Eduardo Granados

Data Frames

The main library I used in this program was Pandas with a little Matplotlib, but mainly Panda. Although I wanted to use Seaborn instead of Matplotlib, but for some reason, some of the graphs didn’t want to work on my terminal, but we working on my Jupyter Notebook so I’m not sure. Although I was still able to do what I wanted, but I would’ve liked to show more with regression lines and other plotting tools. The purpose was I wanted to learn more about the first steps of being a data scientist. See what its like working with big data sets.

Pandas is great when it comes to analyzing data, it can be used in many ways. Pandas can take many forms of data, CSV, SQL Databases, anything that goes on an Excel sheet. One way it could be used is in machine learning. Visualizing the data, then seeing if there are certain patterns or trends to learn from.

**df = pd.DataFrame({"Population": [909976, 8615246, 2872086, 2273305],**

**"State": ["Sweden", "United kingdom", "Italy", "France"]},**

**index=["Stockholm", "London", "Rome", "Paris"])**

Above is a line of code using Pandas. This line is creating a dataframe and setting it equal to **df**. We can see how a dictionary is used for assign values to **Population** and **State**. We can also see that in that same line we set index values for our dataframe.

**df\_pop = pd.read\_csv("european\_cities.csv", delimiter=",",encoding='unicode\_escape', header=0)**

The line of code above is now reading in a larger dataset then setting it equal to **df**.

df\_pop["State"] = df\_pop["State"].apply(lambda x: x.strip())

The line about shows how a lambda function can be applied to pandas.

Documentation link:

https://pandas.pydata.org/docs/