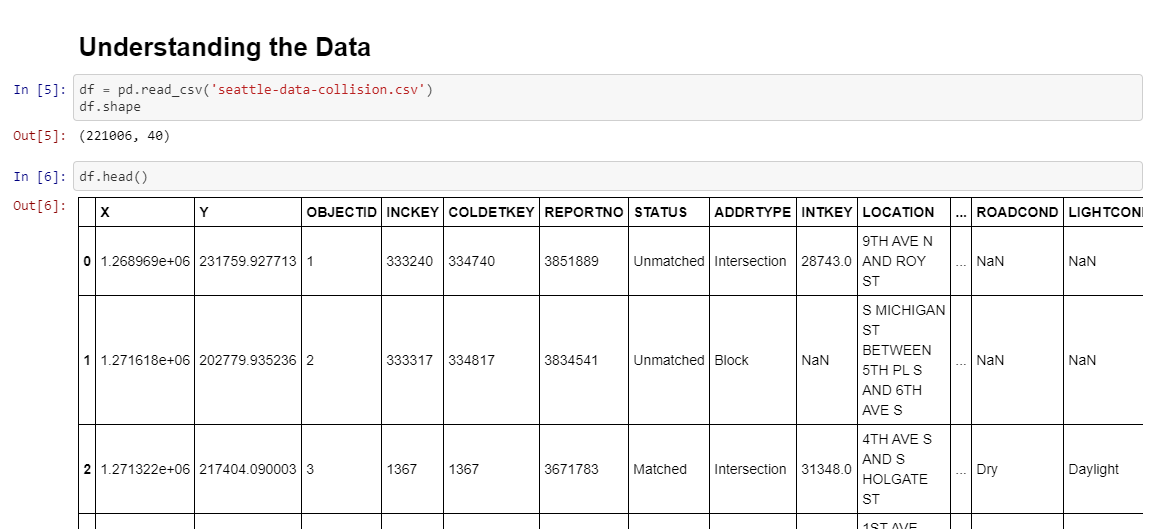
|  |  |
| --- | --- |
|  |  |
| Predicting Severity of accidents |  |
|  |  |
|  | 30th August, 2020Applied Datascience Capstone |
|  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Introduction  * US is one of the busiest countries in terms of traffic and with the increase in the number of vehicles on road, the number of accidents have increased. * In 2018, there were some 12 million vehicles involved in a crash in the United States. With the increase in number of vehicles on the road, the probability of vehicular accidents increases * First responders need to provide the information to the Emergency response team in order for them to provide the appropriate response based on the severity of the accident * If the emergency response team can predict the actual severity of the accident based on the ML algorithm they can dispatch the appropriate teams to the location and may also help in saving some of the lives | |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Seattle Collision Data | * For the implementation of the Machine Learning model, our focus will be on learning the accident data for one city and the model can later on be trained on the data from across the country once we have the data from different cities, counties available in the standardized format. It is very important to standardize the data and clean it to get rid of any anomalous data or data points that can cause our model to break * For our ML model, we are taking dataset made available by the city of Seattle in Washington state. We are grateful to the city of Seattle for making this information public and help the budding data scientist to build a model that can help the people * Accident-Collision data can be downloaded from the following site: <https://data.seattle.gov/Land-Base/Collisions/9kas-rb8d> |  |
|  | Understanding the Data  * Dataset contains 221006 rows and 40 columns * A quick look at the data shows that certain columns are blank and contain null values * Based on the description in the metadata, columns also have a inconsistent values like ADDRTPYE which should only have 3 values – Alley, Block, Intersection has null values flowing in the data * Based on the understanding of the metadata and looking at the data, we had to clean the data and remove columns which were not providing any additional insights and would be helpful in predicting the severity of the collision * Further, details regarding the data and cleansing are present in the notebook: <https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/c3341200-3040-4150-aee1-5dd077bbfbcb/view?access_token=7f89a3a889b486e8220bae7f1e70e5d8072065d3d2e8355c31ed869190451e91> | |  |



# Collisions Data Metadata

