CAPSTONE PROJECT

ON

BOSTON CRIME REPORT

NORTHEASTERN UNIVERSITY

SUBMITTED BY: KARISHMA SIDDHARTH PADALE

Abstract

The Inspectional Services Department (ISD) issues building licenses for development extends inside the City of Boston. Different ventures require various kinds of utilizations, and work can't start until a structure license card is given. Licenses are substantial for a half year beginning the day that they are gotten and may occasionally be reached out for an extra 180 days.

Data extraction:

1. I have downloaded data from publicly available source at

   https://data.boston.gov/dataset/6220d948-eae2-4e4b-8723-2dc8e67722a3/resource/12cb3883-56f5-47de-afa5-3b1cf61b257b/download/tmppj4rb047.csv

2. The data is of Boston crime report which has rows = 404368 and columns = 17
   The data is about the crime scenes in Boston like the crimes monthly, street wise and which area has most crimes.

3. We are answering the following questions:

   How has crime changed over the years?
   Is it possible to predict where or when a crime will be committed?
   Which areas of the city have evolved over this time span?
   In which area most crimes are committed?

Data cleanup

The null values in the column Shooting has been filled up by the following function:

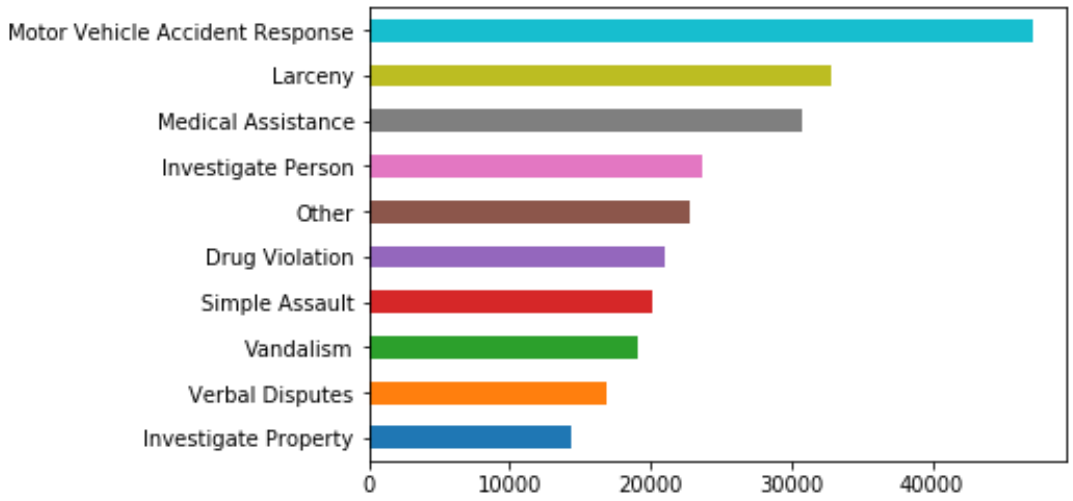
```
new_df = df.fillna(method="pad")

new_df = new_df.dropna(subset=['SHOOTING'],how='any')

new_df.isna().sum()
```

Data visualization and Descriptive Analysis

1. The number of appearances of different types of crimes and sort them in orders.

   Motor Vehicle Accident Response 47090
   Larceny 32826
   Medical Assistance 30721
   Investigate Person 23682
   Other 22746
   Drug Violation 20970
   Simple Assault 20191
   Vandalism 19073
   Verbal Disputes 16823
   Investigate Property 14328
   Name: OFFENSE_CODE_GROUP, dtype: int64

   Analysis: Here, we can see the crime incident that has happened the most frequently in Boston is “Motor Vehicle Accident Response,” and “Larceny” has also been taking place very frequently. Then, I plotted the result for better visualization.



2. Analyzing a Specific Crime:

   I want to specifically analyze larceny in Boston. Hence, I put the part of the data frame that contains larceny into another data frame and called it larceny."

Using the following function we can do that:

```
larceny = df[df.OFFENSE_CODE_GROUP.str.contains('Larceny')]
larceny.head()
```

3. Analyzing Places:

   I want to know the data of crime incidents in different locations of Boston and, more specifically, what places in Boston are more dangerous.

   I used groupby function in Pandas to group the types of criminal locations, and used size function to check the number of entries.

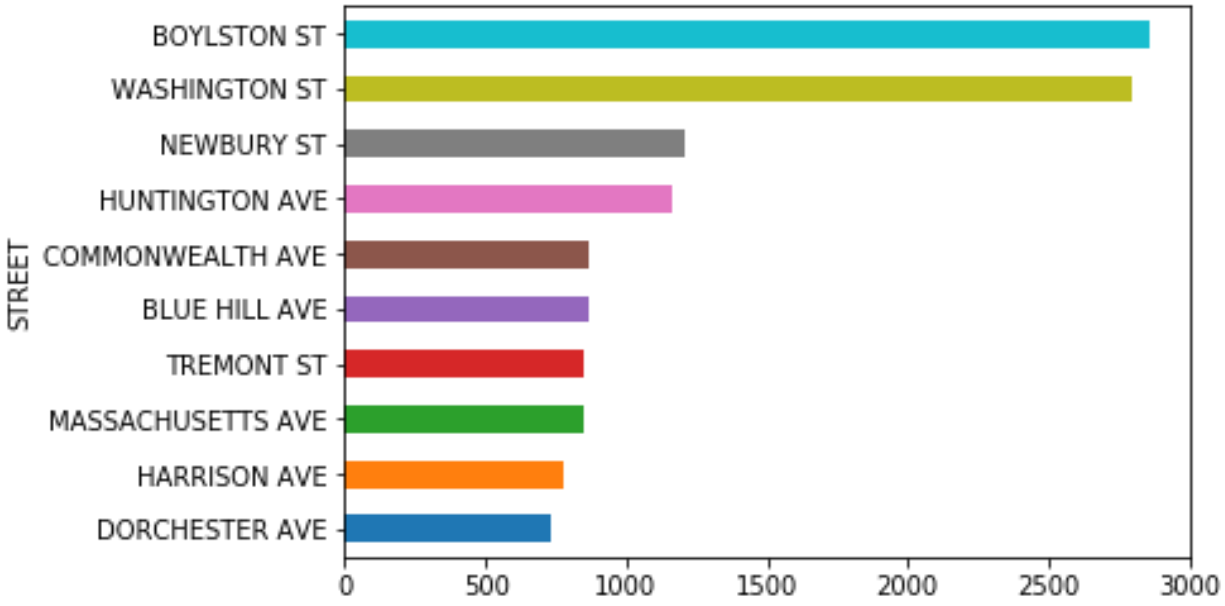
   Using the following functions:

   ```
   larceny.groupby('STREET').size().sort_values(ascending = False)
   ```

4. Top 10 crime reported streets:
   streets = df.groupby([df['STREET'].fillna('NO STREET NAME')])['REPORTING_AREA'].aggregate(np.size).reset_index().sort_values('REPORTING_AREA',ascending = False).head(10)
   streets

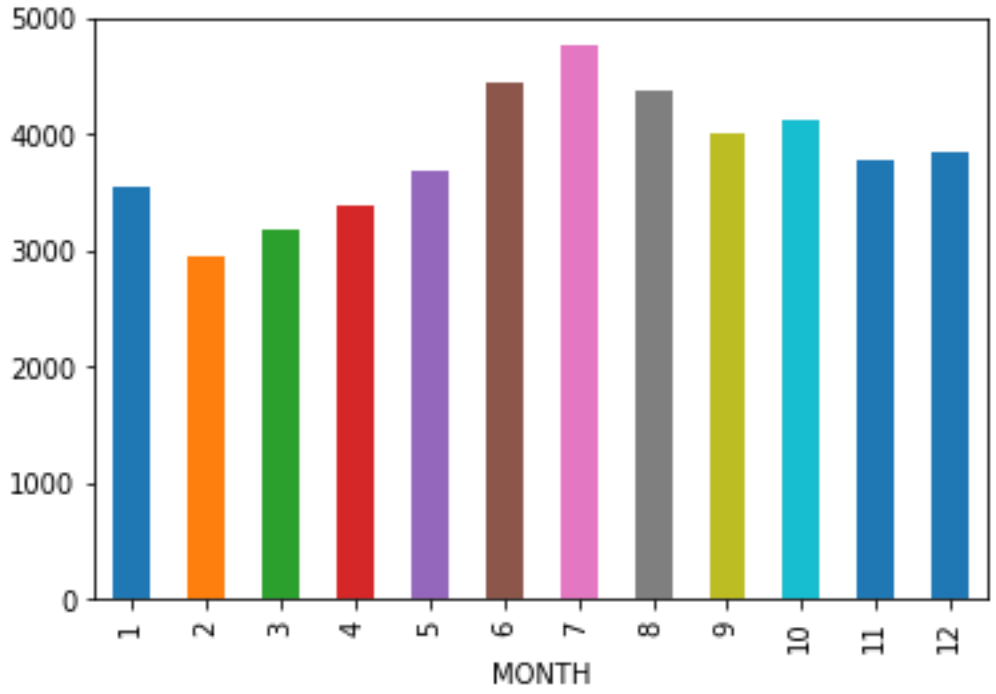
	STREET	REPORTING_AREA
4561	WASHINGTON ST	18319
3209	NO STREET NAME	12012
510	BLUE HILL AVE	9994
577	BOYLSTON ST	9082
1354	DORCHESTER AVE	6407
4353	TREMONT ST	6200
2851	MASSACHUSETTS AVE	6011
2083	HARRISON AVE	5984
845	CENTRE ST	5612
1050	COMMONWEALTH AVE	5259

Looking at the result, we can see the locations in Boston where larceny is more likely to happen are Washington St, Boylston St, and Newbury St.



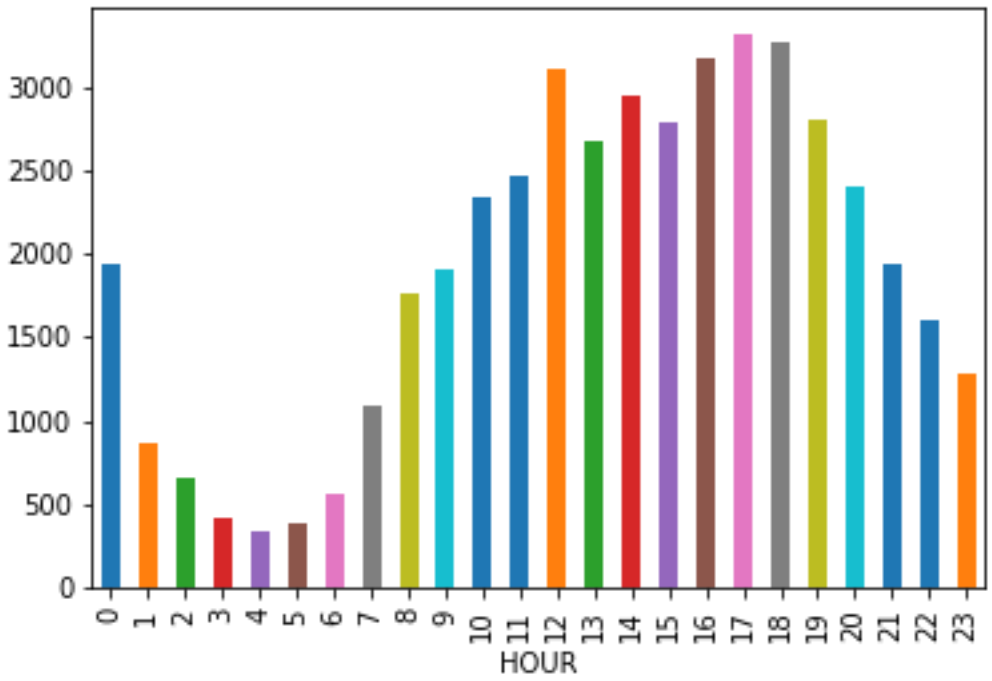
5. Month-wise crime frequency visualization

   I also want to know about the trend of larceny incidents that has been taking place in Boston.

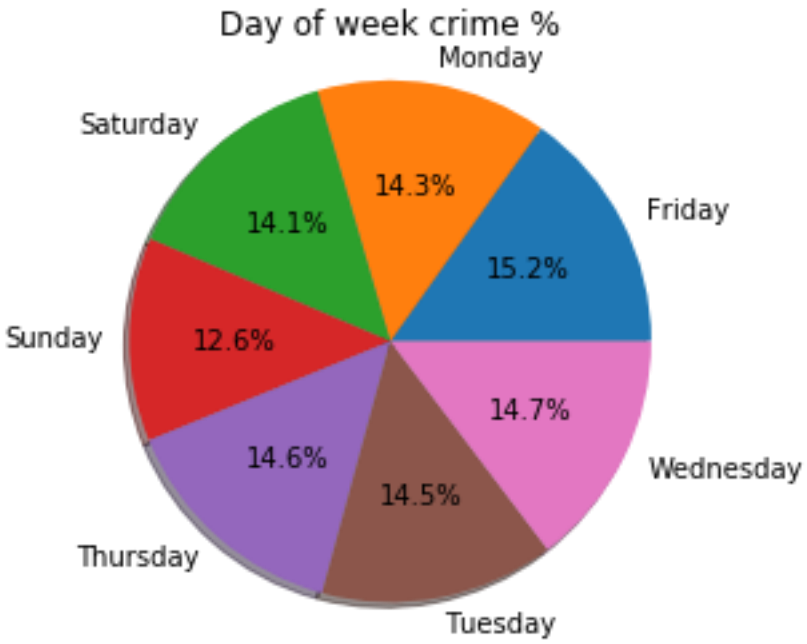


   Based on the bar graph that I computed, larceny happened the most during May, June, and December, whereas September, October, and August appear to be safer.

6. Hourly crime visualization



7. Pie-Chart Visualization of Crime percentage per day of the week



8. Bar Plot visualization of Crimes by street per District