**Relationship between Cities in Maryland**

**Daniel Abebe**

**November 20, 2019**

1. **Introduction**
   1. Background

It’s statistically known that approximately 33,000 to 35,000 persons die due to vehicle accident in United States annually. This number includes any vehicles including 2 or 3 legged motor cycles. The reporting system that supports these reporting is called FARS (Fatality Analysis Reporting System) under DOT (Department of Transportation).Popular Maryland cities neighboring Washington DC and Baltimore city are going to be used for this paper. These cities reside a high number of population and its a center for professionals and people flow from all over the world as a visitor or as a permanent stay. As its diverse area it is, there are a lot of various venues serving the public. This paper will try to relate the venues and fatalities and or similarity between them.

* 1. Problem

Compare Baltimore City vs. Maryland Counties around DC metro area with regard to fatal crashes and venues and observe any relation they might have.

* 1. Interest

It’s the interest of this article to apply machine learning techniques to see if cities are related based on venues, obviously a lot more research can be done if the venues in a city may or may not have fatal crashes though. Not to mention its my interest to select the above mentioned cities because I reside in the DC metro area.

1. **Data acquisition and cleaning**
   1. Data Sources

County and State data is scrapped from here: <http://www.statsamerica.org/CityCountyFinder/Default.aspx>

Fatality crash is from:

<https://www-fars.nhtsa.dot.gov/Main/index.aspx>

* 1. Data Cleaning

1. There are a number of factors or variables that affects to the number of fatalities such as travel speed, weather state traffic lows and so on. But for this paper we are focusing on observing the relationships between cities of interest.
2. **Data cleaning**

The data downloaded from FARS data portal is 2018 fatality. , though FARS encyclopedia latest data they have is for 2017, normally they post data 2 years behind. It was better to take on multiple year data for this project; however thats out of the scope of this project.

Population of State of Maryland is close to 5.8million, Baltimore County is around 800 thousand and Maryland Cities near to Washington DC areas population is close to 2 million. Baltimore and DMV-Maryland cities are around 30miles apart, and the later population size is growing and hence the wealth.

Missing values are all removed if any of the variables are missing.

1. **Feature Selection**

After data cleaning, the features that are selected from the crash dataset are the following:

State, County, Pedestrian, Persons, Month, Road Type, latitude, longitude. A number of variables are left out, not for any reason other than limiting the scope of this project. On the other hand, FARS store masked data points to represent each variable, the location information specifically the latitude and longitude provided for each case helped to retrieve the city and state data.

1. Exploratory Data Analysis References:

The FARS program, In terms of car crash fatalities, person’s variable includes drivers and passengers, and pedestrians on the other hand. Major known highway roads, month the crash occurred with respect to the location point of interest are considered for this analysis.

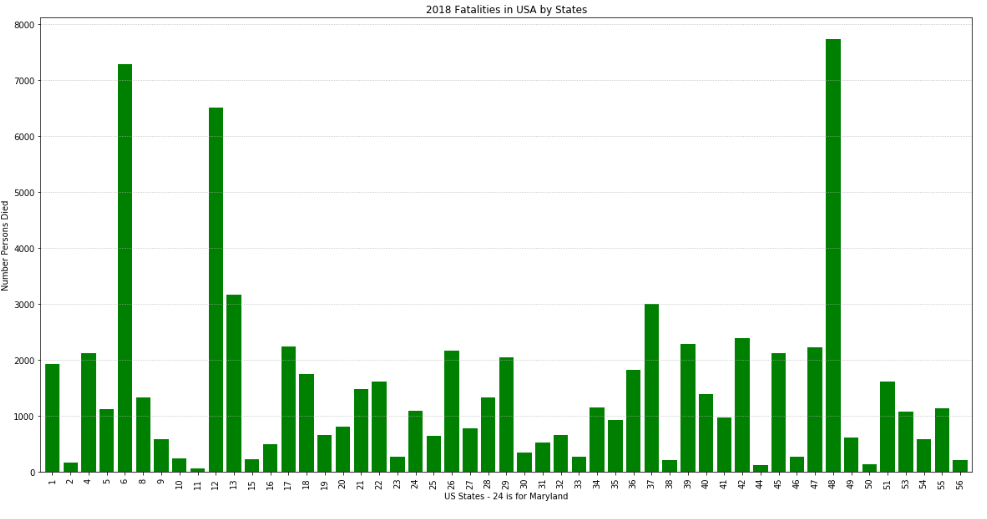
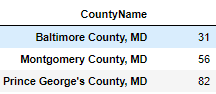


Fig. 1 Number of Fatalities in USA, year 2018

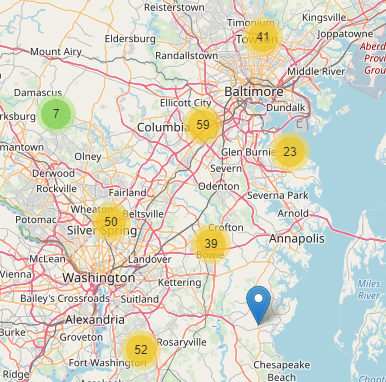
**Number of Cities by County**



The 10 most common Venues on cities from the above Counties.



1. **Crash Data and Analysis:**



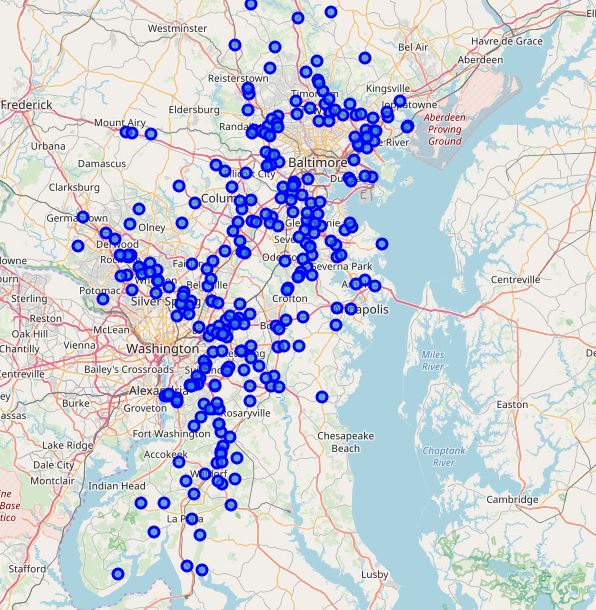
**Number of Fatalities count:**

Prince George’s county: 91

Montgomery County: 50

Baltimore County: 123

**Mapping Fatal accidents in these 3 Counties**



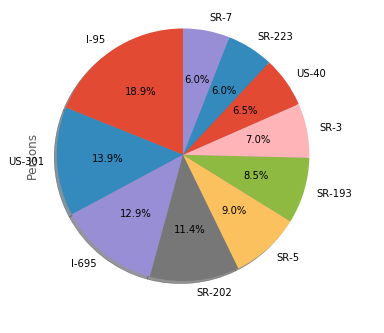
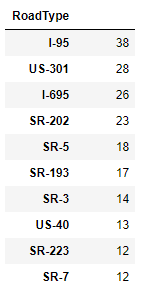
**Relationship between Persons and pedestrian:**



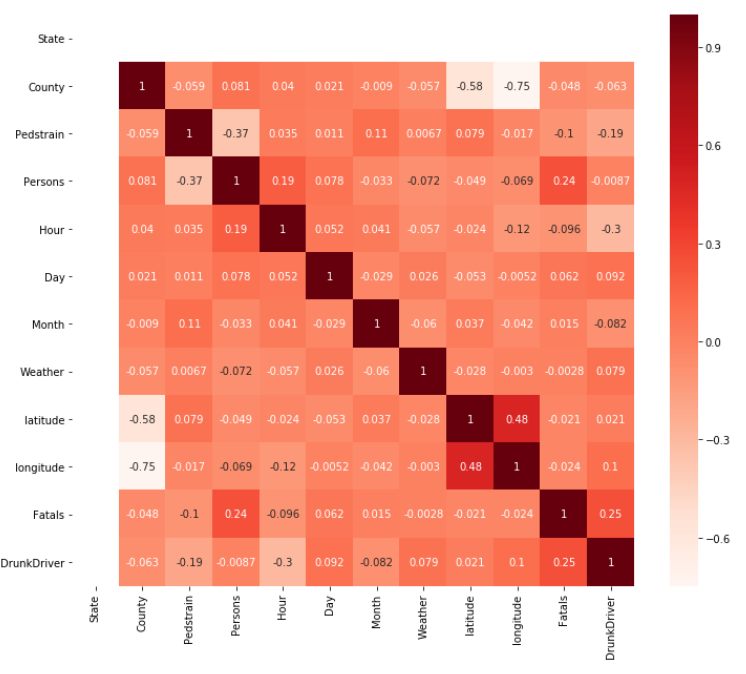
1. **Relationship between Road and fatality counts**

TOP 10 roads by fatality counts around these locations combined.

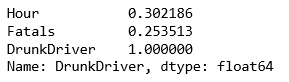
Road type vs. Person fatality.



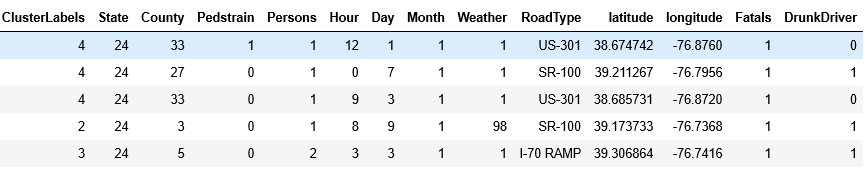
Knowing I-95 as the busiest road in the area, it accounts to the highest percentage of fatalities.

**Pearson correlation for other variables:**

Selecting highly correlated features

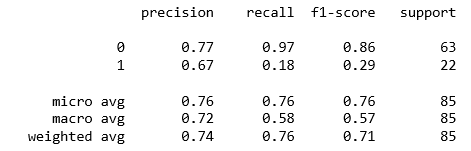


Applied Cluster label:

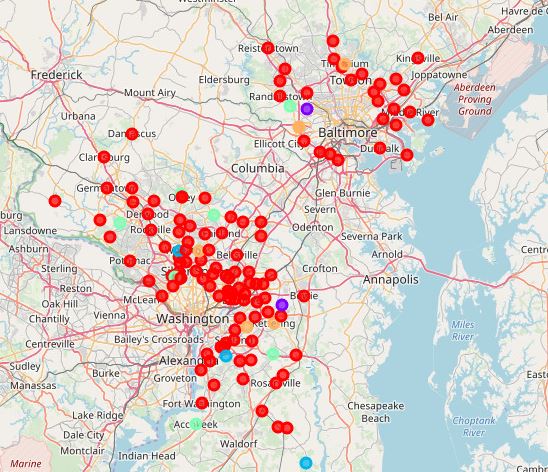


Evaluation

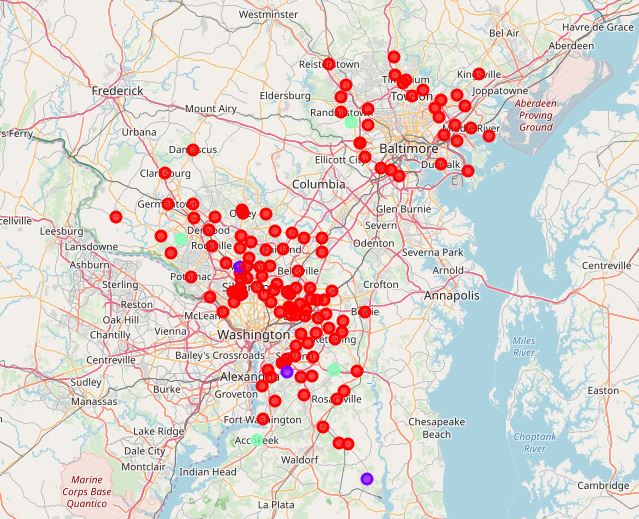
We can check precision, recall,f1-score using classification report



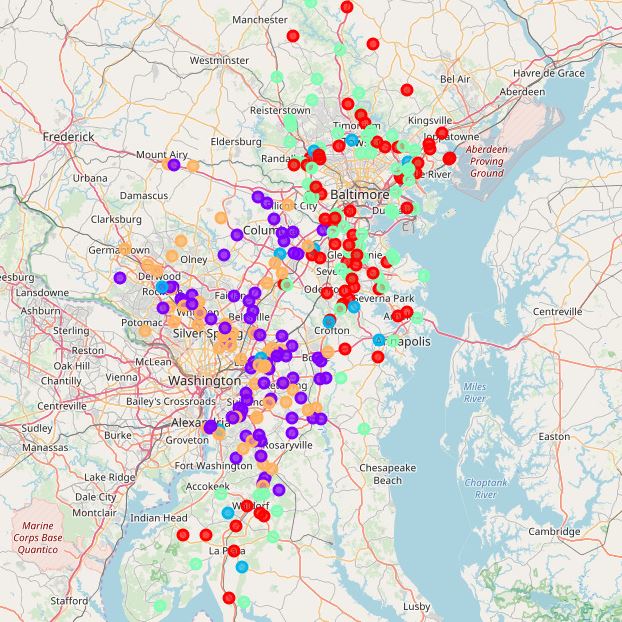
Venues clustering with K=5 applied on 3 counties



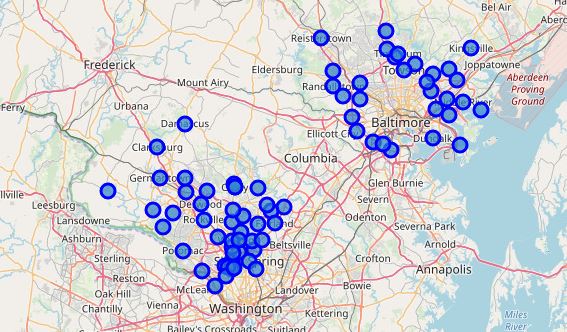
Venues clustering with K=3 applied on 3 counties



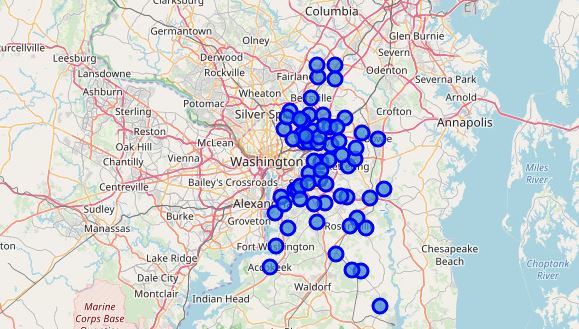
**Fatal Crushes clustering with K=5 applied on 3 counties**



Montgomery and Baltimore City Cities



**Prince George County Cities**



**Conclusions**

In this study, I analyzed the relationship between Venues and Car Crashes at the higher level, and it looks that the crashes are higher and are occurring at the most common venues in cities. These cities are similar in character, though fatalities in Baltimore occur mostly in the outskirt of the city.

**References**

<https://www-fars.nhtsa.dot.gov/Main/index.aspx>

1. <https://www.iihs.org/topics/fatality-statistics/detail/state-by-state>
2. Population size:
3. <http://www.togetherweteach.com/TWTIC/uscityinfo/20md/20md.htm>
4. <https://www.gobankingrates.com/making-money/economy/baltimore-vs-washington-dc/>

Population

<https://www.maryland-demographics.com/counties_by_population>