Denver Crime Data Analysis

MSDS 692 Practicum I

William Shell
Regis University

wshell@regis.com

Denver Crime Analysis Project Stages

Scrub the Data

Model the Data

Obtain the Data

Explore the Data

Interpret the Data

O S E M N

Obtain The Data

Incidents

Date Time stamped data of specific incidents

512,946

Demographics

2010 neighborhood statistics of resident demographics and occupancy

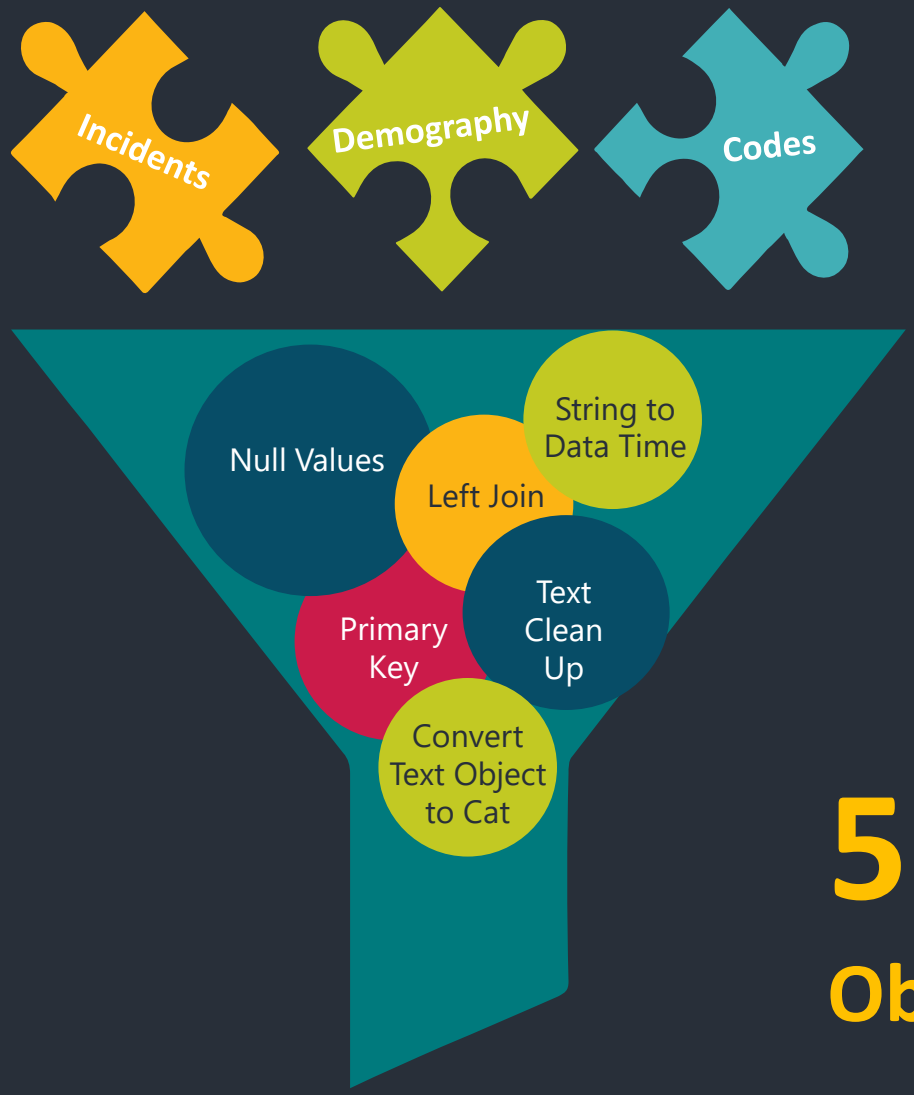
78

Codes

Catalog of criminal and non-criminal incident codes

299

Scrub the Data



Cleaned and Unified Data Set

512,946
Observations

141
Attributes

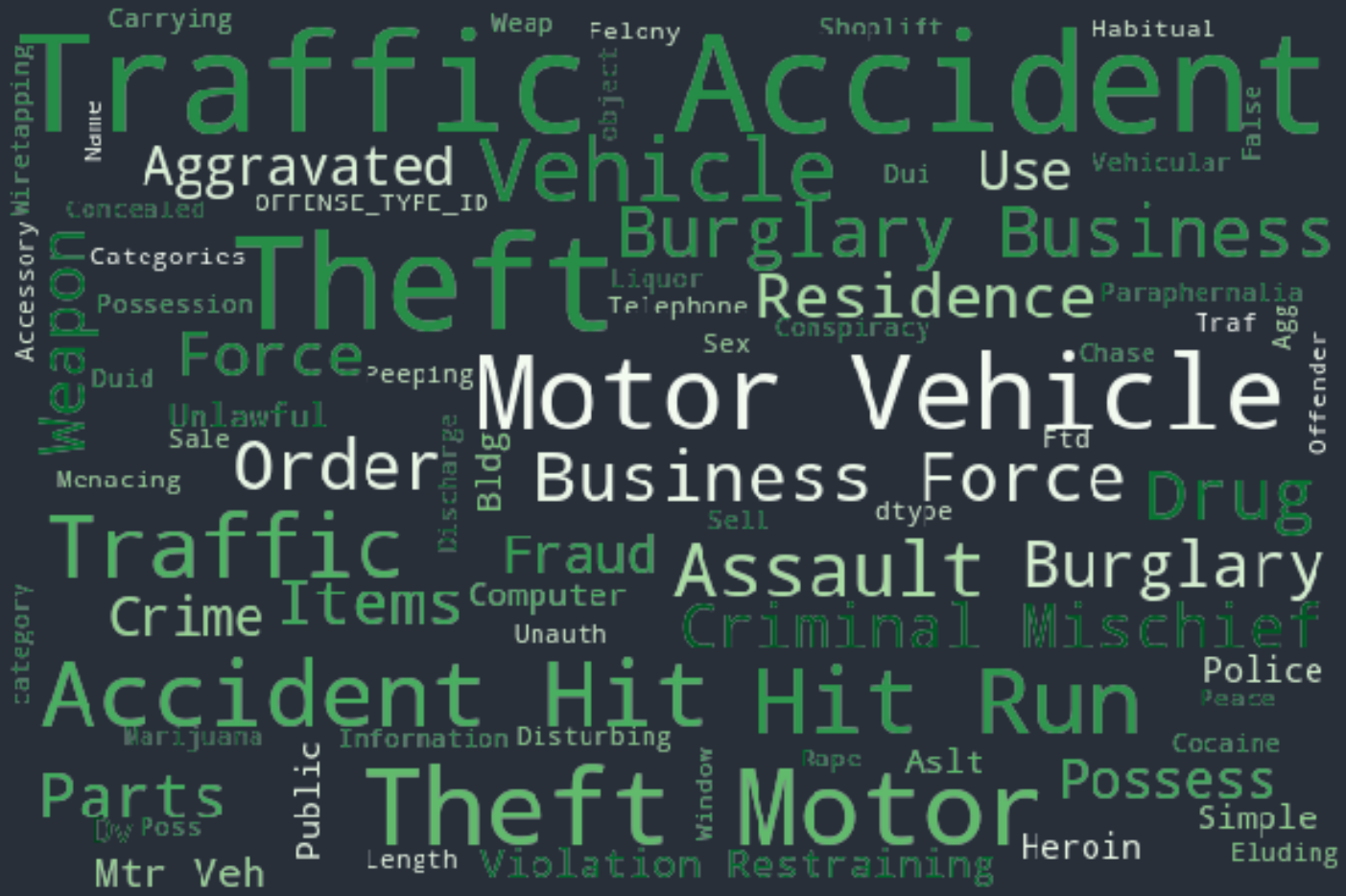
EDA

Time, and Place Incidents

Demographics Associated with Place

Types of Incidents

EDA Types of Incidents



72%

Top 5 Incidents

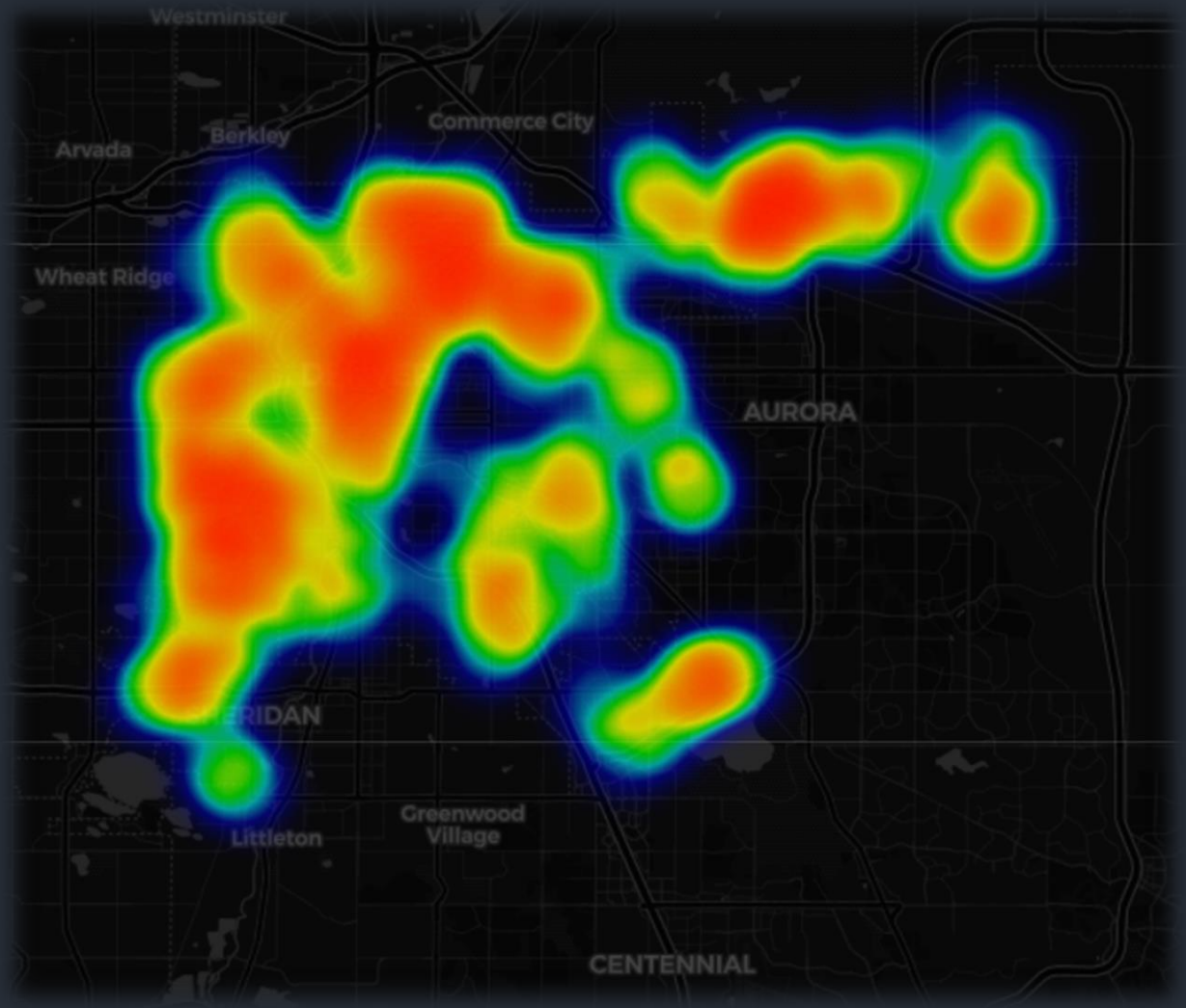
Traffic Accident	133273	26%
All Other Crimes	92841	18%
Public Disorder	53389	10%
Larceny	52027	10%
Theft From Motor Vehicle	38867	8%

EDA Location of Incidents

32%

of Incidents Concentrated in 13% of Neighbourhoods

Five Points	26947	5%
Stapleton	22091	4%
Cbd	19034	4%
Capitol Hill	17707	3%
Montbello	17446	3%
Baker	14050	3%
Lincoln Park	13760	3%
East Colfax	13383	3%
Westwood	12607	2%
Civic Center	12332	2%



EDA Time of Incidents

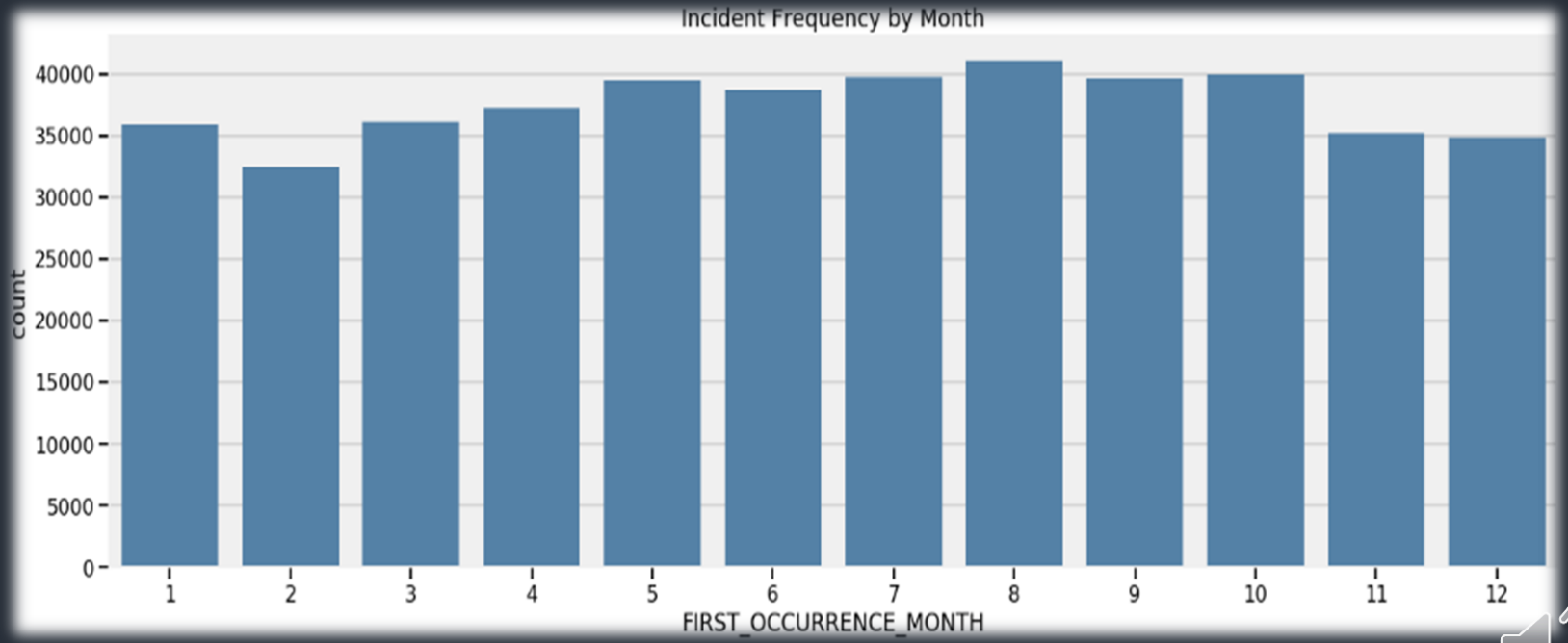
12%

Incident Increase from 2014 – 2019
83K to 93K per year

38%

Incident that Occur
between May and August

Midnight to 4 AM	16%
4 AM to 8 AM	8%
8 AM to 12 PM	18%
12 PM to 4 PM	20%
4 PM to 8 PM	22%
8 PM to Midnight	16%



EDA Demographics

600K

Residents

$62K

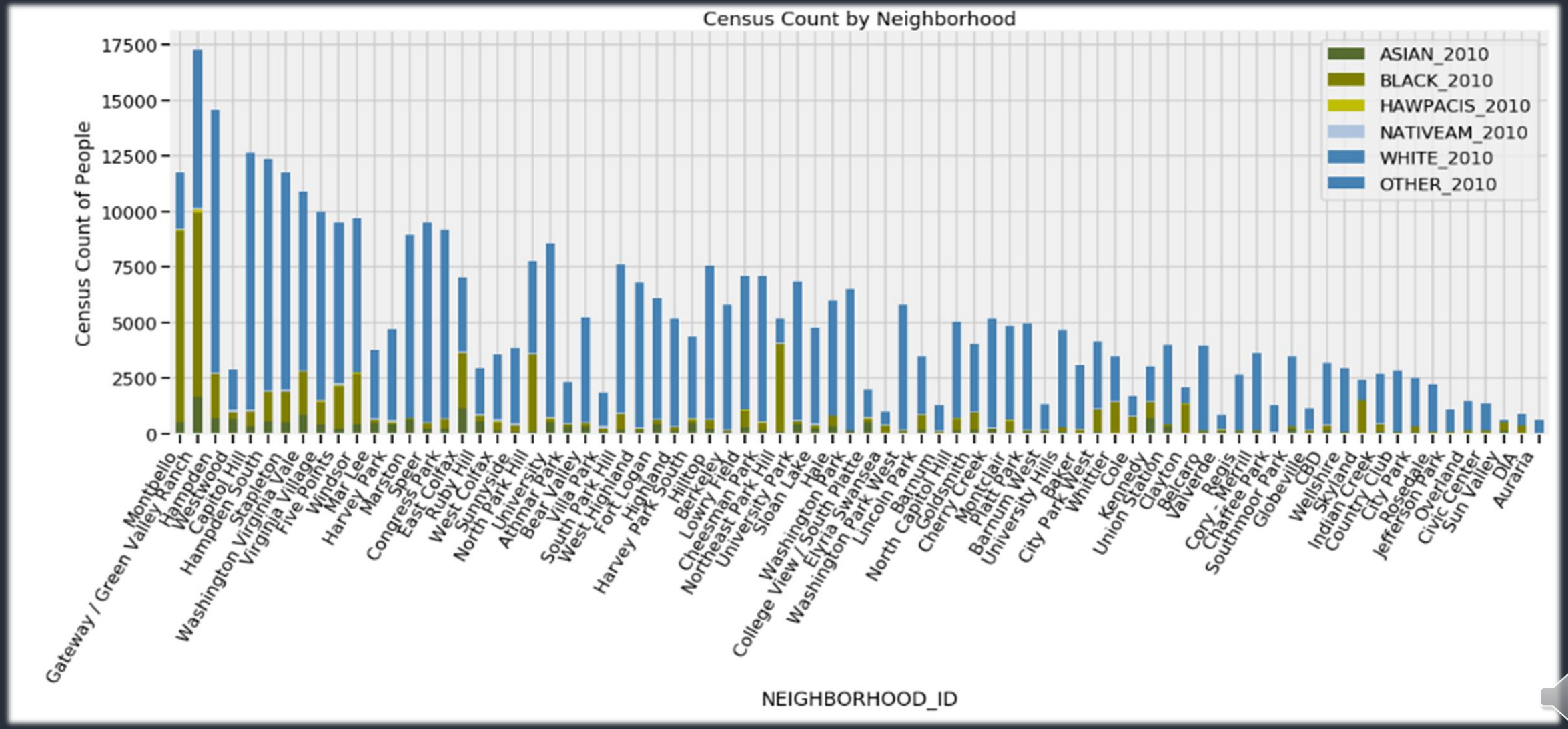
Annual Median Income

286K

Households

Neighborhood	Population	% of Total
Montebello	30348	5%
Gateway Green Valley	22091	4%
Hampden	17547	3%
Westwood	14708	2%
Capitol Hill	14370	2%
Stapleton	13948	2%
Washington Virginia Vale	13030	2%
Virginia Village	12844	2%
Five Points	12712	2%

EDA Demographics



Modelling

1. KNN on IS_CRIME is a simple 0 or 1 classification
2. KNN on INCIDENT_CATEGORY_ID has up to 15 potential classifications
3. Cross Fold validation for best value of K
4. Removed features redundant with crime classifications above (e.g. offense codes etc.

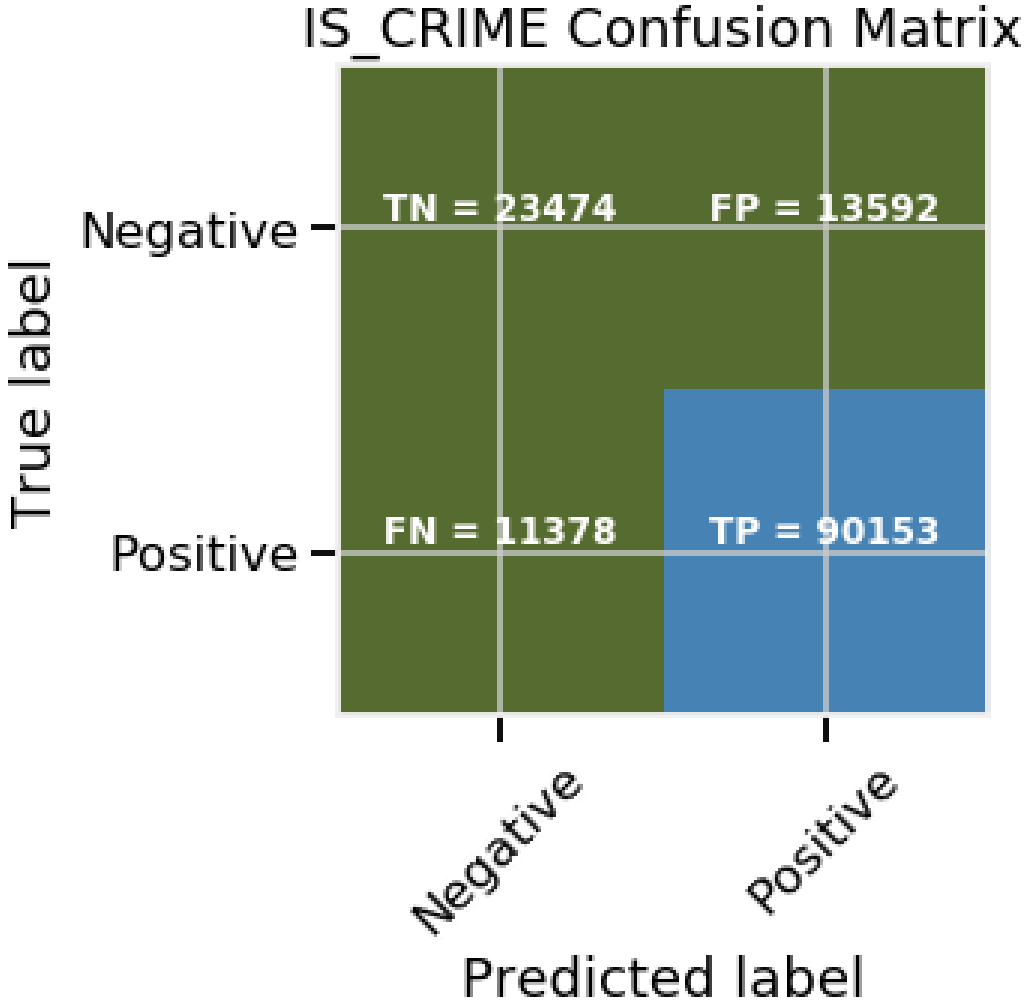
Performance

82%

Model 1 – IS_CRIME

39%

Model 2 – INCIDENT_CATEGORY_ID



Follow On Work

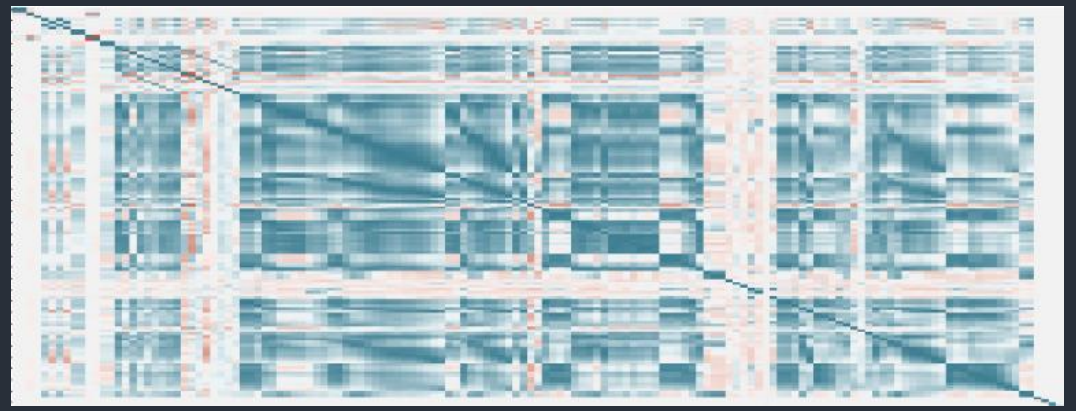
Explore Alternate Classification Models

Deeper Ethnicity and Crime Analysis

Explore Income and Crime Impacts

NEXT STEPS

Deep Dive Correlations



References

Denver Open Data Catalog: Crime. (n.d.). Retrieved from https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime

Dr. Cher Han Lau. (2019, January 10). 5 Steps of a Data Science Project Lifecycle. Retrieved from https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372b492

Generating WordClouds in Python. (n.d.). Retrieved from https://www.datacamp.com/community/tutorials/wordcloud-python

Petrou, T. (2019, October 1). Selecting Subsets of Data in Pandas: Part 1. Retrieved from https://medium.com/dunder-data/selecting-subsets-of-data-in-pandas-6fcd0170be9c

Python | Pandas DataFrame. (n.d.). Retrieved from https://www.geeksforgeeks.org/python-pandas-dataframe/

Sharma, M. (2018, November 3). Data Visualization using Seaborn. Retrieved from https://towardsdatascience.com/data-visualization-using-seaborn-fc24db95a850

Style sheets reference — Matplotlib 3.1.1 documentation. (n.d.). Retrieved from https://matplotlib.org/3.1.1/gallery/style_sheets/style_sheets_reference.html

Summarising, Aggregating, and Grouping data in Python Pandas. (2019, October 6). Retrieved from https://www.shanelynn.ie/summarising-aggregation-and-grouping-data-in-python-pandas/