

**Capstone Project - The Battle of
Neighborhoods
Report**

1. Introduction

1.1 Background

Do people move until they find a place to settle down where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction? Or, do we too often move to a new area without knowing exactly what we're getting into, forcing us to turn tail and run at the first sign of discomfort?

To minimize the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you don't end up wasting your valuable time and money making a move you'll end up regretting. Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.

1.2 Problem

The crime statistics dataset of Boston found on Kaggle has crimes in Boston from 2015 to 2018. The year 2018 being the latest we will be considering the data of that year which is actually old information as of now. The crime rates have changed over time.

This project aims to select the safest Street in Boston based on the total crimes, explore the neighbourhoods of that Street to find the 10 most common venues in each neighbourhood and finally cluster the neighbourhoods using k-mean clustering.

1.3 Interest

Expats who are considering to relocate to Boston will be interested to identify the safest Street in Boston and explore its neighbourhoods and common venues around each neighbourhood.

2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is from Kaggle. data source of the project uses a Boston crime data that shows the crime per Street in Boston. The dataset contains the following columns:

INCIDENT_NUMBER
OFFENSE_CODE
OFFENSE_CODE_GROUP
OFFENSE_DESCRIPTION

DISTRICT
REPORTING_AREA
SHOOTING
OCCURRED_ON_DATE
YEAR HOUR
Lat
Long
Location

2.2 Data Cleaning

The data preparation is done separately. From the Boston crime data, the crimes during the most recent year (2018) are only selected. And all the unnecessary fields are deleted.

MONTH
DAY_OF_WEEK
UCR_PART
STREET

```
df_boston.drop(['OFFENSE_CODE', 'OFFENSE_CODE_GROUP', 'REPORTING_AREA', 'SHOOTING', 'DAY_OF_WEEK', 'UCR_PART', 'HOUR'],
```

```
df_boston.head()
```

	INCIDENT_NUMBER	OFFENSE_DESCRIPTION	DISTRICT	OCCURRED_ON_DATE	YEAR	MONTH	STREET	Lat	Long	Location
0	I182070945	LARCENY ALL OTHERS	D14	2018-09-02 13:00:00	2018	9	LINCOLN ST	42.357791	-71.139371	(42.35779134, -71.13937053)
1	I182070943	VANDALISM	C11	2018-08-21 00:00:00	2018	8	HECLA ST	42.306821	-71.060300	(42.30682138, -71.06030035)
2	I182070941	TOWED MOTOR VEHICLE	D4	2018-09-03 19:27:00	2018	9	CAZENOVE ST	42.346589	-71.072429	(42.34658879, -71.07242943)
3	I182070940	INVESTIGATE PROPERTY	D4	2018-09-03 21:16:00	2018	9	NEWCOMB ST	42.334182	-71.078664	(42.33418175, -71.07866441)
4	I182070938	INVESTIGATE PROPERTY	B3	2018-09-03 21:05:00	2018	9	DELHI ST	42.275365	-71.090361	(42.27536542, -71.09036101)

```
df_boston = pd.read_csv('/Users/adityaghag/Desktop/crime.csv', encoding='unicode_escape')
df_boston.head()
```

	INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE
0	I182070945	619	Larceny	LARCENY ALL OTHERS	D14	806	NaN	2018-09-02 13:00:00
1	I182070943	1432	Vandalism	VANDALISM	C11	347	NaN	2018-08-21 00:00:00
2	I182070941	3410	Towed	TOWED MOTOR VEHICLE	D4	151	NaN	2018-09-03 19:27:00
3	I182070940	3114	Investigate Property	INVESTIGATE PROPERTY	D4	272	NaN	2018-09-03 21:16:00
4	I182070938	3114	Investigate Property	INVESTIGATE PROPERTY	B3	421	NaN	2018-09-03 21:05:00

```
df_boston.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 65685 entries, 0 to 65684
Data columns (total 10 columns):
INCIDENT_NUMBER      65685 non-null object
OFFENSE_DESCRIPTION   65685 non-null object
DISTRICT              65141 non-null object
OCCURRED_ON_DATE      65685 non-null object
YEAR                  65685 non-null int64
MONTH                 65685 non-null int64
STREET                64542 non-null object
Lat                   61464 non-null float64
Long                  61464 non-null float64
Location              65685 non-null object
dtypes: float64(2), int64(2), object(6)
memory usage: 5.0+ MB
```

3. Methodology

3.1 Statistical summary of crimes

The describe function in python is used to get statistics of the Boston crime data, this returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime.

3.2 Analyzing the type of data in the data set

Find below the different types of data types available in the data set.

```
: df_boston.describe()
```

	OFFENSE_CODE	YEAR	MONTH	HOUR	Lat	Long
count	65685.000000	65685.0	65685.000000	65685.000000	61464.000000	61464.000000
mean	2360.655325	2018.0	4.668083	13.091984	42.161178	-70.823006
std	1174.817462	0.0	2.299732	6.265751	2.633861	4.260657
min	111.000000	2018.0	1.000000	0.000000	-1.000000	-71.178674
25%	1102.000000	2018.0	3.000000	9.000000	42.296532	-71.097223
50%	3006.000000	2018.0	5.000000	14.000000	42.325125	-71.077720
75%	3205.000000	2018.0	7.000000	18.000000	42.348343	-71.062498
max	3831.000000	2018.0	9.000000	23.000000	42.395042	-1.000000

4. Results

4.1 Analysing Specific crime in all streets

```
In [6]: # analysing specific crime
larceny = df_boston[df_boston.OFFENSE_CODE_GROUP.str.contains('Larceny')]
larceny.head()
```

	INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	OCCURRED_ON_DATE	YEAR	MOI
0	I182070945	619	Larceny	LARCENY ALL OTHERS	D14	808	2018-09-02 13:00:00	2018	
19	I182070915	614	Larceny From Motor Vehicle	LARCENY THEFT FROM MV - NON-ACCESSORY	B2	181	2018-09-02 18:00:00	2018	
35	I182070893	614	Larceny From Motor Vehicle	LARCENY THEFT FROM MV - NON-ACCESSORY	B3	427	2018-09-03 18:44:00	2018	
36	I182070892	613	Larceny	LARCENY SHOPLIFTING	B2	276	2018-09-03 17:54:00	2018	
45	I182070885	619	Larceny	LARCENY ALL OTHERS	B3	456	2018-09-03 17:17:00	2018	

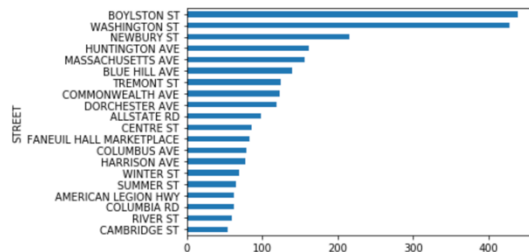
Comparing all streets with the highest crime rate during the year 2018 it is evident that BOYLSTON ST has the highest crimes recorded followed by WASHINGTON ST, NEWBURY ST

```
In [7]: # analysing places
larceny.groupby('STREET').size().sort_values(ascending = False)
```

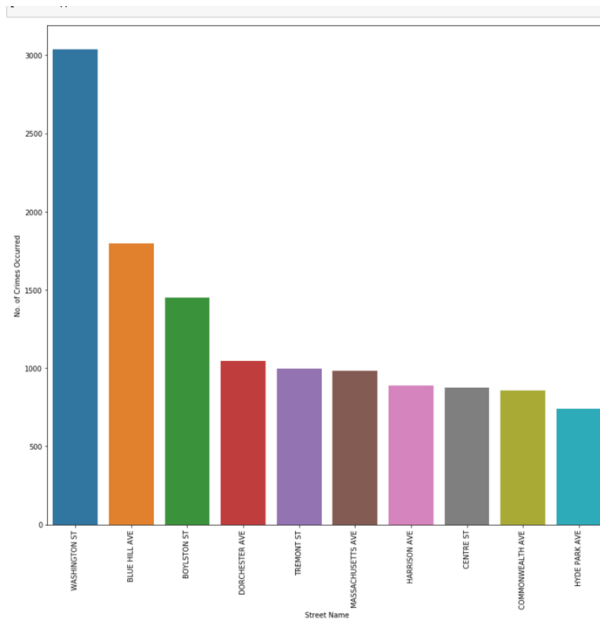
```
Out[7]: STREET
BOYLSTON ST      439
WASHINGTON ST    427
NEWBURY ST       216
HUNTINGTON AVE   161
MASSACHUSETTS AVE 156
...
LEWIS ST          1
LEVERETT AVE      1
LESLIE ST         1
LEROY ST          1
A ST              1
Length: 1388, dtype: int64
```

```
In [8]: # analysing streets where there is more crime for larceny
larceny.groupby('STREET').size().sort_values(ascending = False).head(20).sort_values().plot(kind='barh')
```

```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2779f278>
```



4.2 Analysing streets with more number of crimes



4.Discussion

The aim of this project is to help people who want to relocate to the safest places in boston, expats can choose the neighbourhoods to which they want to relocate based on the most common venues in it. For example, if a person is looking for a neighbourhood with good connectivity and public transportation.

5.Conclusion

This project helps a person get a better understanding of the neighbourhoods with respect to the most common venues in that neighbourhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighbourhood. We have just taken safety as a primary concern to shortlist the safest places in Boston. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist theStreet, such as filtering areas based on a predefined budget.