# Fatal Accidents by Surrounding Locations

BY EDWIN JOHNSON

# Introduction

- Death by accidents is a leading cause of human mortality
- If location and venues at that location have a clear indication as to whether accidents will be fatal
- Objective: Analyze the location data of fatal accidents and analyze the surrounding area for information
- Business question:
  - If an accident is clustered around a certain area it would be in that government officials' best interest to act or improve infrastructure at that location. Individuals as well would be interested in obtaining the knowledge to avoid being another statistic.

# Data

- Data required:
  - Accident data from Toronto where there are clear longitude and latitude data as well as fatalities due to accidents
  - Venue data particularly related to the locations nearby the accident

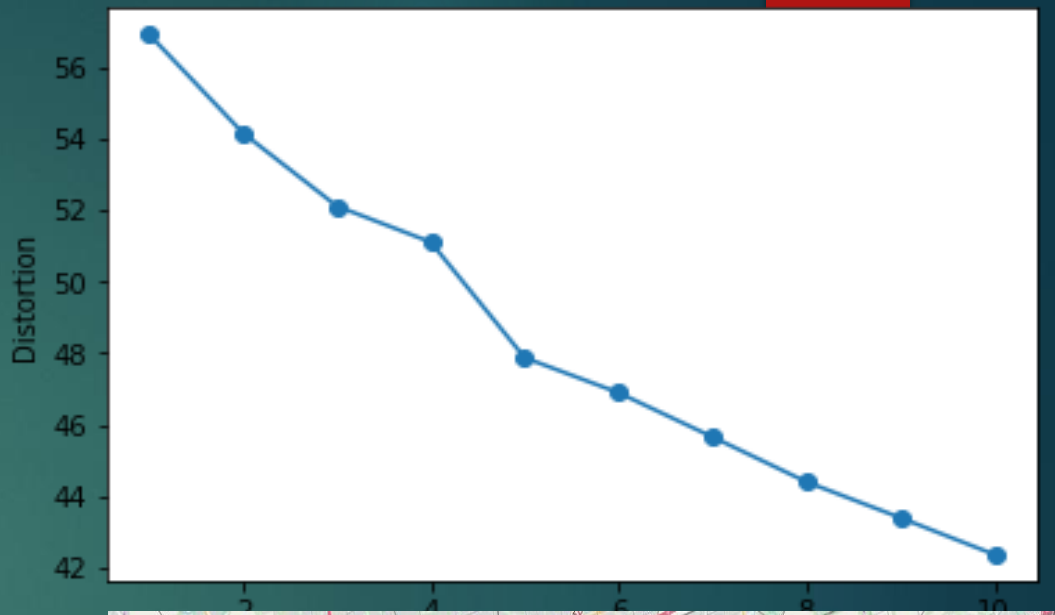
- Sources of data:
  - Kaggle Dataset for Toronto Killed or Severely Injured
  - Foursquare API for Venue Data

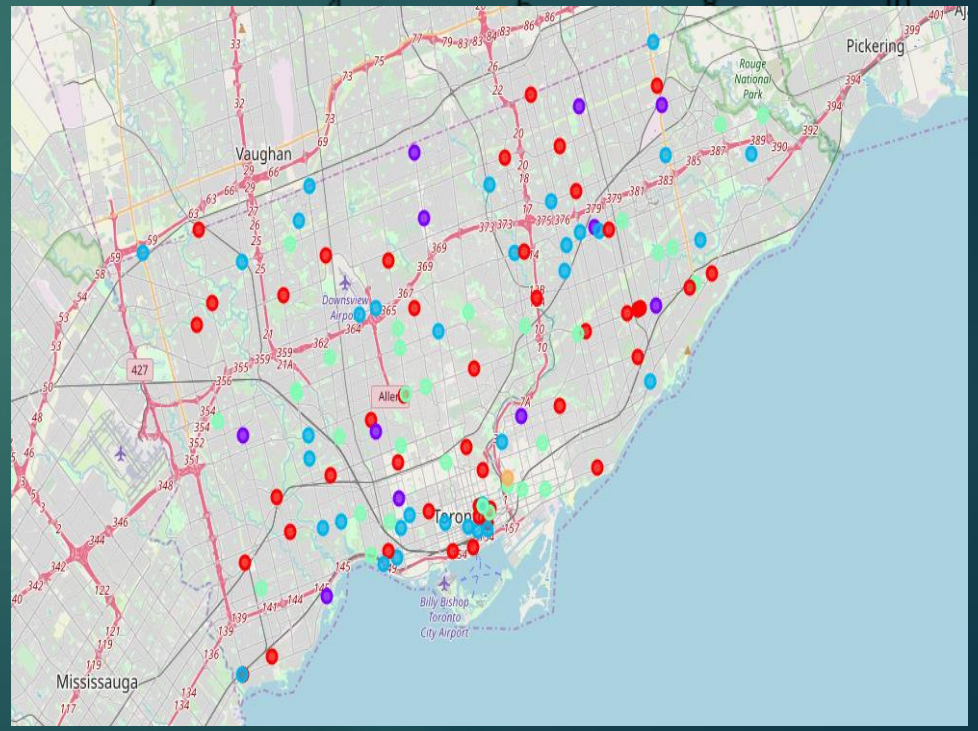
# Methodology

- Importing and cleaning KSI data set for relevant features such as year, longitude, latitude, accident number, and fatality.
- Visualize the accidents
- Using Foursquare API to get venue data
- For each accident place relevant data in the same data frame
- Perform cluster analysis selecting the best K-means using the elbow method
- Visualize the clusters and interpret results

# Results

- Using elbow method picked 5 clusters and used it to cluster data as shown to the right.
- Unfortunately the results were inconclusive the areas and venues where there were fatalities were not seemingly related by any obvious means.



# Discussion

- Surprisingly most of the fatal accidents were not located near public transportation centers or interstates.
- The different clusters show that there are more restaurants and cafes than most other categories with bus stations being one of the fewest seen
- The data almost has an even distribution of fatal accidents for the area.

# Conclusion and Recommendations

- The clustering showed that there was some relationship between the areas however none that was discernable.
- Accuracy of the model could be improved
- A better method would be recommended for future analysis as the venue and area seem to be giving inconclusive results.