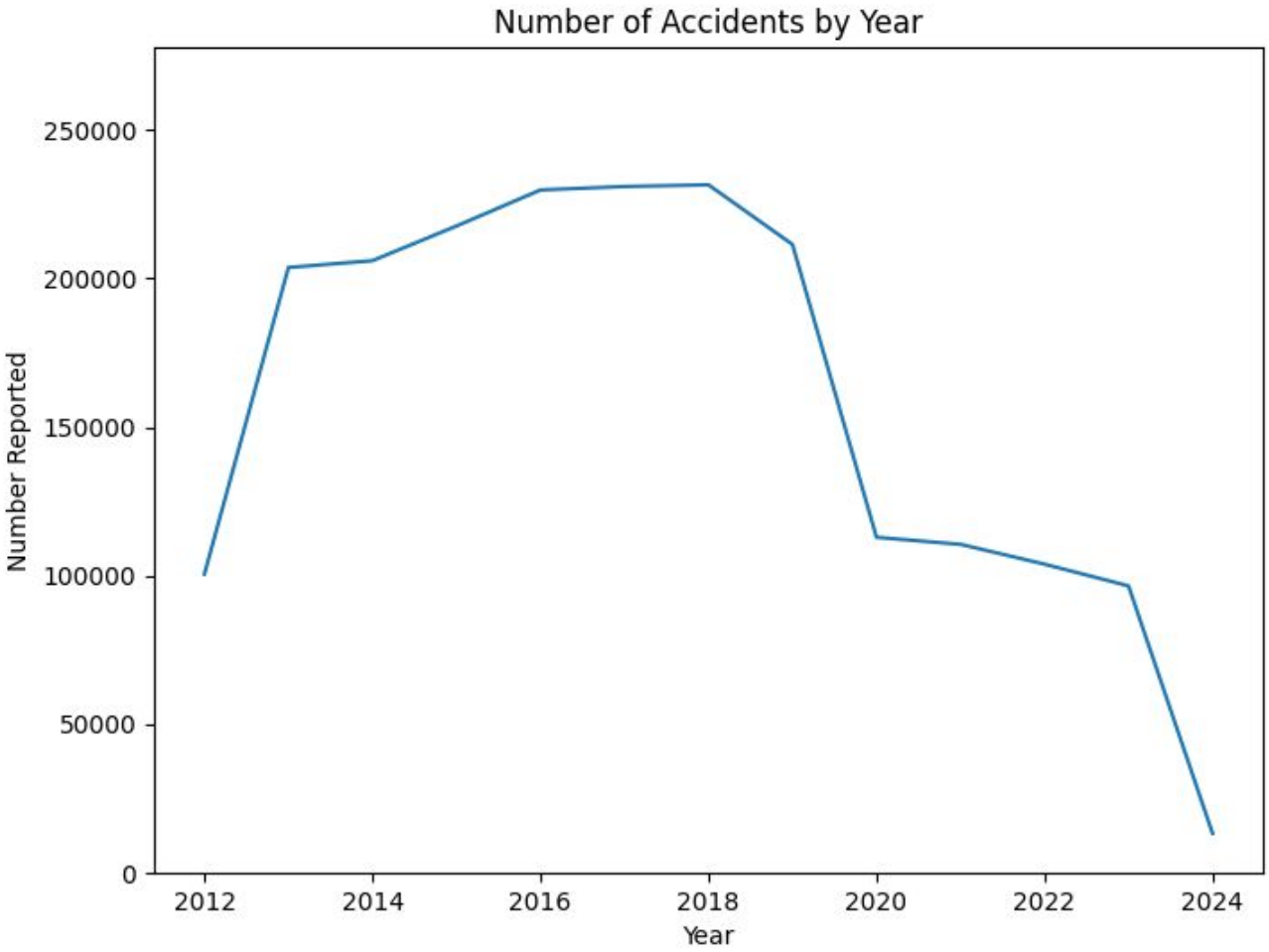
Motor Vehicle Collisions in NYC

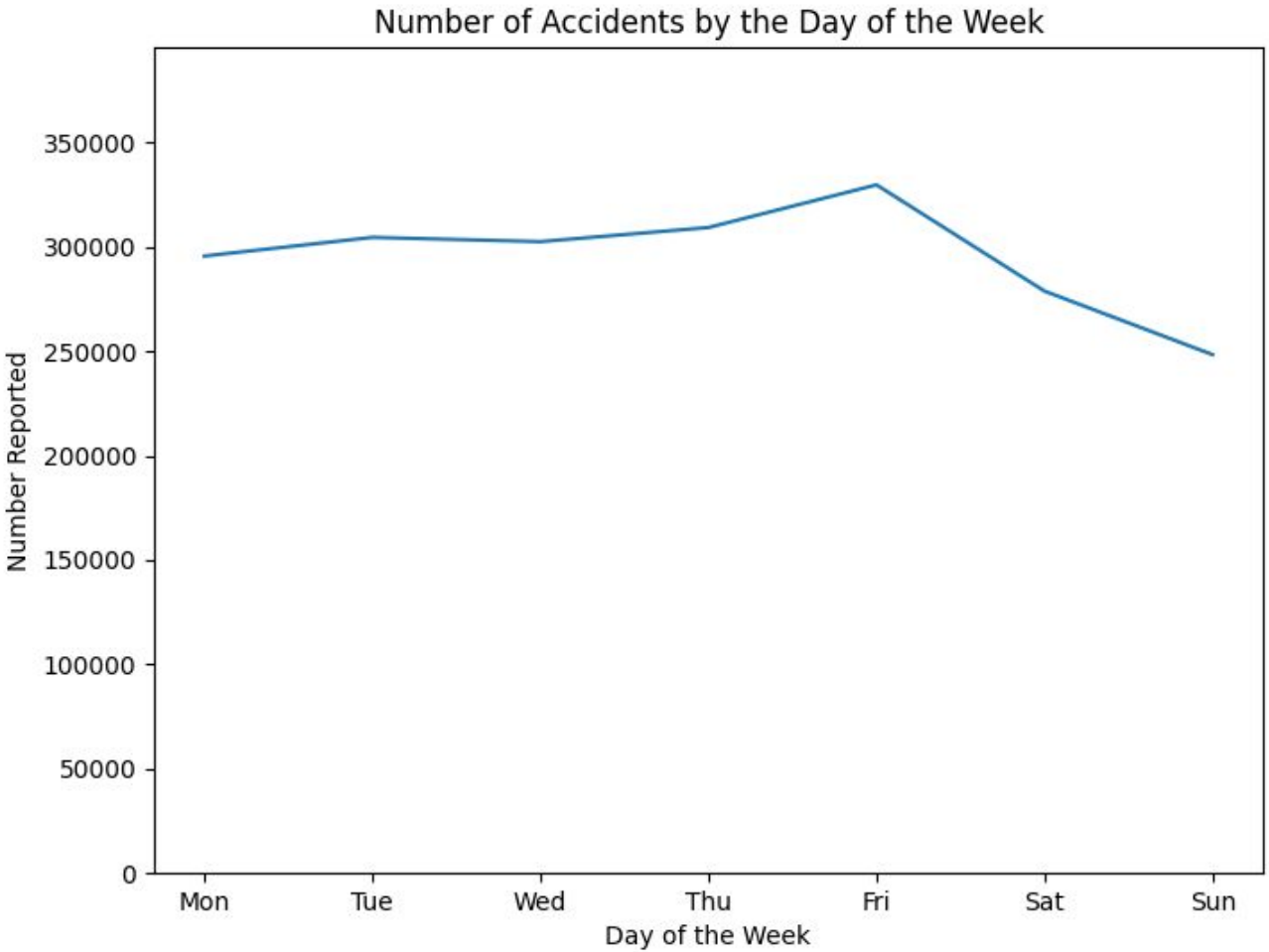
Patrick Hutecker

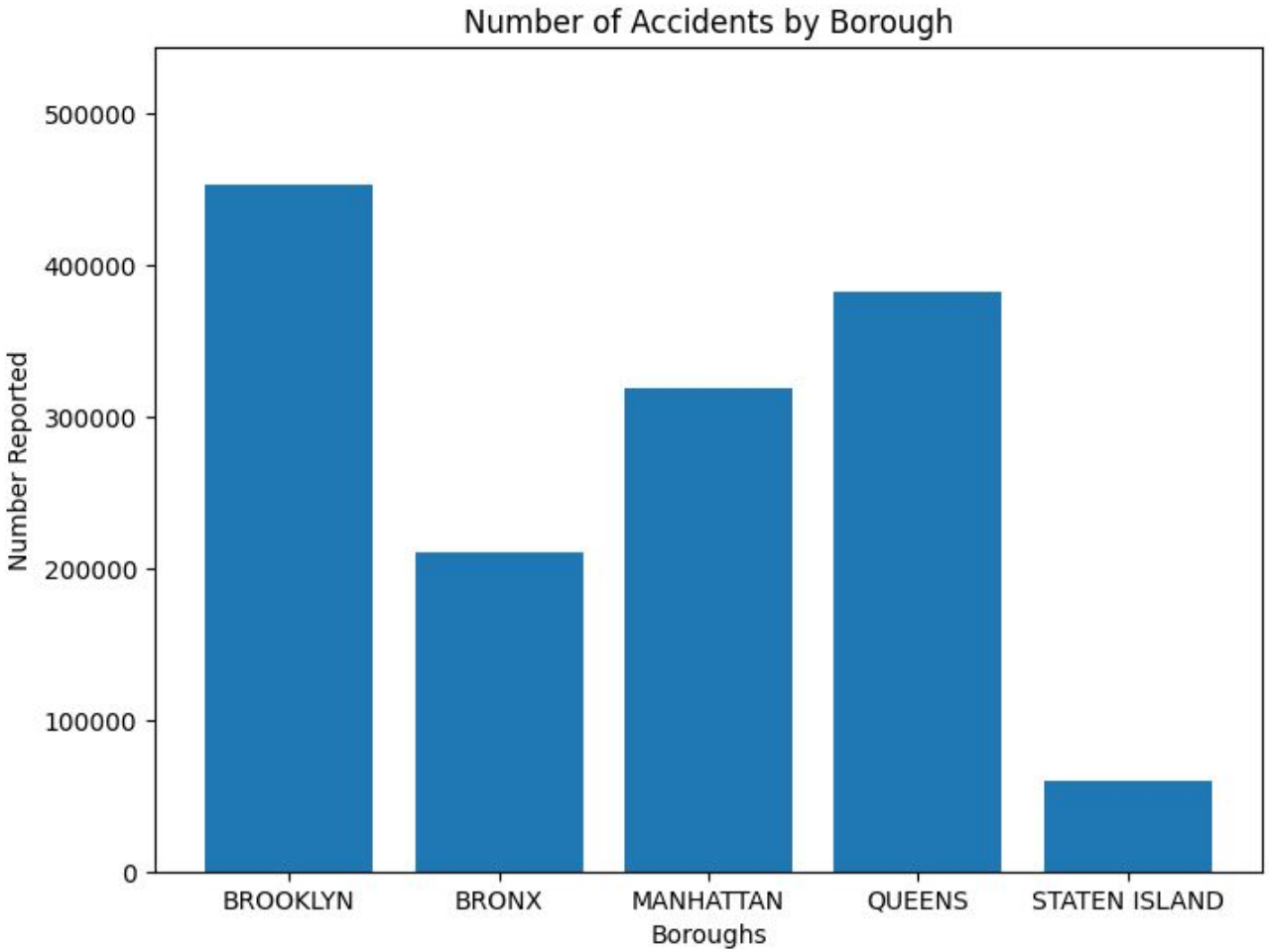
Data Set

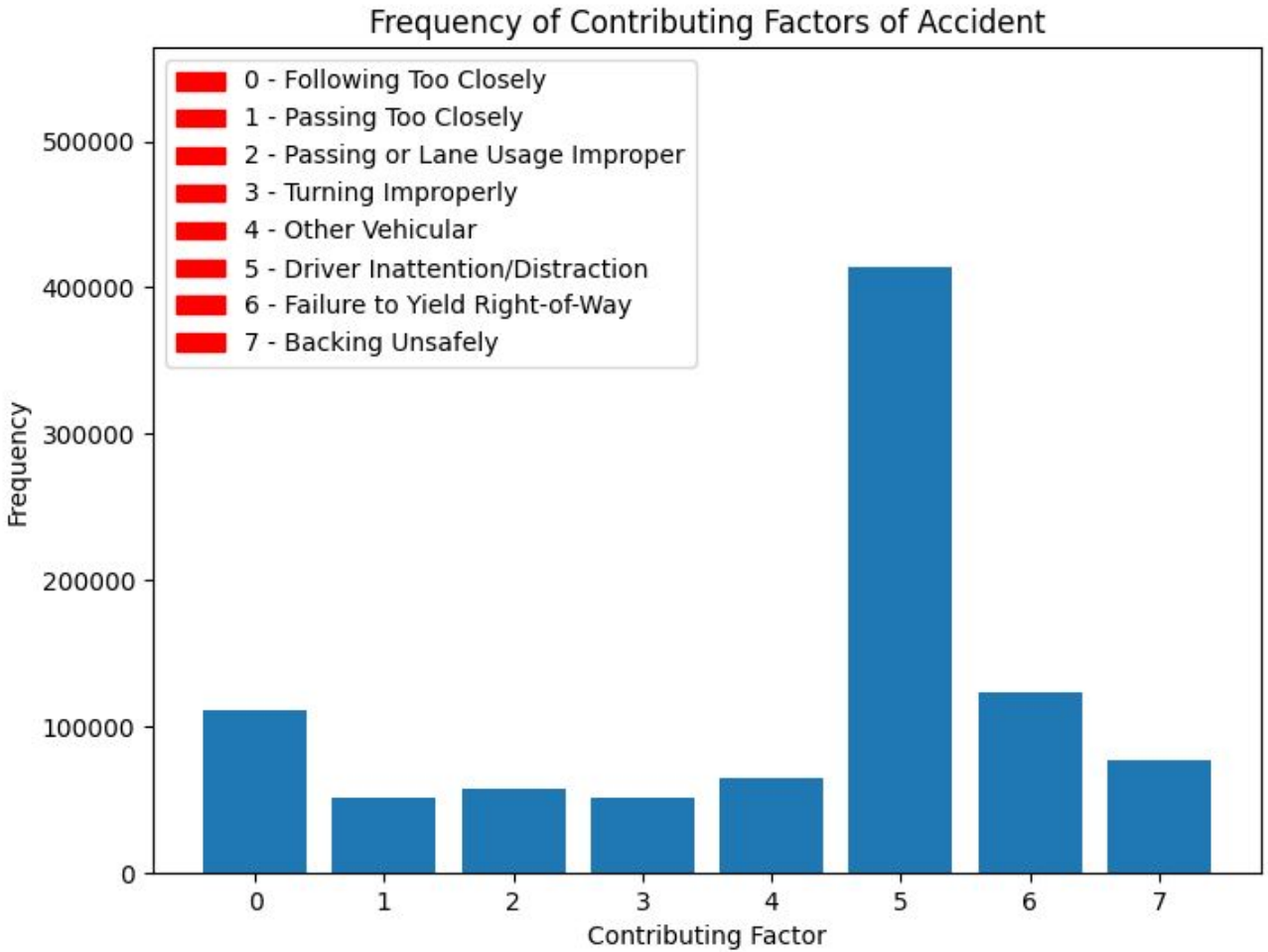
- Motor Vehicle Collisions / Crashes in NYC
- Num data points = **2,069,104**
- Time Range: **2012 - 2024**
- Key Fields
  - Date
  - Location
  - Injuries / Death
  - Cause of accident
  - Vehicle information

Profiling the Data

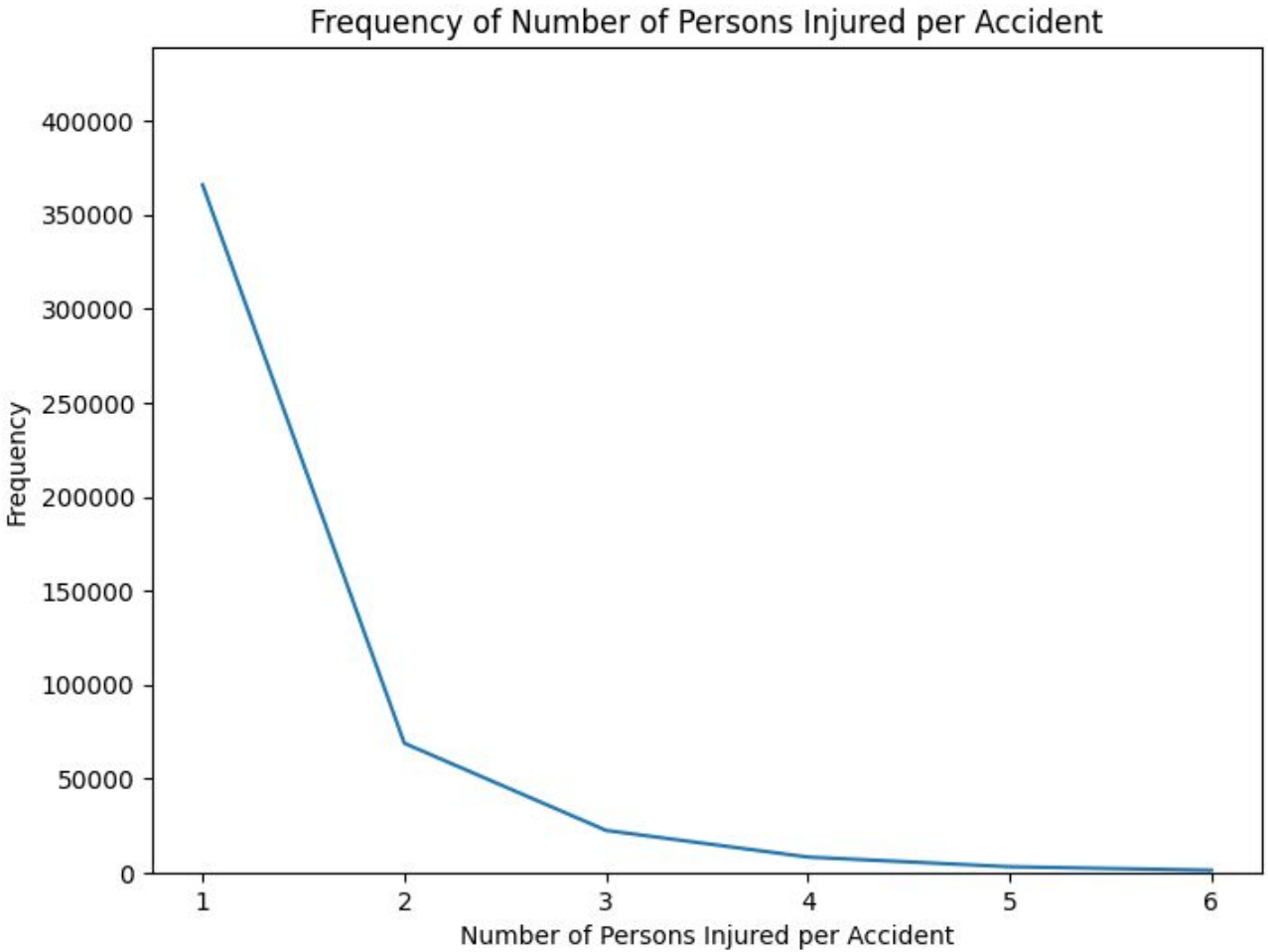

Number of Accidents by Year
250000
200000
150000
100000
50000
0
Number Reported
2012
2014
2016
2018
2020
2022
2024
Year

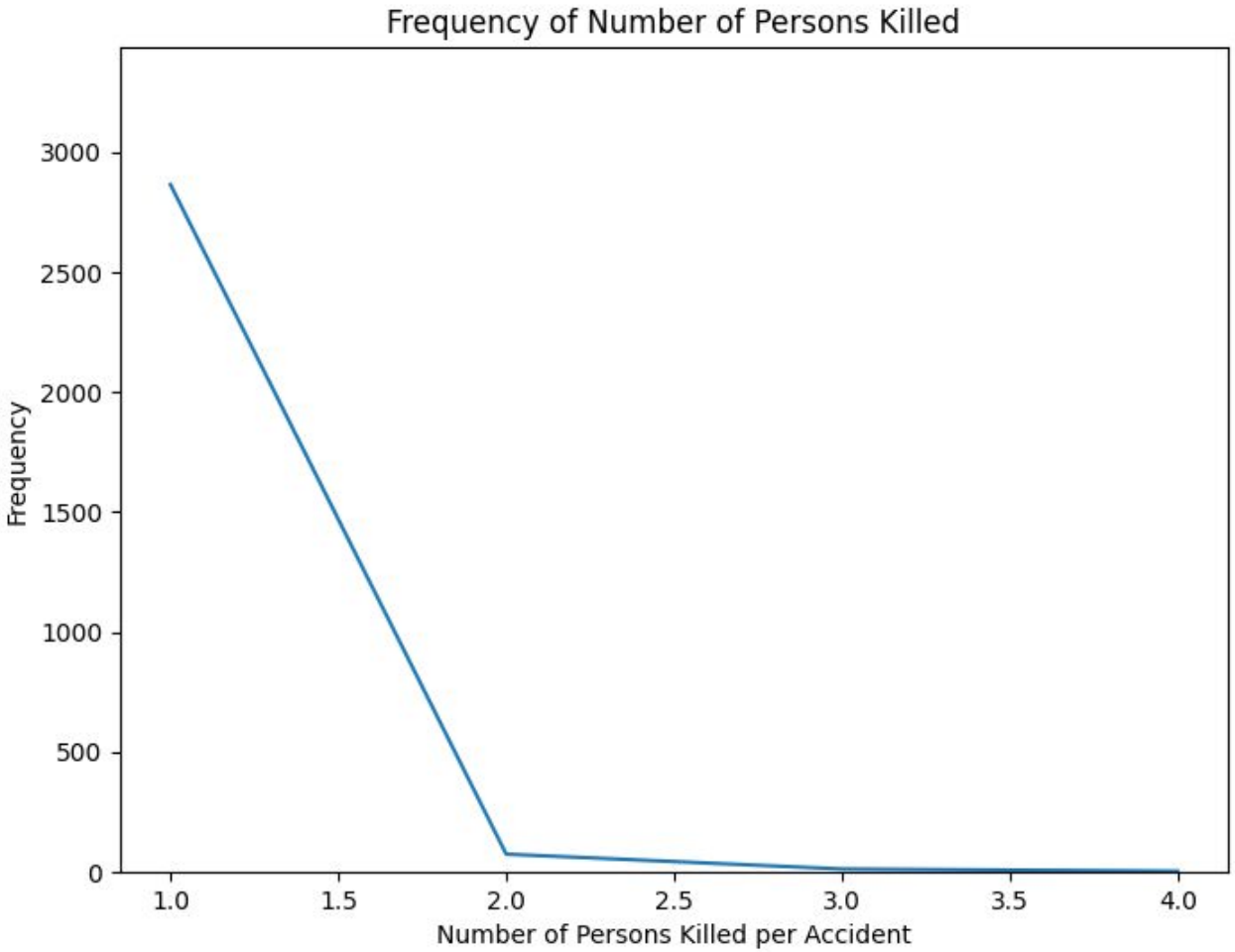

Number of Accidents by the Day of the Week
Number Reported
350000
300000
250000
200000
150000
100000
50000
0
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Day of the Week


Number of Accidents by Borough
500000
400000
300000
200000
100000
0
Number Reported
BROOKLYN
BRONX
MANHATTAN
QUEENS
STATEN ISLAND
Boroughs


Frequency of Contributing Factors of Accident
0 - Following Too Closely
1 - Passing Too Closely
2 - Passing or Lane Usage Improper
3 - Turning Improperly
4 - Other Vehicular
5 - Driver Inattention/Distraction
6 - Failure to Yield Right-of-Way
7 - Backing Unsafely
Frequency
500000
400000
300000
200000
100000
0
0
1
2
3
4
5
6
7
Contributing Factor

Comparing Fields


Frequency of Number of Persons Injured per Accident
400000
350000
300000
250000
200000
150000
100000
50000
0
Frequency
1
2
3
4
5
6
Number of Persons Injured per Accident


Frequency of Number of Persons Killed
Frequency
3000
2500
2000
1500
1000
500
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Number of Persons Killed per Accident

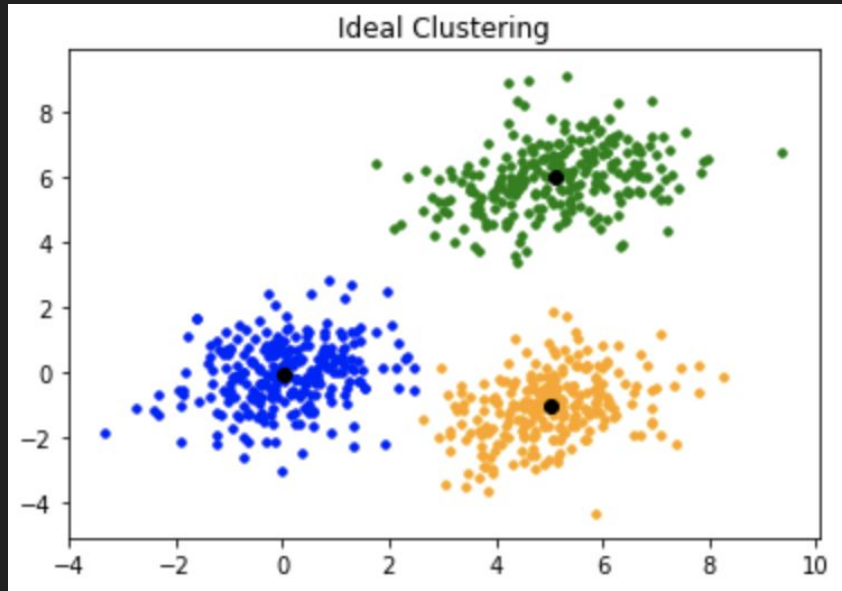
K-means

Factors

- Injuries / Deaths
- **Persons, Pedestrians, Cyclists, Motorists.**

Results

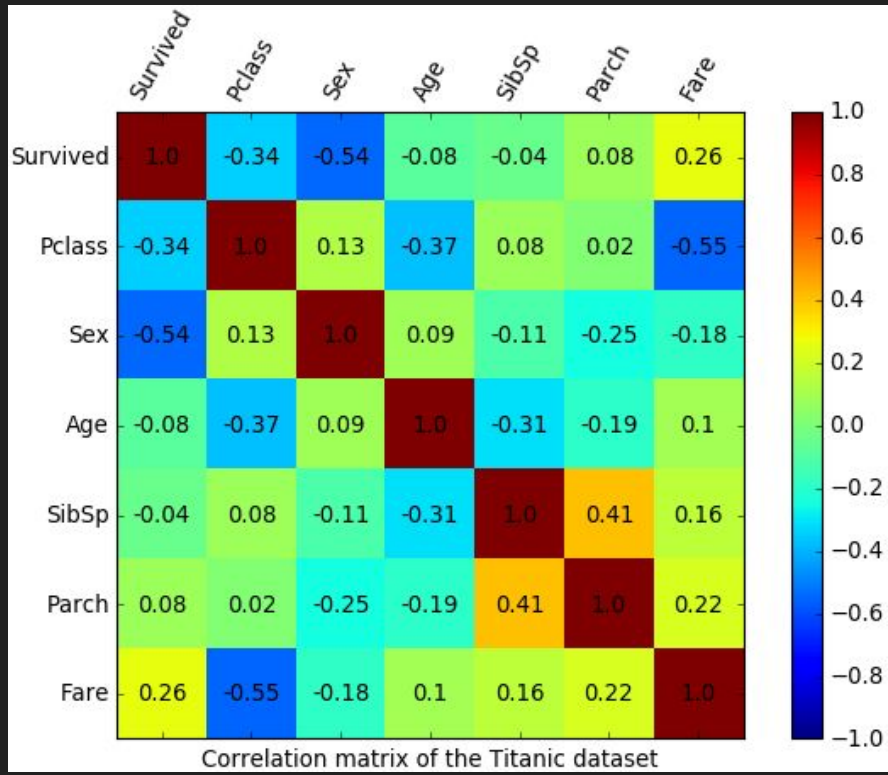
- 1,600,000 no injuries / deaths
- 200,635 Motorist injured
- 111,329 Pedestrian injured



Future Steps

Identify Correlated Fields

- Correlation Matrix
- More in-depth kmeans
- Vectorizing “descriptors”

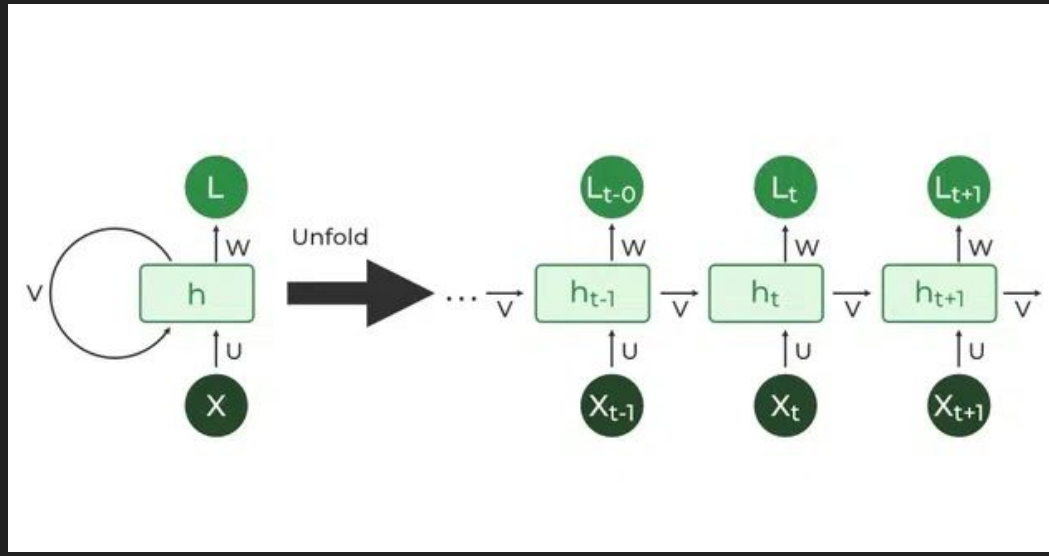


Identify Correlated Fields

- Correlation Matrix
- More in-depth kmeans
- Vectorizing “descriptors”

Build Prediction Models

- LSTM
- Autoencoder
- RNN
- Gradient Descent
- etc.



Questions?