INTRODUCTION

My interest in data and its significance has grown over the past few years. I got into data Analytics / Data Science in 2021 and full-time in 2022 when I started my master's program in Informatics and majored in Data Science at the University of Louisiana at Lafayette. I enjoy making stories with data and gaining insights from it. My adventure into the field of data has had a significant impact on my life. I've been following Avery Smith and a few other people on LinkedIn to find out more about their data experience. I've always been stunned by the feedback from others that took on his challenge, and it inspired me to set a goal for myself to try it out.

Throughout this challenge, I learned about several tools for cleaning data, methods for better data analysis, and various software and its applications. Some of the software and programs used in this challenge included Open Refine, Google Sheets, Tableau, SQL, and Python. I gained more knowledge about the many data career opportunities and the different types of analytics.

The mini-projects we had to complete at the end of each day were the most exciting aspect of the challenge since they encouraged me to learn more about the subject and post something on LinkedIn.

We used a real-world dataset for this challenge that included more than 100,000 rows of data from the crime site in New York City for Q1 of 2018.

WHAT I LEARNED

1. During this challenge, I learned:
2. Different tools can be used to clean messy data.
3. There are a lot more career paths in data than we are aware of
4. Data visualization gives us a clearer idea of what the information means by giving it visual context

From analyzing the data, below are some interesting trends from the NYC Crime Dataset:

1. Misdemeanor has the highest level of offense
2. Brooklyn is the largest borough with an incident of 40,365.

THE PROBLEM

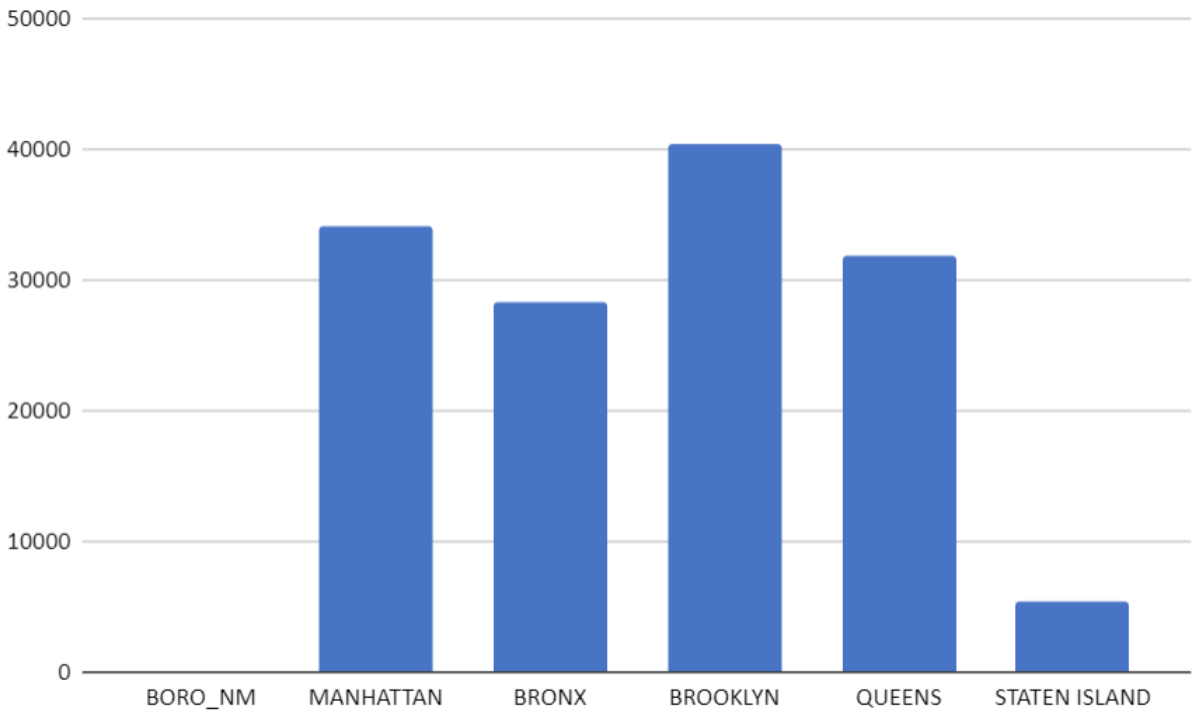
The New York Police Commissioner was concerned about the crime and requested insight from the 2018 Q1 Crime data to address the issue.

THE DATA

This data set has over 100,000 rows and around 36 columns of data. As previously mentioned, the data collection required cleaning because it had some incorrect entries, duplicates, and inconsistent units of measurement. Google Sheets and Open Refine were used to clean the data. Because I was accustomed to using Excel, I used both because I wanted to try them both.

ANALYSIS

Once the data was cleaned efficiently, we were asked to answer some questions required by the commissioner and the NYC Police Department. We started with simply Google Sheet data visualization showing the count of Boroughs with the most crime. The Bar chart below shows that Staten Island has the lowest crime committed.



The next step was to determine which borough had experienced the most criminal activity. After importing the data, I ran the queries on it using SQL in this case. Find the query I used below:

```
SELECT BORO_NM, OFNS_DESC,
Count(*) AS incident_num
FROM "Crime in NY"
GROUP BY BORO_NM, OFNS_DESC
ORDER BY incident_num DESC
Limit 5
```

The result of the query is below:

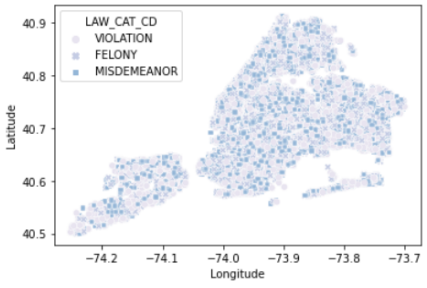
BORO_NM	OFNS_DESC	incident_num
Search	Search	Search
MANHATTAN	PETIT LARCENY	6299
BROOKLYN	PETIT LARCENY	5400
BROOKLYN	HARRASSMENT 2	4705
BRONX	HARRASSMENT 2	3897
MANHATTAN	GRAND LARCENY	3808

We can observe that Manhattan has had the most Petit Larceny incidents, totaling 6,299, followed by Brooklyn, which had 5,400 Petit Larceny incidents.

Another intriguing analysis was the design of a map for the degree of offense: felony, misdemeanor, or violation. This visualization was made with Python's matplotlib library. For this analysis, I personally used Jupiter notebook for Python instead of Google Colab, which was used in the challenge. I made a bar chart that shows the count of boroughs using the Seaborn Library.

I used matplotlib to create a scatterplot with the latitude and longitude of the crime occurrence.

```
In [35]: 1 sns.scatterplot(data=df, x="Longitude", y="Latitude", hue="LAW_CAT_CD", style="LAW_CAT_CD")
Out[35]: <AxesSubplot:xlabel='Longitude', ylabel='Latitude'>
```




```
In [ ]: 1
```

Using Tableau Public to create my dashboard was my favorite part of the project. All thanks to Avery, I began using Tableau because of Avery, who shared his prior project, "Tableau Visualization using your own LinkedIn Data." I created a dashboard with information on different aspects of the crimes found in the NYC Crime Dataset. You may filter it through my dashboard, where I have the map, which displays the longitude and latitude of the borough names, the offense description, the number of incidents based on the borough name, offense statistics by charge, and the ratio of the degree of offense, among other things.

Link to Tableau Dashboard to explore.

This dashboard was built in other to help the commissioner and the police department analyze and monitor the crimes in the NYC boroughs.

CONCLUSION

This challenge was a good experience. I learned a lot that will help me in my journey of becoming a data scientist. 21daystodata provided me with the opportunity to learn about new tools, such as Open Refine and Flourish, that are simple to use and will be useful in my future positions. I also broaden my already-existing knowledge of Python, SQL, and Excel.