COMP20008 : Elements of Data Processing

Assignment : Phase II

Student Name : Hinam Mohit Mehra

Student ID : 653167

1. The Project's title is 'Is Melbourne's Public Transport really reliable ?' Melbourne's Metropolitan Trams Punctuality and Performance is analysed under adverse weather conditions like rainfall and high temperatures. In Phase I, I talked about analysing the performance of both Trams and Trains but since Trains are mostly underground, they are not directly affected by the congestion caused by rain and heat. Hence, now only my focus will only be on Metropolitan Trams.

2. For the investigation, Operational Performance data of Trams for 2015 was requested from PTV since it wasn't readily available from their website. PTV has been using Automatic Vehicle Location(AVL) since 1985 to track the network's schedule adherence. I decided to analyse performance over the entire network of trams on a particular day instead of analysing every tram line separately for this initial investigation. Precipitation and daily maximum temperature data was obtained from Bureau of Meteorology's website using Melbourne's Botanical Gardens as the weather station for precipitation and Melbourne's Olympical park for temperature. The data from all three CSV files was combined into one, with each row depicting the day of the year, rainfall, maximum temperature that day, punctuality percentage of trams, and percentage of the tram timetable delivered. There were only a few missing values (12 days out of 365 days) in the precipitation and temperature data. The same few days were missing from all the other weather stations as well, hence those days were marked as invalid. Apart of that, the rest of the data was complete and didn't need any sanitisation.
The next part of the investigation process was to determine if the data will provide any interesting outcomes. Thus, first a threshold for poor performance was needed to be determined. PTV's yearly goal was to provide 77% of the services on time and deliver 98% of the timetable. PTV defines 'on time' as being not earlier than 1 minute and not later than 5 minutes. Thus, the data was filtered and divided into two sets. The first set being days where less than 77% of the tram services were punctual, the amount of rainfall and maximum temperature that day. And the second set were days where less than 98% of the tram timetable was delivered, with their corresponding precipitation and temperature data. Scatter plots were then used to analyse a relation between weather and poor tram performance. The scatter plots were color coded with each month given a different color. They showed that month's prone to high amounts of rainfall or high temperature showed below than threshold performance. Infact the visualizations showed that on the day when a high 20.8 mm of rainfall was recorded, only 71.70 % of the trams were punctual, and the lowest delivery rate of 94% in the entire year was when the temperature was 41.2°C, the highest temperature recorded that year. This proves that my data wrangling for the initial investigation is on the right track.

3. Initial investigations hinted that there was a direct correlation between poor weather conditions and poor performance of the PTV tram network. Further investigations aim to solidify these investigations by eliminating public holidays, and days when the PTV tram network was disrupted because of construction or accidents. The disruption data will be

obtained from the PTV Timetable API. Analysing the performance of different tram lines separately, and judging them against their importance will provide PTV with interesting results to further optimise their schedules to provide the public with a more reliable network. Further investigations also aims to see the effects of lagged weather effects, like raining heavily on the day before on the performance of public transport. This will prove the reliability or unreliability of the trams further. These investigations are feasible since all the data has already been obtained and sanitised, and since we are only investigating last year's data we are factoring in the improvements that PTV has made over the last couple of years, so we will be providing them will suggestions that haven't already been suggested, and our data processing load is also managed.

4. As an exponentially growing population with an increasingly push towards sustainability and decongesting the cities, it has become very important to understand the reliability of public transport systems and its key drivers. Melbourne is expected to be the biggest Australian city by 2050, and in spite of being the World's largest and oldest Tram network, Melbourne's Tram experience relatively poor reliability. This projects aims to determine the effect of weather on Melbourne's Tram network. We have no hold on the weather but via this investigation we can map down if weather conditions have adverse effects on the punctuality and delivery of the public transport system. Customer satisfaction towards PTV tram network decreased 5.7% from 2014 to 2015, hence it is important to determine the factors that affect performance so that we can further optimise and robust schedules to increase their reliability.

5. The datasets being used for this experiment are :
Operational Performance Reports 2015(Example)
https://goo.gl/cvxAfy
The complete dataset was obtained by emailing PTV.

Melbourne's Rainfall 2015
http://goo.gl/96GKRn

Melbourne's Max Temperature 2015
http://goo.gl/96GKRn

PTV Disruptions API:
https://goo.gl/UW2MI2

In Phase I, I decided to use either the Operational Performance Report or the PTV Timetable API, but now i am using both to eliminate poor performance reports caused by factors that are not related to weather as much as possible.