Final Project Report

Part 1 - Story and Narrative

Link to the dataset	DATASET 1: https://www.kaggle.com/datasets/yasserh/uber-fares-dataset DATASET 2: https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices?resource=download DATASET 3: https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices?resource=download
Example item from the dataset	**DATASET 1** **Key:** 24238194 **Date:** 2015-05-07 19:52:06.0000003 **Fare_amount:** 7.5 **Pickup_longitude:** 73.99981689453125,40 **Pickup_latitude:** 40.73835372924805 **Dropoff_longitue:** -73.99951171875,40 **Dropoff_latitude:** 40.72321701049805 **Passenger_count:** 1 **Pickup_cluster:** **Dropoff_cluster** **DATASET 2** **Distance:**0.44 **Cab_type:**Lyft **Time_stamp:**1544952607890 **Destination:** North Station **Source:** Haymarket Square **Price:** 5 **Surge_multiplier:** 1 **Id:** 424553bb-7174-41ea-aeb4-fe06d4f4b9d7 **Product_id:**lyft_line **name:**Shared **DATASET 3**

	Temp: 42.42 **Location:**Back Bay **Clouds:**1 **Pressure:**1012.14 **Rain:** 0.1228 **Time_stamp:** 1545003901 **Humidity:** 0.77 **Wind:** 11.25
Story you want to deliver	(a story should be in a form of a list of facts, insights, and messages - refer to the lecture slide) Lyft and Uber are two of the most common rideshares. Ridesharing has developed a whole new sector of the transportation industry. It has created a new convenient way to earn some extra cash as a driver, or a new convenient form of last minute transportation for riders. However, the nature of the business can be very unpredictable as it is based on supply and demand of the drivers and the riders. A few questions I considered are: What does the information from the rideshare tell about the locations? Is there a correlation between rides and the day or time? Is there a tendency to different rideshare companies depending on the length of the trip? These are all questions that could be helpful to someone who is a rider or a driver, or just general insight for people who work at the companies. The message I want to deliver are answers to these questions and it will be through visual representations of the data. This can help a driver decide when and where a good place to be when the demand is high, or for a rider to plan ahead for their trips.
Describe your target audience.	(using the questions the lecture slide listed) • **Familiarity with your topic? If not, how do we catch them up?**

	Uber and Lyft are both widely known companies and the idea of taxis and ridesharing are pretty common as well. As a result, not much information is needed to catch people up on the topic. **• Do they care? Why? Why not?** They care **• What do you want them to take away? Key points?** A key point that should be taken away is that the data collected from these rideshares can be used to derive other conclusions about the usage of the app, the locations surrounded and various events that are occurring. This information can be used in planning future rides, trips, and even traffic. **• What do they know about visualization? Are your techniques standard?** For regular customers or employees of these companies, it is most likely their knowledge of visualization is very introductory and not in depth, as a result, the techniques used in the visualizations should be clear, concise and easy to understand. **• How do they encounter your visualization?** This visualization could be accessible for personal use, at the moment, in the future, this could be something implemented by Uber or Lyft themselves. **• Mathematical background? Are you assuming too much? Too little?** I am assuming the audience will have a basic understanding of math, but nothing too advanced or elaborate will be used. **• Device? Mobile phone, computer, print media…** Since Uber and Lyft are accessed as mobile apps, this visualization would be most useful digitally to be accessed on mobile devices like phones, tablets, and laptops.

The goal of your project outcome. And why?	(exploratory vs. explanatory) The goal of the project is exploratory to discover any patterns and trends in people's usage of the services. The product can be used by the company itself to find pitfalls and ways to improve or to gain a better understanding of the market and their demographic.
Narrative structure you plan to use	Martini Structure
Elaborate your choice of narrative structure.	I plan to use the Martini structure. Since the Martini structure begins with guidance first and then exploration. In accordance with this structure the narrative would be more concrete in potential correlations I notice, and trends, and what the data explicitly represents and displays. This would then give the audience the freedom to explore the numbers and visualizations further and draw their own conclusions.
Narrative genre you plan to use	Annotated Chart
Elaborate your choice of narrative genre.	The narrative genre I plan to use is an annotated chart. The visualizations I have would be labeled to explain what they depict to the audience. Since the dataset I have has numerous attributes representing different things, I think an annotated chart would be the best way to represent various attributes together and be able to explain their correlations.

Part 2 - Outline

Plot 1 - Link and Position

Story you want to deliver	The locations in which rides are taken to and from can indicate alot about the environment, the circumstances, and the supply and demand of ridesharing. This plot can help answer, where would a good location be to get a ride from as a rider or to begin giving rides for as a driver. What goes on at these locations and how frequently people will need rides from there.

Specifications on each plot **in the order of how you lay out on your project**	(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story) 1) Task: This chart a) analyzes trend between trips between locations and how often people leave or arrive to those places b) locates where high ride frequency locations are at 2) Attributes: Pickup location, dropoff location 3) Marks: point mark, line mark 4) Channels: - Corresponding latitude and longitude for different points/locations - line marks indicating the trips between different locations - ~~Thickness of line marks to indicate frequency of rides too and from the location~~ 5) How this plot adds to the story: My visualization aims to deliver the characteristics of the trips that occur at these locations. The plot will provide more insight into the trends and concentrations of rides on certain locations.
Elaborate the choice of their marks and channels for each vis	Line marks - represent the trips too and from a location and between locations, Line channels - thickness representing quantity of trips Point marks - represent locations where trips are started from or end at Point channels - vertical and horizontal positioning by longitude and latitude

Plot 2 - Attribute

Story you want to deliver	Could there be correlation between price, distance, and the respective company for rides? Do Uber rides get more expensive over distance? Do Lyft rides? Which would be best for short distance trips or more cost efficient trips?
Specifications on each plot **in the order of how you lay out on your project**	(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story) 1) Task: This chart a) analyzes trends between ride price, distance and company b) Locates correlation with those trends 2) Attributes: price, distance, brand 3) Marks: bar 4) Channels: - Color for rideshare brand/company - Aligned vertically for price - Aligned horizontally for ride distance 5) How this plot adds to the story: My visualizations aim to deliver ride statistics for various rides under corresponding weather conditions. This plot will provide more specific insights on ride statistics and the weather in which they took place.
Elaborate the choice of their marks and channels for each vis	Bar mark - represent the number of passengers for each company, and each quantity of riders Bar channel - color for company, height for price

~~Plot 3 - Attribute~~

Story you want to deliver	Could there be a correlation between the length, time, or size of a ride depending on the weather? This can give further insight on when a driver can deicide to go on the clock and give rides, or for a rider to see when they can rely on a ride.
Specifications on each plot **in the order of how you lay out on your project**	(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story) 6) Task: This chart a) analyzes trends between ride size, ride company, and length of ride b) Locates correlation with each of these attributes 7) Attributes: temperature, passengers, length of ride, brand 8) Marks: point 9) Channels: - Color for rideshare brand/company - Aligned vertically for quantity of distance of the ride - Aligned horizontally for ride size 10) How this plot adds to the story: My visualizations aim to deliver ride statistics for various rides. This plot will provide more specific insights on distance of rides and how many passengers are included based on the distance.
Elaborate the choice of their marks and channels for each vis	Point mark - represent ride instances Point channel - vertical position by distance of trip, horizontal position by number of passengers in the ride

Following sample answer about a single plot shows how detailed your answers to part 2 should be.

1. Plot 1
   6) Task: This chart a) analyzes trend between Height and Weight of patients with heart diseases and b) locates outliers within the patients
   7) Attributes: Height, Weight
   8) Marks: point mark
   9) Channels:
      - aligned vertical position channel for Height
      - aligned horizontal channel for Weight
   10) How this plot adds to the story:
   My visualizations aim to deliver health characteristics of patients with heart disease. This plot will provide more specific insights on Height and Weight.