Project Report

This shiny app project start from the EDA of Bus and Train data from MBTA. As I mentioned in my EDA report, the aim is to find ways to improve the accuracy of Google Maps estimated time cost of direction. To reach the goal, I first do the EDA part based on the data download from MBTA official website that includes 2021 September to 2022 October. During the EDA process, I found that the bus and train can have an extreme delay at some date. To figure out the cause of this problem, I first tried to find the weather forecast information during those days and found that it's true that extreme weather can be the influence factor of the late, but in the five examples I selected, only one is caused by the sudden sleet. On the rest dates, the weather looks normal, at least no large differences with the rest date during that week. So the next step, I try to search if there is any news about what happened to those routes on that date but there is no relative result. So I made a hypothesis that this can be caused by some accidental car accidents, or traffic jams, which won't be reported as news so I can not find them in the search engine. Also, I mentioned that based on my personal experiences, a large passenger flow amount can cause a delay in the bus. The drivers need to wait more time at each stop to let passengers get on and off.

When it comes to the shiny app part, I found that I can't find the data of those influence factors, so I can not reflect those factors on the map, or calculate how they would affect the cost of time the users will spend. After discussing it with my friend, I think it would be better if I choose the departure time and the seasons as the effects. So in the shiny app, you can see that I choose the data that includes the departure times and seasons based on different dates. I didn't find the same data set for the bus and ferry, so I can't combine the data sets and do it as showing the users the differences between different transportation ways on the same map. Instead, I choose to ask the users to select the train route they want to take, the departure time, and the season they want to take the trains. After their selection, I will show the stops that have the train on the map. For the second and third pages, I show them the table of train and bus information based on their selection, so if they want to have more specific information, they can check the tables.

So in conclusion, I think there are multiple reasons I met such problems. First, it's hard to find the relation between the influence factors I chose and harder to find ways to quantify those effects. Secondly, my code skills can not support me to represent my idea well. To make improvements, I think first and the most important thing is to keep practicing. If I can find a good data set and quantified correlations, it will be easier to deal with this assignment.