Data Wrangling with Baltimore City 911 Report Data

Vrushali Vishal Patil

Task 1

skimpy summary

Data Summary

dataframe	Values
Number of rows	292761
Number of columns	16

Data Types

Column Type	Count
string	12
float64	3
int64	1

number

column_name	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Longitude	1635	0.56	-77	0.044	-82	-77	-77	-77	-76	
Latitude	1635	0.56	39	0.03	38	39	39	39	40	
Location 1	292761	100	nan	nan	nan	nan	nan	nan	nan	
Total Incidents	0	0	1	0	1	1	1	1	1	

string

column_name	NA	NA %	words per row	total words
CrimeDate	0	0	1	292761
CrimeTime	26	0.01	1	292735
CrimeCode	0	0	1	292761
Location	1336	0.46	3.4	1000942
Description	0	0	1.8	539328
Inside/Outside	33760	11.53	0.88	259001
Weapon	231229	78.98	0.21	61532
Post	2376	0.81	0.99	290385
District	0	0	1	292761
Neighborhood	10504	3.59	1.6	461736
Premise	34012	11.62	1.3	367241
vri_name1	258177	88.19	0.15	42654

End

p0 represents minimum value, p50 denotes median value and p100 is the maximum value for particular column.

Task 1

Column	Total Unique Values	Top 3 Categories
CrimeDate	2300	{'04/27/2015': 421, '06/05/2016': 255, '12/20/2018': 212}
CrimeTime	1475	{'18:00:00': 6772, '17:00:00': 6463, '12:00:00': 6037}
CrimeCode	81	{'4E': 48061, '6D': 38977, '5A': 25351}
Location	27064	{'1500 RUSSELL ST': 847, '3500 BOSTON ST': 708, '2400 FREDERICK AVE': 558}
Description	14	{'LARCENY': 65508, 'COMMON ASSAULT': 48061, 'BURGLARY': 42237}
Inside/Outside	4	{'O': 128776, 'I': 124454, 'Outside': 5113}
Weapon	5	{'FIREARM': 29036, 'OTHER': 17313, 'KNIFE': 10023}
Post	133	{'114': 5227, '913': 4977, '111': 4915}
District	10	{'NORTHEAST': 44207, 'SOUTHEAST': 41816, 'CENTRAL': 33508}
Neighborhood	278	{'DOWNTOWN': 9312, 'FRANKFORD': 6922, 'BELAIR-EDISON': 6206}
Premise	125	{'STREET': 104827, 'ROW/TOWNHOUSE-OCC': 54077, 'APT/CONDO - OCCUPIED': 11689}
vri_name1	8	{'Tri-District': 6607, 'Western': 5825, 'Central': 5420}

```
Overall missingness: 18.52%
```

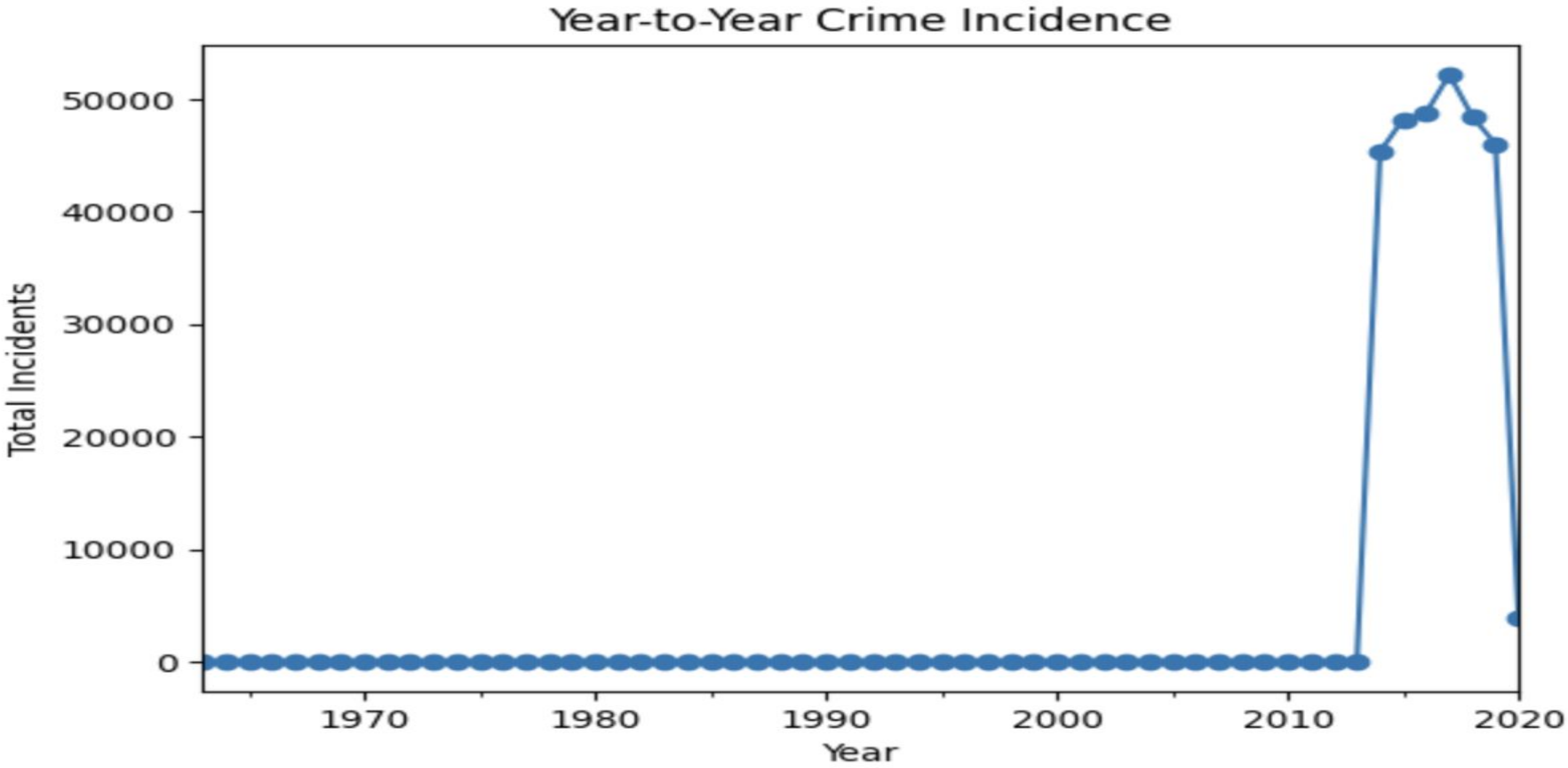
Table represents the all discrete attributes along with total number of unique values present in that column and top 3 categories with their count.
18.52% rows of the dataset has missing values.

Observations

- The 'Location 1' column lacks values throughout the dataset.
- Numerous rows contain missing values.
- Some columns, such as 'CrimeDate' and 'CrimeTime,' possess incorrect data types ('object') .
- Categorical variables, like 'Inside/Outside,' exhibit inconsistent labels (e.g., 'I' and 'O').
- To address these issues, a thorough data cleaning process is required, involving the handling of missing values, correction of data types, and standardization of categorical labels.

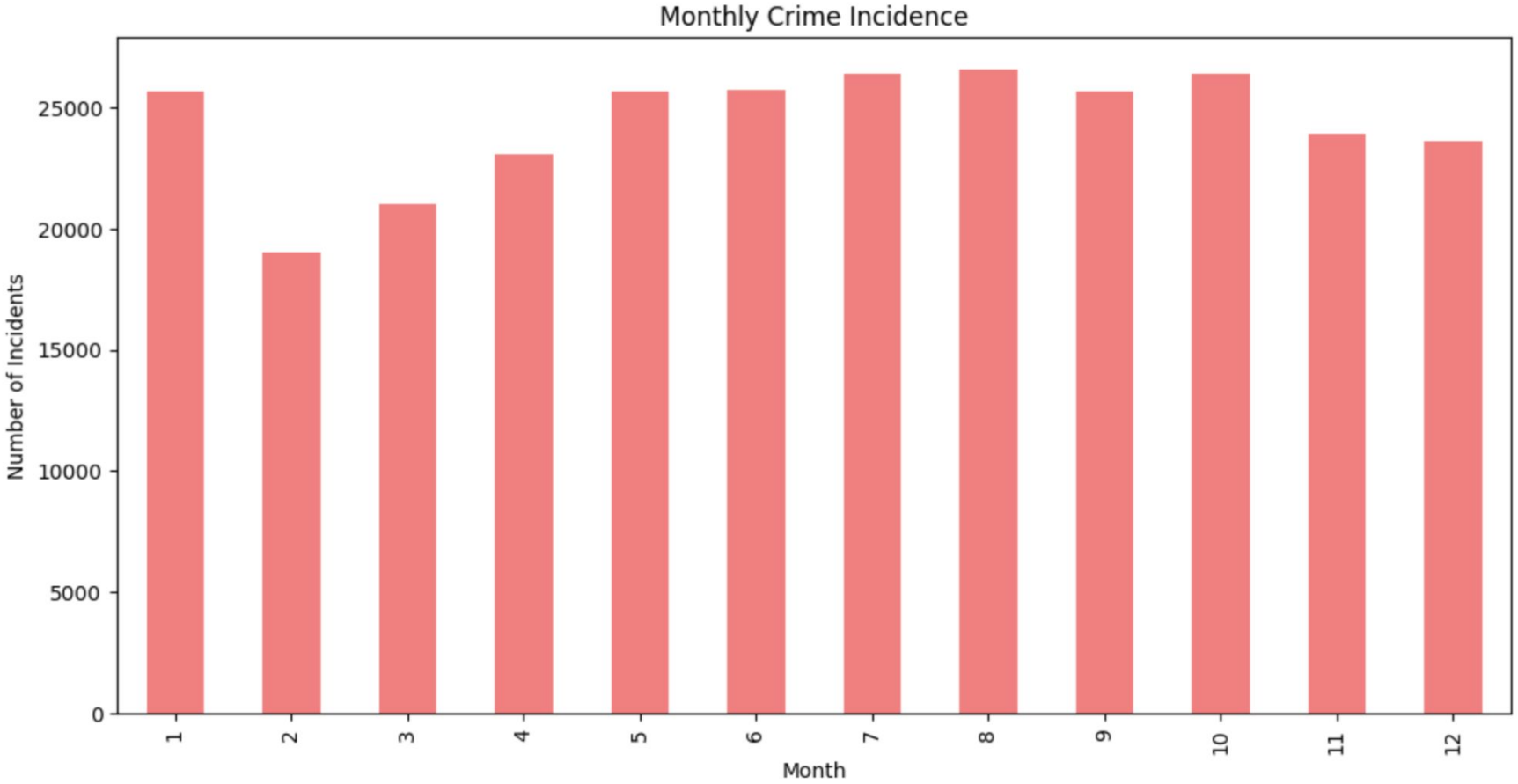
Task 2

The frequency of crime incidents has exhibited an upward trend since the year 2014.



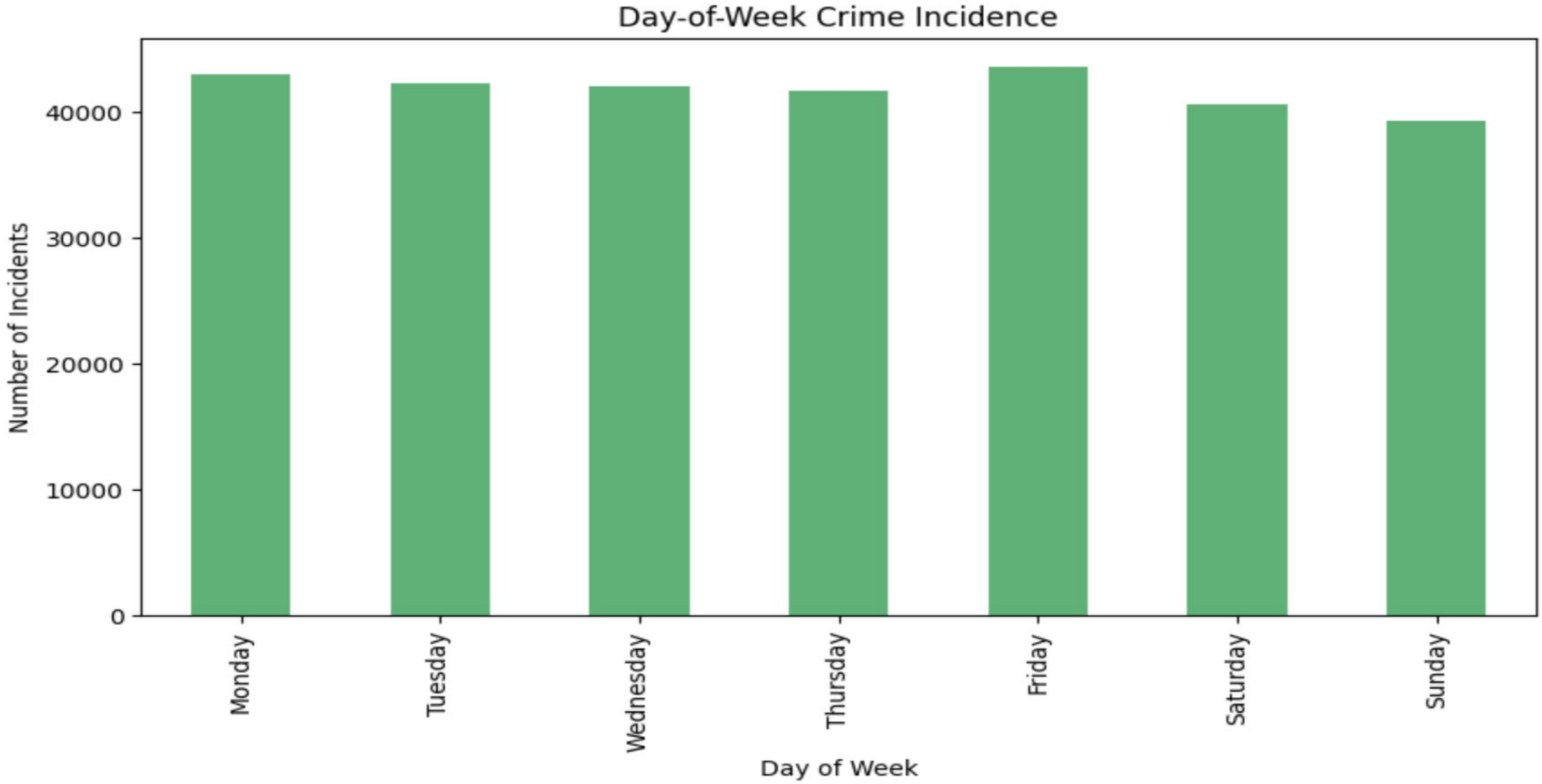
Task 2

The number of crime incidents remains relatively consistent across all months.



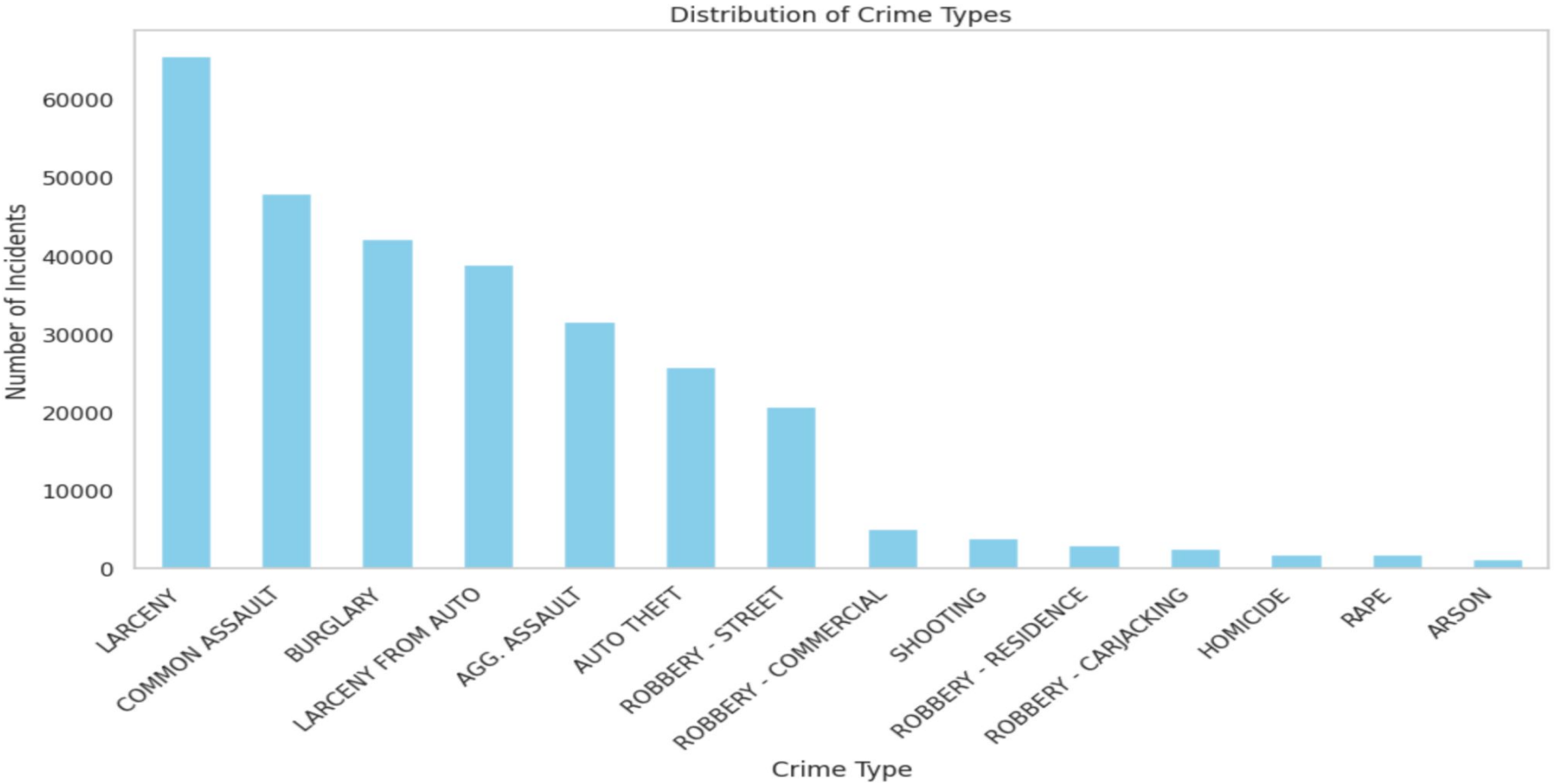
Task 2

The number of crime incidents remains relatively consistent across all days of the week.



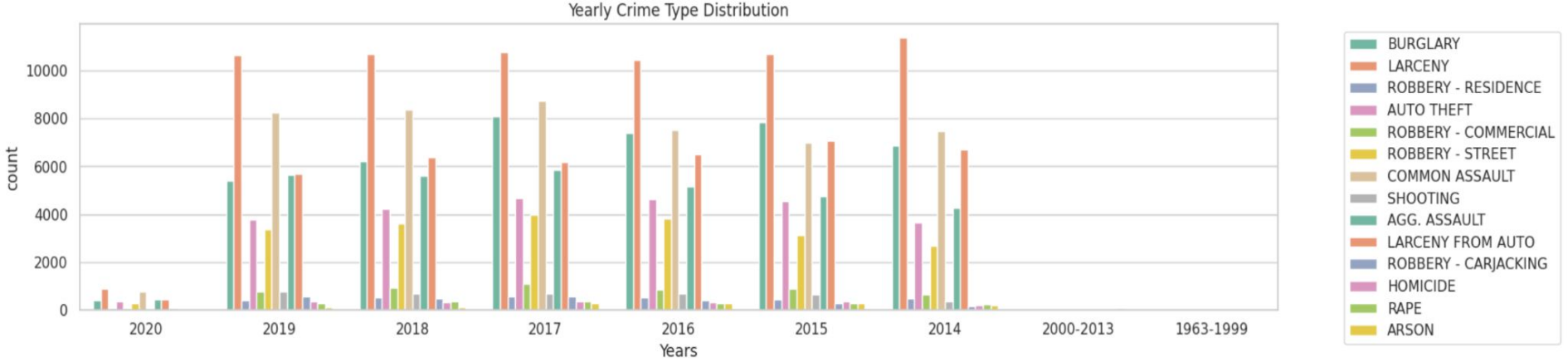
Task 3

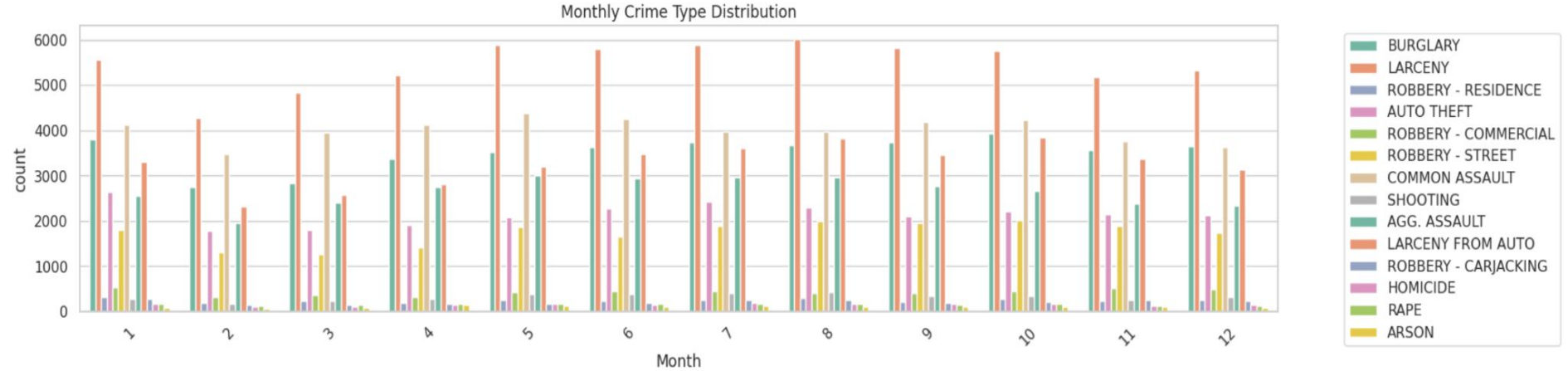
Larceny stands out as the most prevalent type of crime observed, while arson,rape, homicide rank as the least frequent ones.



Task 4

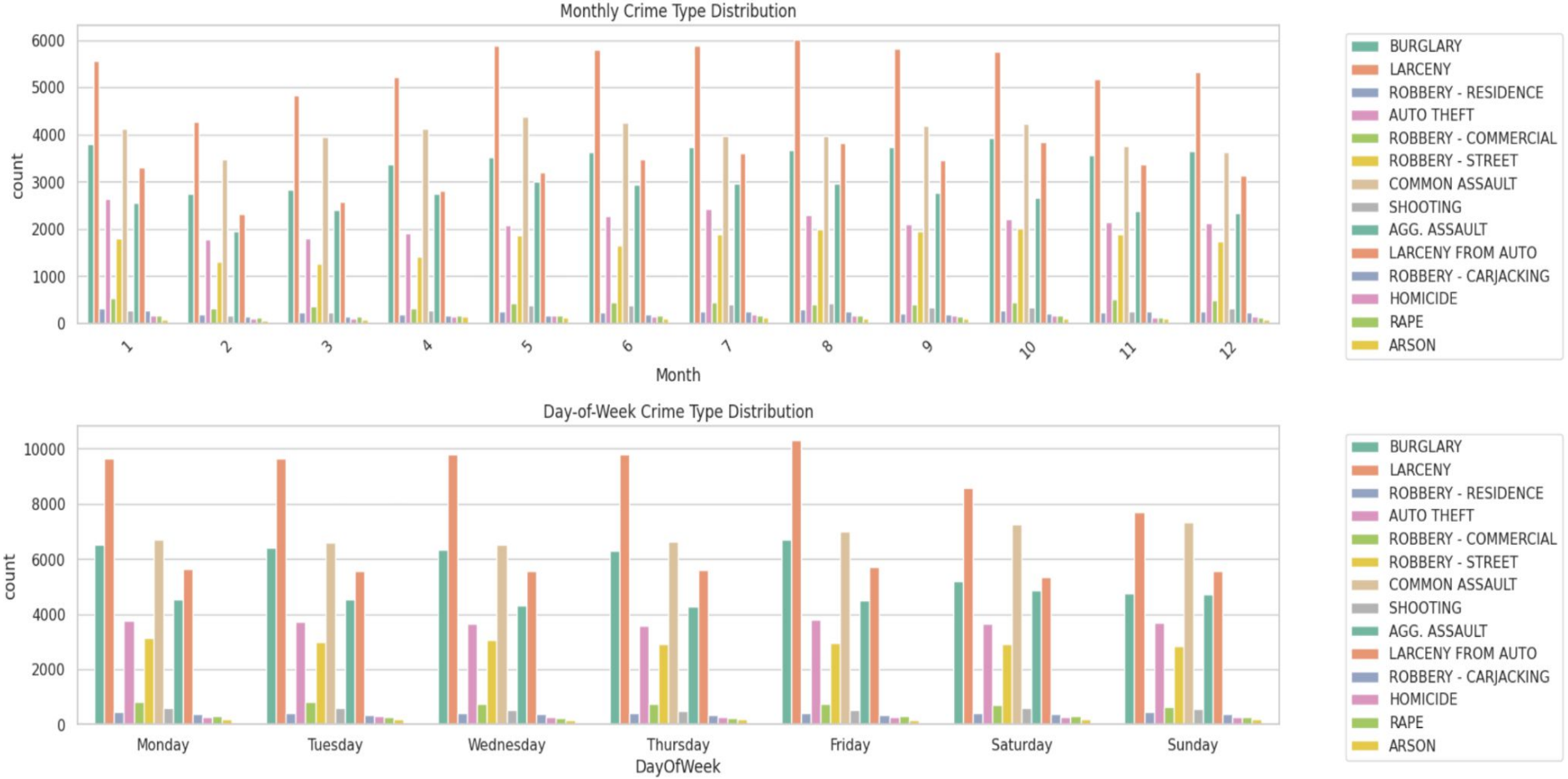
The average count of larceny incidents for the years spanning from 2014 to 2020 surpasses 10,000, and the average monthly larceny count exceeds 5000.





Task 4

The average monthly count of larceny incidents surpasses 5000, whereas the daily count for the same exceeds 8000.

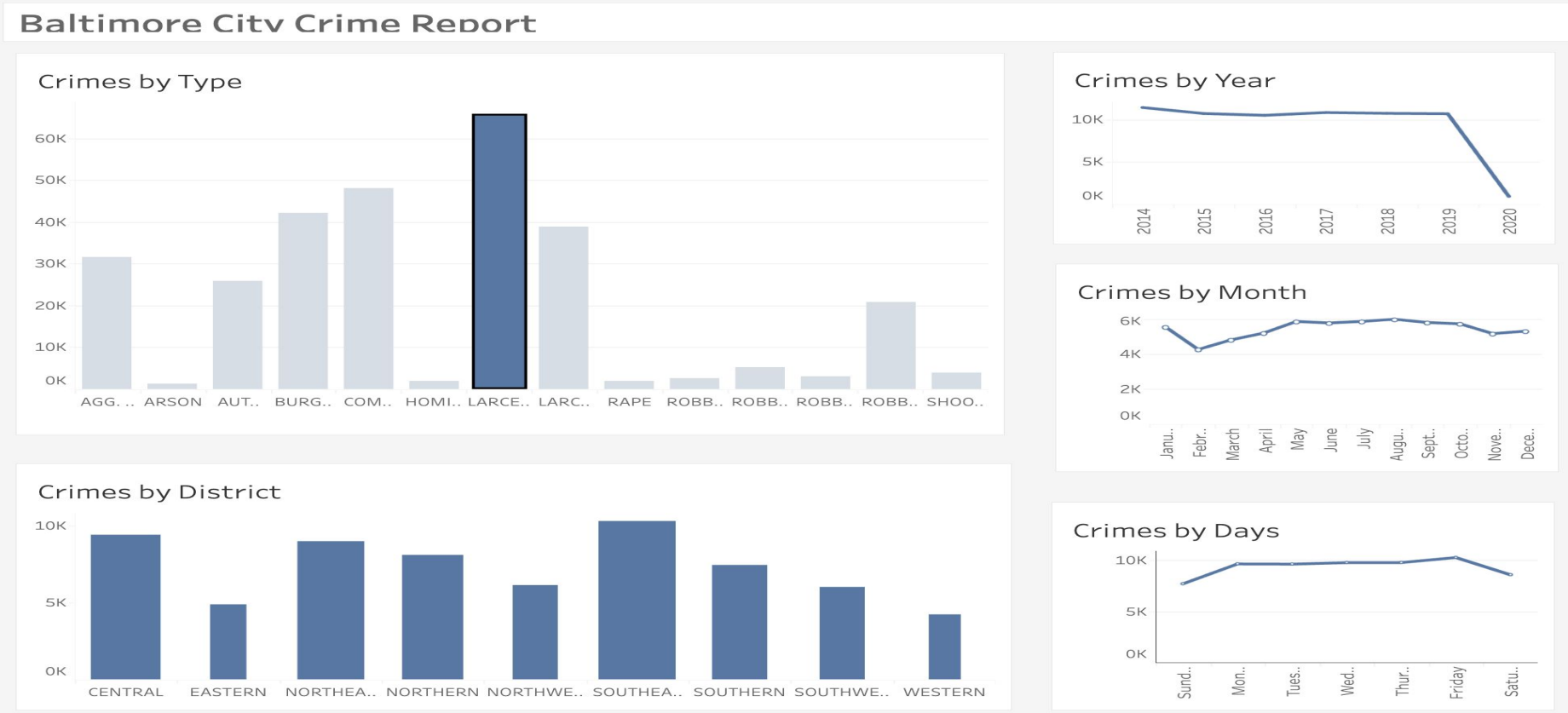


Task 5

The left-hand interactive dashboard displays crime incidents categorized by district (geographical region) and crime types. On the right-hand side, the dashboard provides a breakdown of crime incidents on a yearly, monthly, and daily basis.

Task 5

If the user selects the "Larceny" crime type, the remaining charts are dynamically updated to display larceny-related incidents based on geographical regions, sorted by year, month, and day.



Task 5

If the user chooses the Southeast region, all other charts are promptly updated to display crime data specific to the Southeast region, organized by year, month, and day.

Task 5

The user has the option to click on any specific year, month, or day to retrieve a report detailing the type and district of crimes that occurred during that particular time-frame.

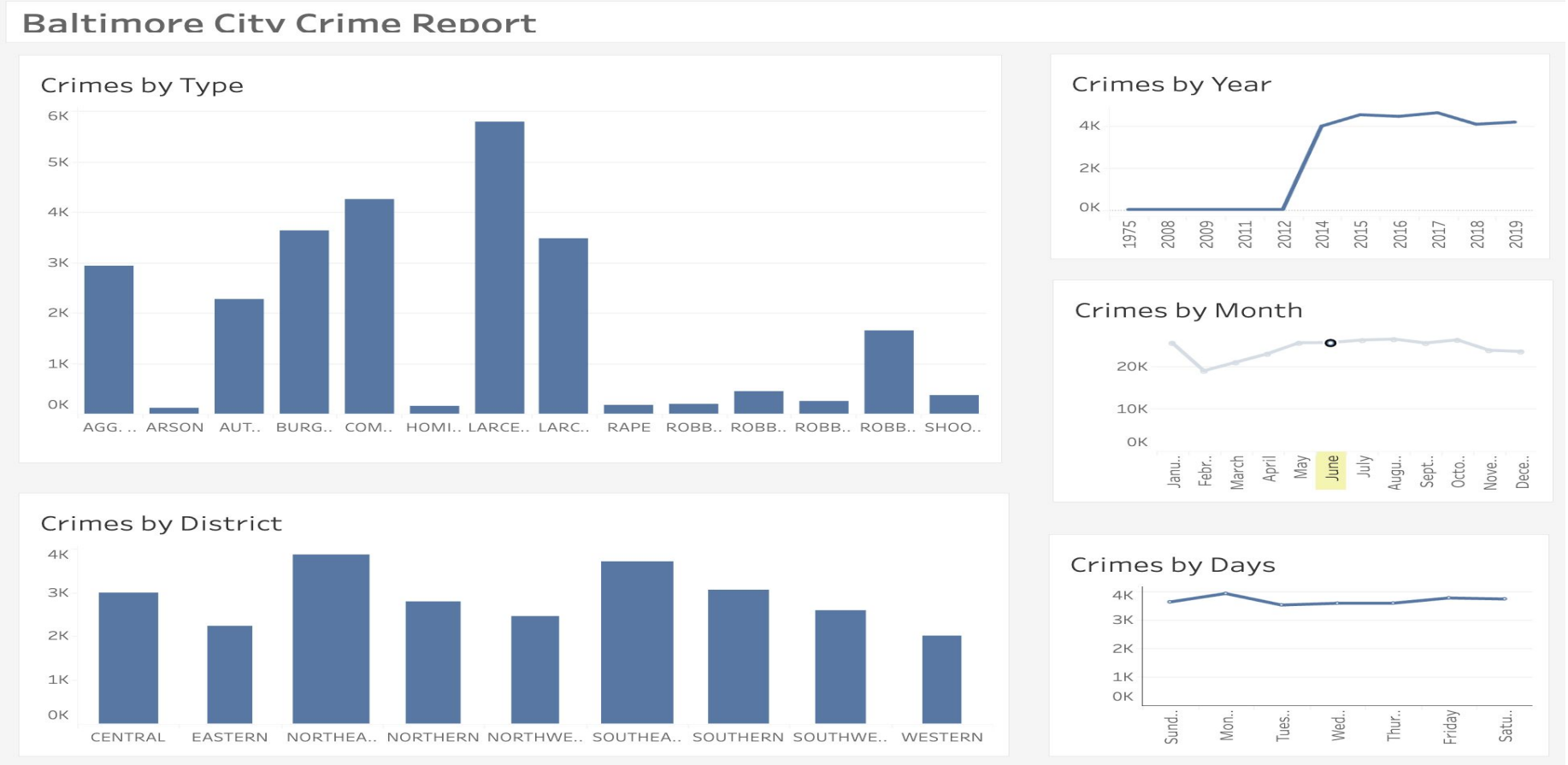


Tableau Dashboard Link

https://prod-useast-b.online.tableau.com/t/vrushalipatil/views/BaltimoreCityCrimeReport/Dashboard1/7c78e2a3-c8e9-42ec-afc1-04c8a584d376/e254a9f8-b8e0-4a98-907a-89a3e82e4b78

Thank you!