Segmenting and Clustering Crimes Against Females in New York City

Amr Elrasad

Outline

- Introduction
- Data
- Methodology
- Results & Discussion
- Conclusion

Introduction

- This project is utilizing the power of data science and machine learning to infer a safety maps of the New York City for women.
- The project uses the data of incidents reported to New York police department to analyse the relation between police precincts locations and the sex of the victim.
- In this project, I focus on female sex to be able to come up with safety map showing how danger or safe each precinct is in New York City.

Business Interest

- The project idea is interesting for many stakeholders such as:
- Mobile apps development industry: this looks like a good app on mobile phone for each woman to avoid danger spots in the city. I can be also helpful for tourists.
- Police department: the information inferred by this study will help the police department to better re-distribute and organize the patrol service
- Employers: this would help employers better understands the risks of their business locations in relation to female employees.
- Business Owners: they would utilize this information to direct their advertisement towards women if their business location is in the nearest areas.
- Housing for females: the information might help planning for female students/employee housing developments to choose best spots of such projects.

Methodology

- Data Analysis
- Foursquare API
- Clustering

Methodology- Data Analysis

- Read data frame
- Choose relevant information
- Choose victim sex
- One hot encoding
- Scoring
- Adding empty clusters
- Cleaning repetitions
- Grouping

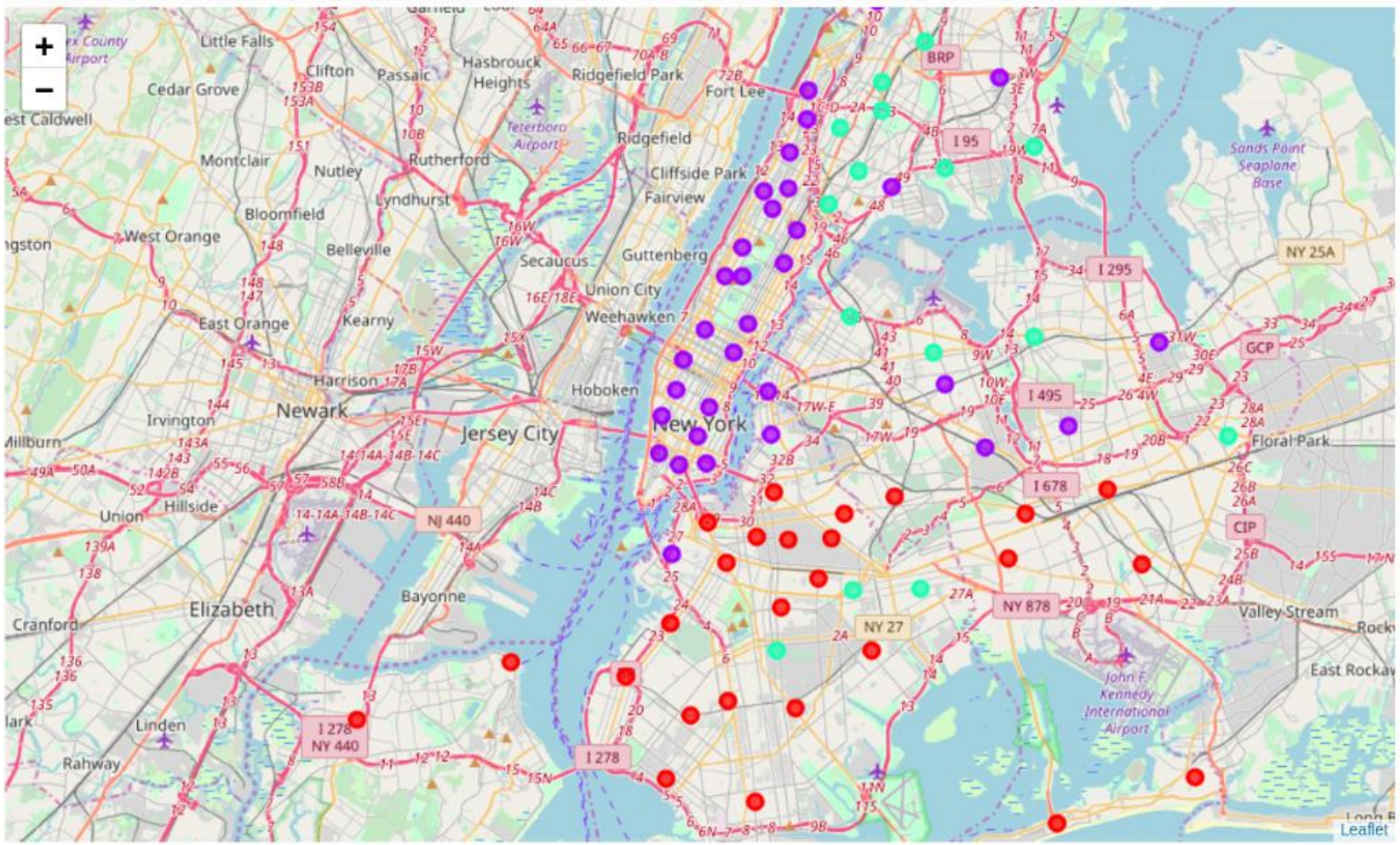
Methodology- Foursquare API

- Foursquare API is used to get the location coordinates of each precinct.
- A list of all of them is created on my Foursquare account.
- I get list id query to get list details.
- A loop is used to parse list and nested dictionaries to get each list item location and updates the corresponding data frames.

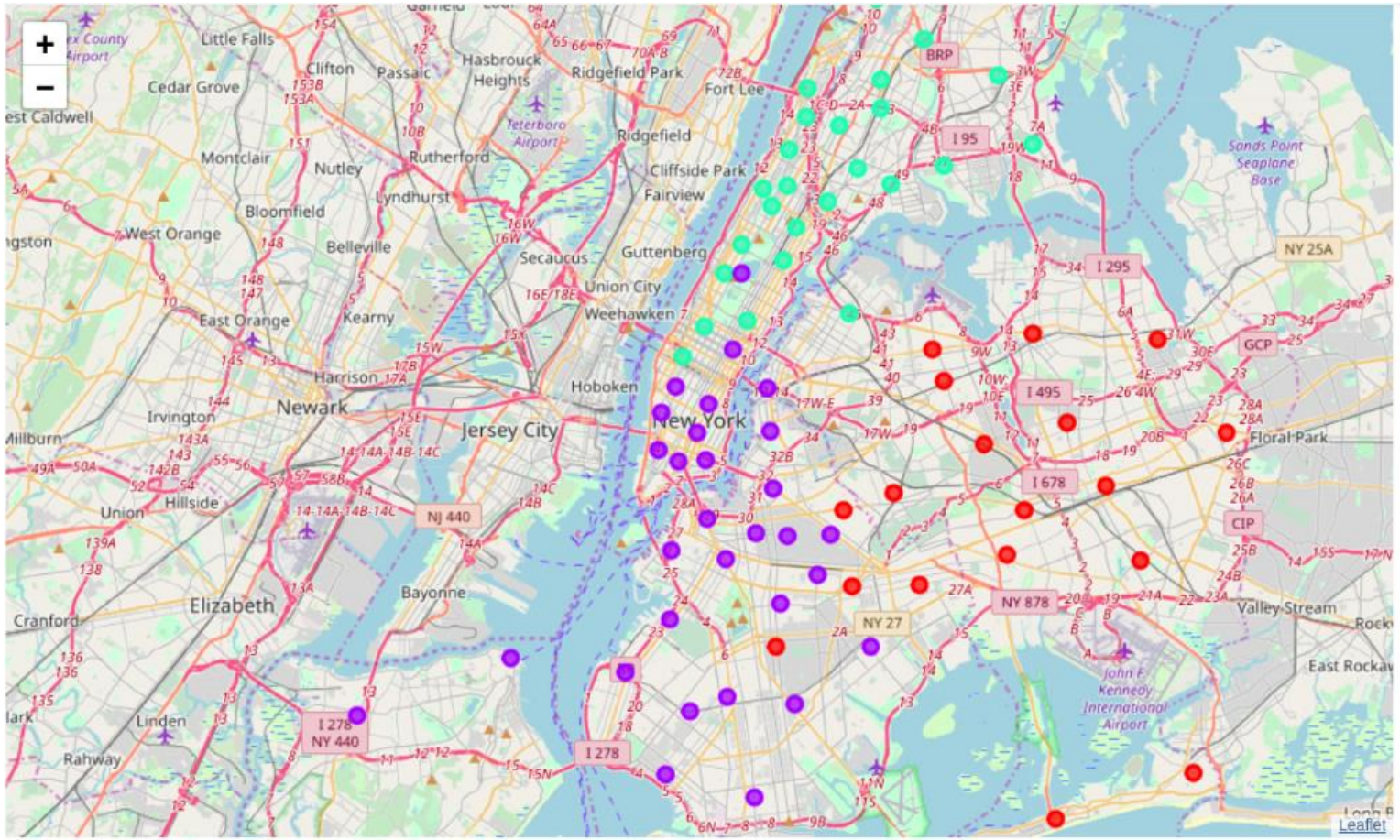
Methodology- Clustering

- I used two clustering approaches:
- K-Means: This algorithm helps clustering data into desired number of clusters and helps infer the relation between data samples.
- DBSCAN: This algorithm can give insights about outliers and can help decide appropriate number of clusters.
- More details are in the results section

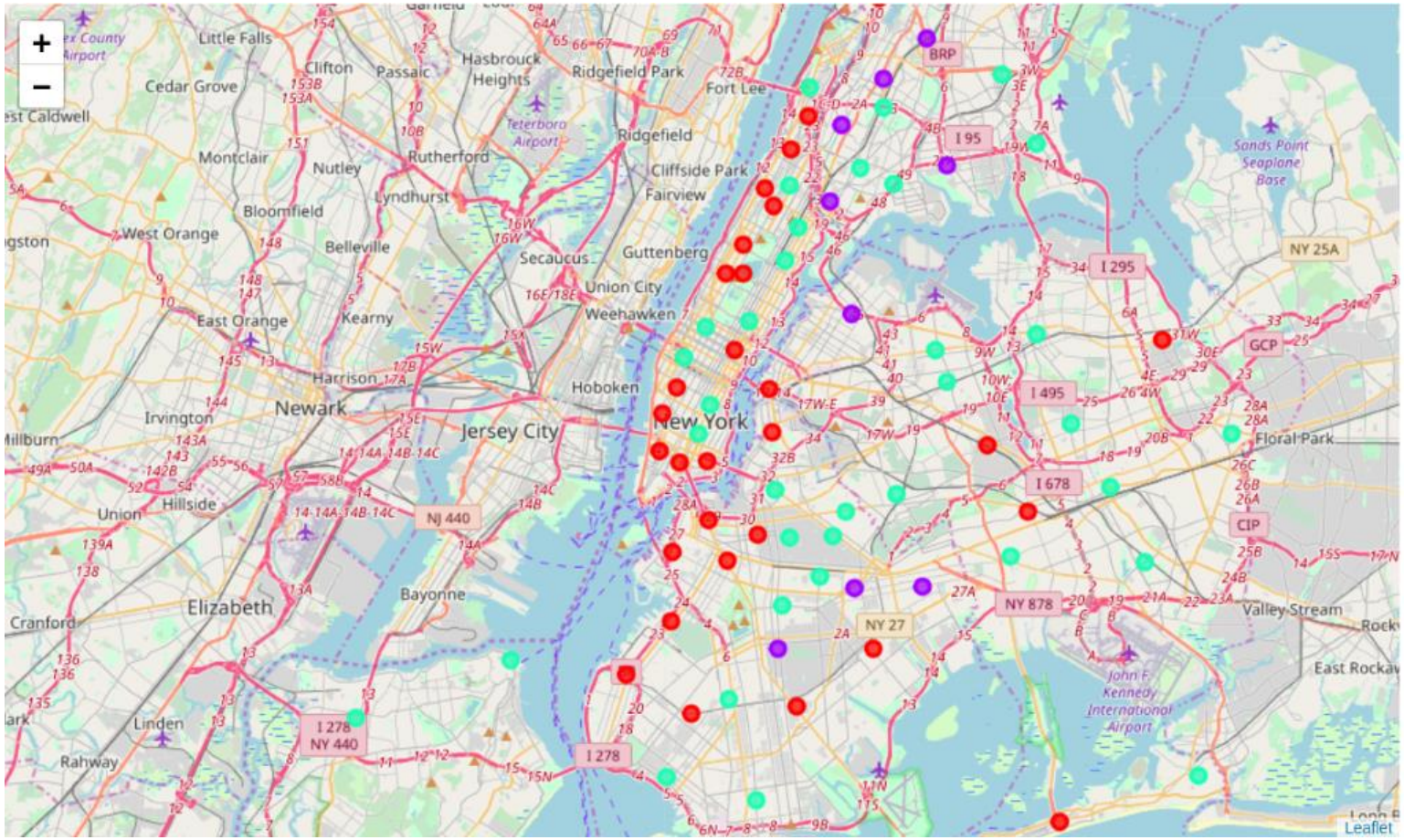
Results- K-means I



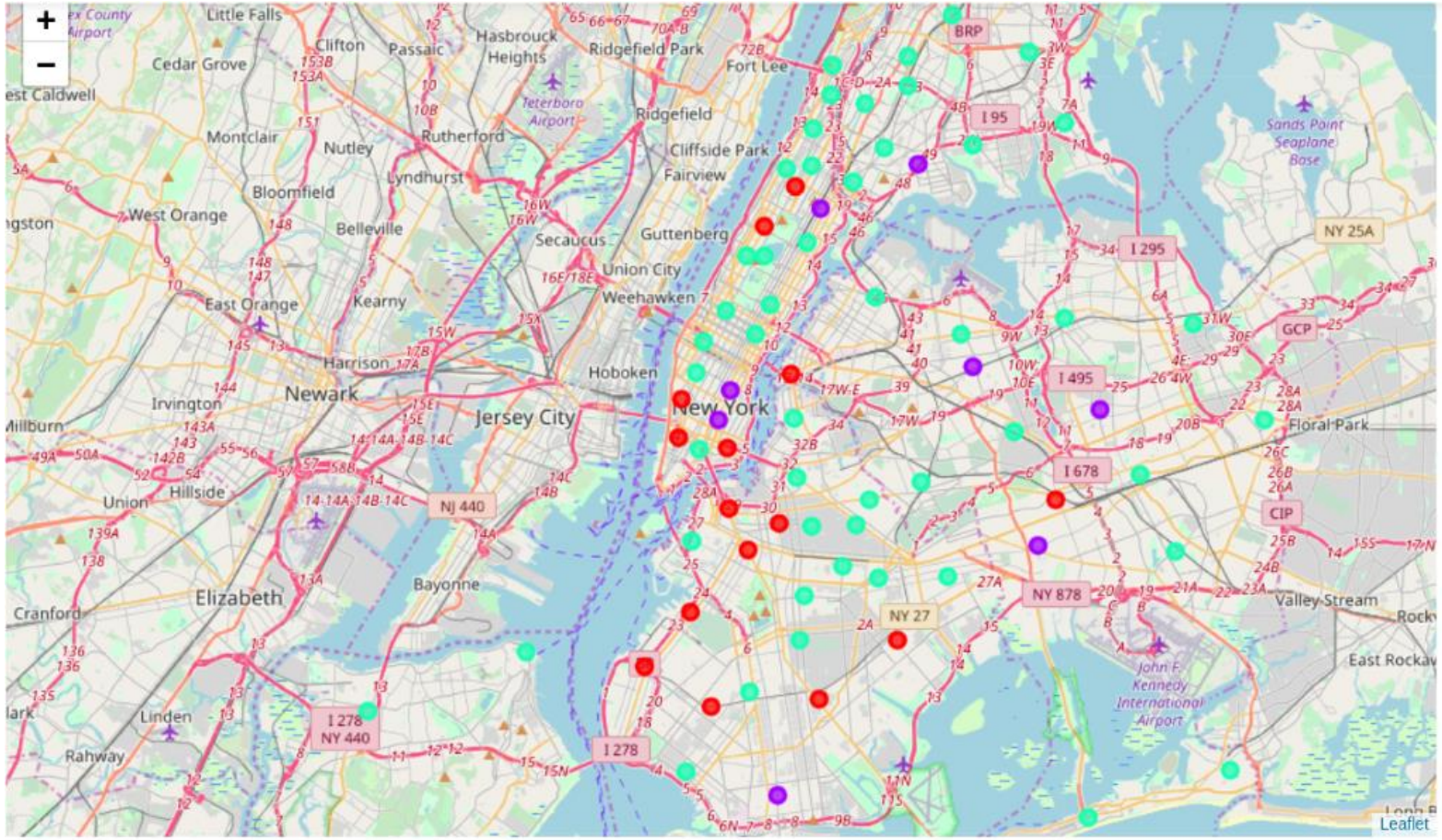
Results- K-means II



Results- K-means III



Results- DBSCAN



Conclusion

- This study provided insights about crime against women in New York City. The study used the geographical locations, the crime category for clustering
- Four different experiments were made. The four outcomes can be of potential interests for many stakeholders
  and governmental agencies.
- I hope this study will provide safety of each woman in New York and would help increase crime avoidance and prevention rates.