COURSERA CAPSTONE PROJECT: NEIGHBORHOOD OPPORTUNITY FOR A SECURITY SERVICE IN DENVER, COLORADO

Fabian A Williams
November 2019

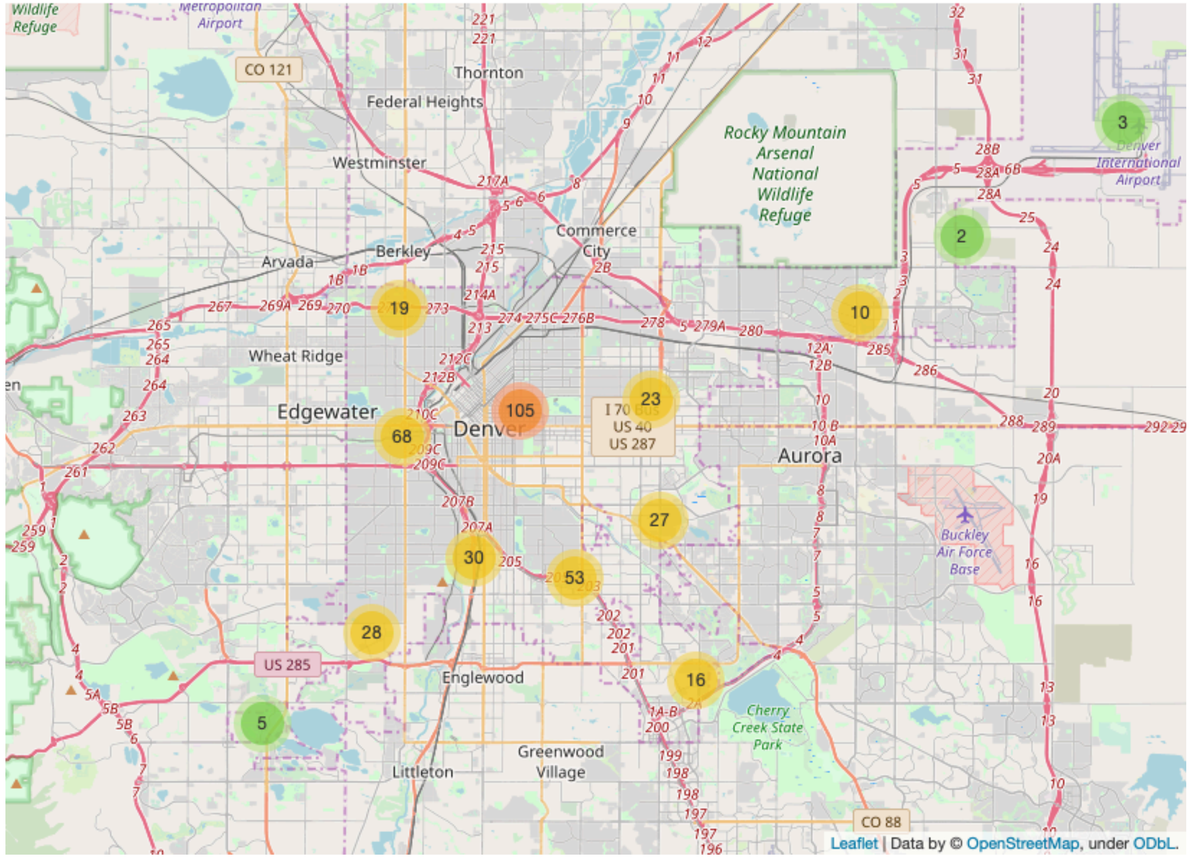
TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The success and final outcome of this project required a lot of documentation review and practice which was provided by IBM Data Science learning and development team using the Coursera learning platform. I am extremely grateful to have gotten this support all throughout the IBM Profession Data Science Certificate course.

To my online learning colleagues I say thank you for reviewing and commenting on my course submissions.

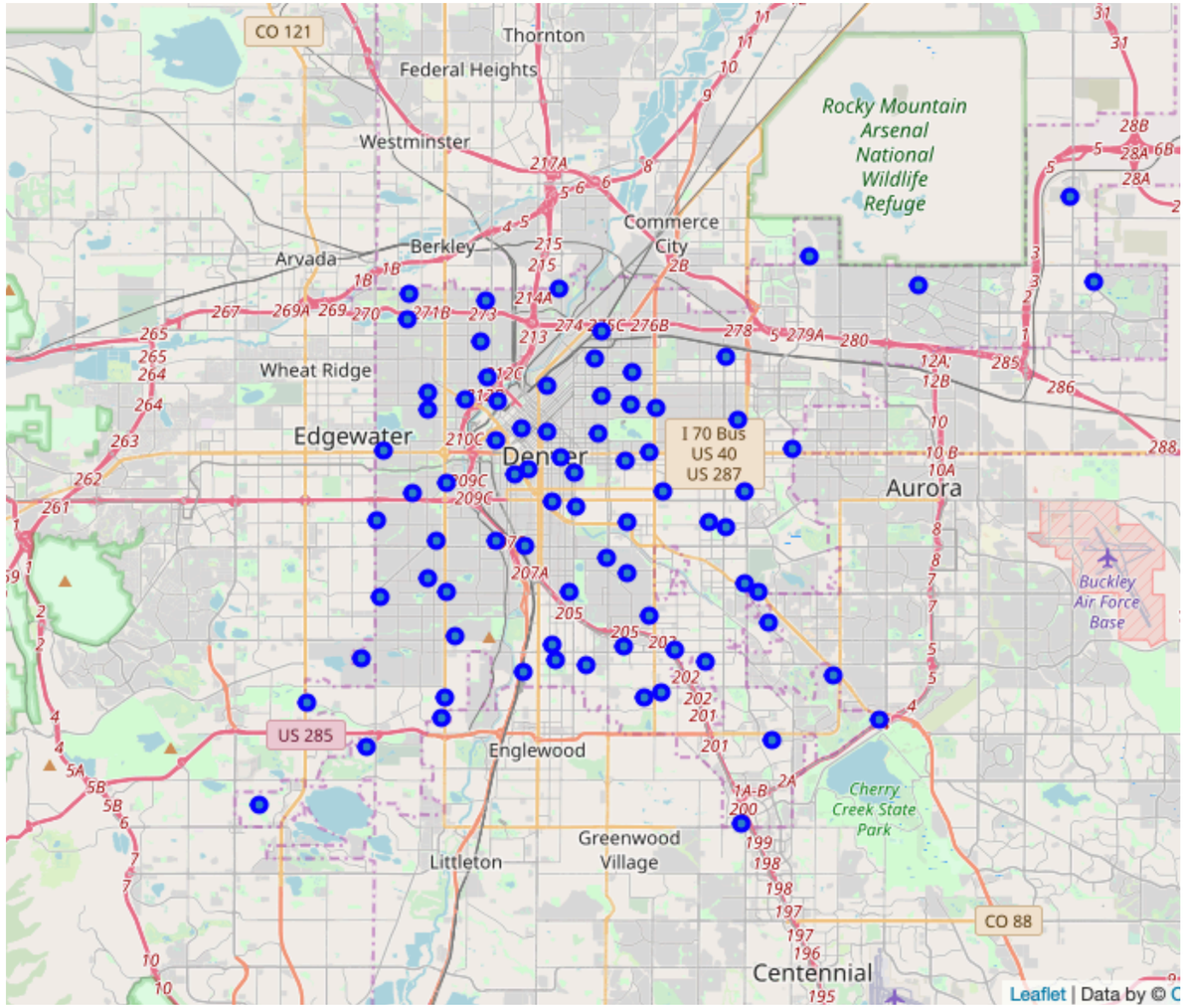
INTRODUCTION



Denver, Colorado Crime map

- Denver, is the capital of the U.S. state of Colorado. Denver is located in the South Platte River Valley on the western edge of the High Plains just east of the Front Range of the Rocky Mountains.
- Denver is ranked as a Beta world city by the Globalization and World Cities Research Network. With an estimated population of 716,492 in 2018, Denver is the 19th-most populous U.S. city, and with a 19.38% increase since the 2010 United States Census, it has been one of the fastest-growing major cities in the United States.
- Though rated the best place to live, can Denver, CO considered a safe place to operate a business.

THE BUSINESS PROBLEM



BURGLARY CRIME by NEIGHBORHOOD

- Investors in a security firm would like to perform a study of the neighborhood crimes within Denver, CO.
- The company's intent is to identify neighborhoods with burglary crime, across multiple business locations/venue categories.
- The crimes within a city are varied, one of the key features of the security service is the installation of closed circuit camera grid over the neighborhoods.
- The segmentation of these businesses will also form apart of the project planning process, as all neighborhoods will not be launched at the same time.

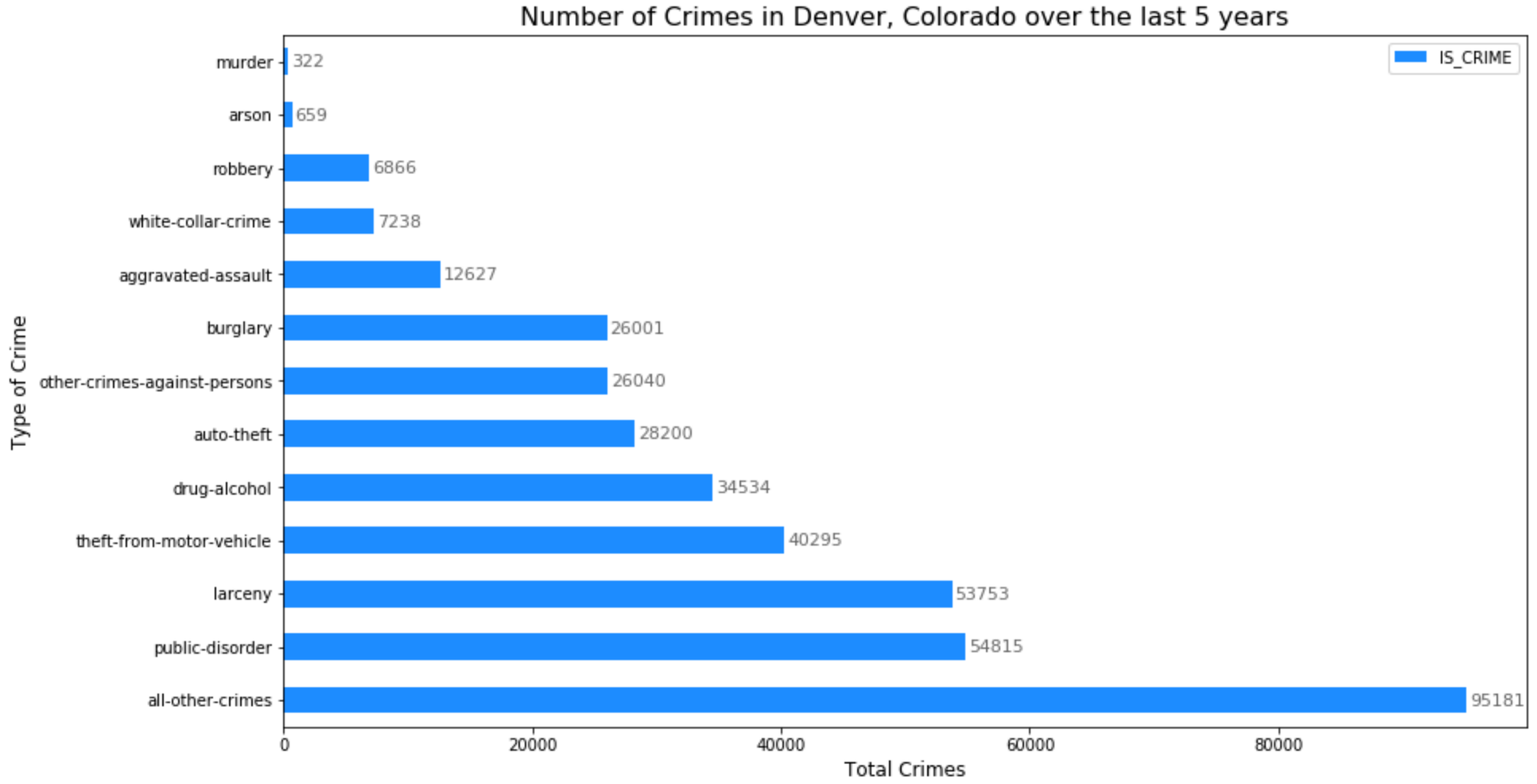
```
INCIDENT_ID                int64
OFFENSE_ID                 int64
OFFENSE_CODE               int64
OFFENSE_CODE_EXTENSION     int64
OFFENSE_TYPE_ID           object
OFFENSE_CATEGORY_ID       object
FIRST_OCCURRENCE_DATE     object
LAST_OCCURRENCE_DATE      object
REPORTED_DATE             object
INCIDENT_ADDRESS          object
GEO_X                    float64
GEO_Y                    float64
GEO_LON                  float64
GEO_LAT                  float64
DISTRICT_ID                int64
PRECINCT_ID                int64
NEIGHBORHOOD_ID           object
IS_CRIME                   int64
IS_TRAFFIC                 int64
dtype: object
```

DATA TABLE COLUMNS

DATA DESCRIPTION

- The databases accessed for this project was from www.denvergov.org. This dataset includes criminal offenses in the City and County of Denver for the previous five calendar years plus the current year to date.
- Not all the columns of this table is required, for the purpose of this report the relevant columns are:
  - OFFENCE_CATEGORY_ID: the type of crime
  - GEO_LON: longitude
  - GEO_LAT: latitude
  - NEIGHBORHOOD_ID: name of the neighborhood
  - IS_CRIME: crime committed using a boolean values 1 or 0
- The data in these columns will be wrangled to provide an initial view of the burglary data which will then be used with the FourSquare API to explore the world around where the crimes have been committed.

DATA DESCRIPTION CONT'D



Crimes in Denver, colorado

METHODOLOGY

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	athmar-park	Vietnamese Restaurant	Chinese Restaurant	Thai Restaurant	Grocery Store	Sandwich Place
1	auraria	American Restaurant	Theater	Fast Food Restaurant	Brewery	Basketball Stadium
2	baker	Bar	Chinese Restaurant	Breakfast Spot	Coffee Shop	Shipping Store
3	barnum	Fast Food Restaurant	Bakery	Vietnamese Restaurant	Mexican Restaurant	Dim Sum Restaurant
4	barnum-west	Convenience Store	Donut Shop	American Restaurant	Mexican Restaurant	Liquor Store

Top 5 rows of Venue Clusters Table

```
# Run k-means to cluster the neighborhood into 6 clusters

# set number of clusters
kclusters = 5

denver_grouped_clustering = denver_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(denver_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

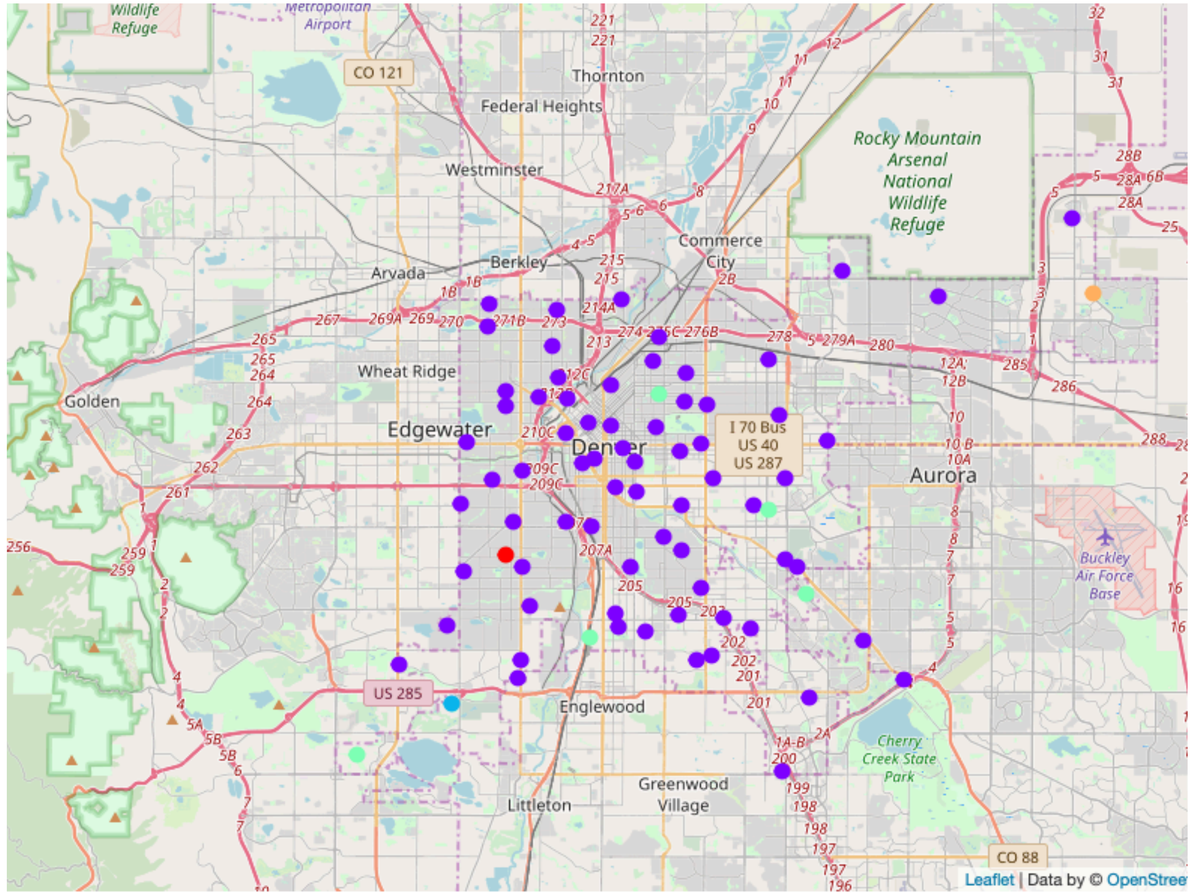
```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1], dtype=int32)
```

K means Algorithm, $K = 5$

	NEIGHBORHOOD_ID	GEO_LON	GEO_LAT	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	athmar-park	-105.024240	39.697324	1	Vietnamese Restaurant	Chinese Restaurant	Thai Restaurant	Grocery Store	Sandwich Place
1	auraria	-105.004869	39.744035	1	American Restaurant	Theater	Fast Food Restaurant	Brewery	Basketball Stadium
2	baker	-104.992622	39.711534	1	Bar	Chinese Restaurant	Breakfast Spot	Coffee Shop	Shipping Store
3	barnum	-105.028408	39.712805	1	Fast Food Restaurant	Bakery	Vietnamese Restaurant	Mexican Restaurant	Dim Sum Restaurant
4	barnum-west	-105.052573	39.719299	1	Convenience Store	Donut Shop	American Restaurant	Mexican Restaurant	Liquor Store

Top 5 rows of data after K means Algorithm

- This analysis was done using Python. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Python has built in libraries that will be installed in a Python Notebook to perform the analysis.
- The database is imported using Pandas, and an initial view of data was created using Matplotlib, a horizontal bar chart. The data was wrangled to identify burglary crimes and then grouped by neighborhood. The data frame included the longitude and attitude of burglary crimes. The data extracted using from the Denver, CO.org crime database will be merged with FourSquare API data to explore the world around where burglary crimes have been committed.
- Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data.
- K Means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group.



Cluster Map

RESULTS

- Cluster 1: Red
  - The 1st most common venue in this neighborhood is a Seafood Restaurant, followed by Women's Store, Flower Shop, Fishing Spot and Fish Market. Only one(1) neighborhood associated with this cluster.
- Cluster 2: Purple
  - Largest cluster with 1st most common venue ranging from Vietnamese and American Restaurants, Pharmacy, Convenience Stores, Liqour Store, Bars and others; followed by Chinese, Mexican and Asian Restaurants, Spa, Stadium and Coffee Shop. This cluster seems to be include well established venues across the city centre. Seventy(70) neighborhoods identified associated with this cluster.
- Cluster 3: Orange
  - The 1st most common venue in this neighborhood is Park, followed by Women's Store, Donut Shop, Flea Market and Fishing Spot. Only one(1) neighborhood associated with this cluster.
- Cluster 4: Pale Green
  - The 1st most common venue are Parks, followed by varying common venues from Trail, Lake, Dog Run and Fishing Spot. Five(5) neighborhoods associated with this cluster.
- Cluster 5: Blue
  - The 1st most common venue Gym, followed by Women's Store, Dog Run, Fish Market and Fishing Spot. Only one(1) neighborhood associated with this cluster.

CONCLUSION

- The Denver, CO Burglary data when clustered are generally dispersed across the geography. The clusters do not overlap, however there are clusters that only have one(1) neighborhood.
- Cluster 4, primarily consist of outdoor venues whilst Cluster 2 is associated with venues found in a busy city centre. This analysis provides context for the security company to establish a business to secure these venues.
- The analysis could be further used for other categories of crime in Denver, CO. The company may develop additional services based on the analysis of the other crime data.
- **Cluster 2, would be the likely opportunity to begin deploying the security service. This cluster has common venues that is associated with a city centre and likely to affected by burglary crimes.**

REFERENCES

- https://en.wikipedia.org/wiki/Denver
- www.denvergov.org
- https://labs.cognitiveclass.ai/tools/jupyterlab/lab/tree/labs/coursera/ML0101EN/ML0101EN-Clus-K-Means-Customer-Seg-py-v1.ipynb
- https://labs.cognitiveclass.ai/tools/jupyterlab/lab/tree/labs/DV0101EN/DV0101EN-3-5-1-Generating-Maps-in-Python-py-v2.0.ipynb
- https://matplotlib.org
- https://python-visualization.github.io/folium/modules.html