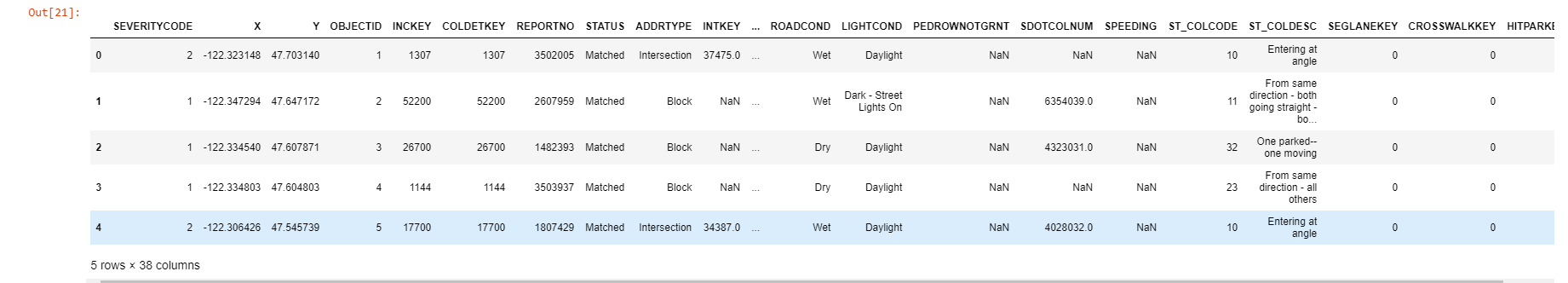
**Coursera Capstone Project: Correlation Between Light Available and Accidents**

Lyba Mughees

August 29th,2020

**Introduction**

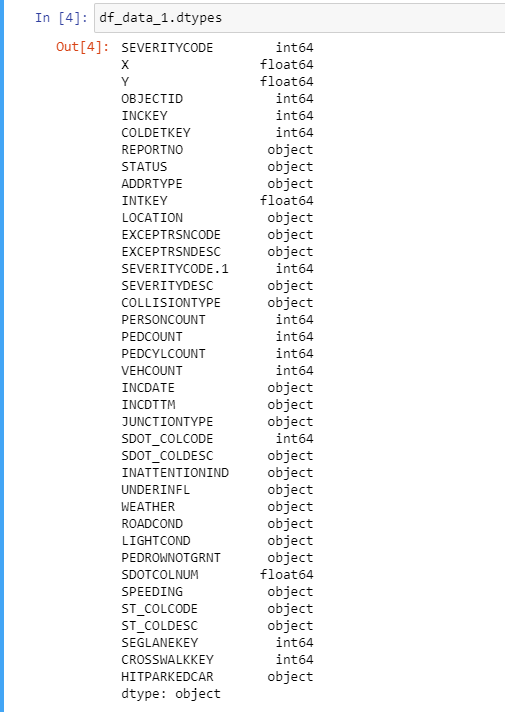
Road accidents are very common for most drivers. A lot of the times, these road accidents are not that severe or life threatening. Over the past few years, the number of road accidents have been increasing with a higher number of road accidents over the past few years. This is having a negative impact on the society since the cost of the injuries is increasing. About 1.35 million people die in road crashes worldwide annually. On top of the deaths, about 50 million people suffer serious but non-life-threatening injuries from road crashes each year. These numbers are very unfortunate and high. On top of the injuries, it costs the government 10.7 billion dollars in medical costs for accidents. This is a huge concern for the society and there needs to be a way to reduce this cost. I chose to explore the data in the attribute of the amount of accidents that occur in areas labeled as “LIGHTCOND”. The reason I chose to explore this data was because there can be an easier fix to the problem if the reason for the accidents was darkness.



**Methodology**

For the data analysis of the project, IBM Watson studio was used for the notebook and Github was used for the repository. The cvs data file was uploaded directly onto the notebook I used in Watson Studio. Along with the data that was analysed, there were four Python libraries used. Those libraries were Pandas, Numpy, Seaborn and Matplotlib.

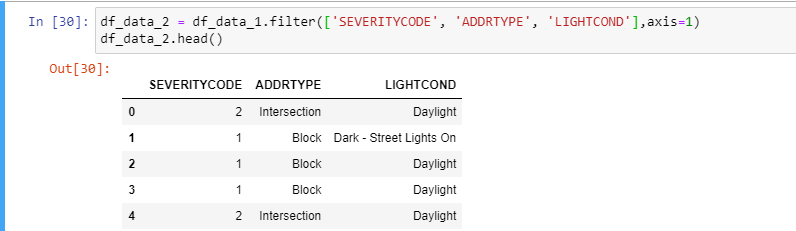
After the data from the csv was uploaded onto the Notebook, I used the function df.types to figure out the data types for the data in the spreadsheet. I was mainly focussing on whether the light conditions have a factor on the severity of the accidents. So, for this I was focussing on Severity Code, Light Condition and Address Type. I noticed that severity code was int64 while light condition and address type were both objects.



After I determined the data types, I ran a value count to light conditions and address type to see where the most accidents occur since alleys are usually dark and intersections are well lit.

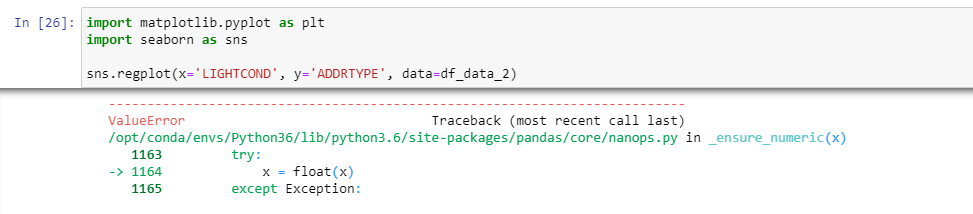


To make my analysis easier, I created a new data frame with only the data I was interested in.

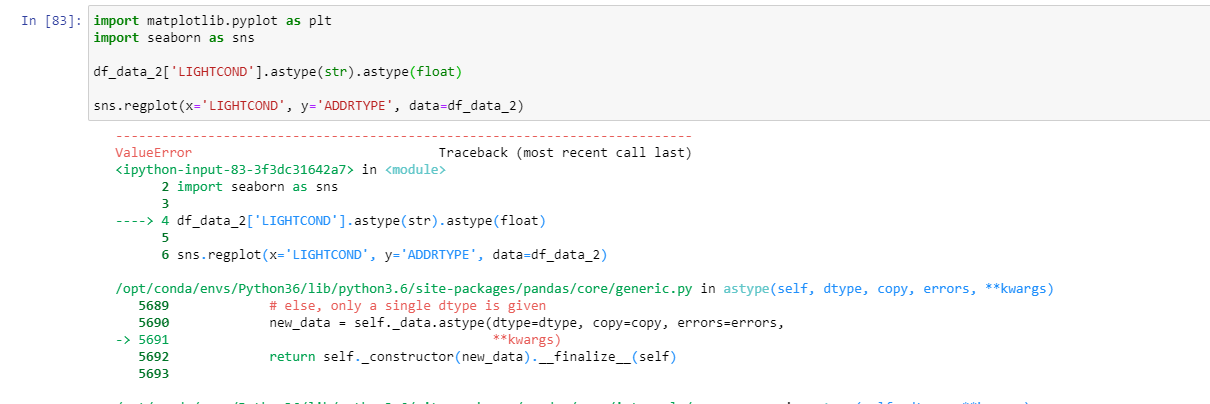


**Results**

At first I decided to use sns.regplot to determine the relationship between the light conditions and address type. When I ran this code, there was an error that occurred in the data type.



Due to this error, I decided to take a different approach. I tried switching “LightCond” to float. This also resulted in an error.



A reasoning behind this correlation analysis not working could be because the accidents based on the light conditions and address type have a relatively low correlation. Meaning that the light condition has no effect on where the accidents occur.

**Discussions**

Based on these results, a recommendation that can be taken is to use different variables that may provide a larger correlation like speeding and severity. This will determine if the severity of the accident is worse based on the speed that it occurs at.

**Conclusions**

The analysis did not turn out the way expected. The amount of light available in an area did not affect the number of accidents that occurred in that area. Since I chose to investigate the amount of light in the area, I had to stick with it through out the project. The data shows there is not a high correlation of lighting conditions and address type. If I had chosen a different variable, maybe I would have had a higher correlation.