IBM DATA SCIENCE CAPSTONE REPORT

CAR ACCIDENT SEVERITY in SEATTLE

INTRODUCTION

Car accidents in Seattle happen at all times, but if the main causes of accidents are determined, advance warning or mitigating methods can be performed. For example, certain intersections may be more susceptible to accidents due to heavy usage or the way they are constructed. As a result, better street lights can be added (only protected left and right turns) or traffic personnel can be used to direct the cars. If it is determined that accidents occur the majority of a time a driver is speeding, has a high blood alcohol level, or was not paying attention, the data can be used as evidence for enacting harsher laws and regulations. In addition, the data can be advertised to the public to show them the consequences of driving under these conditions. There are also uncontrollable factors such as weather and road conditions. If certain patterns are discovered to cause many accidents, local government can know when to send alerts to the public to drive more cautiously or even avoid the roads entirely.

TARGET AUDIENCE

1. The Seattle administration,
   By targeting areas prone to areas to speeding accidents, interventions such as speed bumps, street light, traffic personnel etc. can be put in place to reduce accidents.
2. Rescue groups and emergency services in Seattle,
   By having enough data on the crash one can predict the severity and therefore take action more quickly potentially saving lives.
3. Car Insurance Companies,
   Areas where parked cars are prone to getting damaged. Owners in those localities may be asked to pay more premium on their car insurance.

DATA

The data comes from collision and accident reports in Seattle during the years 2004-present. It was collected by the Seattle Police Department and Traffic Records department. The data has 38 independent variables and 194,673 records and will be used to identify the key variables that cause accidents. We will use SEVERITYCODE as our dependent variable Y, and try different combinations of independent variables X to get the result. Since the observations are quite large, we may need to filter out the

missing value and delete the unrelated columns first. Then we can select the factor which may have more impact on the accidents, such as address type, weather, road condition, and light condition.