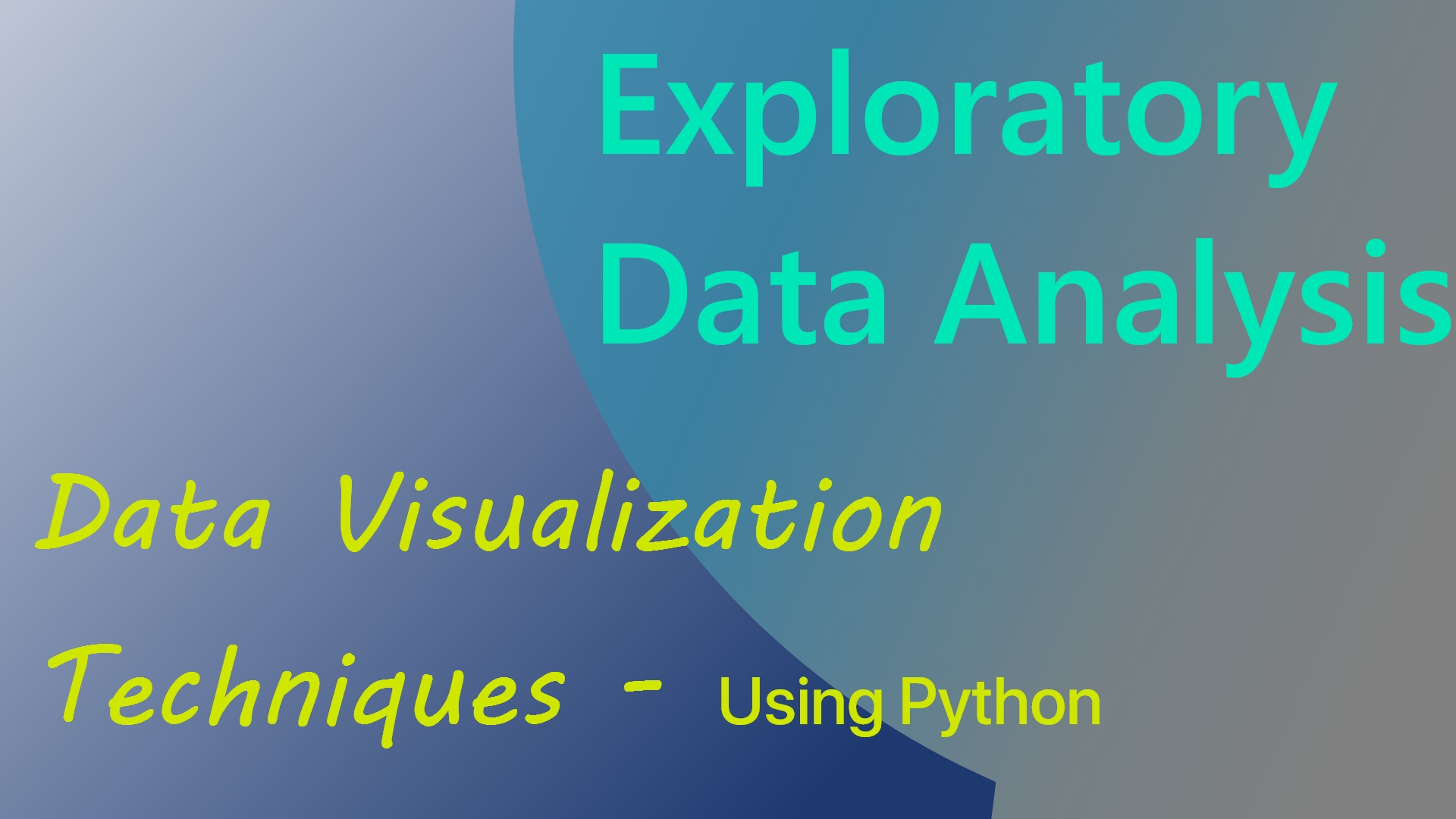
**Exploratory Data Analysis Using Data Visualization Techniques**



Exploratory Data Analysis or **EDA** refers to the process of analyzing and summarizing datasets to gain insights into the data. This helps in understanding the data and identify patterns, relationships and anomalies.

EDA is a crucial step in any data analysis project. Some of the techniques used in EDA include Python, R, SQL, Excel

The objectives of EDA are to:

* Enable unexpected discoveries in the data
* Suggest hypotheses about the causes of observed phenomena
* Assess assumptions on which statistical inference will be based
* Support the selection of appropriate statistical tools and techniques
* Provide a basis for further data collection through surveys or experiments

To ease more on analyzing data Python is discussed more in details.

**Why Python for EDA?**

* Python is a general yet powerful and versatile programming language widely used in Data Analysis.
* Python is rich in libraries and tools that make it easy to perform the EDA tasks. To mention these libraries are Pandas, NumPy, Matplotlib, Seaborn and Plotly – to be discussed further.
* Python is easy to learn and to be used in EDA tasks hence it is an ideal choice for beginners and experts alike.

**Introduction to Data Visualization**

Data Visualization helps in the communication of insights derived from complex datasets.

Python is preferred to do the task since it comes in handy with libraries that enables a Data Analyst or Data Scientist to create visualization for the data prepared.

The libraries are **matplotlib,** **seaborn,** and **plotly** help convey the findings effectively.

Matplotlib helps to create a 2d plot in Python.

Plotly helps to offer interactive capabilities to visualizions.

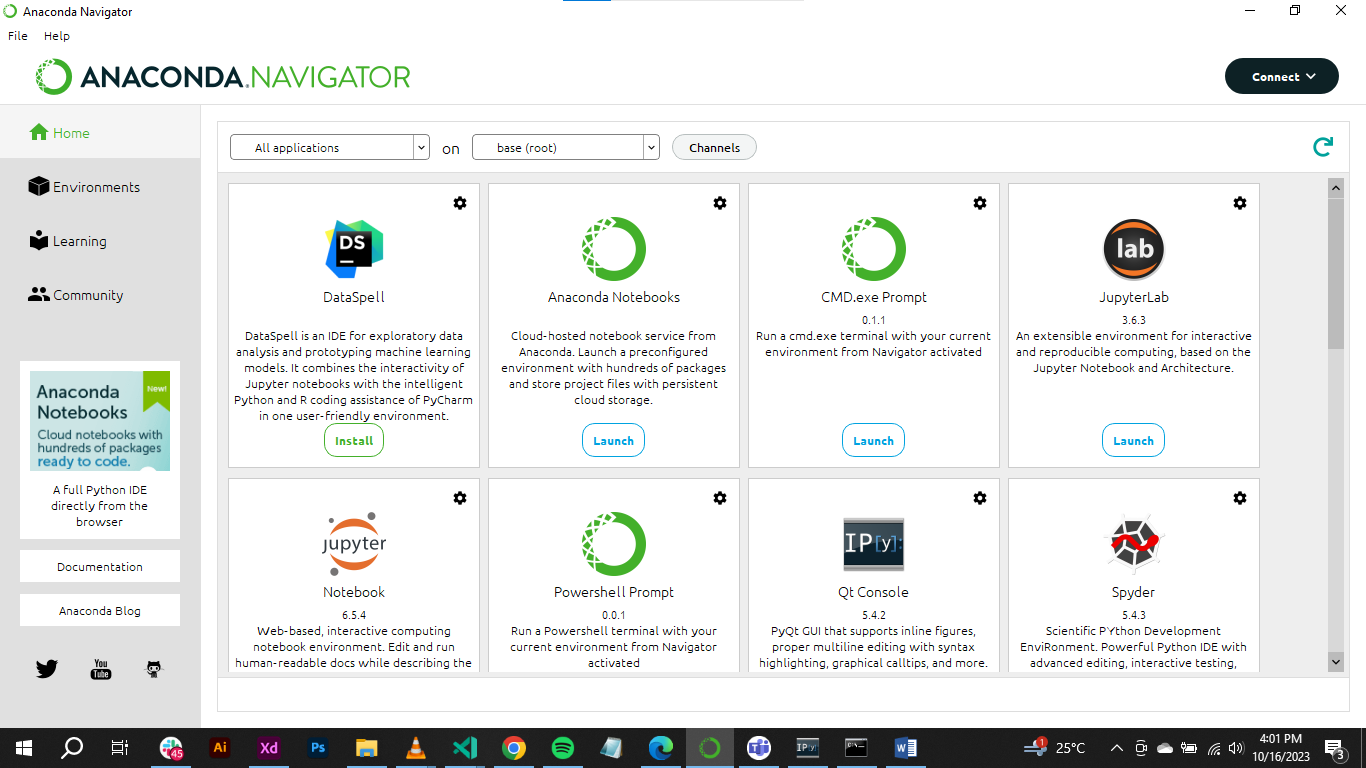
**Steps in EDA**

To conduct EDA appropriately there are steps to be followed to ensure the Data Scientist in the end has a clean data. The steps are:

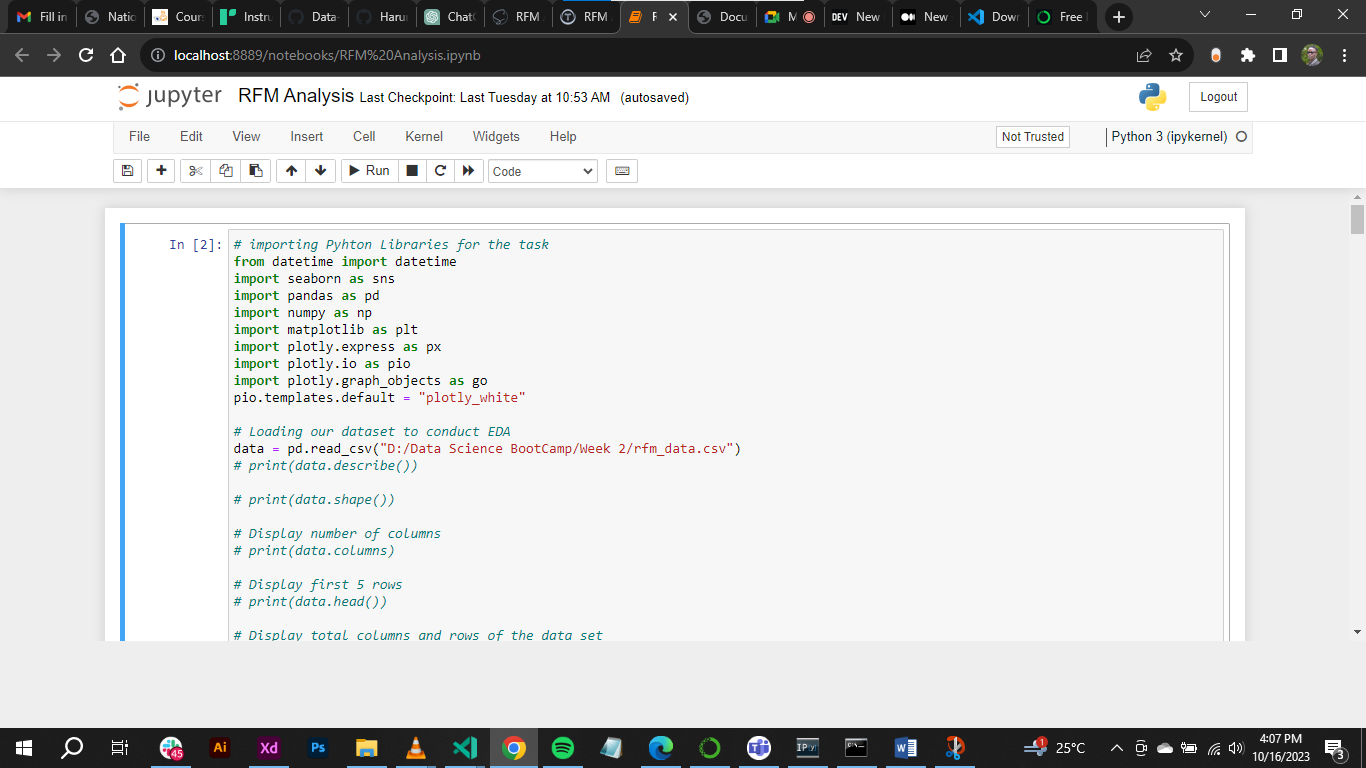
Before we dive into EDA lets prepare our environment.

Note: I used VS Code studio and Anaconda to conduct the Exploratory Data Analysis for the dataset provided. Also, your PC needs to have a core i5 processor and at least 8GB RAM due to high data analysis processing power requirements.

Download VS code [**here**](https://code.visualstudio.com/download/) and Anaconda [**here.**](https://www.anaconda.com/download)

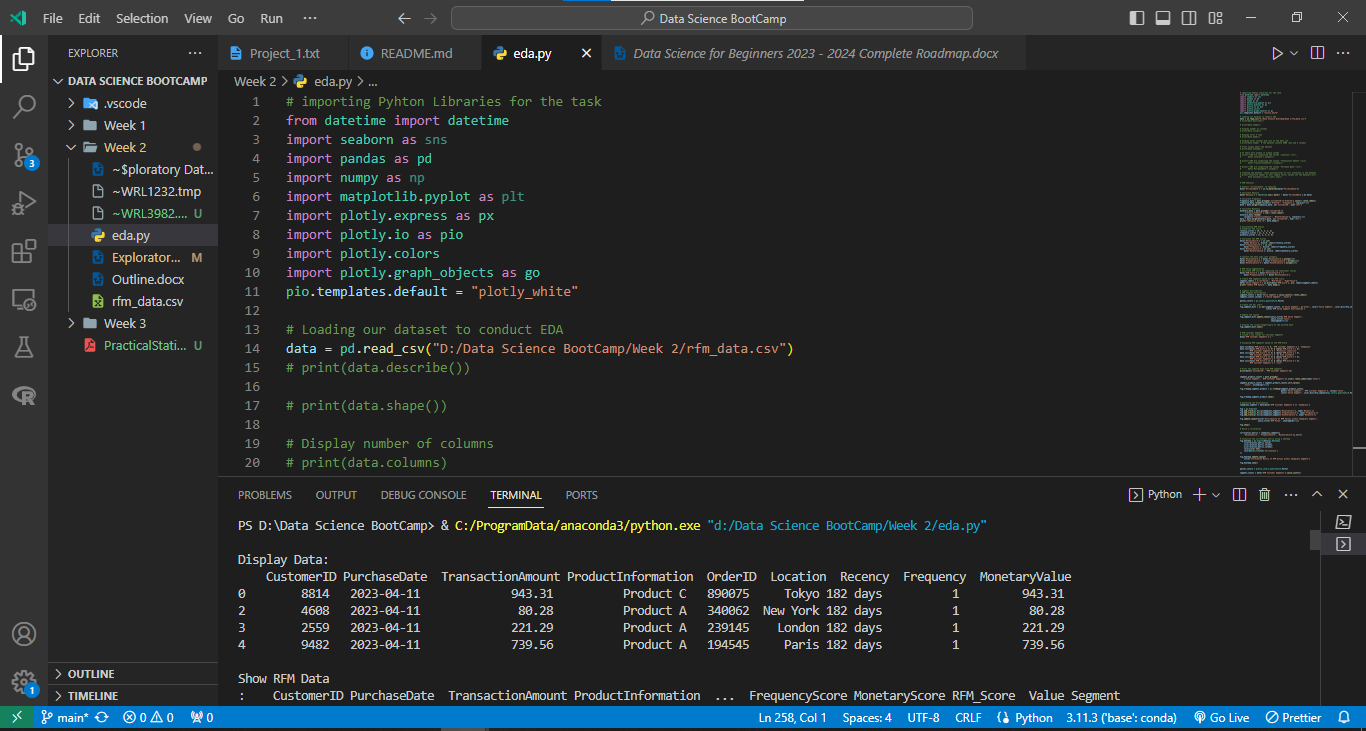
After Downloading Anaconda and installing the interface should appear as follows: 

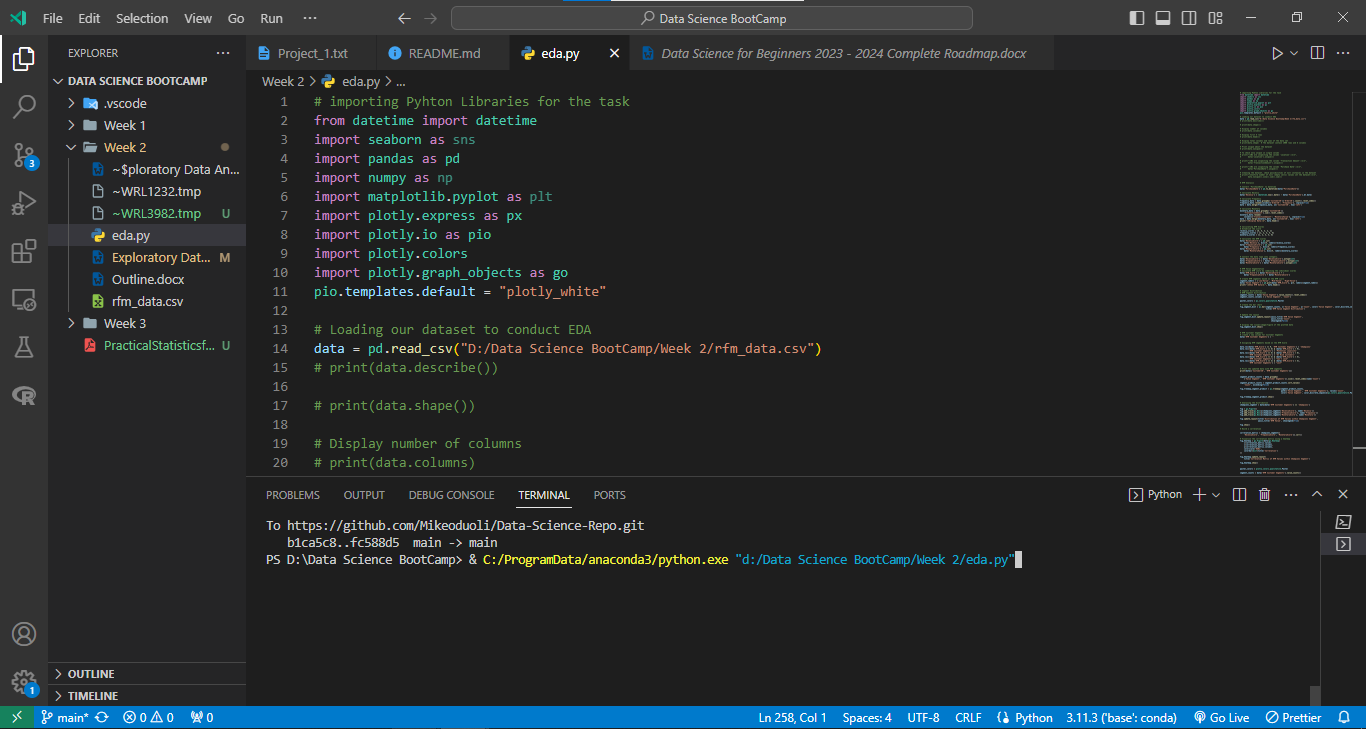
Now click **Launch** - ***Notebook(Jupyter)***. The interface of Jupyter is as shown below.



Now the remaining part is to conduct the EDA process.

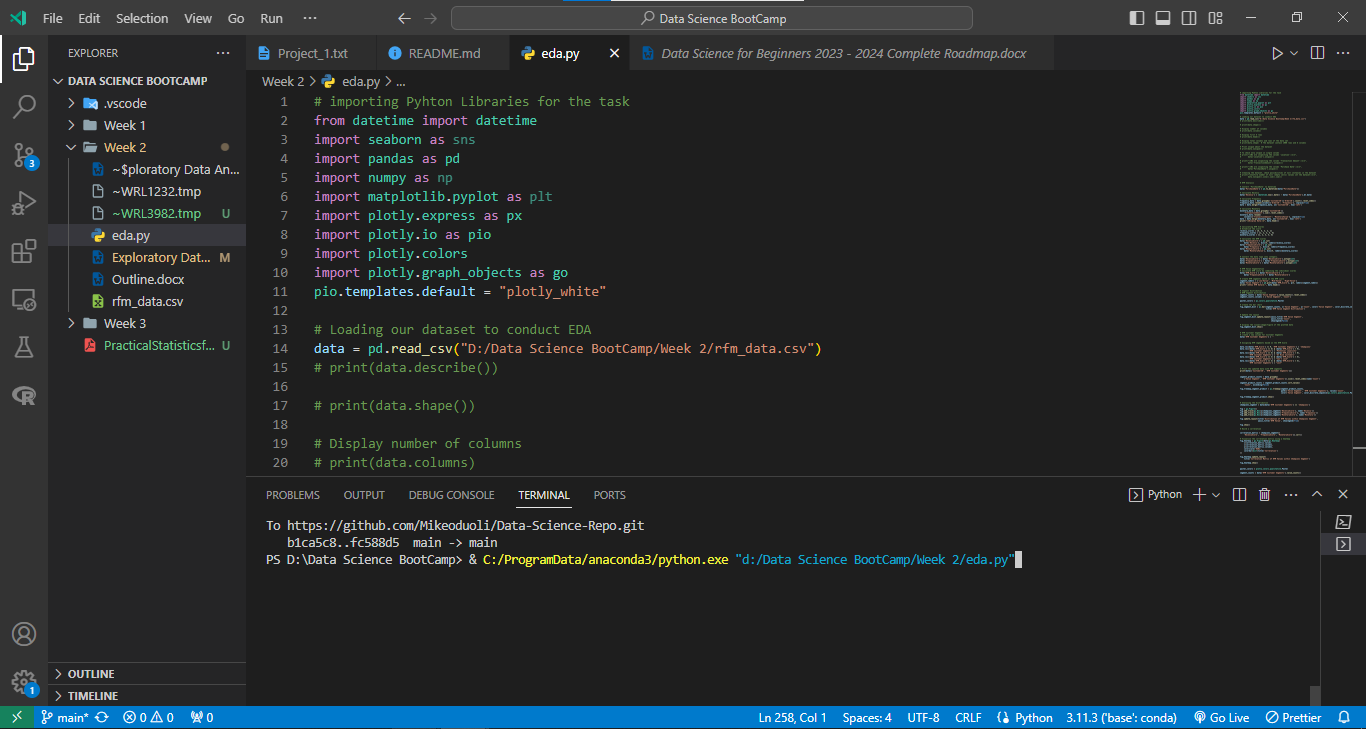
To Use VS Code – after downloading. Create a folder for your project.



To run the Python file open terminal via keyboard shortcut **CTRL + `** (called back-tick) or locate the 3 dots on tab bar. 

Run this command **C:/ProgramData/anaconda3/python.exe "d:/Data Science BootCamp/Week 2/eda.py" (replace path with your own path to the file where scripts are written)**

1. Understand the data at hand – the dataset to be analyzed. This can be done by first importing the Python Libraries.



1. Clean the data using the imported libraries. This is done by inspecting the data.

Dirty data can contain null values, duplicates, inconsistent column names, spaces in between the data.

1. Then analyze the relationship in data variables.