

1. Business Understanding

1.1. Problem

Car accidents are a worry for the whole world. As city population increases day by day, so does the number of cars in cities and so does the pollution. That is why the last years one of the politics campaigning spots had been new measures to reduce CO₂ emissions. Cities like Seattle are having success in promoting greener solutions of transport (promoting bike use, investing on public transport infrastructure and promoting pedestrian areas, among others), but now they need to assure that are safe. The aim of these project is to study the possible causes related with pedestrian-involved and bike-involved accidents, identifying the measures that need to be taken to reduce collisions amount and its severity and, finally, to develop a predictive model to evaluate severity.

1.2. Background

The study will be focused on Seattle traffic data. According Seattle Department of Transportation, while city population is growing by a 1.8% as yearly rate from 2011, average daily traffic remains stable. Which means a new trend in transportation usage, clearly reducing the use of cars, while the rates of transit ridership and going-by-walk have not stopped of growing (Transit ridership presents a 14.3% yearly growth average, followed by Pedestrian with an 8.3% yearly growth average from 2011). However, bike usage has not register remarkable growth during past years, neither facilities have been yet deployed for them. This drive us to an increase of pedestrian collisions by 9.5% by year. Also bike-involved collision has raised from 2011 to 2017 by 1.5% yearly growth average, even if not related with any traffic increase. This insights make us aware that for having success implementing future measures to reduce pollution in cities, first we need to think in new parallel measures for increase safety for these new users.

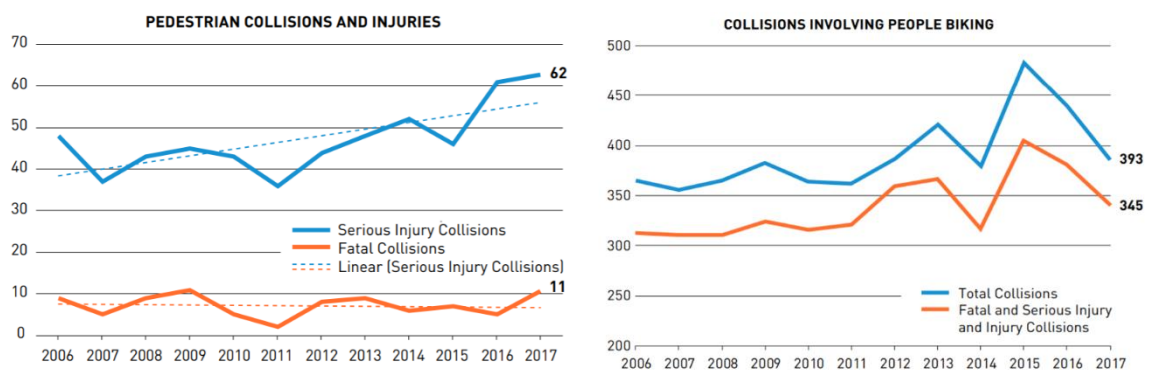


Figure 1 and 2: Pedestrian involved collisions & Collision involving people biking from Traffic Report 2018 published by Seattle Department Of Traffic

1.3. Stakeholders

Reduction in severity and amount of collisions is interesting for Public Development Authority as for the Department of Transport and Departments of Environment, to foster their policies.

2. Data Understanding

2.1. Data Description

The dataset provided by the course has a total of 38 columns x 194,673 rows. There is a column called Severity Code, with specific values 1 or 2, meaning property damage-only collision or injury collision. By developing a machine learning model we will try to predict these value, based on other features given as location, day, weather, road condition or light condition.

As the dataset show duplicated columns, blank cells and irrelevant data, first step is data cleaning.