Report on Capstone Project

Topic: Battle of Neighbourhood for helping group of hospitals to open up Emergency Units

Business problem: To help bit chain of hospitals help identify the areas or locations to open new emergency care units

1.Introduction

1.1Background

One of the most important reason for the death of the victim in the case of accidents is the lack of emergency care available in the nearby areas of accident. If the proper and prompt emergency care and attention can be provided to such patients, there is 100% chance for them to recover and chances of death will be drastically reduced. The main cause for these accidents are due to vehicle collision and there are various reasons such as inattention of the driver, outside car distraction, pavement slippery are a few to name. Whatever may the cause, the main reason for the cause of death is unavailability of emergency services in the localities in the neighbourhood of accidents. The hospitals has the social responsibility of opening these emergency units which will help the accident victims to be addressed with immediate action with care.

1.2 Problem

Data that contain the information about the vehicle collisions, the reasons and the number of time collisions occurred in the specific area can be used to find the areas which are more prone to accidents and the recommendation can be made to the chain of hospitals who are looking for opening new emergency units in the nearby locations.

1.3 Interest

Definitely the big group or chains of hospitals would be interested in getting these exact locations identified for opening up their new emergency units, which will definitely help in saving the lives of many accident prone victims.

The approach followed by this group of hospitals is to locate the nearby accident prone areas and zones based on the previous history of such data. Followed by locating such zones in one city, it can be similarly applied to the neighbouring cities or states with the battle of neighbourhood technique.

The suitable historical data analysed will be able to provide the details of the locations which are more prone to accident and thereby suitable for setting

up emergency care units. In this project such an attempt is made to analyse the existing data where the number of collisions of the vehicles are more and battle of neighbourhood logic is applied to help the group of hospitals.

2. Data Collection and Cleaning:

2.1 Data Sources:

Vehicle collision data will be cleaned and analysed to identify the locations with maximum number of accidents and the contributing reasons.

The first important data source is the one which contains the details of vehicle collisions across various cities and is available in the link below: https://data.cityofnewyork.us/resource/qiz3-axqb.json

The second data source contains the geo location information of NewYork and is available in the link below:

https://geo.nyu.edu/catalog/nyu_2451_34572

2.2 Feature selection and data cleaning and exploration

The data set contains various features important among them are accident date, accident time, borough, collision id, contributing factors, longitude, latitude, number of cyclists, motorists, pedestrians injured, killed, vehicle types. Many of the entries are NaN, in the first round, all such entries are cleaned up in multiple rounds and the below table displays the contents.

Table1:

	accident_date	accident_time	borough	collision_id	contributing_factor_vehicle_1	contributing_factor_vehicle_2	contributing_factor_vehicle_3	contribu
0	2019-04-18T00:00:00.000	2019-11-29 20:57:00	BRONX	4117220	Traffic Control Disregarded	Unspecified	Unspecified	
1	2019-05-10T00:00:00.000	2019-11-29 08:58:00	MANHATTAN	4129801	Passing Too Closely	Unspecified	NaN	
2	2019-04-29T00:00:00.000	2019-11-29 18:30:00	BRONX	4123153	Unspecified	Unspecified	NaN	
	2019-05-	2019-11-29						

The location wise data with number of collisions are derived and is displayed in the below table

Table 2:

```
BROOKLYN          215
QUEENS            181
BRONX             120
MANHATTAN         118
STATEN ISLAND      13
Name: borough, dtype: int64
```

The max number of occurrences of vehicle collision occurs in Brooklyn, Queens, Bronx and Manhattan.

In these areas, the reason for the vehicle is listed in the table below.

Table 3:

	borough	contributing_factor_vehicle_1
0	BRONX	Traffic Control Disregarded
1	MANHATTAN	Passing Too Closely
2	BRONX	Unspecified
6	BROOKLYN	Driver Inattention/Distraction
7	MANHATTAN	Other Vehicular
8	BRONX	View Obstructed/Limited
9	BROOKLYN	Unspecified
10	BROOKLYN	Following Too Closely
11	BRONX	Unsafe Lane Changing
13	BROOKLYN	Driver Inattention/Distraction
14	MANHATTAN	Unspecified

The reasons for the occurrence of vehicle collision is listed in the table below.

Table 4:

```
Driver Inattention/Distraction                238
Unspecified                                   222
Following Too Closely                          94
Failure to Yield Right-of-Way                  71
Passing Too Closely                            49
Passing or Lane Usage Improper                 47
Backing Unsafely                               42
Other Vehicular                                32
Reaction to Uninvolved Vehicle                 29
Unsafe Lane Changing                           28
Turning Improperly                             26
Traffic Control Disregarded                    22
Unsafe Speed                                   18
Driver Inexperience                            17
Alcohol Involvement                            10
Pavement Slippery                               8
```

Also area wise contributing factors are analysed in the locations of max number of vehicle collision occurrences.

Now the geolocation information is obtained by using foursquare APIs and the respective information is tabulated, example table for Bronx neighbourhood is shown in the below table.

Table 5:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

The respective data of neighbourhood is collected for its details for all the top listed vehicle collision occurrences and can now be used to recommend to the group of hospitals for opening their emergency units.

In summary, this project is aimed to understand the vehicle collision data and geo location data to identify the areas with maximum number of vehicle collision occurrence, which can be used by group of hospitals for opening up their emergency units.