Capstone Project - The Battle of Neighborhoods: Darko Knezevic

Introduction: Business Problem:

Boston is the capital of the U.S. state of Massachusetts. It is a city with a growing economy and a great place to acquire an education. Some of the most prestigious schools within the U.S.A are located in Boston. A lot of large companies are seeking to place their firms in Boston. It is a global city that is placed among the top 30 most economically powerful cities in the world. Boston's colleges and universities exert a significant impact on the regional economy. Boston attracts hundreds of thousands of students from around the world, who contribute billions of dollars annually to the city's economy. The city is considered highly innovative for a variety of reasons, including the presence of academia, access to venture capital, and the presence of many high-tech companies. Tourism also composes a large part of Boston's economy.

However, the amount of crime needs to be understood for Boston to continue to thrive as a major hub for innovation, tourism, academia, and employment. If the Boston Police Department knew more about which areas to concentrate on, then Boston's citizen, tourists, and students would feel much safer. Certain areas within the city could be targeted for more protection by the Boston Police Department based on data.

This project aims to select the safest and least safest areas in Boston based on the total crimes, explore the venues in these areas and finally cluster the areas using k-mean clustering. This report will be targeted to the Boston Police Department who are looking to analyze crimes incidents, as the city continues to be a leader in some many fields. The crime statistics will provide an insight for the Boston Police Department towards their approach in dealing with different areas of city. The most common venues in the safest areas will be explored.

Data:

Based on definition of our problem, factors that will influence our decision are:

- The total number of crimes committed in each major area during the most recent year available. For example, acquiring data from the most recent year available will be done to do analysis.
- The most common venues in each of the areas that are the safest.

Following data sources will be needed to extract/ generate the required information:

- Preprocessing a real-world data set from Kaggle showing the Boston Crimes from the last few years. Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

- Creating a new dataset of the areas that are the safest in Boston and generating their co-ordinates.

The Boston crime file provides information about the following:

- INCIDENT_NUMBER
- OFFENSE_CODE
- OFFENSE_CODE_GROUP
- OFFENSE_DESCRIPTION
- DISTRICT
- REPORTING_AREA
- SHOOTING
- OCCURRED_ON_DATE
- YEAR
- MONTH
- DAY_OF_WEEK
- HOUR
- UCR_PART
- STREET
- LATITUDE
- LONGITUDE
- Location

Data Set URL: https://www.kaggle.com/AnalyzeBoston/crimes-in-boston

Current Boston police Districts Map: