

Objetivo

Como desarrollares cuando nos toca trabajar con datos, nos vendrá a la mente lenguajes como Python, R o Julia, los cuales en los últimos años su comunidad ha creado una serie de herramientas que facilitan este trabajo. Sin embargo, la comunidad de F# & .NET no se han quedado con las manos cruzadas y han generado herramientas inspiradas en las existentes en los otros lenguajes.

En este taller el participante aprenderá y conocerá las herramientas, así como librerías que permiten a los desarrolladores de .NET a trabajar con datos. Trabajar con los datos que recolectamos de las aplicaciones, hoy en día nos permite conocer un poco mas a los usuarios y poder hacer su experiencia más agradable, así también permite a las empresas detectar diferentes factores acerca de algún producto o servicio que estas ofrecen.

Como sabemos en un proyecto de Ciencia de Datos existen diferentes fases, por lo que en este taller nos enfocaremos en la fase del Análisis exploratorio de los datos, tal vez en un futuro trabajaremos sobre la fase de Modelado.

Requisitos

Para trabajar en este taller, debemos contar con las siguientes herramientas:

- [.NET CLI](#)
- [Anaconda](#) o el plugin de [.NET Interactive Notebooks](#)
- [VS Code](#)
- Conocimiento [C#](#) o [F#](#)
- Terminal

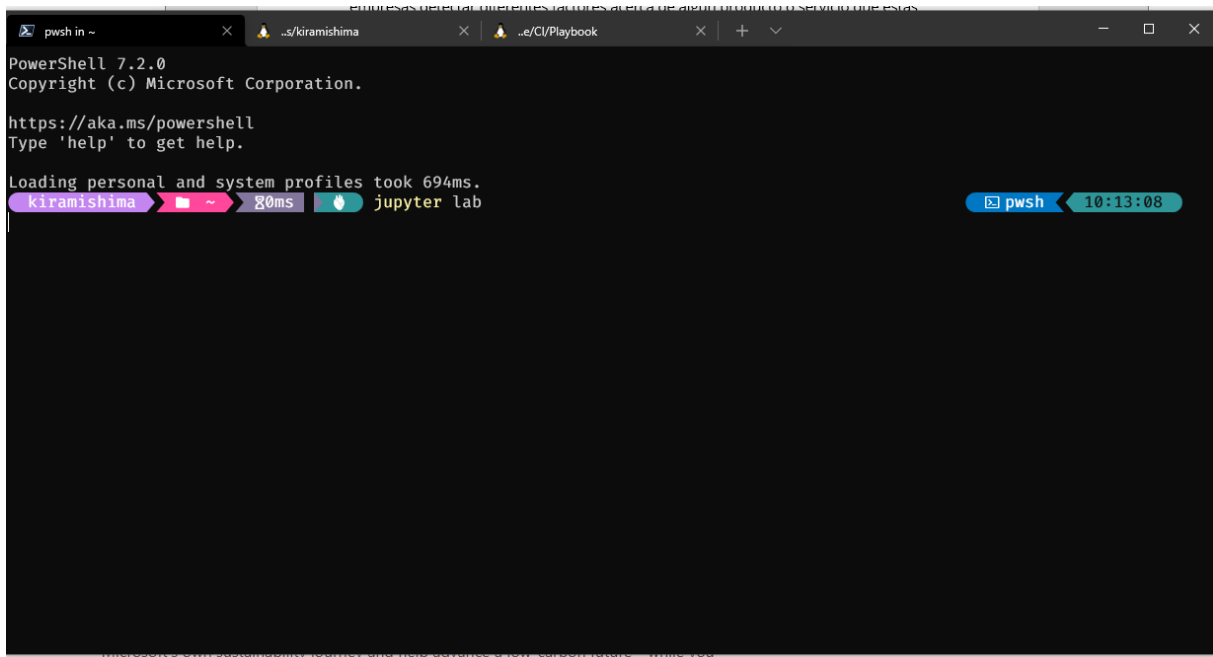
Una vez que instalemos lo necesario es hora de codear

Paso 1 – Crear un notebook de Jupyter o un .NET Interactive Notebook

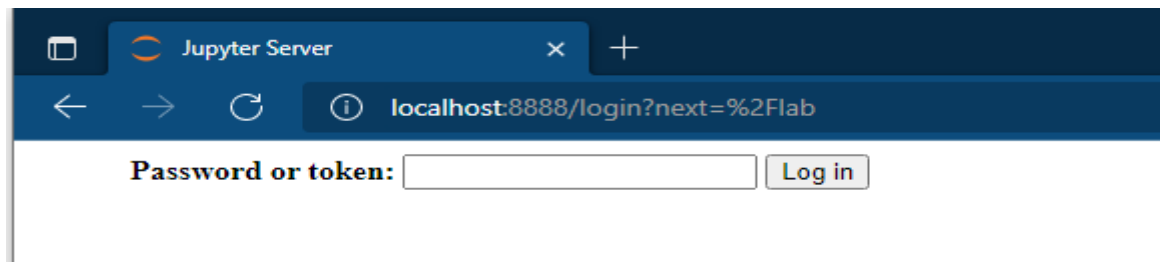
- **Notebook con Jupyter**

Para este paso debemos abrir una ventana de terminal, y ejecutar el comando **jupyter lab** , una vez que se ejecute, nos dará una dirección IP. Si es la primera vez que se ejecuta nos dará un token, el cual debemos de copiar y usarlo cuando se solicite.

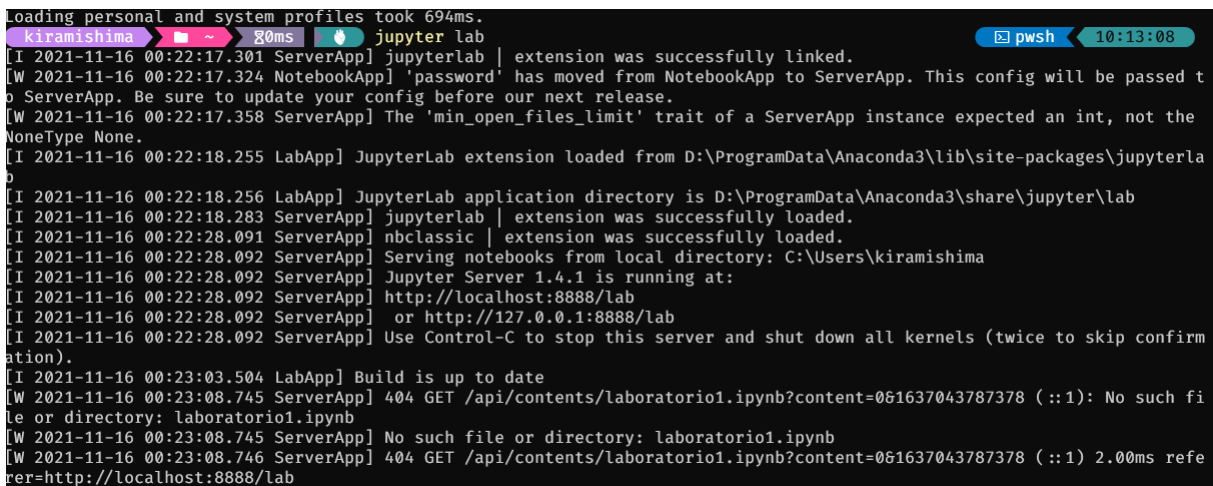
Taller : Notebooks con .NET



Abrimos una ventana de navegador, pegamos la dirección de localhost de la instancia de jupyter y si nos pide un token, pegamos el que nos proporciona Jupyter al levantar el servidor. En caso de que aparezca crear usuario o contraseña, creamos uno y ese será el que esta ligado a nuestra instancia cada vez que ejecutemos el comando.

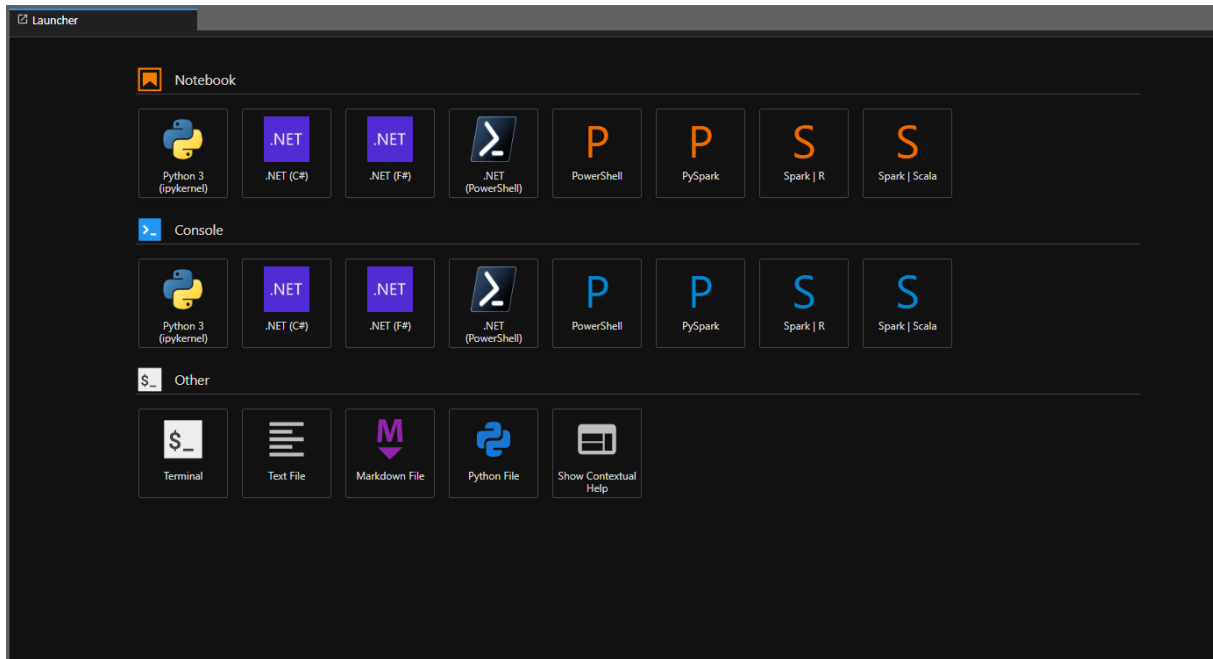


En caso de algún error de no obtener el token o recordar el password asignado, de clic [aquí](#).

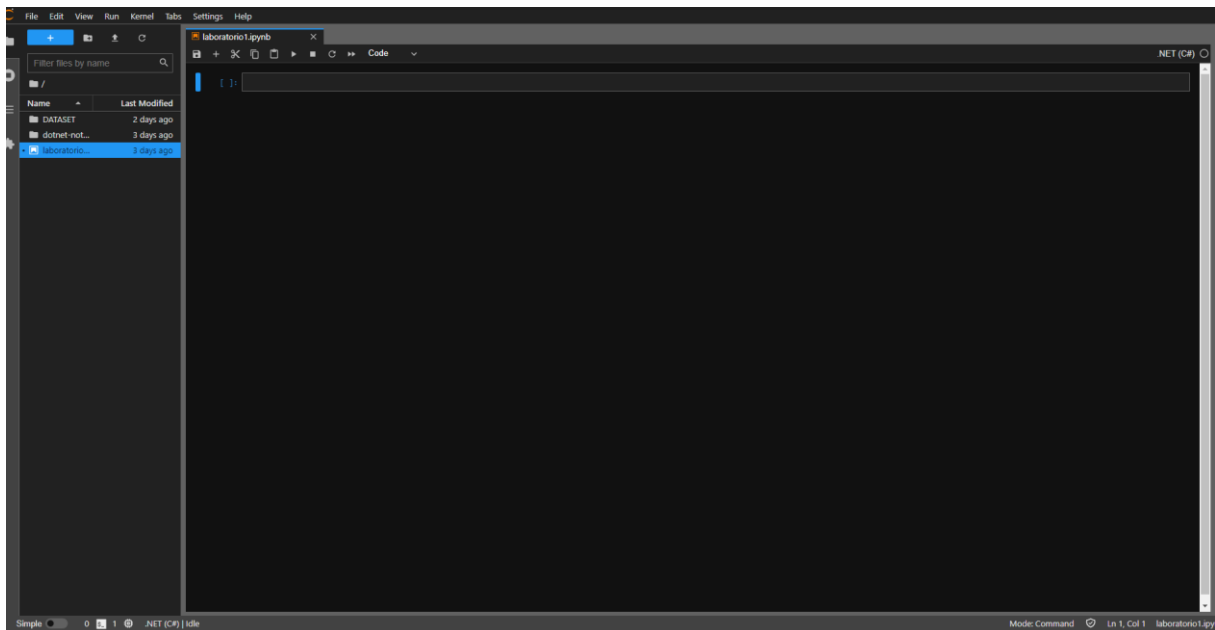


Taller : Notebooks con .NET

Una vez pasando ese filtro de seguridad, veremos la página principal de Jupyter, tenemos 3 secciones: notebooks, console & other. Creamos un nuevo notebook, aquí dependerá si manejan C# o F#.

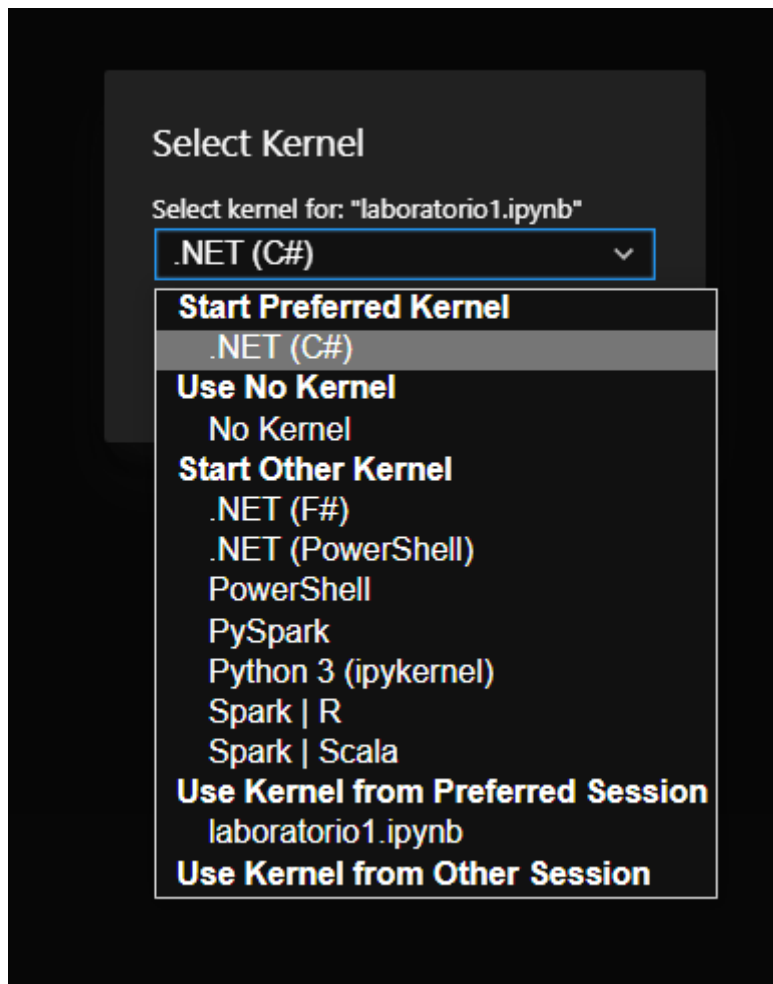
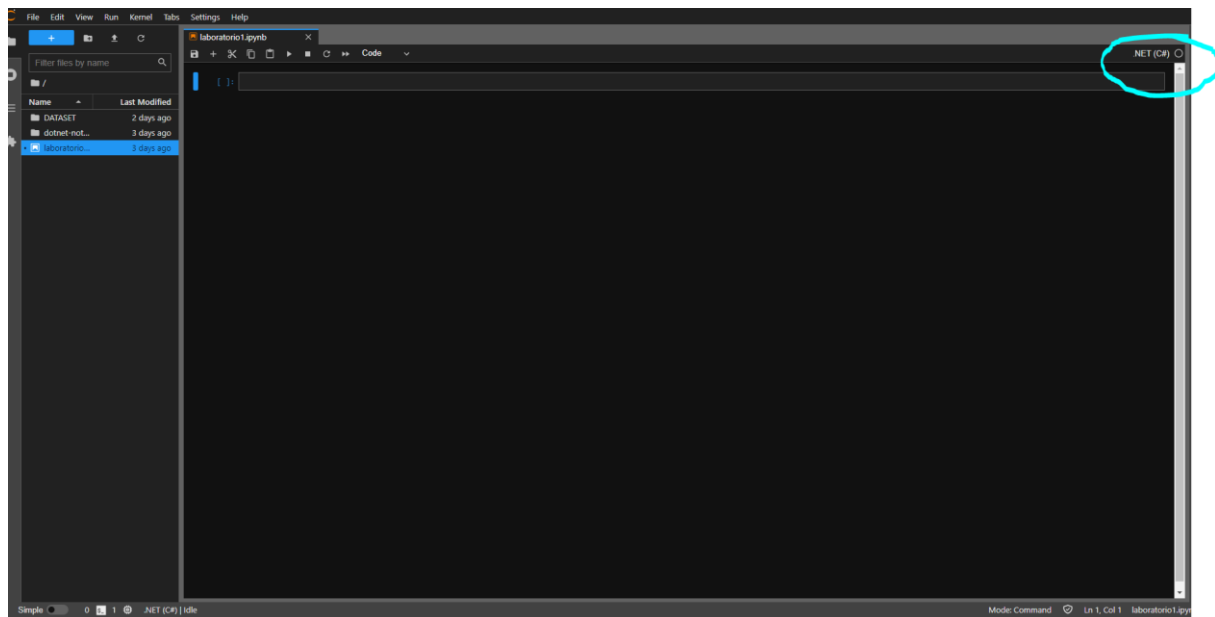


En este caso he creado un notebook con el kernel de F# , dentro se he creado el directorio DATASET, el cual tendrá el archivo data que simulara nuestra Base de datos.



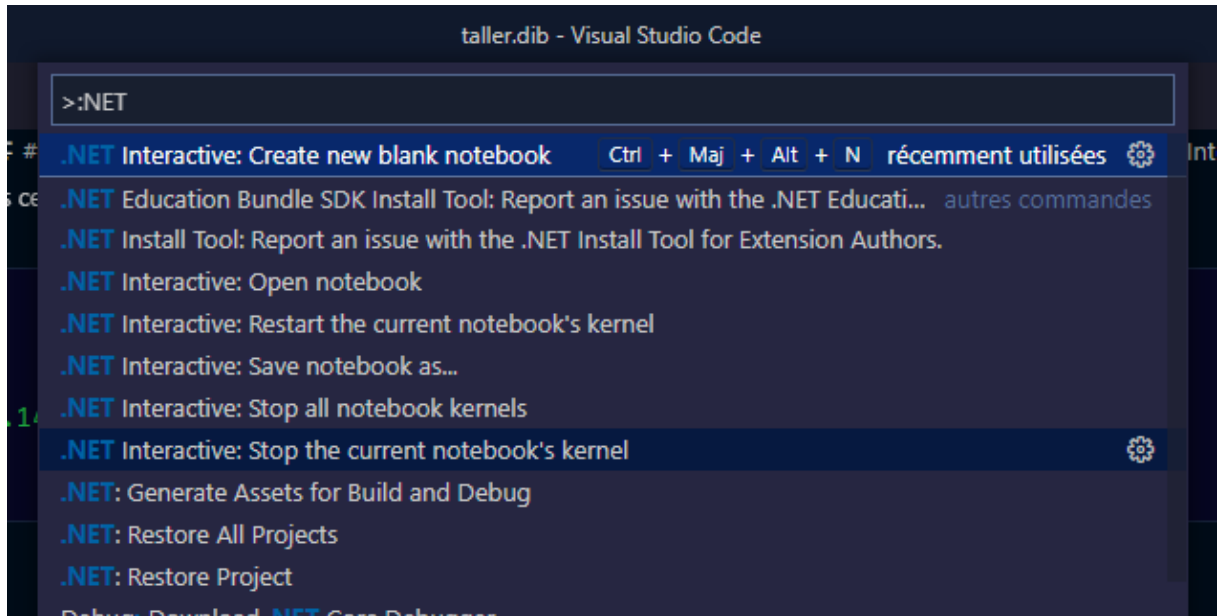
En dado caso que queramos cambiar de kernel en la parte superior izquierda sobre el kernel , le damos clic y eso no abrirá una ventana para cambiar el kernel de nuestro notebook.

Taller : Notebooks con .NET

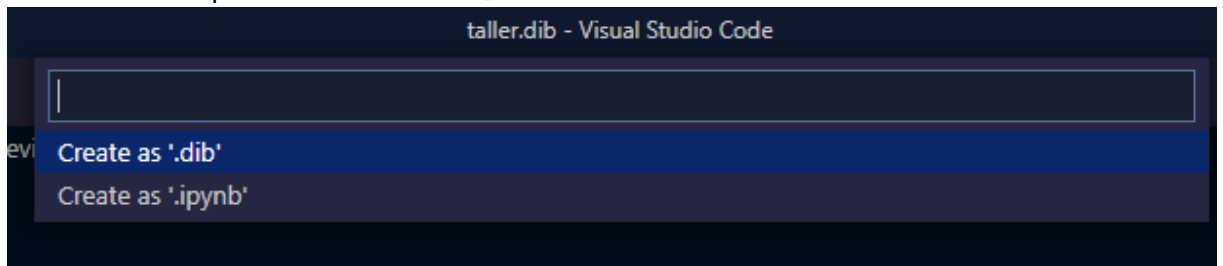


- **Notebook en VS Code con el plugin de .NET It Notebook**

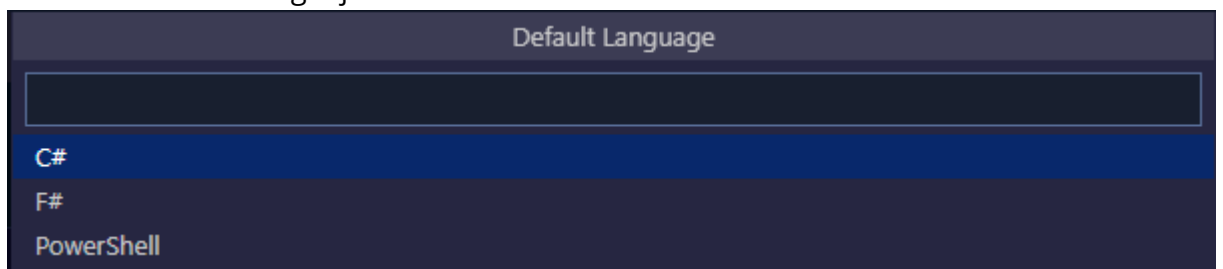
Para crear un notebook usando VS Code, oprimimos la combinación de teclas ALT+SHIFT+P para desplegar le menú de opciones. Escribimos .NET y nos saldrá un listado del cual seleccionamos **Create new blank notebook**.



Esto iniciara el proceso de creación, seleccionamos Create as '.dib'

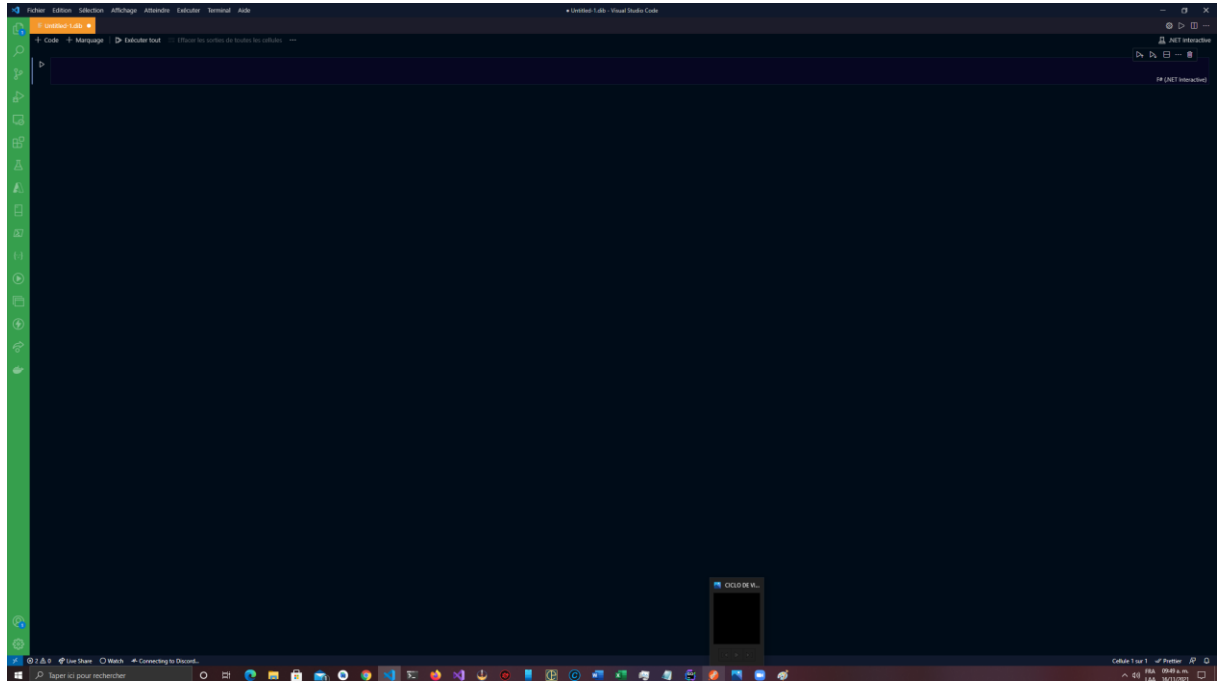


Seleccionamos el lenguaje: C# o F#



Taller : Notebooks con .NET

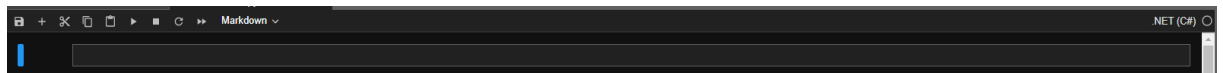
Esto creara el notebook, le damos guardar y seleccionamos donde guardar dentro de nuestro equipo, donde guardemos este notebook, crearemos el directorio DATASET el cual tendrá el archivo data que simulara nuestra Base de datos.



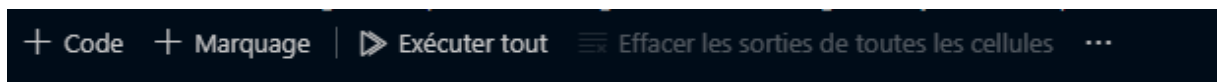
Paso 2 – Celdas Markdown & Ejecución

Este paso solo describiremos un poco la diferencia entre las celdas de ejecución y de Markdown, ya que hay ocasiones en que el participante llega a confundirse al querer ejecutar su código en la celda.

Si trabajamos con Jupyter, tendremos una barra de herramientas internamente en el notebook. Tendremos un icono de Guardar, Agregar Celda, Cortar Celda, Copiar Celda, Pegar Celda, Ejecutar la celda seleccionada, Detener, Reiniciar el Kernel y Reiniciar el Kernel y ejecutar todo el notebook, y finalmente un select con las opciones de Code o Markdown.



En Jupyter por default al crear una nueva celda en el notebook serán de tipo Code, pero si lo que deseamos es agregar una descripción de lo que trata el notebook debemos cambiarla a tipo Markdown seleccionando la celda y seleccionando el tipo en el select de la barra de herramientas interna del notebook.



En VS Code es más fácil de crear los dos tipos de celda, ya que el menú es mas sencillo, tenemos la opción de Code, Markdown, Ejecutar todo & Limpiar las celdas.

Paso 3 – Problematica: Empresa Tailwind Sports & Redes Sociales

La empresa **Tailwind Sports SA de CV**, tiene 7 años en el mercado de la venta de ropa deportiva en línea, en ese tiempo ha logrado posesionarse como una empresa seria y comprometida con la calidad y precio de sus mercancías. Fue fundada por Barry Allen y Pedro Maximoff, quienes vieron la tendencia e innovación del comercio en línea, como una oportunidad para poder generar ingresos.

Los primeros años, fueron años de adaptación y reconstrucción continua tanto del diseño de la mercancía como de la tecnología, ya que ambos factores tienden a un cambio constante que demandan una presencia dentro del negocio.

A lo largo de un año ha visto que sus ventas tienen una tendencia alcista, pero también ha notado que hay un porcentaje de devoluciones importante, la empresa desea inferir a que se debe este porcentaje, para ello lo contrata a usted y su equipo para que con base en los datos proporcionados por la compañía, se pueda determinar la causa a dichas devoluciones y puedan tener la

seguridad en la comercialización por éste medio y le pide, realiza un Análisis de la información (Analítica de negocios) para interpretar los resultados.

La empresa a lo largo de sus años ya establecida cuenta con diferentes canales de distribución para la venta, utiliza dos medios de colocación de la mercancía, el propio, que consiste en su página Web y los medios necesarios para concertar la venta, además cuenta con una red de personas que se dedican a promocionar la mercancía por un porcentaje como comisión, estas personas se les denomina afiliados y están integrados en su mayoría por influencer.

Los medios de distribución, los integra las plataformas Facebook, YouTube, Instagram y el Website.

De acuerdo a nuestra base de datos, los ingresos los conforma las compras, que dan origen a las ventas, pero además a este rubro lo integran las devoluciones que a lo largo del año 2017 han sido representativas. Siendo este punto un aspecto importante a considerar ya que desde el punto de vista económico, las devoluciones impactan de manera directa en el objetivo económico de las organizaciones, haciendo que los ingresos presentados en la base de datos se vean disminuidos porque por un lado las ventas generan entradas de dinero real y por otro lado las devoluciones tienen un doble impacto, el primero tiene que ver con las mismas ventas, al devolverse la mercancía ésta ya no se puede comercializar de la misma forma y por otro lado esto implica la devolución del dinero que ya se contemplaba como parte del efectivo que a la empresa le da liquidez para seguir operando con su ciclo financiero.

Los clientes de la organización están clasificados en dos tipos: clientes frecuentes o repetidos y clientes nuevos, además la mercancía está dirigida al sexo femenino y masculino. Por tanto, los productos que se comercializan van enfocados al género en específico.

Los artículos que se comercializan son: pantalones, chamarras, camisas, playeras, blusa, calcetines, leggings, bra, entre otros...

La base de datos está compuesta por:

Orden	Fecha	Medio	Vendedor	Plataforma	Comisión	Tipo Orden	Tipo de Cliente	Sexo	Categoría	Producto	Ingres
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Mujeres	Chamarras y Sudaderas	Sudadera Slouch Crop top		\$42.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Mujeres	Leggings	Sin costuras alta cintura a 3/4		\$54.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Hombre	Interior	Camiseta Onyx Imperial		\$50.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Mujeres	Camisetas y Tops	Camiseta esencial		\$24.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Hombre	Pantalones	Pants Fallout		\$40.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Mujeres	Bra deportivo	Bra bloque tonal		\$35.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Mujeres	Chamarras y Sudaderas	Sudadera Slouch		\$42.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Hombre	Interior	Camiseta Onyx Real		\$40.00
63	20/01/2017	Influencer	reginaacosta314	Youtube	30% Compra	Repetido	Hombre	Sudaderas y chamarras	Chamarras de pista Gravity		\$55.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Hombre	Interior	Shorts Interiores Element		\$35.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Mujeres	Camisetas y Tops	Camiseta Carbon mesh		\$34.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Mujeres	Leggings	Sin costuras alta cintura		\$54.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Hombre	Pantalones	Pants Onyx		\$48.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Mujeres	Shorts	Shorts Flex		\$34.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Hombre	Pantalones	Pants Onyx		\$48.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Hombre	Pantalones	Pants de Entrenamiento Reactivo		\$40.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Mujeres	Ropa Interior	Tanga sin costuras		\$15.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Hombre	Camisetas	Camiseta Respiro		\$45.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Mujeres	Leggings	Leggings Ombre sin Costuras		\$40.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Hombre	Camisetas	Camiseta Heather		\$50.00
79	28/01/2017	Propio	Publicidad Face	Facebook	0% Compra	Repetido	Hombre	Sudaderas y chamarras	Sudadera con capucha Imperial		\$60.00

Nuestro objetivo para este taller será presentar un mini reporte de acuerdo al análisis de la información en donde se presente:

- Importe de la venta total
- Integración de las ventas totales (compras & devoluciones)
- Importe y porcentaje de las devoluciones totales
- Graficar las ventas mensuales
- Comportamiento de las ventas en las diferentes plataformas

Laboratorio:

Para ver poder trabajar con nuestra información, vamos a utilizar las siguientes librerías:

- **Fsharp.Data**
- **Deedle**
- **Plotly.NET**
- **Newtonsoft.Json**

1. Cargando Datos a un Dataframe de Deedle

Para cargar nuestro CSV a un dataframe, usaremos Deedle, para ello debemos utilizar la sintaxis:

F#

```
Frame.ReadCsv(archivo.csv, separators=";", hasHeaders=true)
```

C#

```
Frame.ReadCsv(archivo.csv, separators:";", hasHeaders:true)
```

Para consultar los 5 primeros registros de nuestro Dataframe, utilizaremos la propiedad **Rows**, el cual contiene el método **Between**, al cual le pasamos el índice desde donde queremos comenzar y el índice al cual queremos llegar y ejecutamos el método **Print** para ver la respuesta.

```
df.Rows.Between(0, 5).Print()
```

Para ver información sobre las columnas que cargamos a nuestro Dataframe, y el tipo de dato, usamos la propiedad **Columns** y su propiedad **Keys**.

C# / F#

```
df.Columns.Keys
```

2. Operaciones con Dataframes: **Stats**

Deedle posee la clase Stats con la cual contiene las operaciones de estadística mas populares: sum, median, average, quantile, variance, etc.

F#

```
df.["Columna"] |> Stats.sum
```

C#

```
Stats.sum (df["Columna"])
```

3. **Filtrar & Agrupar** en un Dataframe

Para filtrar información sobre un dataframe de Deedle , podemos usar el método FilterRowsBy el cual recibe la columna a la cual deseamos aplicarle el filtro y el valor a filtrar, este nos regresara un dataframe con los valores filtrado al cual por ejemplo le podemos aplicar alguna operación después.

F# / C#

```
df.FilterRowsBy("Columna", "Valor a Filtrar")
```

También podemos usar llamar directamente desde el módulo Frame, el método filterRowValues, que nos permite pasarle una función y filtrar sobre mas columnas. En C# podemos usar Where que es similar a Linq.

F#

```
df |> Frame.filterRowValues (fun row -> row.["Columna1"] = "Valor1" ||  
row.["Columna2"] = "Valor2")
```

C#

```
df.Rows.Where(kvp => kvp.Value.GetAs<tipo de dato>(" Columna1") == " Valor1" ||  
kvp.Value.GetAs< tipo de dato >(" Columna2 ") == " Valor2");
```

4. Graficas con Plotly.NET

Para graficar nuestros resultados, podemos usar Plotly, librería la cual contiene varios tipos de gráficas, las cuales son fáciles de crear. A continuación, se muestran la sintaxis básica para la gráfica de columnas o de barras como comúnmente se les conoce.

F#

open Plotly.NET

```
let x = [| 10; 20; 30; 40; 50 |]
```

```
let y = [| "A"; "B"; "C"; "D"; "E" |]
```

```
Chart.Column(y2,x2)
```

C#

```
using Plotly.NET;
```

```
using Microsoft.FSharp.Core; // use this for less verbose and more helpful intellisense
```

```
using Plotly.NET.LayoutObjects;
```

```
var x2 = new Int16[] { 10, 20, 30, 40, 50 };
```

```
var y = new string[] { "A", "B", "C", "D", "E"};
```

```
Trace trace = new Trace("column");
```

```
trace.SetValue("x", x2);
```

```
trace.SetValue("y", y);
```

```
trace.SetValue("name", "Hello from C#");
```

```
GenericChart
```

```
.ofTraceObject(true,trace)
```

```
.Show();
```