Capstone Project - The Battle of Neighborhoods
Report

Name: Madhumitha Ramesh babu

1. Introduction

1.1 Background:

The average American moves about eleven times in their lifetime. This brings us to the question: Do people move until they find a place to settle down where they truly feel happy, or do our wants and needs change over time, prompting us to eventually leave a town we once called home for a new area that will bring us satisfaction? Or, do we too often move to a new area without knowing exactly what we're getting into, forcing us to turn tail and run at the first sign of discomfort? To minimize the chances of this happening, we should always do proper research when planning our next move in life. Consider the following factors when picking a new place to live so you don't end up wasting your valuable time and money making a move, you'll end up regretting. Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.

1.2 Problem:

The crime statistics dataset of Chicago found on Kaggle has crimes in each Boroughs of Chicago from 2014 to 2015. The year 2015 being the latest we will be considering the data of that year which is actually old information as of now. The crime rates in each borough may have changed over time. This project aims to select the safest borough in Chicago based on the total crimes, explore the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhood's using k-mean clustering.

1.3 Interest:

Expats who are considering to relocate to Chicago will be interested to identify the safest borough in Chicago and explore its neighborhood's and common venues around each neighborhood.

2. Data Acquisition and Cleaning
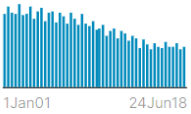
2.1 Data Acquisition

The data acquired for this project is a combination of data for one particular location. The data source of the project uses a Chicago crime data that shows the crime per borough in Chicago. The dataset contains the following columns:

- ID
- Case Number
- Date

- Block
- IUCR
- Primary Type
- Description
- Location Description
- Arrest
- Domestic
- Beat
- District
- Ward
- Community Area
- FBI Code
- X Coordinate
- Y Coordinate
- Year
- Updated On
- Latitude

2.2 Data Cleaning

The data preparation is done by taking the 2014-2015 data collected and the major categories of the crime are pivoted to get the total crimes per borough as per the category.

	ID	Case Number	Date	Block	IUCR		Primary Type	
		6635448 unique values	1Jan01 — 24Jun18	59590 unique values	0820 0486 Other (400)	8% 8% 84%	THEFT BATTERY Other (33)	21% 18% 61%
⚠ Whoops, something went wrong loading your data.								
1	10000092	HY189866	03/18/2015 07:44:00 PM	047XX W OHIO ST	041A		BATTERY	
2	10000094	HY190059	03/18/2015 11:00:00 PM	066XX S MARSHFIELD AVE	4625		OTHER OFFENSE	
3	10000095	HY190052	03/18/2015 10:45:00 PM	044XX S LAKE PARK AVE	0486		BATTERY	
4	10000096	HY190054	03/18/2015 10:30:00 PM	051XX S MICHIGAN AVE	0460		BATTERY	
5	10000097	HY189976	03/18/2015 09:00:00 PM	047XX W ADAMS ST	031A		ROBBERY	
6	10000098	HY190032	03/18/2015 10:00:00 PM	049XX S DREXEL BLVD	0460		BATTERY	
7	10000099	HY190047	03/18/2015 11:00:00 PM	070XX S MORGAN ST	0486		BATTERY	

Data will be cleaned in order to find out which area was most affected and how was it reduced in the different years and what type of crime is most likely to occur is also analyzed.

3.Methodology
3.1Exploratory Data Analysis

3.1.1 Statistical summary of crimes
The describe function in python is used to get statistics of the Chicago crime data, this returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime.

3.1.2 Analyzing the type of data in the data set
Find below the different types of data types available in the data set.

```
Data columns (total 22 columns):
ID                    263770 non-null int64
Case Number           263770 non-null object
Date                  263770 non-null object
Block                 263770 non-null object
IUCR                  263770 non-null object
Primary Type          263770 non-null object
Description           263770 non-null object
Location Description   263427 non-null object
Arrest                263770 non-null bool
Domestic              263770 non-null bool
Beat                  263770 non-null int64
District              263770 non-null float64
Ward                  263768 non-null float64
Community Area        263770 non-null float64
FBI Code              263770 non-null object
X Coordinate          257888 non-null float64
Y Coordinate          257888 non-null float64
Year                  263770 non-null int64
Updated On            263770 non-null object
Latitude              257888 non-null float64
Longitude             257888 non-null float64
Location              257888 non-null object
dtypes: bool(2), float64(7), int64(3), object(10)
memory usage: 40.8+ MB
```

4.Discussion
The aim of this project is to help people who want to relocate to the safest borough in Chicago, expats can choose the neighborhoods to which they want to relocate based on the most common venues in it. For example, if a person is looking for a neighborhood with good connectivity and public transportation.

5.Conclusion
This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood. We have just taken safety as a primary concern to shortlist the safest borough of Chicago. The future of this project includes taking other factors such as cost of living in the areas into consideration to shortlist the borough, such as filtering areas based on a predefined budget.