

BUS240F Information Visualization

Uber vs Lyft rideshare data

Oct.11th, 2022

Pavan Kumar Swamy Padamati,
Yashodhan Damle & Ziyu Fan





Contents

- ✓ Introduction
- ✓ Data Overview
- ✓ Data Preparation
- ✓ Figures
- ✓ Pros/Cons
- ✓ Q/A



Introduction



Source

The dataset is from Kaggle (link: <https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma>). We failed to reach the author of the dataset for the original data, but we will keep on trying

Information

This dataset includes detailed observations of uber and Lyft rides from November to December 2018 in the central Boston area, including the length of rides, prices, weather conditions, etc.

Potential Users

Uber & Lyft company; Other rideshare companies; Boston transportation departments; Frequent user of Uber/Lyft

Data Overview

Data size

There are 57 columns, 693072 observations in total. The data set is in csv format



Key variables

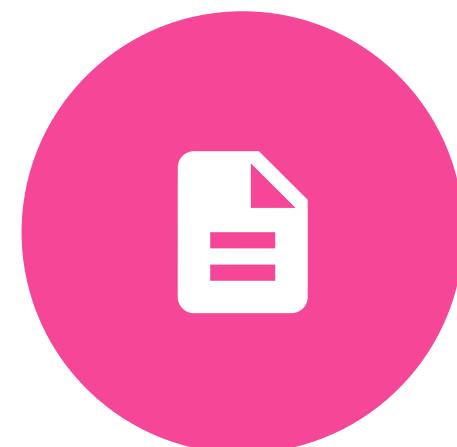
Hour, day, month; price, distance, destination; cab_type, product_id; Temperature, short_summary

Data Structure

chr (11): id, datetime, timezone, source, destination, cab_type, product_id, name, short_summary, long_summary. etc.,
dbl (46): timestamp, hour, day, month, price, distance, surge_multiplier, latitude, longitude, temperature, apparent temperature, etc.,..



Data Preparation



Original Data

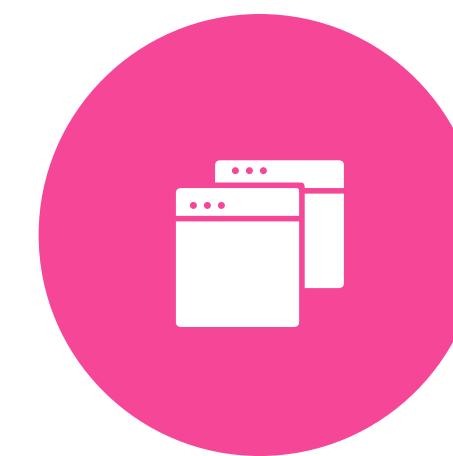
This is a large dataset, so we need to clean up and select observations and variables



Data Cleaning

Removed rows with NA columns while making a correlation matrix.

Also we have removed unwanted columns which we are not using for analysis, including the detailed UV_index, visibility, windspeed, etc.



New dataset

how does it look like now?
Rows: 693071 Columns: 20
And considered only data with no null,NA while making correlation

Figures

- ✓ Market Share of Uber Vs Lyft [Nov'18-Dec'18]
- ✓ Average Ride Fare vs Cab Type
- ✓ Revenue Per Cab Type
- ✓ Market Share of different cab lines
- ✓ Average Ride Price vs Hour of the Day
- ✓ Price per mile vs Climate
- ✓ Climate vs Number of cab rides



Uber vs Lyft graphs that represent key findings

Figure1:

Market Share of Uber Vs Lyft

[Nov'18- Dec'18]

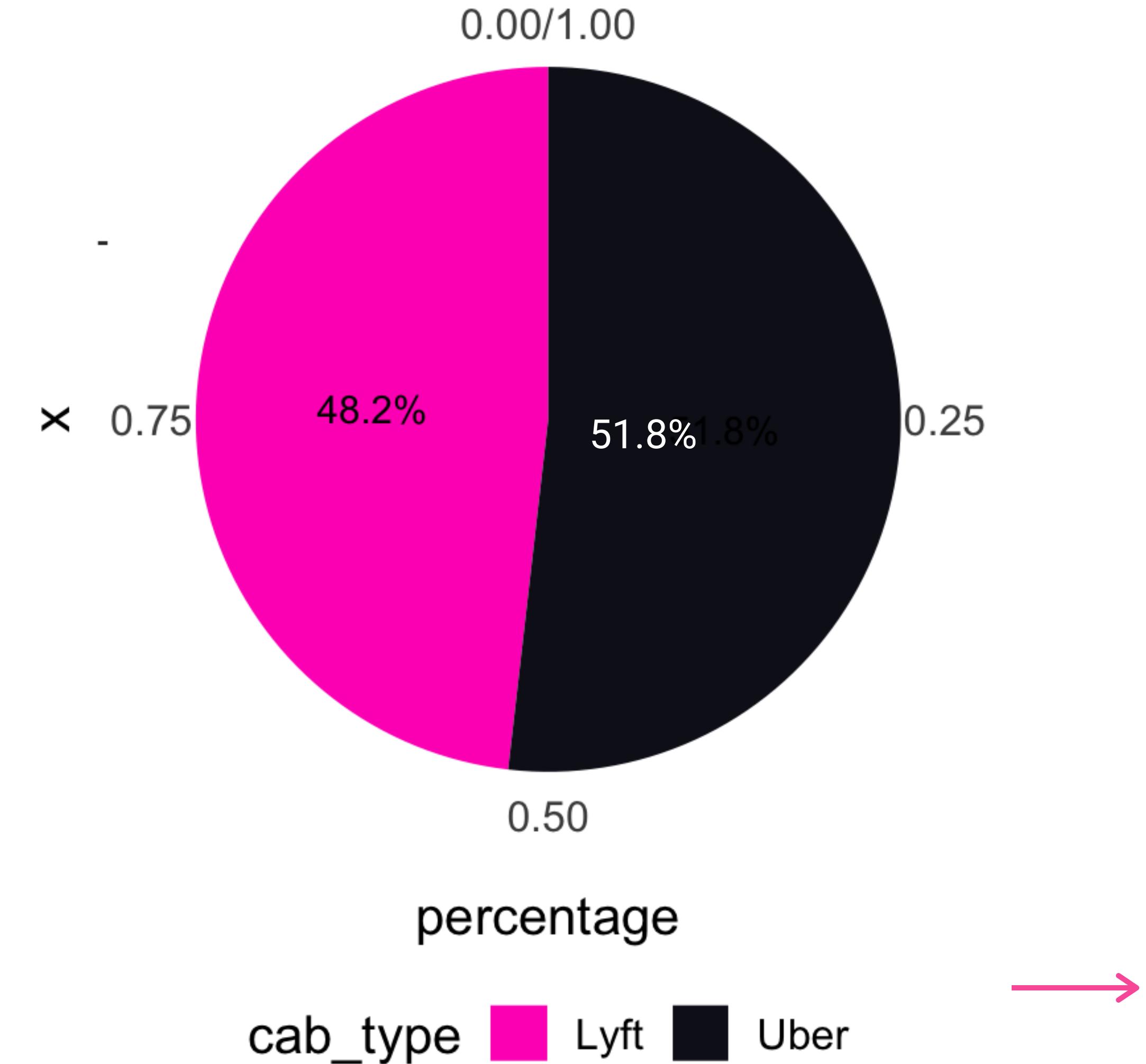


Figure2: Average Ride Fare vs Cab Type

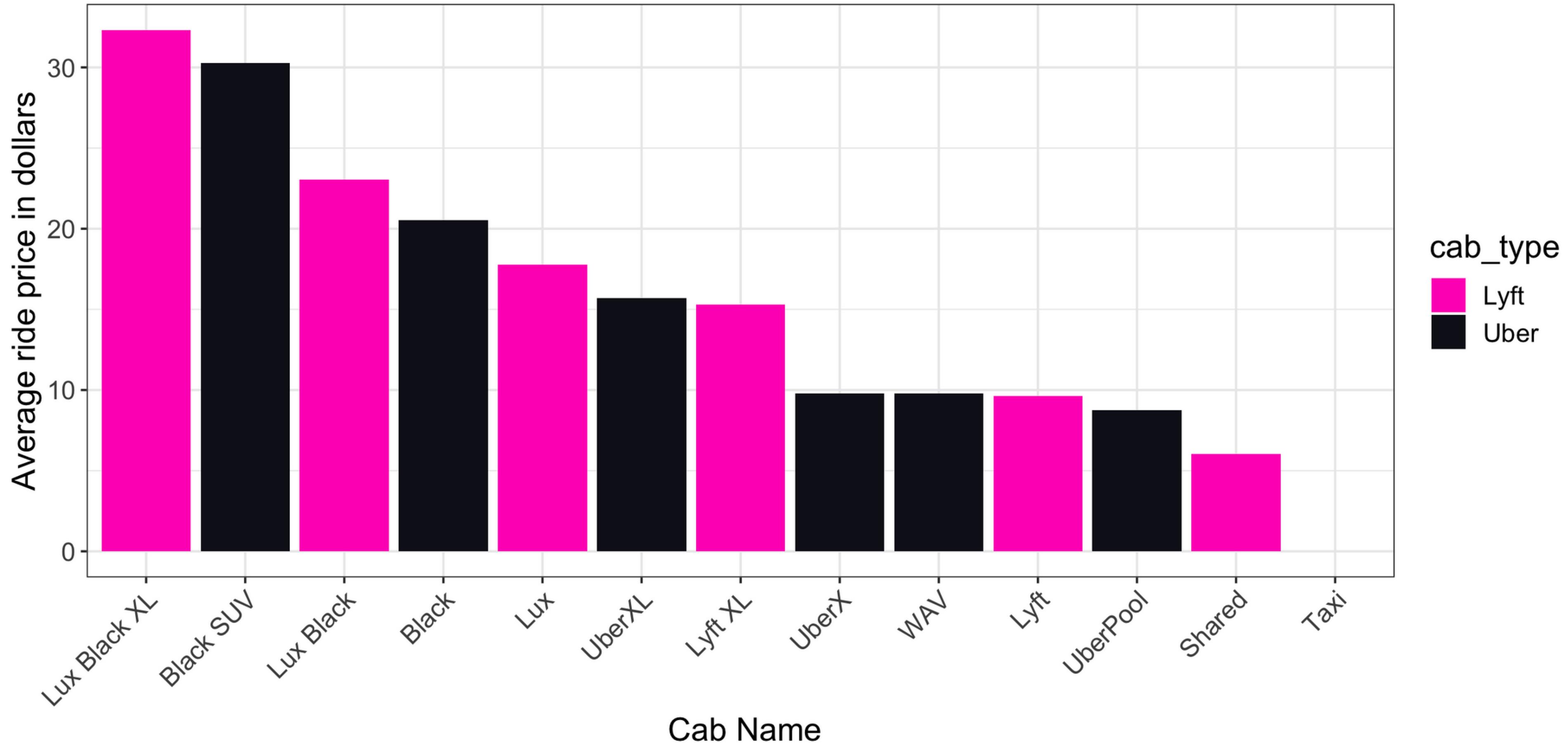


Figure3: Revenue Per Cab Type

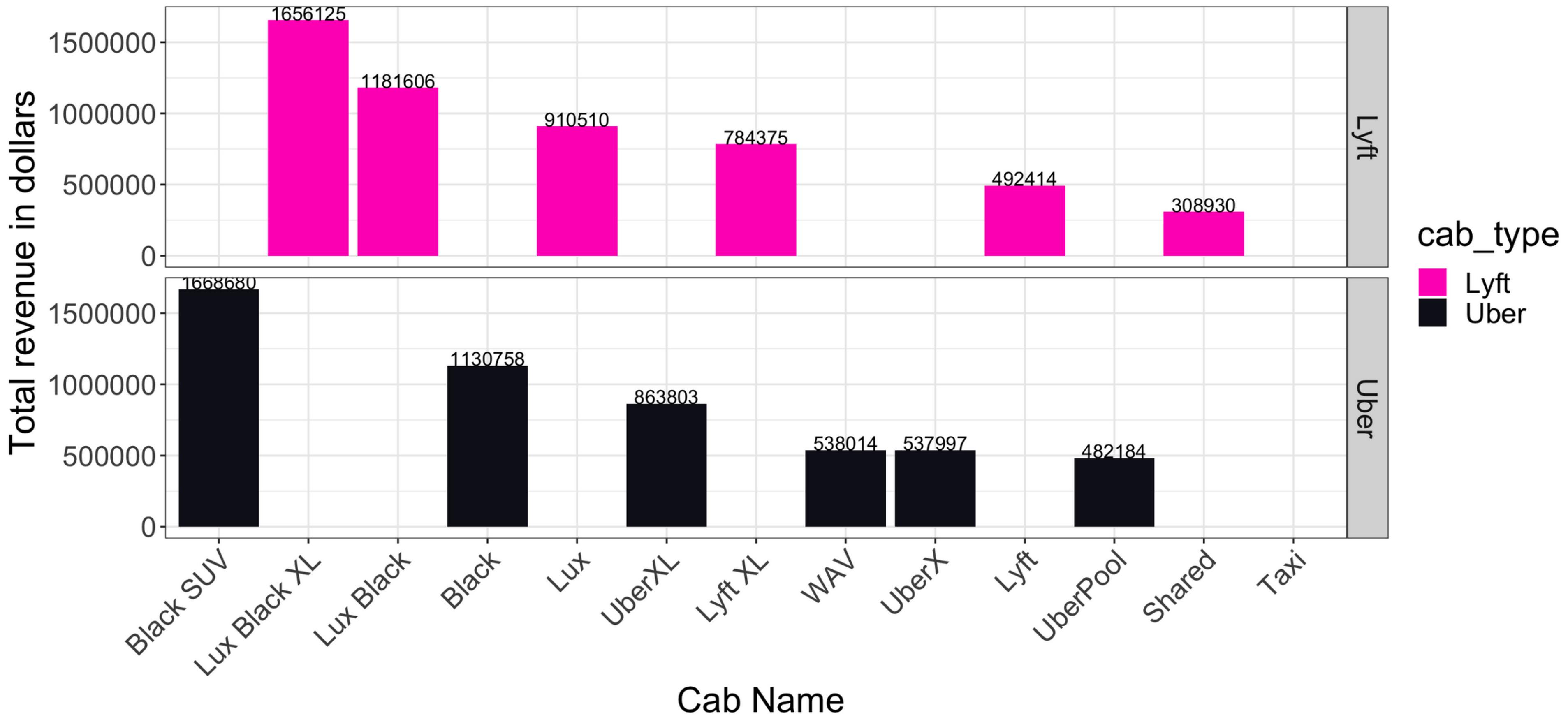


Figure4: Market Share of different cab lines

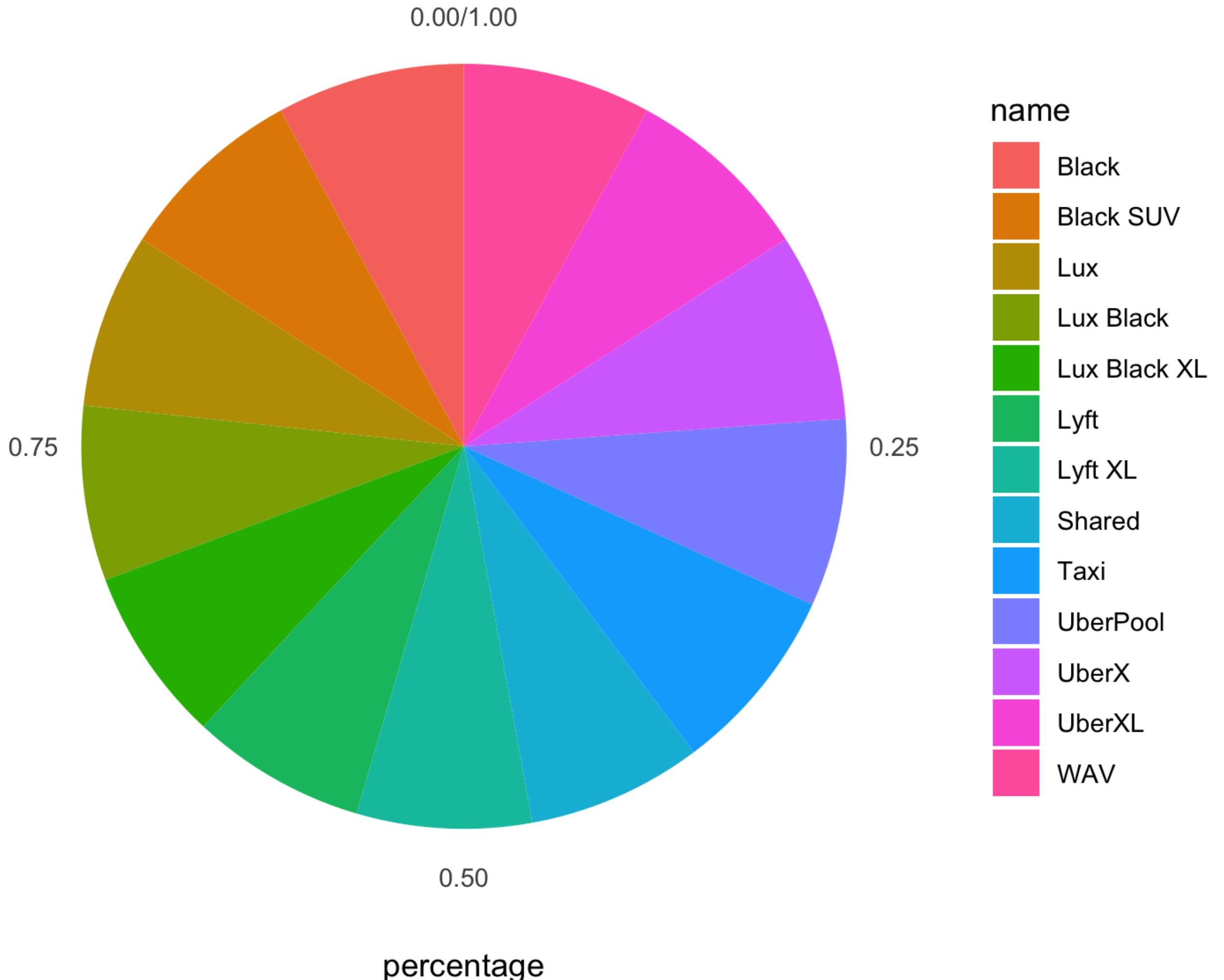


Figure5: Average Ride Price vs Hour of the Day

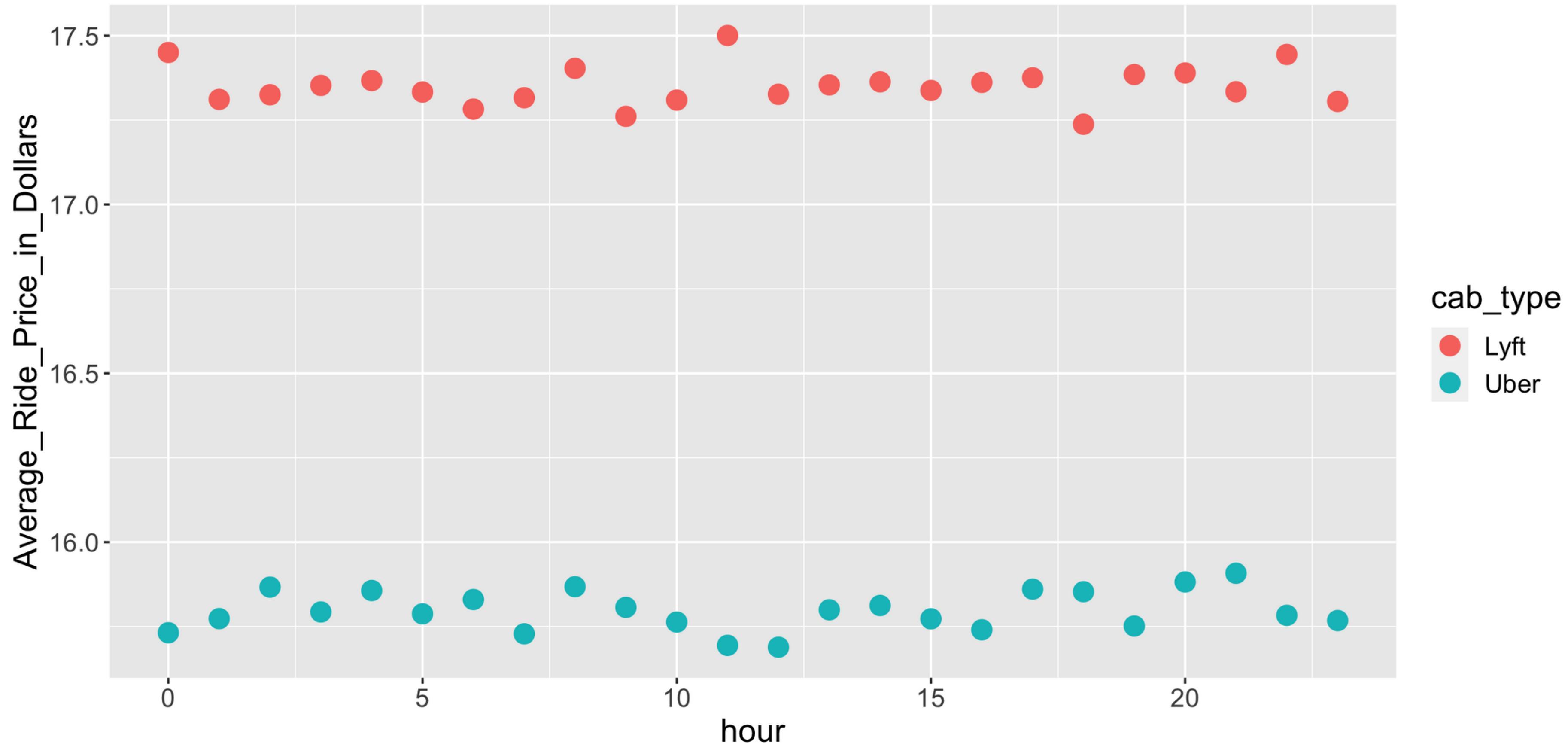


Figure6: Price per mile vs Climate

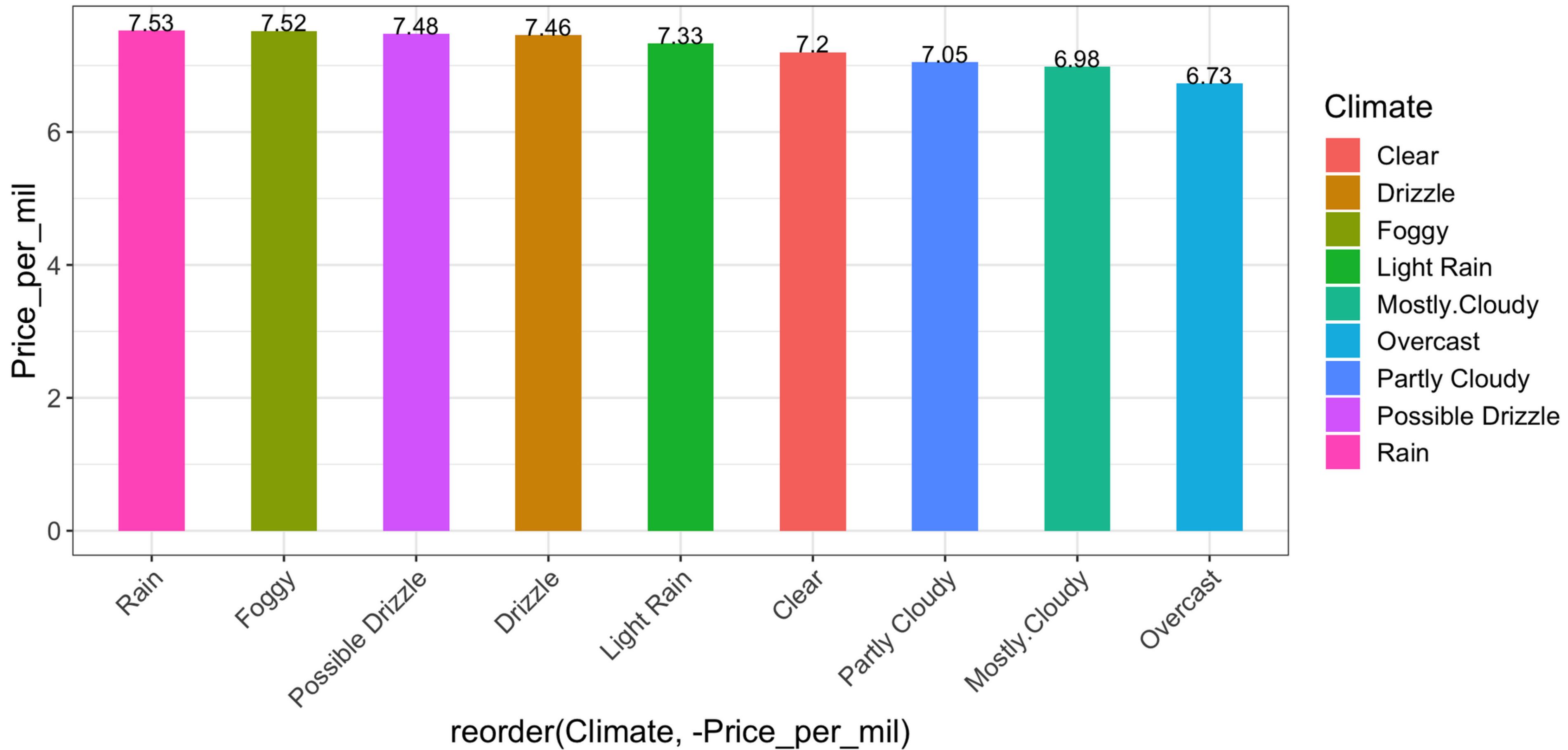
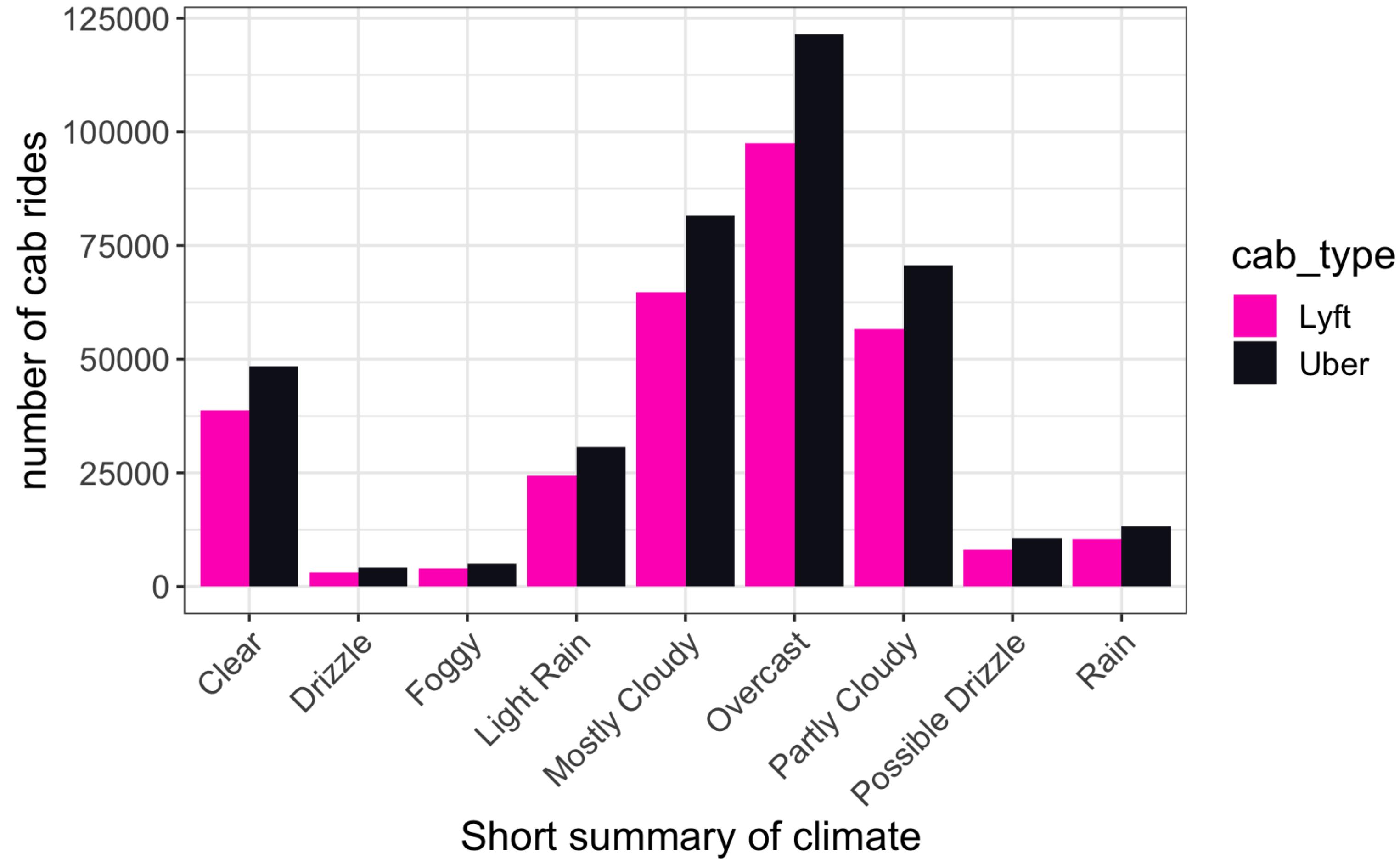


Figure7: Climate vs Number of cab rides



Pros & Cons

Pros

- The dataset is large and inclusive with detailed information on price variations as per climatic conditions, which gives us some interesting insights for further research into this area.
- The dataset could have many potential users, like Uber & Lyft companies, new taxi companies, rideshare app users, for Marketing and Advertising purposes.

Cons

- This dataset only includes the data from November and December of 2018, which is a short period to base our conclusions off.
- This dataset only includes the central area of Boston

References

Dataset link

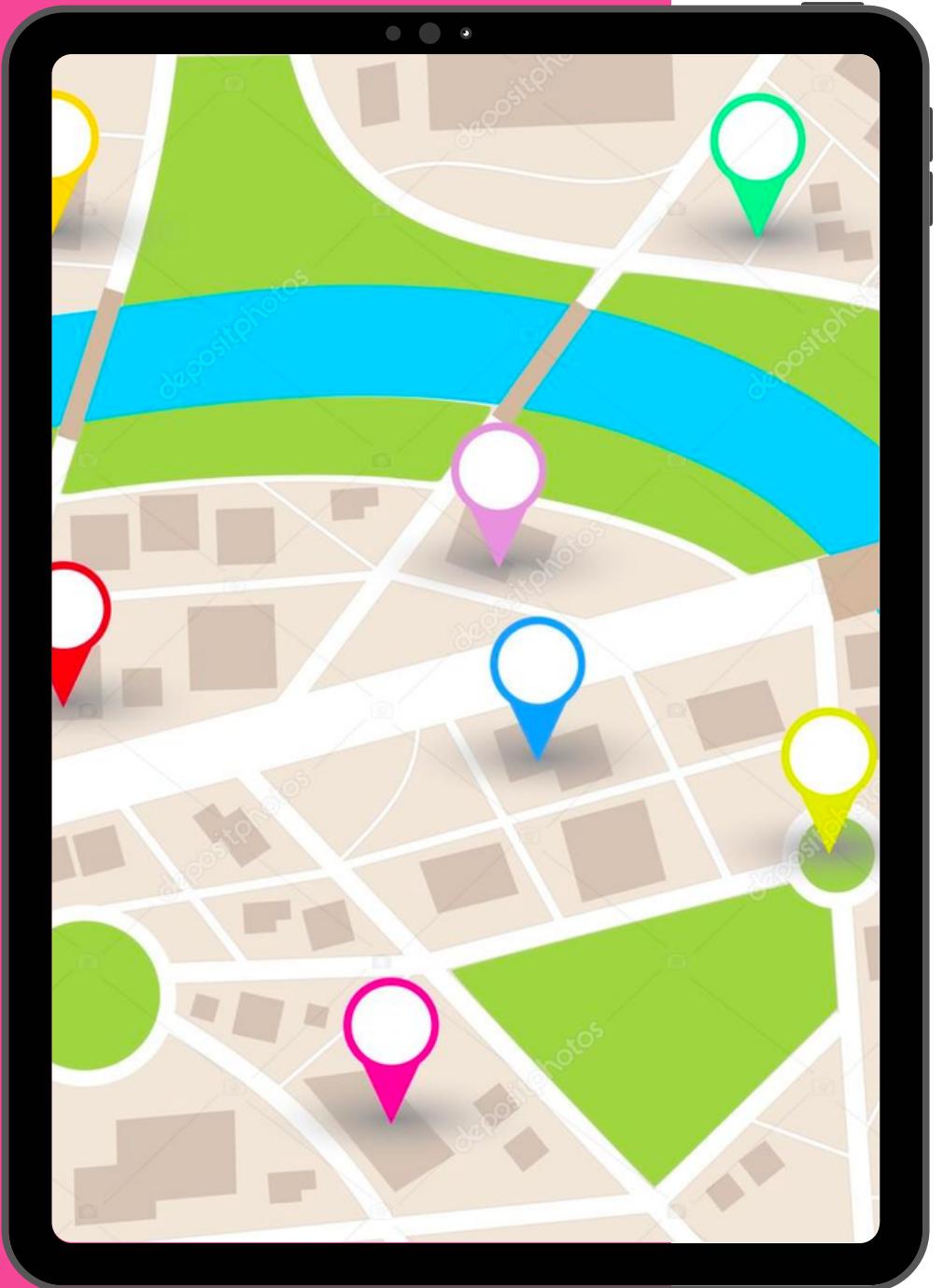
<https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma>

Image Uber vs Lyft

<https://www.ridester.com/uber-vs-lyft/>

Image Boston Map

<https://stock.adobe.com/search?k=boston+map>



Questions?

Thank you for listening!