ETINOSA EGHAGHE

National College of Ireland,
MSc in Data Analytics
New York City Shooting Incident
X23138548@sudent.ncirl.ie

Personal Task Journal

1. Did the preliminary analyse compare contrast and evaluated rate of crime in New York City and tried to analyse preference area regard to crime rate.
2. Choose to analyse crime and decided to select data regarding shooting in NYPD
3. The dataset was in Json format taking from DATA.GOV and it a record of shooting in New York City from 2006 to 2022
   Dataset Link: https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic
4. Created a function to import a JSON file from my device to python.
5. Create a data base in MongoDB Compass to be later use for collection of data and name it nypd-shooting-incident.
6. Programmatically connected to MongoDB from Jupiter notebook. Using Connection String of MongoDB connected and tried to insert Json dataset was not successful by insert many functions, so created a loop and then inserted Json data using insert many functions of Mongo Client libraries. Using collection checked if the data was present in database.
7. Using find function fetched the data back from MongoDB and add to data frame.
8. Transformation process begins, data cleaning and exploration. For project this involves
   - checking the shape, head, and columns of the data
   - checking and removing missing, null, unknown value and duplicates rows
   - dropping of columns not needed
   - changed and converting of variable types.
9. next connect to localhost PostgreSQL server using psycopg2 sqlalchemy libraries from Python, by creating a connection string which include creating cursor. Table, Convert Data Frame to SQL and insert into the table. After storing the connection is closed
10. Data imported back from PostgreSQL and fetched into data frame for visualisation.
11. Visualisation process using seaborn, NumPy pandas, plotly. express, matplotlib.pyplot, the visualization included plotting of px.scatter_mapbox, Scatter plot on a map, Plot a pie, bar plot
12. Converted to CSV file and save in my device.
13. Further analysis perform that is applying the knowledge from machine leaning. A multiple **linear regression model** built with some of the variable to get a more insight.
14. In conclusion Brooklyn has the highest susceptible to shooting incident. In summer between end of May to ending of August in New York consistently experiences high danger due to increase in shooting incident in these months. For further analysis analyst can include all feature and analyse deeper to get more insight, this can enable a better understanding between incident and population trends.
15. Contributed to writing part of the report.