**Preliminary report**

**Team Members**

Xu Wu 109607168

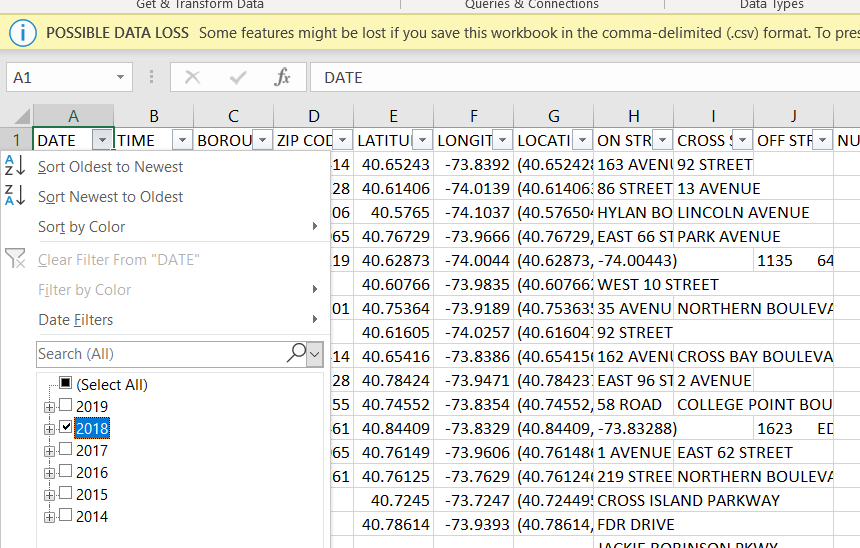
Chongxuan Yin 110445665

**Overview**

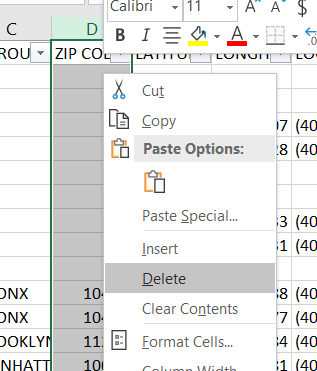
This is the data from <https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95> which contains all motor vehicle collisions in NYC that have occurred in the past few years. The data we want has too many rows, missing values, unrelated attributes, and so on. Thus, we first do the data cleaning and reduction for the data. Then, we make a large NYC map, and combined with the latitude and longitude we get in the data to draw points on the map. In order to distinguish these points, we add different colors to the points in the different boroughs. Last, we add mouse over point function to show the information of the collision.

**Data Cleaning and Reduction**

1. Because we only care about the data last year, we open the data by Excel and filter the data for 2018.

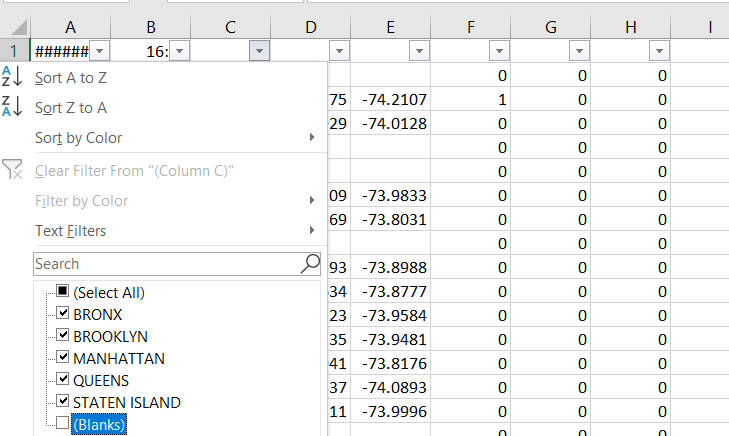


1. Remove the attributes that we don’t care about or contain too many missing values by deleting these columns in EXCEL.

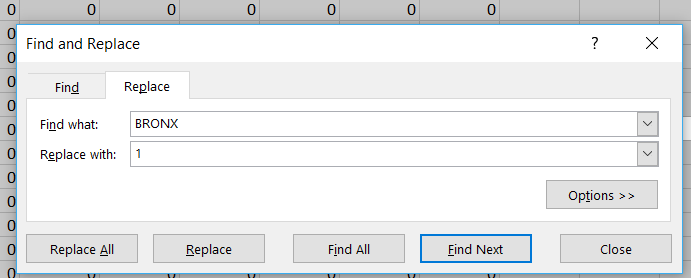


The attributes we delete are ZIP CODE, LOCATION(we already have LATITUDE and LONGITUDE), ON STREET NAME, CROSS STREET NAME, OFF STREET NAME, CONTRIBUTING FACTOR VEHICLE 1, CONTRIBUTING FACTOR VEHICLE 2, CONTRIBUTING FACTOR VEHICLE 3, CONTRIBUTING FACTOR VEHICLE 4, CONTRIBUTING FACTOR VEHICLE 5, UNIQUE KEY, VEHICLE TYPE CODE 1, VEHICLE TYPE CODE 2, VEHICLE TYPE CODE 3, VEHICLE TYPE CODE 4, and VEHICLE TYPE CODE 5.

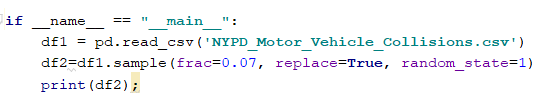
1. Since we have a large number of data, we remove the rows that do not have all the attribute values by filtering and deleting the rows in EXCEL.



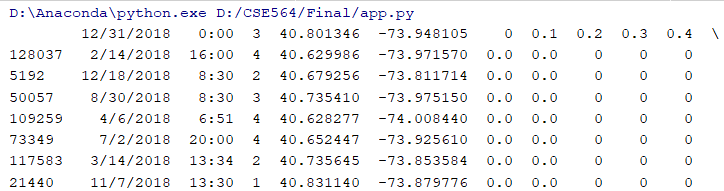
1. Translate all the types of the attribute values into numbers. For example, if the value of attribute BOROUGH is BRONX, we can translate it as 1.



1. Implement random sampling. Then, reduce the data to around 10000 rows.

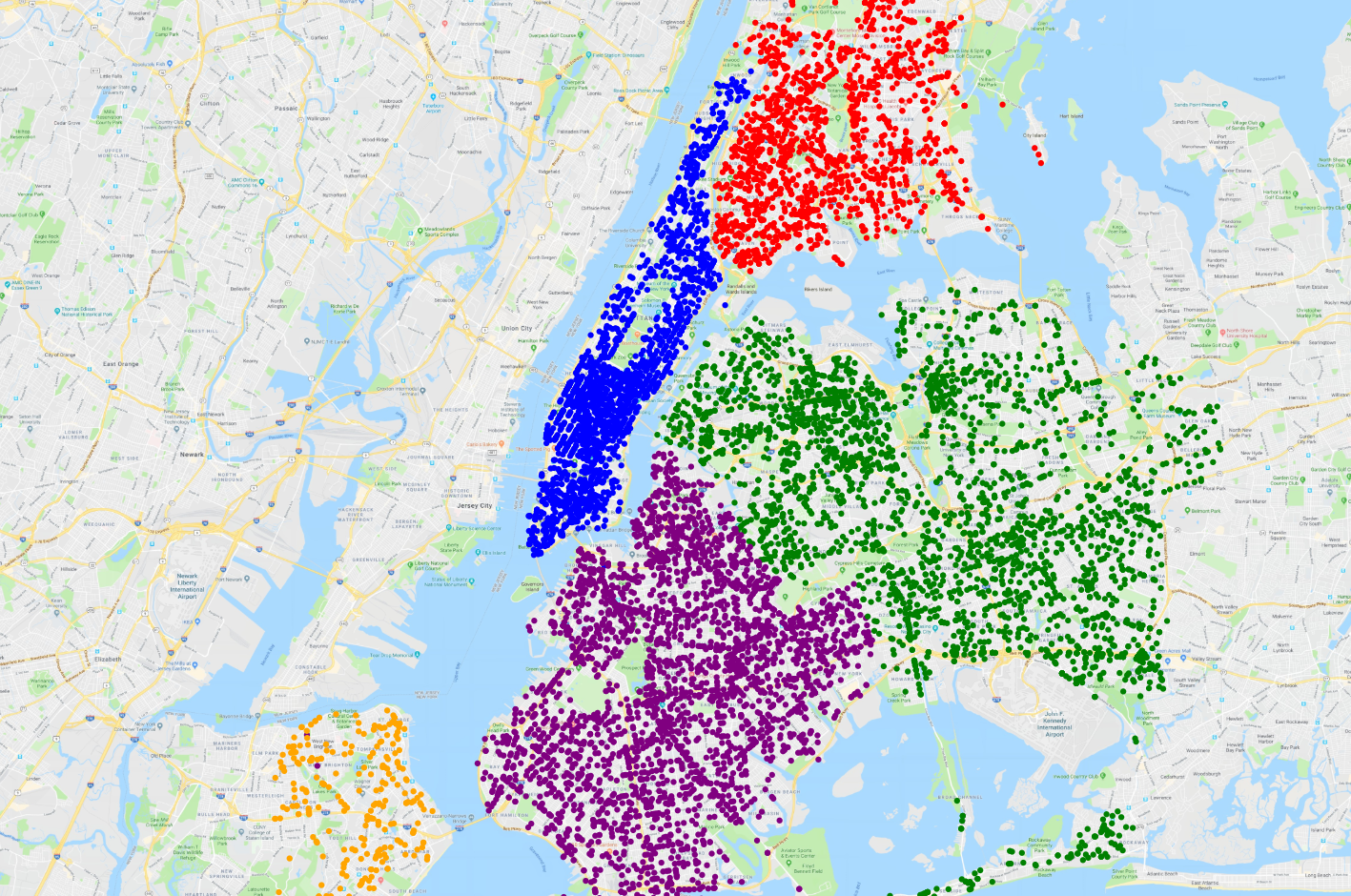


Result:





**Layout**

1. We need a NYC map as a background image to show exact location of each point, such as which part of road. At beginning, we consider using D3 to draw this map. However, it is difficult to find every road information. Then we decide get a large road map from google map.
2. Mapping point to map by Latitude and longitude.
3. Points in different boroughs have different color.
4. Mouse over a point will show the date and time of the collision.

