ACRS Report Type

In this data we didn’t find anything unusual. There were three categories reported which consisted of Fatal Crash, Injury, and Property Damage Crash.

Agency Name

When we looked at the Agency Name feature, we found that there were more than 4 agencies involved; therefore, we decided to keep the top three agencies and gather the rest into an additional category named Other. Montgomery County Police was the top agency involved in the reports.

Bad Weather

In this plot, we can find the 4 values we kept for the weather. Each value is associated with a location value in which we plotted in the latitude and longitude map with its own color value.

Circumstance

In the Circumstance feature we found that most categories had the word N/A included so we went ahead and stripped that away. Once our data was cleaned, we clearly see that other than the N/A’s, the circumstances that followed had to do with water/moist.

Collision type

Under the Collision feature, we found that there were multiple combinations of vehicles so we decided to gather them in similar categories. We decided to gather all those accidents with any involvement turning left under the Left category and the same for Right, and Sideswipe. We kept the top two categories and gathered the Unknown’s and N/A’s with the category Other.

Crashed Date

Cross Street Name

This feature had a large amount of values so we decided as our first step to shorten the data to top 20. Once we did this we were able to visualize that the top value was that of an empty address. As a final result to this data we gathered the top 4 streets along with the Other category involving the empty addresses.

Cross Street Type

In this feature we saw that it only had a small amount of values so we decided to keep most of them. However, we also decided to join the Unknown and the Empty value into the Other category.

Driver At Fault

In this feature we saw that there was nothing unusual. Only three values are present in this feature.

Driver Distracted

When analyzing this data, we realized that the number one reason for accidents was when the driver was not distracted. We went ahead and kept the top three reasons for the accident and gathered the rest into the Other category. This was done because the other values had very little effect in this data.

Driver Substance Abuse

In this feature we had a little over 10 values so we shrunk it down to 3. We kept the value Non detected as it was the number one value, followed by the other value which contained the unknown and N/A values. For the final value we gathered all the other values as we realized that they all fall under the driving under the influence category.

Driver License

This feature had nothing unusual. We kept all the values with the exception of the empty value, we named it Other.

Injury Severity

This feature had 5 values in which the No apparent injury was the most reported.

Light

In the Light feature we decided to gather the valued into three categories. We kept the value under Daylight and we gathered those with some light under one category. Those values under unknown and N/A were added into the other category

Location

In this plot, we can find the values of latitude(y-value) and longitude(x-values) plotted with a red circle on the map.

Municipality

In the Municipality feature we can see that the only significant values are the top 4.