In the literature review assignment, as a group we reviewed 8 papers in total and I mainly focused on two papers which are **Uber Related Data Analysis using Machine Learning** and **Identifying the Hotspots in Urban Areas Using Taxi GPS Trajectories**.

Below are my preparation outline for Goal, Challenges and Result after reading both of them, I hope this can prove the work of mine during this assignment, the important approach part is mainly included in the group assignment:

**Uber Related Data Analysis using Machine Learning**:

Goal:

The goal of this paper, Uber Related Data Analysis using Machine Learning is to predict pickup of cab from a coordinated cluster of points so that we can bridge the supply demand gap of cab services, deduce ETA and optimize the route selection. The test example used in this paper is New York City but we can try this method in any city. It aims to construct a system that dispatches cabs to clusters(areas) to wait thus reducing the wait time of both ends which are drivers and passengers.

Challenges:

The main challenges mentioned in this paper are predicting cab pickups efficiently, handling large dataset and reducing forecasting errors and overfitting when dealing with extensive trip data. So a data visualization before we actually use the approach and build the pipeline can be helpful to understand the data and try our best to reduce the noise and make the algorithm more efficient.

Result:

The paper used data visualization with dot graphs and heatmaps showing the frequency of trips during different times and locations of predicted results of New York City and it showed that the use of unsupervised learning via k-means clustering is efficient since it showed common hotspots in New York City.

**Identifying the Hotspots in Urban Areas Using Taxi GPS Trajectories**

Goal:

The main goal of this paper is to identify hotspots areas in Lanzhou city using taxi data. Urban hotspots are areas with high levels of human activities so pick-up cab service demand is relatively high. So arranging more cab service in those areas will increase the working efficiency.

Challenges:

This project faces multiple challenges, including the need to process extensive taxi trajectory data of longitude, latitude and times of pick-up and drop-off. Issues of data cleaning, outlier detection and matching GPS data to road network topology are also things that need to be considered.

Result:

The paper presented the results by using a heatmap that shows the the level of business of each grid that showed in the city map.