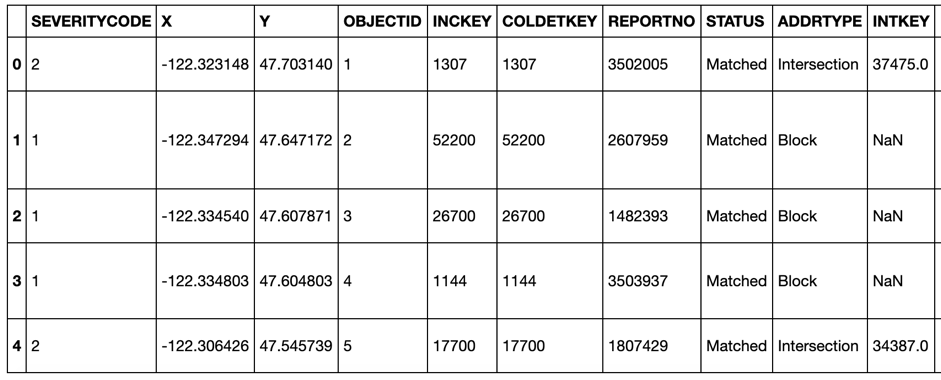
**Predicting Traffic Accident Severity in Seattle**

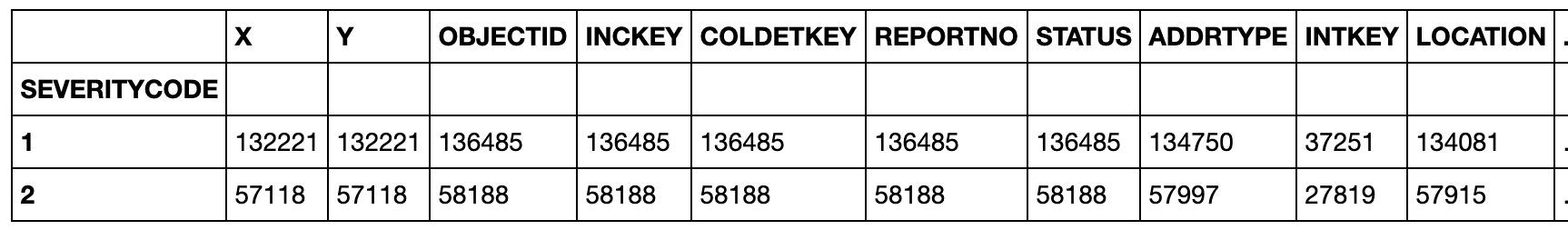
1. **Introduction**  
     
   Our client, Business X, has retained our services to help them predict traffic accident severity in Seattle, where they currently operate. Business X recently began revamping their employee safety program and discovered that their employees were involved in an alarming amount of severe traffic accidents. Business X would like to mitigate this problem and has identified three potential countermeasures:
   1. Allow employees to work from home on rainy days to avoid hazardous road conditions.
   2. Deploy an application onto their employees’ cell phones to prevent distracted driving.
   3. Shift their operating hours by two hours from 9am-5pm to 7am-3pm so that their employees can commute before the morning and evening rush hours.

Business X will make their decision on which countermeasure to choose based on which of the three has the highest probability to reduce the severity of traffic accidents. Thus, we will build a machine learning model to predict the severity of traffic accidents in and around Seattle and use the results to advise Business X as to which countermeasure is likely to be the most successful at reducing the frequency of severe traffic accidents for their employees.

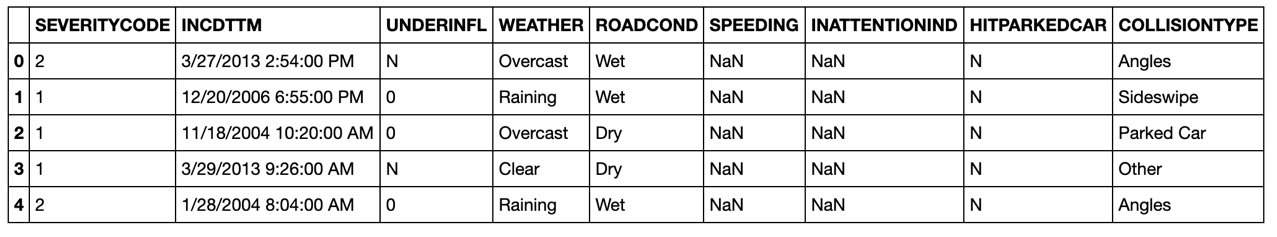
1. **Data**  
     
   The data we will use to build and train our machine learning model comes from the Seattle Department of Transportation and contains information related to traffic collisions and accidents in and around the Seattle area. Critically, this dataset includes a severity classifier (SEVERITYCODE) for each traffic accident (1 for minor vs. 2 for severe) as well as attributes related to Business X’s three potential countermeasures, namely WEATHER for weather conditions at the time of the accident, the time the accident occurred (INCDTTM), and whether or not the accident was due to an inattentive driver (INATTENTIONIND). In all, the dataset contains 37 attributes, some of which will be used to refine the dataset while others will not be relevant or necessary for our analysis.



The dataset initially includes information on 194673 accidents, 132221 of which are labeled as minor (1) and 57118 as severe (2). The dataset will therefore need to be balanced in preparation for building the machine learning model to avoid bias.



The relevant attributes for our analysis and machine learning model include those that allow us to eliminate confounding variables on accident severity, such as those caused while driving under the influence (UNDERINFL) or speeding (SPEEDING), in addition to those directly related to Business X’s problem, as described above.



1. **Methodology**
2. **Results**
3. **Discussion**
4. **Conclusion**