **MI GMSRI** puedes realizar la actividad.
4. En el caso de problemas **MySQL y phpmyadmin** con **MI GMSRI**, pásala a csv con **sqlite** y trabaja con dos ficheros g

## REPRESENTACIONES GRÁFICAS

1. Deberás incluir al menos 5 tipos de gráficos diferentes de los resultados procesados.
2. De los 5, se valorará la elección de al menos 1 que no se haya visto en clase.
3. Alguno de los gráficos deberá incluir datos categóricos y numéricos.
4. Se valorará la elaboración de hipótesis a partir de las conclusiones obtenidas de los gráficos, por ejemplo, *“el número de ventas es mayor en tal ciudad...”* o *“el tipo de pago que prevalece es tal....”*