UTSA Data Analytics & Visualization Bootcamp

Group 1

Uber Ride Patterns in Extreme Weather

Group 1 Team Members

Barbie Hurtado

Brian Paniagua

Walaa Youssef

Nathan Farley

Project Information

Project Scope

The aim of this project is to analyze the relationship between weather conditions and Uber ride usage, especially during extreme weather conditions. The problem we address is whether weather conditions, ranging from subzero temperatures to scorching heat, influence the frequency of ride-sharing usage.

Data to be Analyzed

- Compare the Uber rides vs the weather in New York City
- Compare the following data:
  - Ride demand
  - Weather conditions
  - Geographic distribution in NYC.

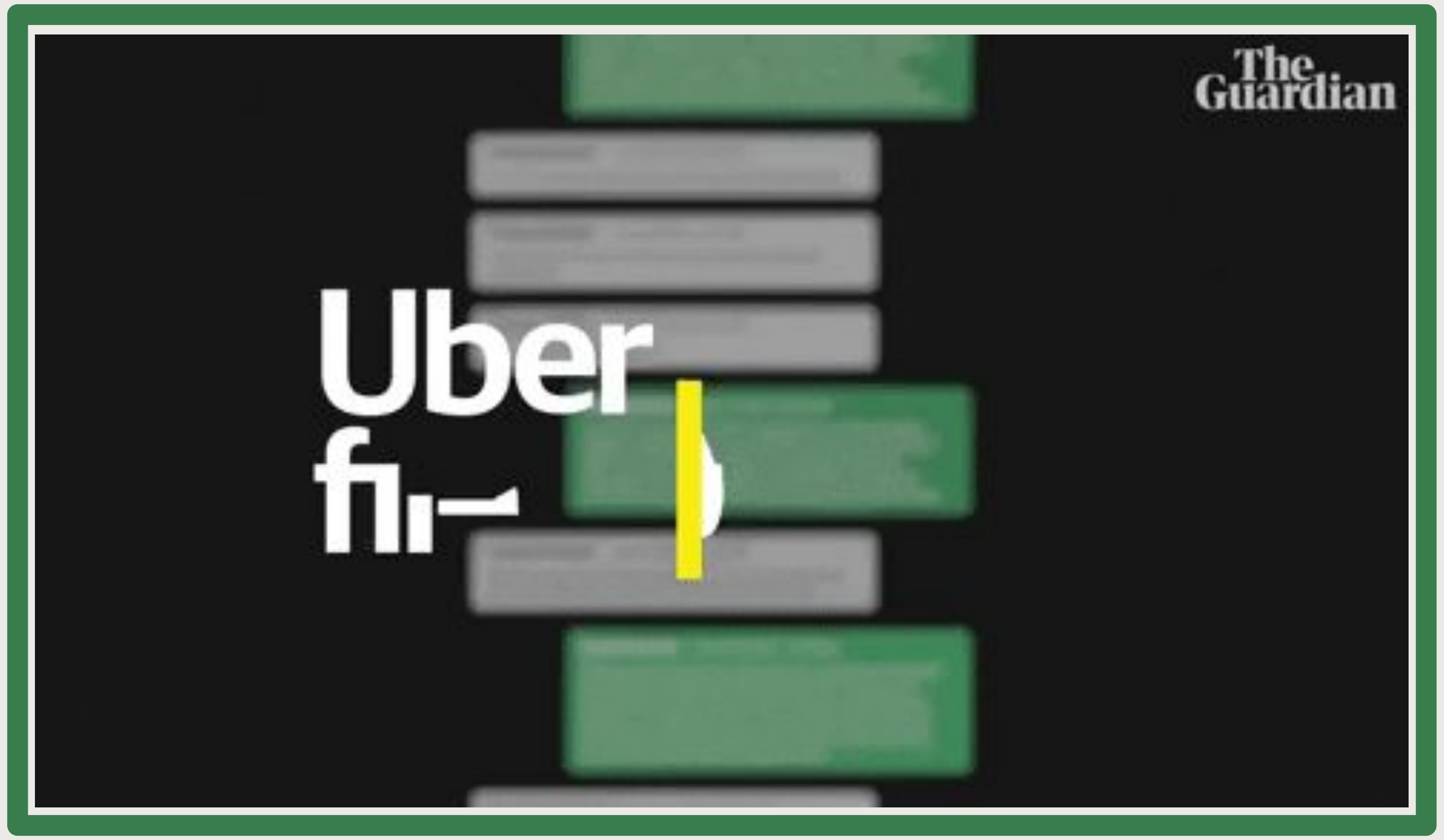
Hypothesis

- **Hypothesis:** Extreme weather conditions significantly affect the demand and geographic distribution of rides for Uber and other FHV companies.
- **Null Hypothesis:** Extreme weather conditions have no significant impact on the demand and geographic distribution of rides for Uber and other FHV companies.

Questions

1. Correlation Study: Is there a specific correlation between weather factors (e.g., temperature, precipitation) and the number of rides provided by Uber and other FHV companies?

    a. Demand Impact: How does weather influence ride demand for Uber and other FHV companies in New York?
    b. Strategic Planning: Can correlations between weather and ride frequency be leveraged to plan driver promotions and increase ride shares?



What Would Be Accomplished After the Analysis?

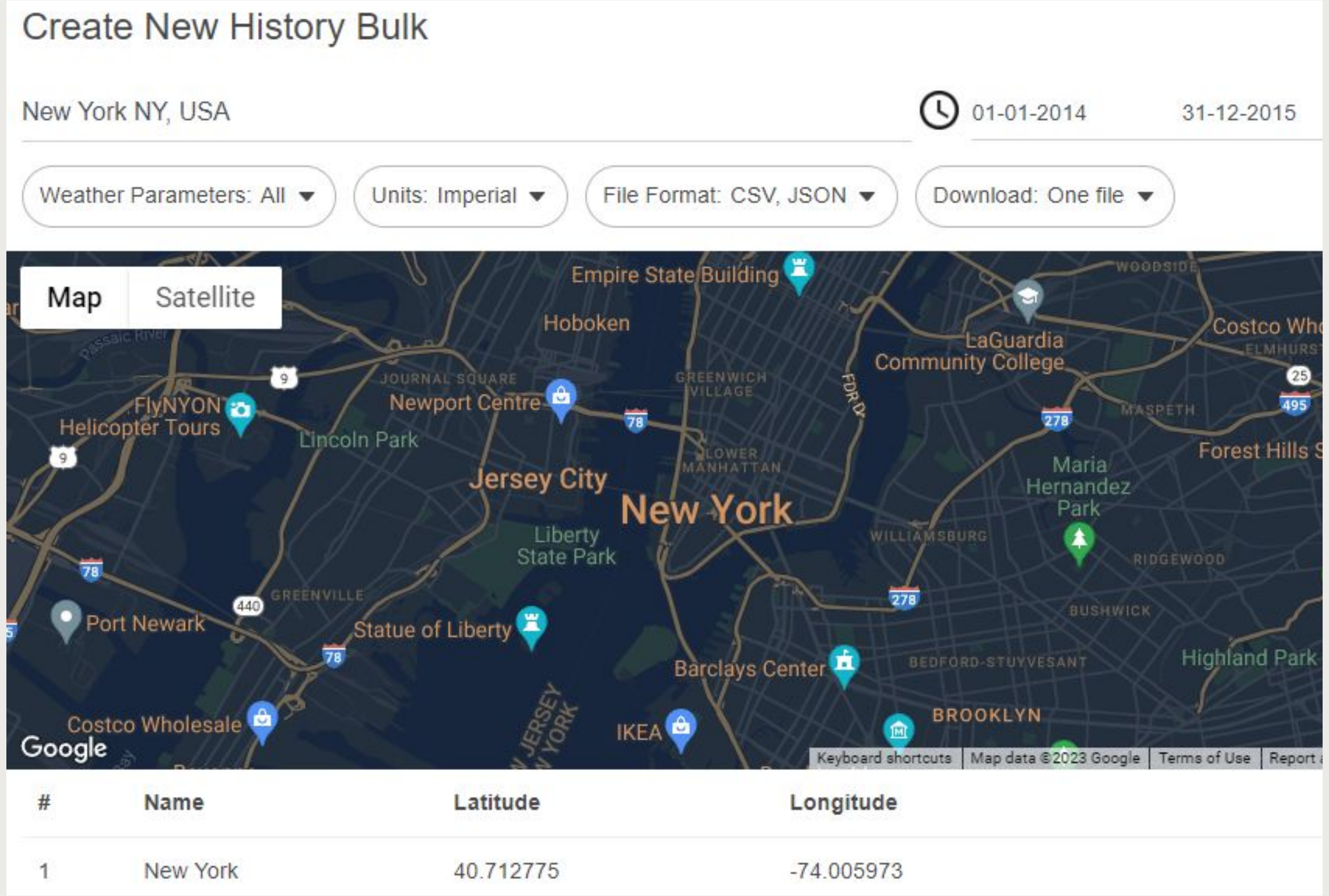
This analysis is anticipated to reveal significant insights into how weather influences ride-sharing. It may inform optimized service strategies, contribute to urban transportation planning, and influence marketing and promotional tactics for FHV companies during extreme weather conditions.

Project Data Sources

Data Retrieval

.csv data:
Historical weather data

https://home.openweathermap.org/history_bulks/



Data Retrieval

.csv data:
Uber Pickups in New York City

https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city?resource=download&select=Uber-Jan-Feb-FOIL.csv

Data Cleanup

2014 Uber Data

```
Before Data Cleaning
Number of rows for each column
Base    4534327
Date    4534327
Time    4534327
dtype: int64

Percent of Data:
100%

After Data Cleaning
Number of rows for each column
Base    937192
Date    937192
Time    937192
dtype: int64

Percent of Data:
20.67%
```

2015 Uber Data

```
Before Data Cleaning
Number of rows for each column
Base    4525316
Date    4525316
Time    4525316
dtype: int64

Percent of Data:
100%

After Data Cleaning
Number of rows for each column
Base    4525316
Date    4525316
Time    4525316
dtype: int64

Percent of Data:
100.0%
```

Data Cleanup

2014 Uber Rides Summary Table

Base	Total Trips	Maximum Trips	Minimum Trips	Start Date	End Date	Total Days	Average Trips/Day	Most Active Day	Least Active Day
B02512	205673	2186	397	2014-04-01	2014-09-30	183	1123.896175	2014-04-30	2014-07-05
B02598	1393113	13484	3557	2014-04-01	2014-09-30	183	7612.639344	2014-04-30	2014-07-05
B02617	1458853	16472	1985	2014-04-01	2014-09-30	183	7971.874317	2014-09-05	2014-04-20
B02682	1212789	13327	2882	2014-04-01	2014-09-30	183	6627.262295	2014-04-30	2014-07-05
B02764	263899	9632	114	2014-04-01	2014-09-30	183	1442.071038	2014-09-27	2014-07-05

2015 Uber Rides Summary Table

Base	Total Trips	Maximum Trips	Minimum Trips	Start Date	End Date	Total Days	Average Trips/Day	Most Active Day	Least Active Day
B02617	185902	2446	320	2015-01-01	2015-06-30	181	1027.082873	2015-02-13	2015-04-04
B02598	655439	11443	1272	2015-01-01	2015-06-30	181	3621.209945	2015-02-13	2015-03-17
B02682	834533	15129	1331	2015-01-01	2015-06-30	181	4610.679558	2015-02-14	2015-04-14
B02764	778959	14474	1358	2015-01-01	2015-06-30	181	4303.640884	2015-02-14	2015-03-03
B02512	1698371	33953	1379	2015-01-01	2015-06-30	181	9383.265193	2015-02-14	2015-03-08
B02765	353017	7365	929	2015-01-01	2015-06-30	181	1950.370166	2015-02-20	2015-01-27
B02835	17160	1056	264	2015-01-01	2015-06-30	181	94.806630	2015-06-27	2015-06-08
B02836	1935	97	39	2015-01-01	2015-06-30	181	10.690608	2015-06-27	2015-06-03

Data Cleanup

Weather Data by Times

	Date	time_HMS	temp	feels_like	temp_min	temp_max	humidity	wind_speed	Cloudiness	weather_description
0	2014-01-01	0:00:00	32.23	20.43	30.16	32.43	48	19.55	0	sky is clear
1	2014-01-01	1:00:00	30.78	18.18	30.15	31.26	50	23.00	0	sky is clear
2	2014-01-01	2:00:00	30.27	18.57	28.36	30.94	48	17.27	0	sky is clear
3	2014-01-01	3:00:00	28.78	16.18	26.56	29.95	46	20.69	0	sky is clear
4	2014-01-01	4:00:00	27.64	15.04	26.24	28.94	44	19.55	0	sky is clear

Weather Summary

	Date	Mean_Temp	Mean_Feels_Like	Min_Temp	Max_Temp	Mean_Humidity	Mean_Wind_Speed	Mean_Cloudiness	Weather_Description
0	2014-01-01	28.941667	19.074167	22.96	34.95	48.958333	12.807917	13.125000	sky is clear
1	2014-01-02	29.206538	18.154615	22.96	34.14	71.576923	16.800000	93.846154	overcast clouds
2	2014-01-03	16.499286	3.899286	10.04	26.55	75.642857	20.561786	75.714286	snow
3	2014-01-04	15.930417	5.985417	6.76	28.11	50.083333	8.940000	3.125000	sky is clear
4	2014-01-05	30.174138	22.742759	18.95	37.11	76.068966	8.063793	59.310345	sky is clear

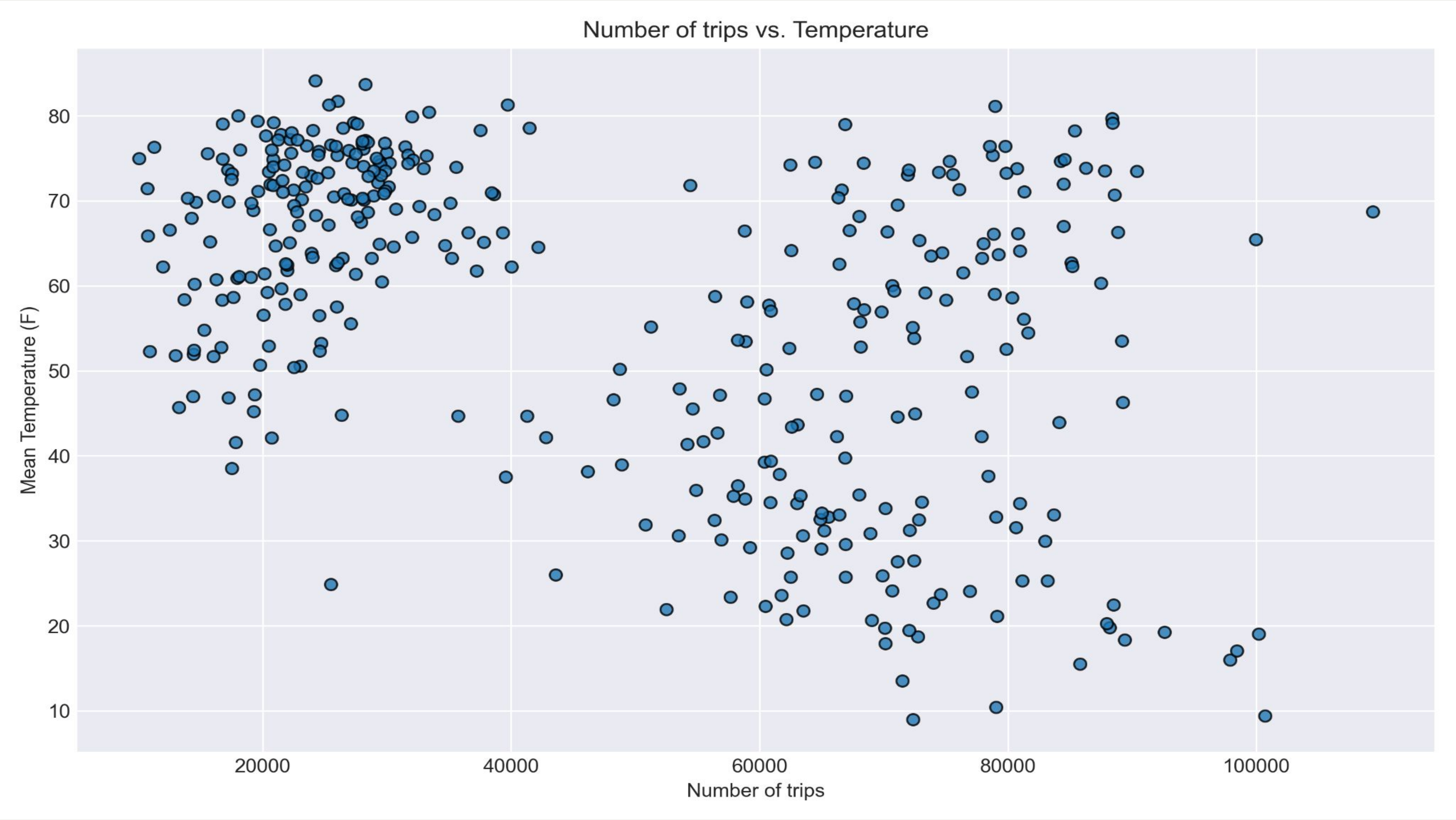
Data Cleanup

Weather and Ride Data Integration

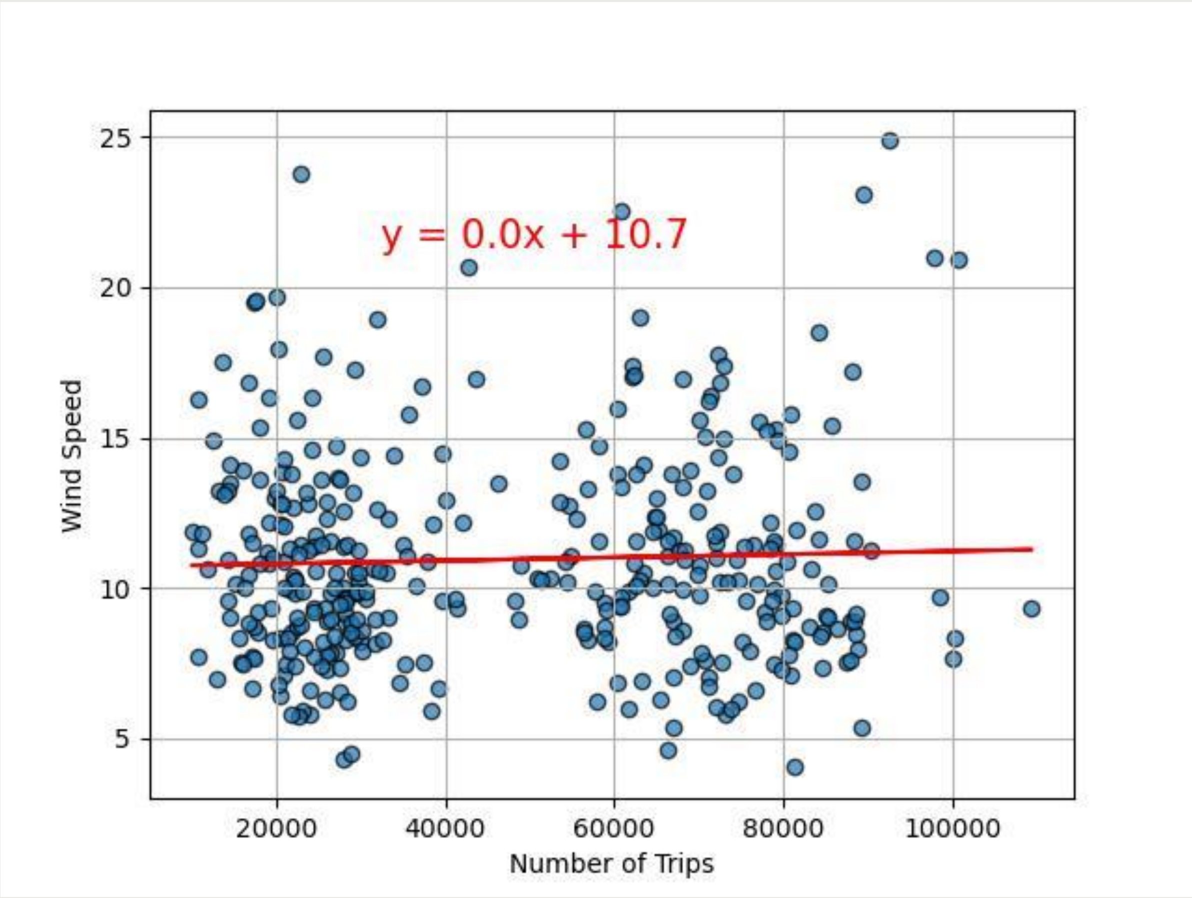
Date	Number of trips	Mean_Temp	Mean_Feels_Like	Min_Temp	Max_Temp	Mean_Humidity	Mean_Wind_Speed	Mean_Cloudiness	Weather_Description
2014-04-01	14376	46.994583	41.735417	35.94	61.03	47.708333	10.919167	0.833333	sky is clear
2014-04-02	17230	46.833333	43.668333	39.99	53.94	66.458333	6.657917	28.958333	sky is clear
2014-04-03	20482	52.919167	50.192083	42.01	67.35	50.791667	6.417083	0.000000	sky is clear
2014-04-04	26353	44.803077	38.980385	39.99	55.74	72.807692	11.565000	92.307692	light rain
2014-04-05	19273	45.201429	39.788929	39.00	55.36	71.392857	16.367857	77.321429	mist
...	...	...	...	...	...	...	...	...	...
2015-06-26	86287	73.835833	73.507083	68.99	80.49	55.083333	8.668750	62.500000	broken clouds
2015-06-27	109381	68.684400	68.369600	60.51	76.01	66.280000	9.359600	69.600000	overcast clouds
2015-06-28	80811	66.128611	66.526111	59.43	75.40	86.861111	15.771111	99.305556	mist
2015-06-29	66346	70.360417	69.815000	62.55	77.47	57.833333	11.585833	58.958333	overcast clouds
2015-06-30	72018	73.613333	73.310833	65.66	81.79	54.875000	10.987083	14.375000	sky is clear

rows × 10 columns

Mean Temp vs. Trips

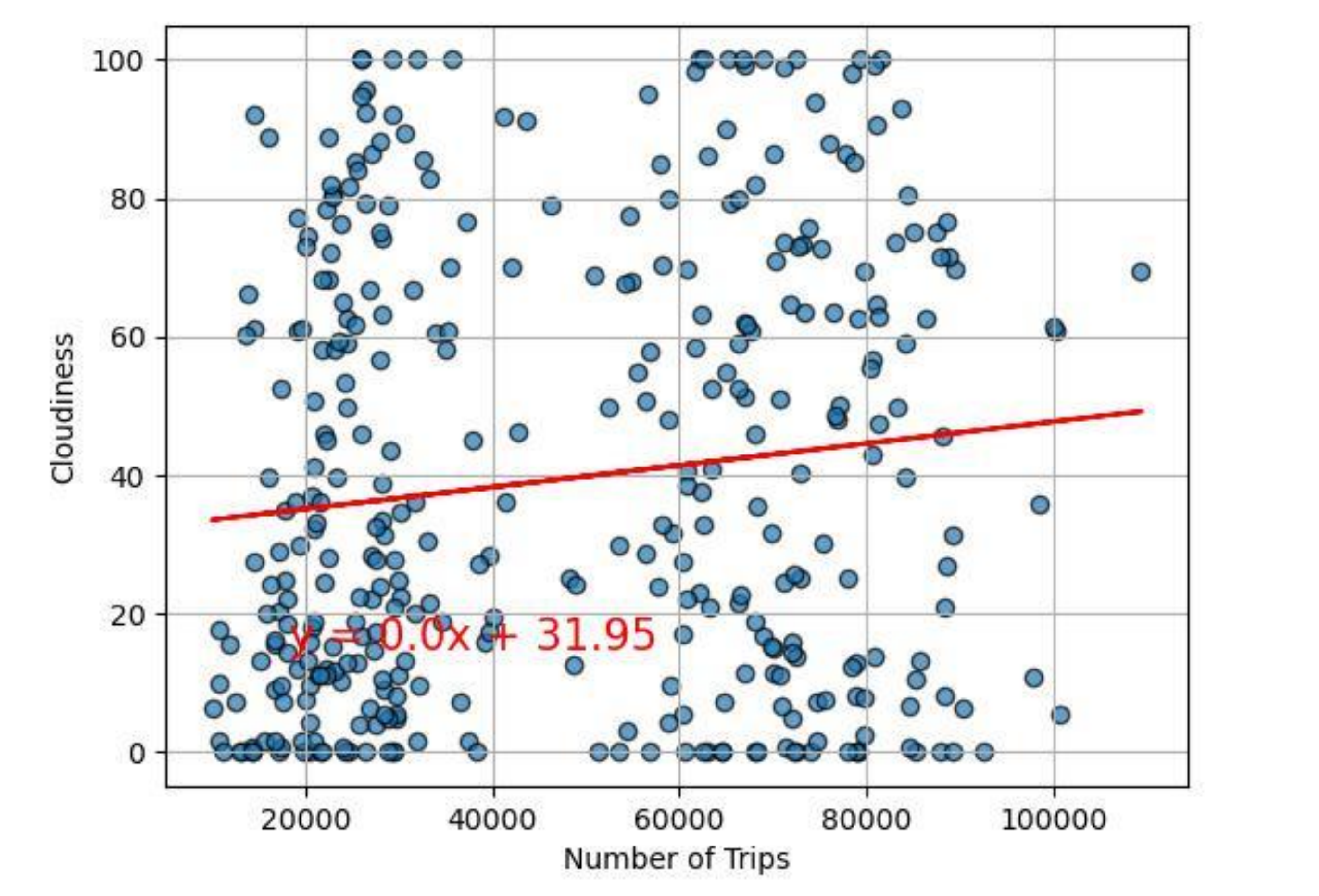


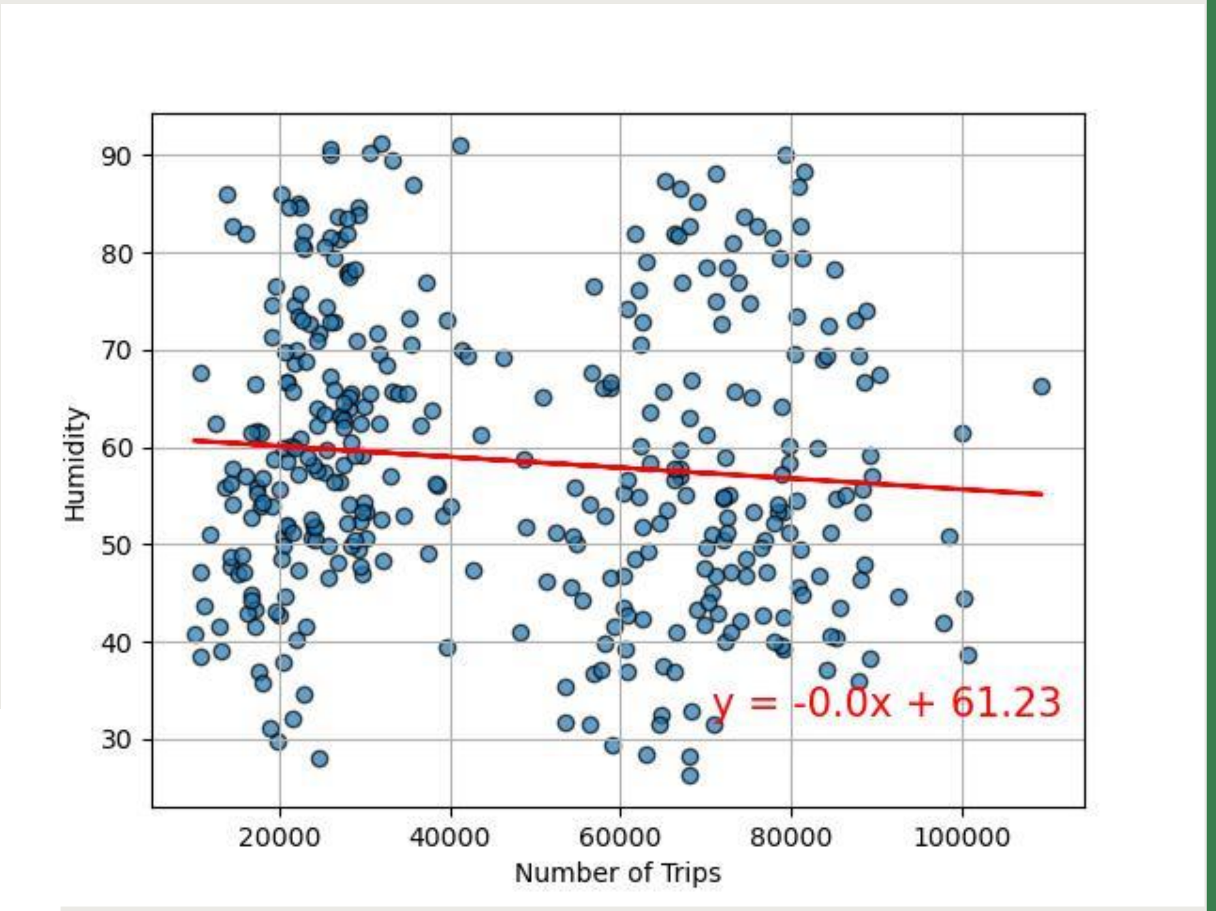
Weather Conditions vs. Uber Trips



Wind vs. Trips

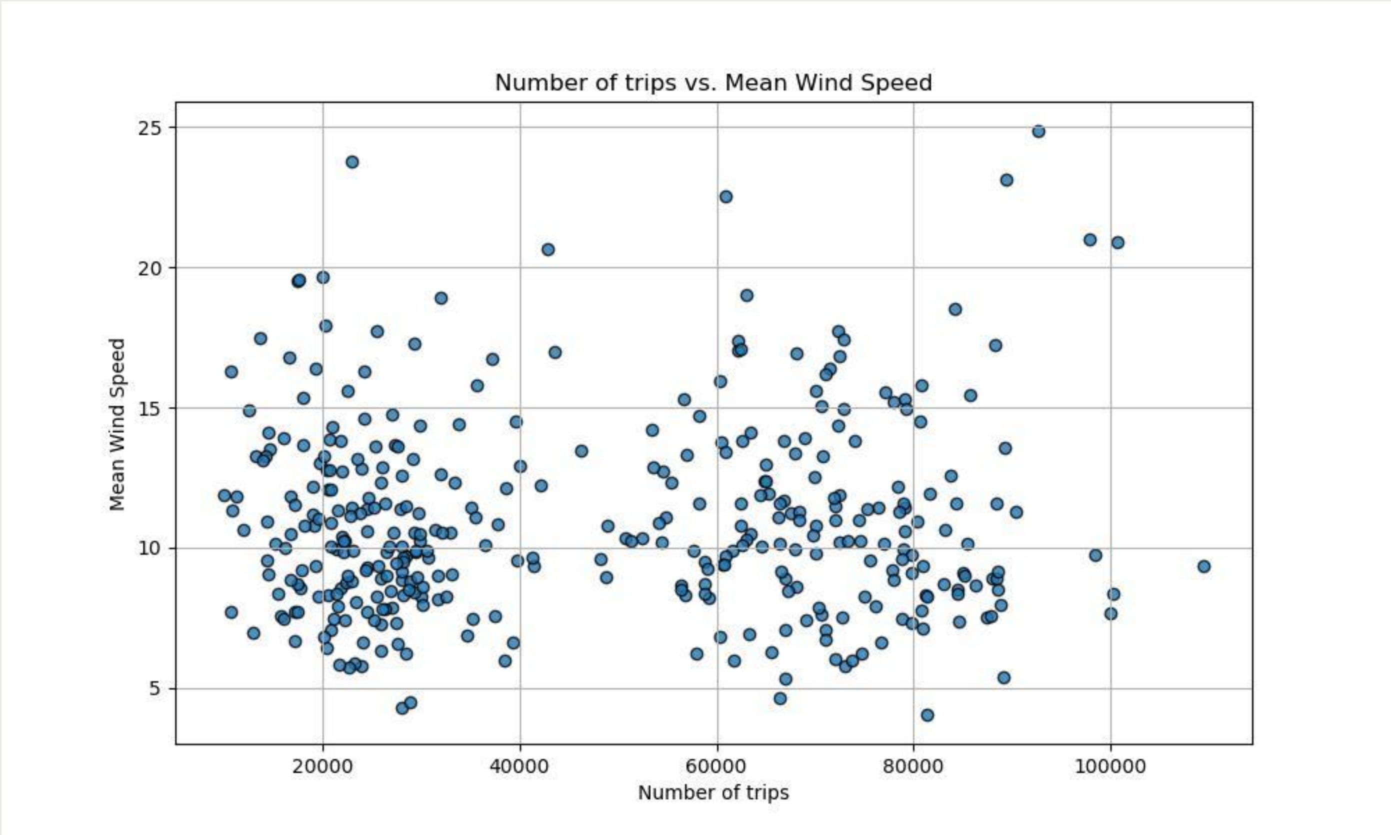
Cloudiness vs. Trips

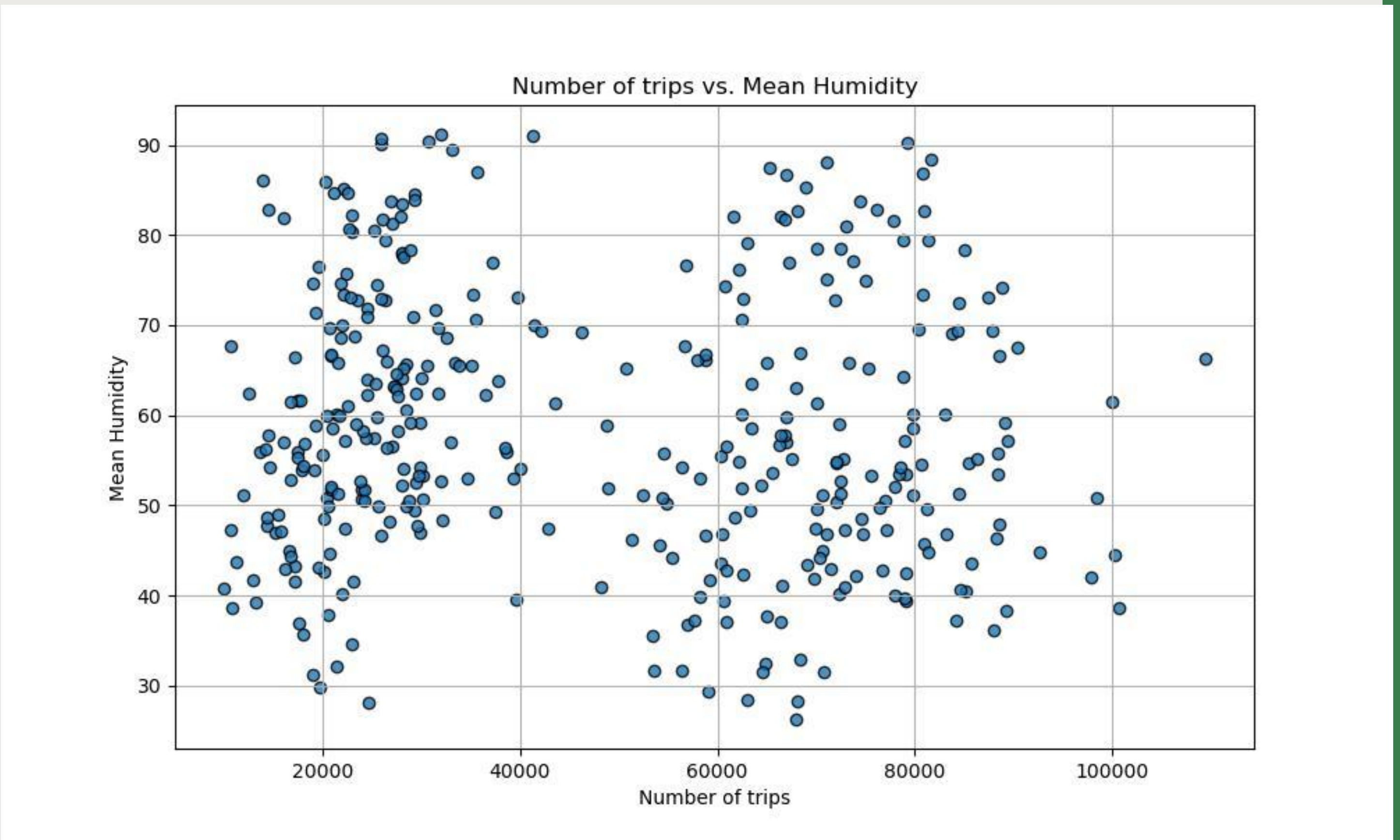




Humidity vs. Trips

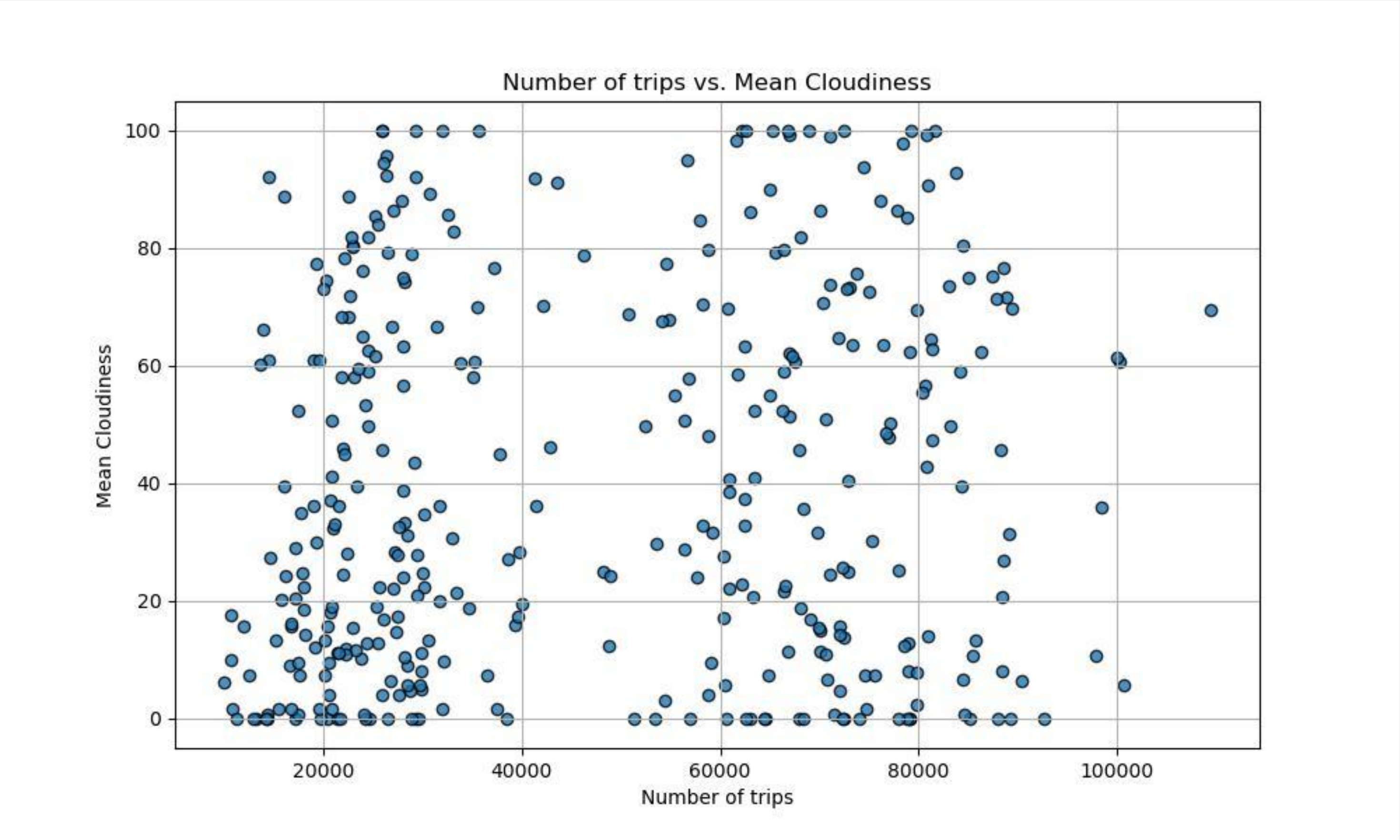
Weather Conditions vs. Uber Trips





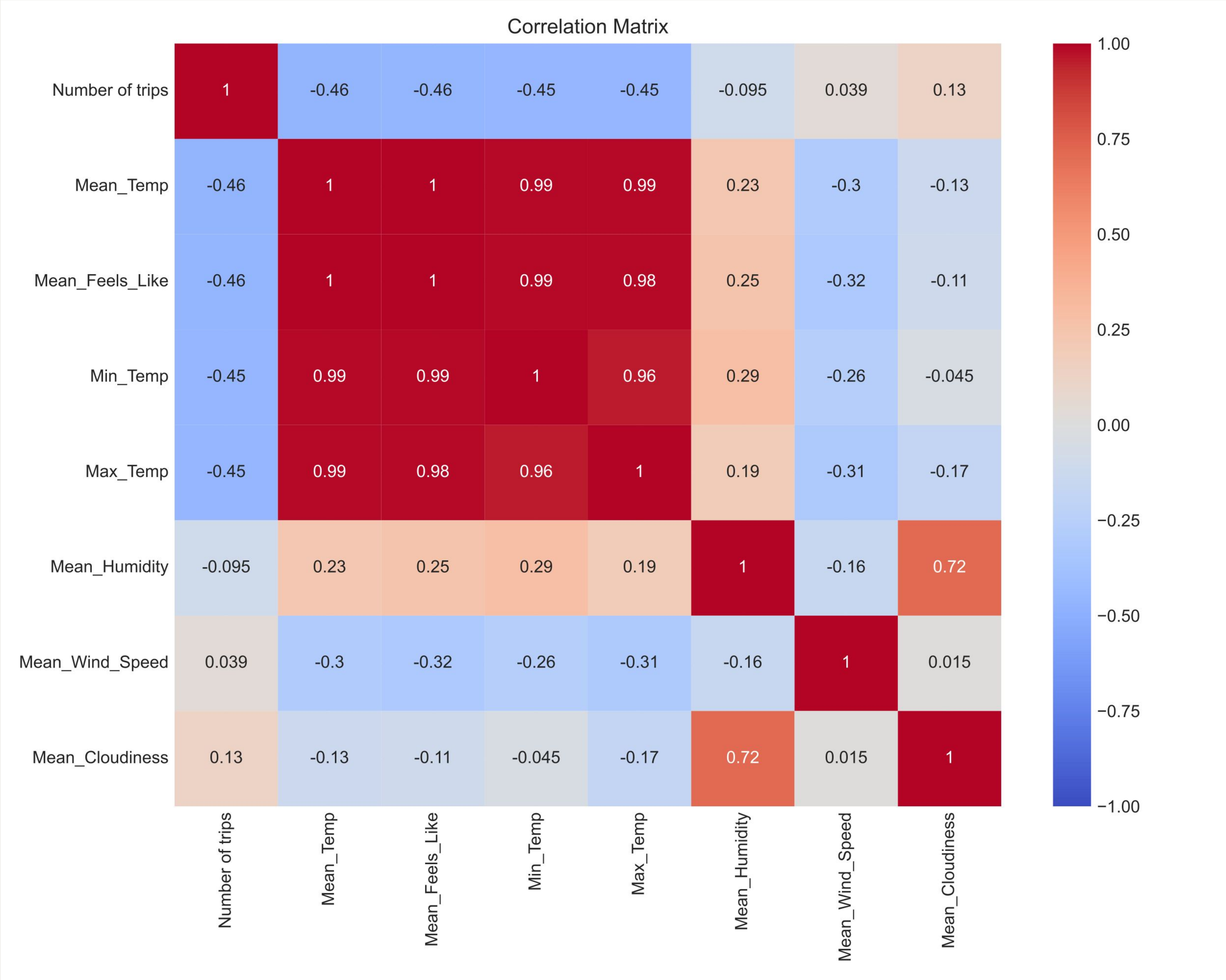
Mean Cloudiness vs. Trips

Mean Wind vs. Trips



Mean Humidity vs. Trips

Correlation Matrix



Interpretation & Recommendations:

The most prominent relationship is between temperature and the number of trips. As temperatures rise, the number of trips tend to decrease. This could be due to people choosing alternative modes of transportation, walking, or avoiding outings during hot days. Humidity and cloudiness also show relationships with the number of trips, but these correlations are weaker, suggesting other factors might be at play.

Given this correlation data, Uber and other FHV companies might want to:

Promote Services on Colder Days: Since there is a negative correlation between temperature and the number of trips, it may be strategic to offer promotions or discounts during colder days to incentivize more users.

Optimize Fleet Deployment: Knowing that there are fewer trips on hotter days, companies can optimize the number of drivers they have on the road during such conditions.

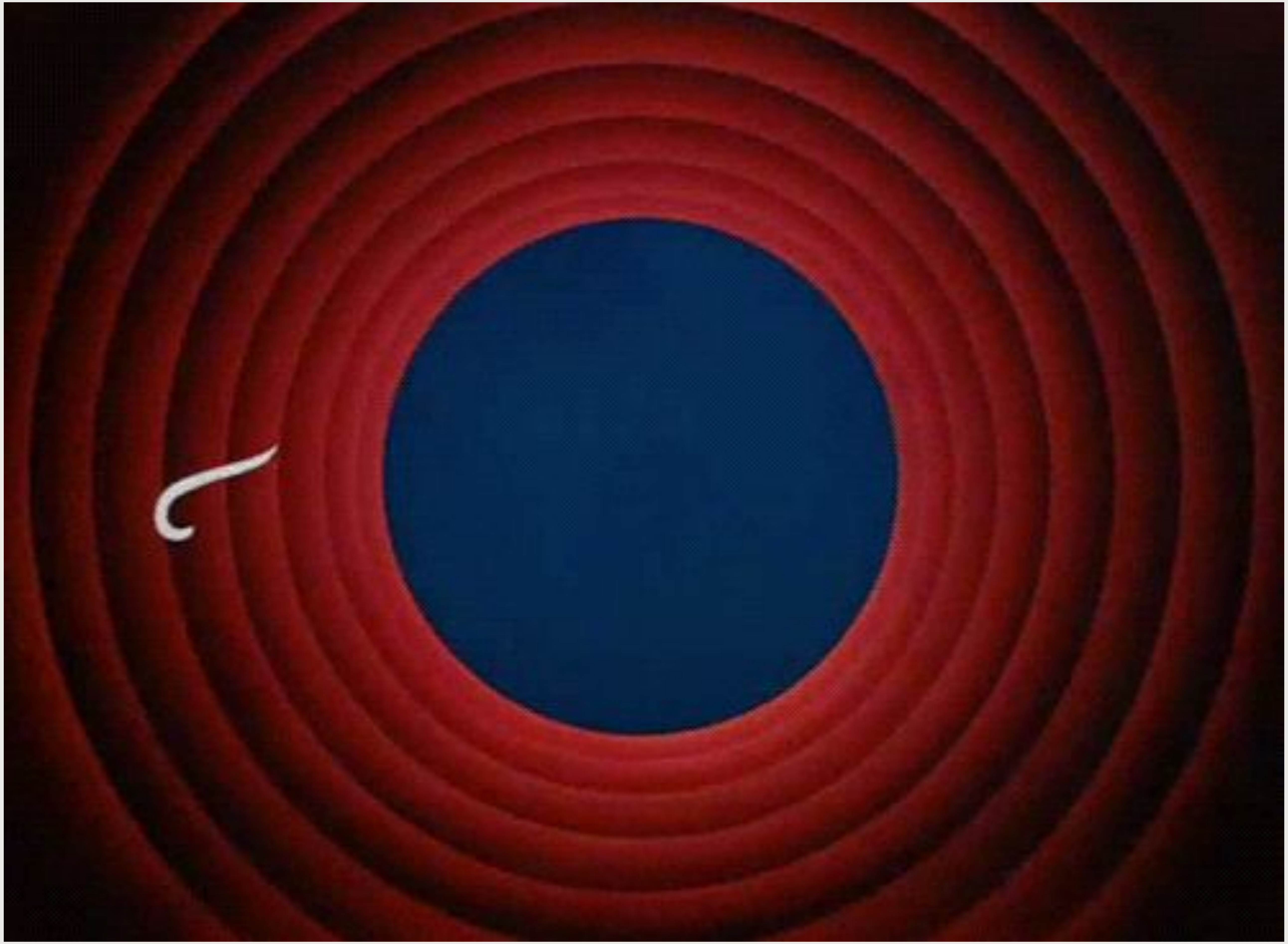
Consider Humidity and Cloudiness: While the correlations here are weaker, these weather parameters still influence ride demand. It might be worth delving deeper to understand if there are specific thresholds where humidity or cloudiness dramatically affects trip numbers.

Limitations

- Absence of an accessible API for the required data, leaving us reliant on the sole available dataset at our disposal.
- Within our Uber data, a noticeable gap emerges from April to September 2014, followed by a leap to January to June 2015, causing the omission of the entire Fall season.
- Our dataset was confined solely to NYC, preventing us from accessing data from all boroughs.
- The project's timeframe was restricted to class hours only, due to the demanding professional commitments of all team members.

Conclusion

The hypothesis that extreme weather conditions significantly affect the demand of rides for Uber and other FHV companies was proven by this analysis.




QUESTIONS?