**Research work**

**Project Overview:**

This project is a cooperative project between our company and Shenzhen University, which is based on the data of public transportation trip to analyze the basic situation of Shenzhen OD traffic flow. I have some results attached.

**Data Sources:**

Data provided by the Shenzhen rail transportation company. The main types of data include subway and bus swipe record, subway and bus GPS data.

**Project Ideas:**

1. Combined the bus swipe record and subway stations data into a file;
2. Extract the data of the first station per passenger per day as the site of their possible resident address;
3. Only swipe record with time intervals larger than 3hrs are counted, and the place of the second swipe is regarded as the work site during weekday; swipe record with time intervals larger than 6hrs are counted, and the place of the second swipe is regarded as business site during weekend;
4. Using a month's data to cluster the working site, resident site and business site, combine all sites within 1km into one site, the resident site with the highest frequency is regarded as resident site, the business site with the highest frequency is regarded as the business site and the business site with the highest frequency is regarded as business site,
5. Though a Collaborative Filtering model to look back in history record data and find the most similar scenario to the present one by measuring the cosine similarity between the explaining vector.

**Achievements:**

1. Integrated subway and bus swipe record and deduced origin-destination time and geographical position.
2. Calculate transit flow of subway and metro stations，and categorize employment and residential areas through the data between morning peak 7:00-10:00 and evening peak 18:00-20:00 in the weekday and categorize business and residential areas through the weekend data.
3. Through the data from each site, developed a real-time dynamic station heat map of Shenzhen City, and established a highly accurate collaborative filtering model to predict traffic flows.

In my zip file，under the web folder, there are two samples html, including station-heat. html and subway-od. html. All of these are made by Java and E-charts software. Data are processed through Java and diagrams are painted through E-charts. In the subway-od. html, the line between point and point represents the morning peak of OD flow in one day. When you click on the line, the traffic flow will show. In the station-heat. html, red represents the place of residence and green represents the place of employment. These are team work, and I am mainly responsible for coordinating the manager to establish the project ideas and visualizing the latter part of the expression.