Recomendación Ignacio Anguita:

* Cruzar el dataset de precios con ventas🡪 cogiendo el de ventas y hacer una trasposición. Al poder transponerlos cambiaremos los d0, d1, d2, d3… por la fecha actual y conociendo la fecha actual y la semana a la que pertenece esa fecha tendremos información suficiente para cruzarla con el dataset de prices.
* Al cruzarlo con el dataset de prices, podremos crear nuevas columnas que sean: el nombre del ítem y tienda y la revenue que se generó por día. Al tener la revenue que se generó por día, luego lo podemos agrupar por mes y entonces sí que podemos tener un dataset (que podemos subir a BI) donde tenemos las revenues (es decir, las ventas que se han realizado por item y tienda).
* Su recomendación es hacer datasets pequeños e ir subiéndolos a BI.
* Por ejemplo, si queremos hacer una página donde se vean las ventas por tienda, podríamos hacer un dataset más pequeño donde tendremos para cada tienda las ventas por año.
* El ejercicio está pensado para que lo hagamos con Python y con pandas.
* Lo más importante es cruzar precios con unidades vendidas para tener revenues en €. Una vez tengamos eso, lo agrupamos por meses y empezamos a trabajar con ello.
* Para el dashborard lo interesante es ver cuánto se ha vendido por año por cada ítem y encontrar qué ítems se venden más en global, por tienda, ver en que tiendas se venden más. Para eso tenemos que ir creando pequeños datasets.
* Dashboard para septiembre: una primera página introductoria donde se vea la situación de la empresa: cuanto se ha vendido por año, cuánto se vende por tienda, cuánto se vende por departamento, con algún filtro donde se pueda ir jugando.
* Otro dashboard donde sea más por tienda y hacer un slicer donde tú escojas la tienda y otro donde escojas el año y se pueda ver los productos más vendidos, menos vendidos y una serie de gráficos que indiquen la distribución.
* Lo que nos interesa es el cruce de precios por unidades vendidas.
* Analizar los ítems vendidos por tiendas y ver qué productos se venden más.
* Centrar títulos de los dashboards, poner colores, etc ya que esos dashboards son los que usaremos para la presentación.
* Se puede presentar la pantalla en Power BI solicitando la versión de prueba aunque no es un problema que no se pueda poner en versión de presentación.
* Hay un botón para exportarlo a pdf🡪 enviarle ese pdf a mitades de agosto a Ignacio.
* En septiembre deberíamos tener acabado la parte del dashboard y la parte del clustering (segunda parte).
* En el clustering tenemos que encontrar puntos con características comunes. En este caso tenemos que encontrar productos con características similares. De momento nos podríamos crear un dataset donde cada fila fuese un producto y cada columna fuesen cosas relevantes de ese producto, como por ejemplo: el precio medio o el nº de items vendidos o nº de ítems vendidos entre semana y fin de semana.

Cada fila tendría que ser un producto (no producto-tienda, si no, producto). Con ese dataset luego aplicaremos los modelos de ML y ver como se produce el clustering. (Seguramente nos juntará los de precio bajo o los de muchas unidades vendidas) El trabajo interesante será coger ese dataset y añadir el cluster al que pertenece y lo que tendremos que hacer es graficar en power BI e intentar entender como se han creado esos clusters. Por ejemplo, si yo he puesto la variable precio, veremos como está distribuida la variable precio medio en cada cluster.

Después, les podremos dar nombre a esos clusters, por ejemplo: los productos con precio alto y muchas ventas serán los productos estrella, mientras que los productos con precios bajos y pocas ventas serán productos “desastre”.

* El tema de time series lo hablaremos en septiembre.
* Año,mes, producto, tienda y revenue es lo mínimo para empezar a trabajar