



Anexo A: Datos de Olist

La siguiente es la relación de los archivos provistos por el equipo de ingeniería de datos de la empresa Olist.

Tabla 1 Archivos del E-commerce Olist para estudio de retrasos en entregas.

Archivo	Descripción	Formato
olist_customers_dataset	Datos de identificación de los clientes y la ubicación del domicilio que tienen registrado con Olist para recibir pedidos.	xlsx
olist_orders_dataset	Datos de identificación de las órdenes realizadas por los clientes, su estatus de envío y la cronología de entrega en el domicilio del cliente.	csv
olist_geolocation_dataset	Datos de geolocalización de códigos postales de zonas de Brasil.	csv
olist_order_items_dataset	Contiene la relación de artículos contenidos en las órdenes de los clientes	csv
olist_order_payments_dataset	Contiene la relación de pagos que cada cliente hizo en sus órdenes por medio de pago	csv
states_abbreviations	Contiene la relación de nombres de los estados de Brasil y sus abreviaciones.	json

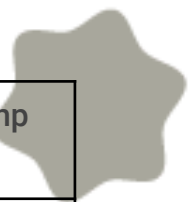
A su vez, el equipo de Olist también proporcionó los diccionarios de datos y el diagrama de entidad-relación correspondiente que describe cómo se interrelacionan las variables contenidas en dichos archivos.

Tabla 2 Diccionario de datos del archivo olist_customers_dataset

Archivo	Descripción	Formato
customer_id	Código identificador del cliente.	Código Alfanumérico
customer_unique_id	Código único de identificador del cliente.	Código Alfanumérico
customer_zip_code_prefix	Código postal del domicilio de cliente, formado por cinco dígitos	Código Numérico (texto)
customer_city	Nombre de la ciudad donde se ubica el domicilio del cliente	Texto
customer_state	Abreviación del estado en donde se ubica el domicilio del cliente	Texto

Tabla 3 Diccionario de datos del archivo olist_orders_dataset

Archivo	Descripción	Formato
order_id	Código identificador de la orden	Código Alfanumérico
customer_id	Código identificador del cliente.	Código Alfanumérico
order_status	Estatus de la orden.	Texto
order_purchase_timestamp	Fecha en la que cliente realizó la compra.	Timestamp
order_approved_at	Fecha en que se aprobó la compra.	Timestamp
order_delivered_carrier_date	Fecha en la que el retailer entregó el paquete de envío al transportista.	Timestamp
order_delivered_customer_date	Fecha real en la que llegó el pedido al domicilio del cliente.	Timestamp



order_estimated_delivery_date	Fecha estimada de la llegada del pedido.	Timestamp
distance_distribution_center	Distancia del domicilio del cliente al centro de distribución más cercano (en kilómetros).	Número real

Tabla 4 Diccionario de datos del archivo olist_geolocation_dataset

Archivo	Descripción	Formato
geolocation_zip_code_prefix	Código postal de la ciudad, formado por cinco dígitos	Código numérico (texto)
geolocation_lat	Latitud donde se ubica la ciudad	Coordenada
geolocation_lng	Longitud donde se ubica la ciudad	Coordenada
geolocation_city	Nombre de la ciudad	Texto
geolocation_state	Nombre del estado	Texto

Tabla 5 Diccionario de datos del archivo olist_order_items_dataset

Archivo	Descripción	Formato
order_id	Código identificador de la orden	Código Alfanumérico
order_item_id	Numeración del artículo dentro de la orden	Número entero
product_id	Identificador del producto	Código Alfanumérico
seller_id	Identificador del vendedor	Código Alfanumérico
shipping_limit_date	Fecha límite para realizar el envío	Timestamp
price	Valor monetario del artículo	Número real
freight_value	Gastos de envío del artículo	Número real

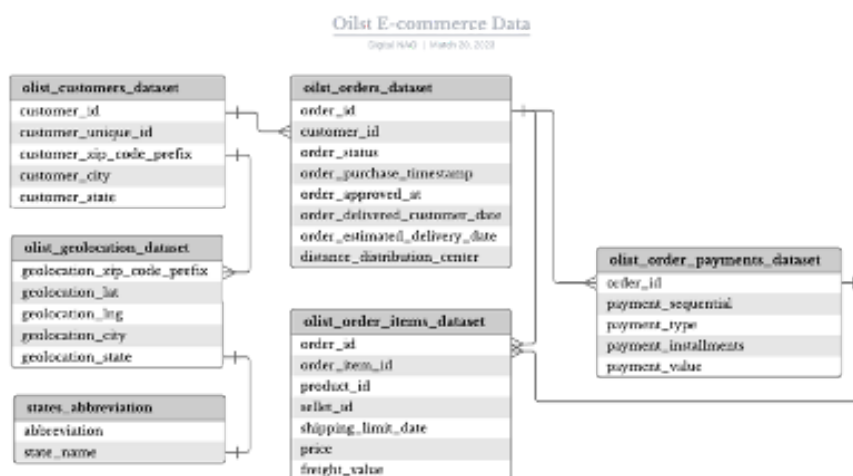
Tabla 6 Diccionario de datos del archivo olist_order_payments_dataset

Archivo	Descripción	Formato
order_id	Código identificador de la orden.	Código Alfanumérico
payment_sequential	Numeración del método de pago usado en la orden.	Número entero
payment_type	Tipo de método de pago usado.	Texto
payment_installments	Se refiere al número de pago que el cliente ha realizado, con relación a las parcialidades permitidas por su método de pago.	Número entero
payment_value	Valor monetario del pago	Número real

Tabla 7 Diccionario de datos del archivo states_abbreviations

Archivo	Descripción	Formato
abbreviation	Abreviatura del nombre de la ciudad.	Texto
state_name	Nombre del estado en donde se encuentra la ciudad correspondiente.	Texto

Figura 1: Diagrama de las relaciones entre los campos de los archivos de Oilst





Anexo B: Trabajando con Jupyter Notebooks

Los archivos de trabajo de este reto están escritos en Jupyter Notebooks (con extensión .ipynb) del proyecto Jupyter (Jupyter Project, 2023), los cuales permiten la ejecución interactiva de código en Python, incluyendo la visualización de imágenes, escritura de texto científico y muchas otras funcionalidades. En este sentido, recomendamos su uso a través del editor de código VSCode, para lo cual describiremos una ruta de instalación y configuración.

Las siguientes instrucciones asumen que la computadora de trabajo tiene instalado Python, o bien, alguna distribución de paquetes de Python para cómputo científico como Anaconda[1]:

- **Paso 1:** Instalar VSCode (<https://code.visualstudio.com/download>)
- **Paso 2:** Una vez instalado, abrimos desde VSCode la carpeta de trabajo donde se localizan los archivos de trabajo con extensión .ipynb. (Archivo > Abrir Carpeta > carpeta de nuestros archivos).
- **Paso 3:** En la barra de herramientas situada en la parte izquierda del editor, seleccionamos el botón de extensiones e instalamos la extensión Jupyter (Figura 2).
- **Paso 4:** Con el mismo proceso, instalamos la extensión Python IntelliSense Pylance de VSCode (Figura 3).

Figura 2

Instalación de la extensión Jupyter en VScode correspondiente al Paso 3.

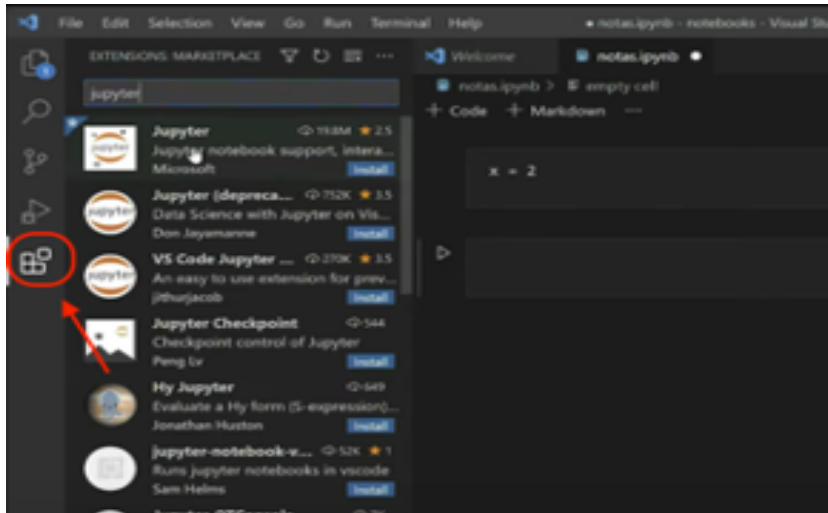
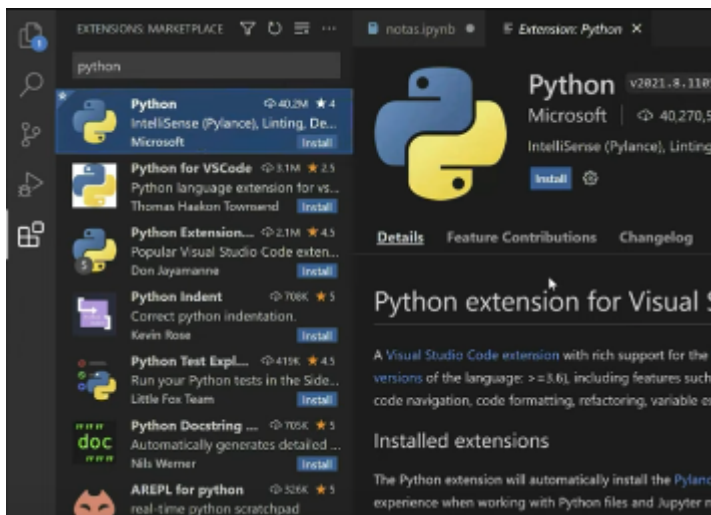


Figura 3

Instalación de la extensión Jupyter en VScode correspondiente al Paso 3.



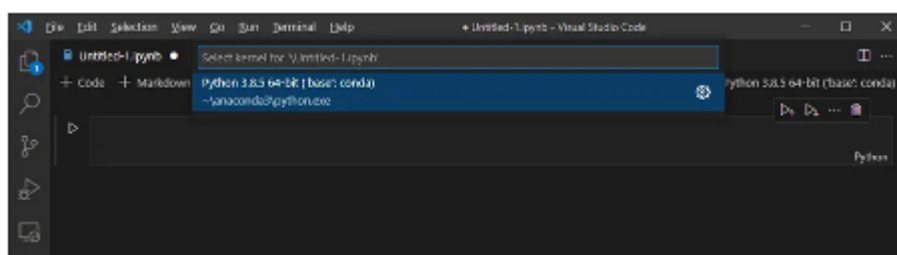
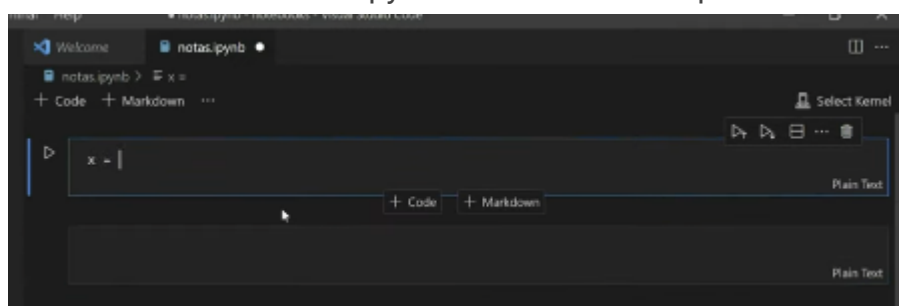
- **Paso 5:** Reiniciamos el editor de código (dar click en ventana OK que sale en parte inferior o cerrar y abrir manualmente).
- **Paso 6:** Tras reiniciar, podemos abrir y visualizar los notebooks de tipo .ipynb. Ahora seleccionamos cualquiera de los archivos de trabajo y demos click en el botón Select Kernel, donde debemos seleccionar el ambiente de Python con el que vamos a trabajar (Figura 7). Nota: si en este punto obtenemos una advertencia acerca de que es necesario

instalar el paquete ipykernel, sólo debemos aceptar las advertencias para instalarlo.

- **Paso 7:** Ahora podremos ejecutar celdas de los Jupyter Notebook de manera interactiva. Para más información consulte la documentación de Jupyter Notebooks en la página de VSCode: <https://code.visualstudio.com/docs/datascience/jupyter-notebooks>

Figura 8

Selección del Kernel de Jupyter en VSCode correspondiente al Paso 6.



Recurso C

Descarga la siguiente carpeta [Recursos externos DN_COM_58](#), para obtener todos los recursos que necesitarás para completar el reto.

Nota. Con base en el documento [La samba del Data Analyst DN_COM_58_material](#), guíate para avanzar en las evidencias de cada ciclo.