

# Kabhi Khushi Kabhie Cab



**Presented By:**

Ankit Kumar

Noyonica Chatterjee

Surabhi Tannu

# Problem Statement

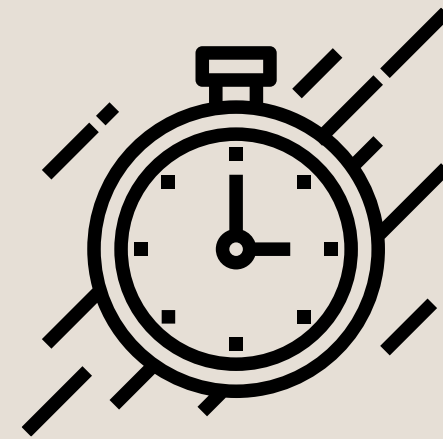
Addressing the uncertainties in cab pricing, availability, and traffic-related delays to enhance users' overall transportation experience

# Solutions



## **Predictive Pricing Model**

Predicting minimum cab ride prices based on routes and times for cost-effective traveler choices



## **Cab Availability Forecasting**

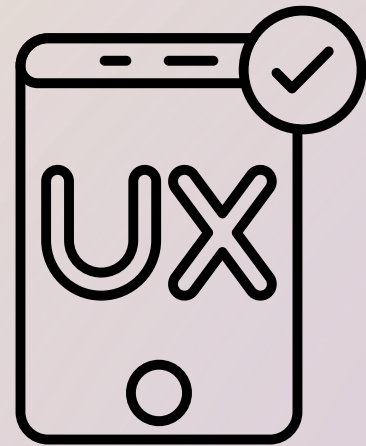
Estimating cab waiting times to reduce uncertainty and enhance user convenience



## **Traffic Condition Prediction**

Forecasting traffic congestion to aid efficient travel planning and reduce delays

# Potential Applications and Impact



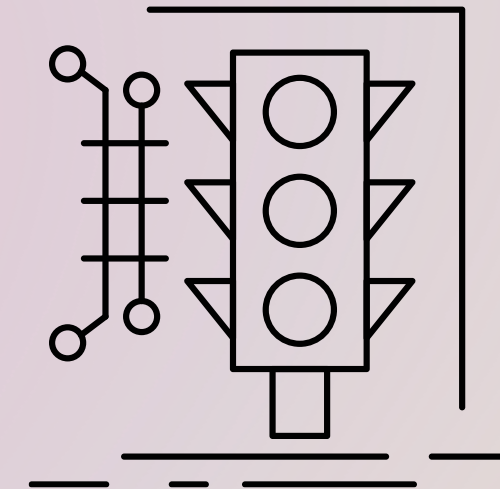
## Enhanced User Experience

Would enhance convenience and cost-efficiency for cab users, leading to increased customer satisfaction



## Smart Traffic Prediction

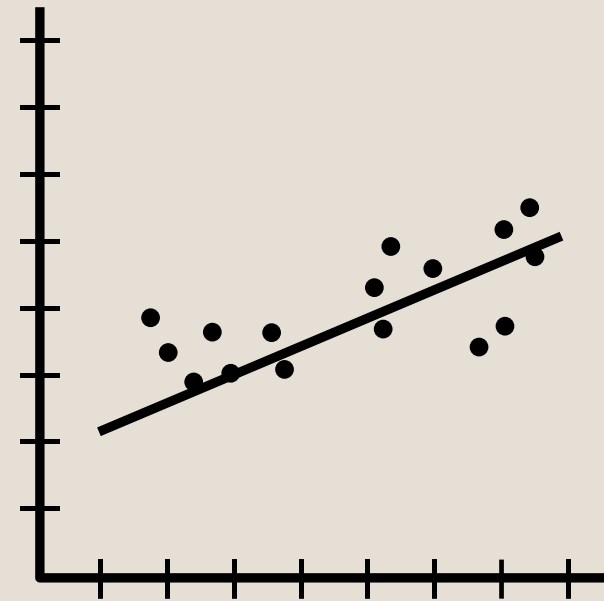
Traffic prediction data can aid in time efficient time experience.





# Literature Review

variables
passenger_count
fare_amount
pickup_datetime
pickup_longitude
pickup_latitude
dropoff_longitude
dropoff_latitude



The company uses historical data and machine-learning models combined to predict ETA.

It uses two layered neural network to predict ETA.

$$A_{ij} = \frac{\exp(QK^T / \sqrt{d})_{i,j}}{\sum_{j=1}^L \exp(QK^T / \sqrt{d})_{i,j}}$$

$$Q = X_{emb} W_q$$

$$K = X_{emb} W_k$$

$$V = X_{emb} W_v$$

Method	Accuracy Percentage)	(in
Random Forest	79.7864	
Linear Regression	71.8255	

Fig 9. Accuracy

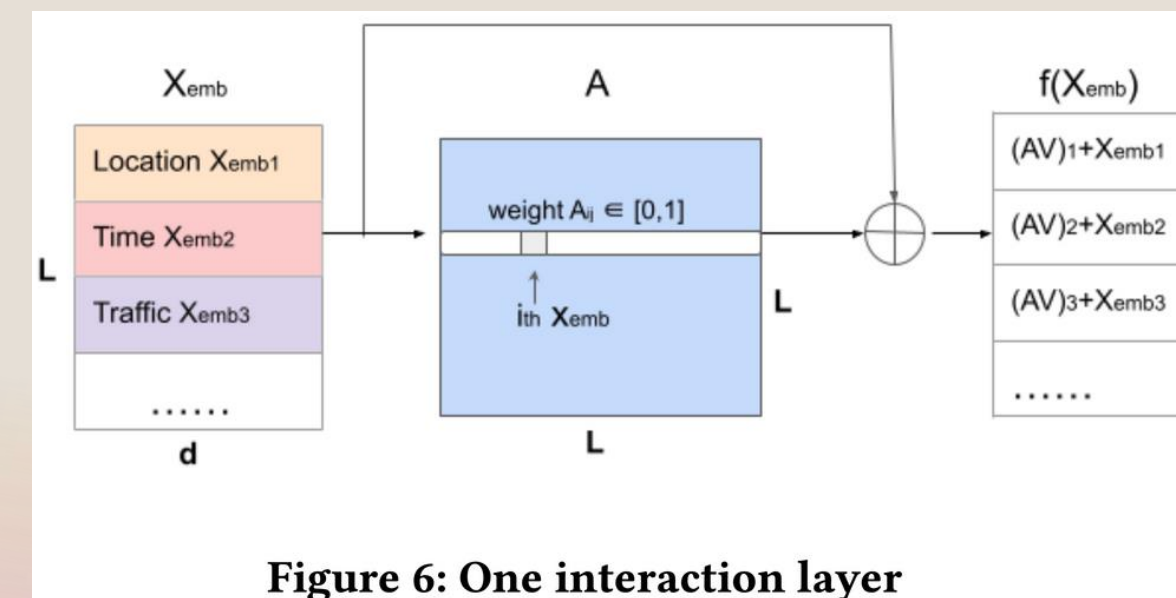
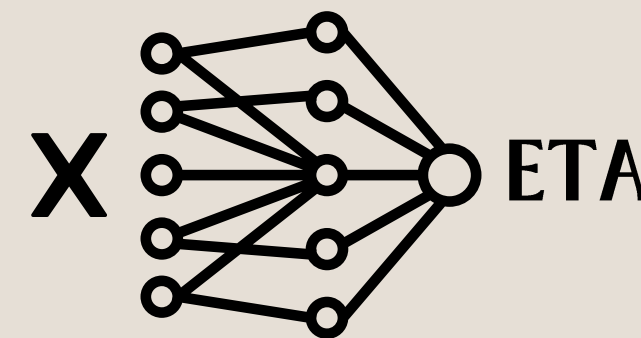


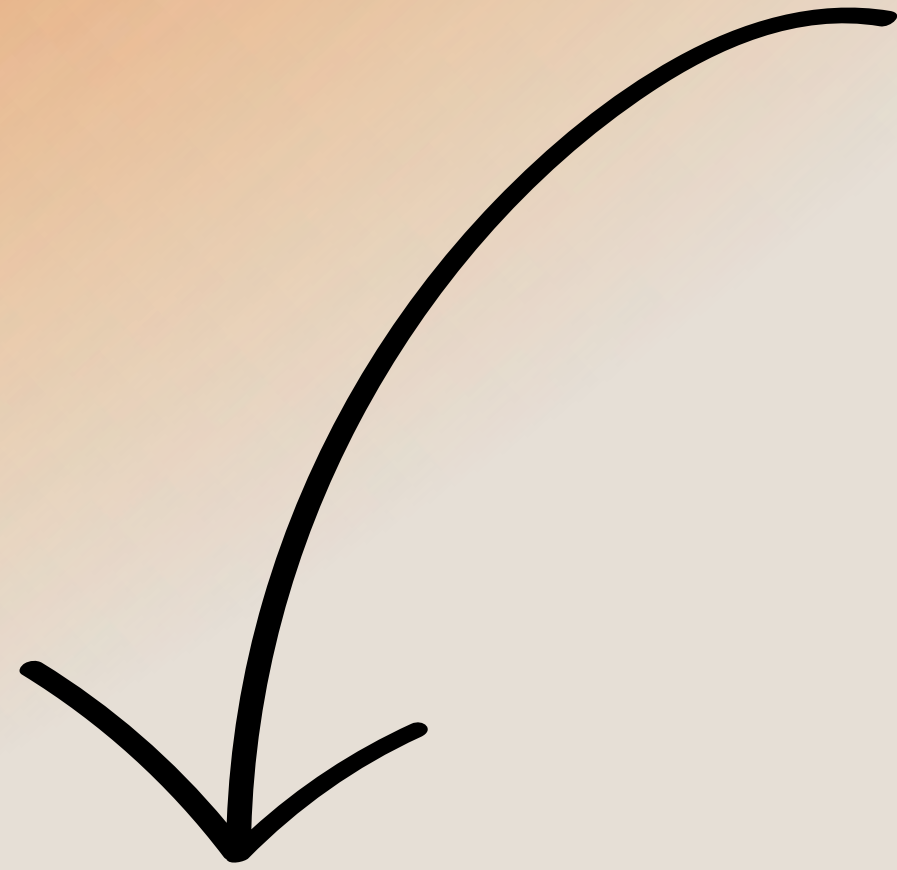
Figure 6: One interaction layer



[Real time prediction of Cab Fair using Machine Learning](#)  
(Dr. A. Pravin, et al.)

[DeepETA: An ETA Post-processing System at Scale](#)  
(Uber Technologies Inc.)

# Scraping the Web



**Get a ride**

Plaksha University

Elante Mall

Today ▼    Now ▼

**Choose a ride**

**Recommended**

**UberGo** 1  
7 mins away • 2:09 am  
Affordable, compact rides  
**₹284.68** 16% off  
₹324.68

**UberXL** 1  
10 mins away • 2:12 am  
Affordable, SUV rides  
**₹599.15**

**MOTO** 1  
6 mins away • 2:08 am  
Send packages only  
**₹189.98**

**Economy**

**Premier** 1  
9 mins away • 2:11 am  
Comfortable sedans, top-quality drivers  
**₹319.80** 16% off  
₹359.80

	Cab_Type	Cab_Price	Arrival_Time	Pick_Up	Destination	Current Time	Current Day
330818	UberXLPerson1	₹232.09	7 mins away•20:18	Plaksha University	Sector 82	03/10/2023, 20:01:04	Tuesday
188085	UberXLPerson1	₹609.93	9 mins away•21:02	Plaksha University	Railway Station Chandigarh (Panchkula Side)	20/09/2023, 20:13:49	Wednesday
44997	PremierPerson1	₹202.01	Unavailable	Plaksha University	Chandigarh Airport Departure Terminal (IXC)	12/09/2023, 0:17:05	Tuesday
218293	UberGoPerson1	₹174.22	5 mins away•15:07	Sector 82	Plaksha University	22/09/2023, 14:52:05	Friday
231396	PremierPerson1	₹316.27	8 mins away•7:06	Plaksha University	Railway Station Chandigarh (Panchkula Side)	23/09/2023, 06:28:19	Saturday

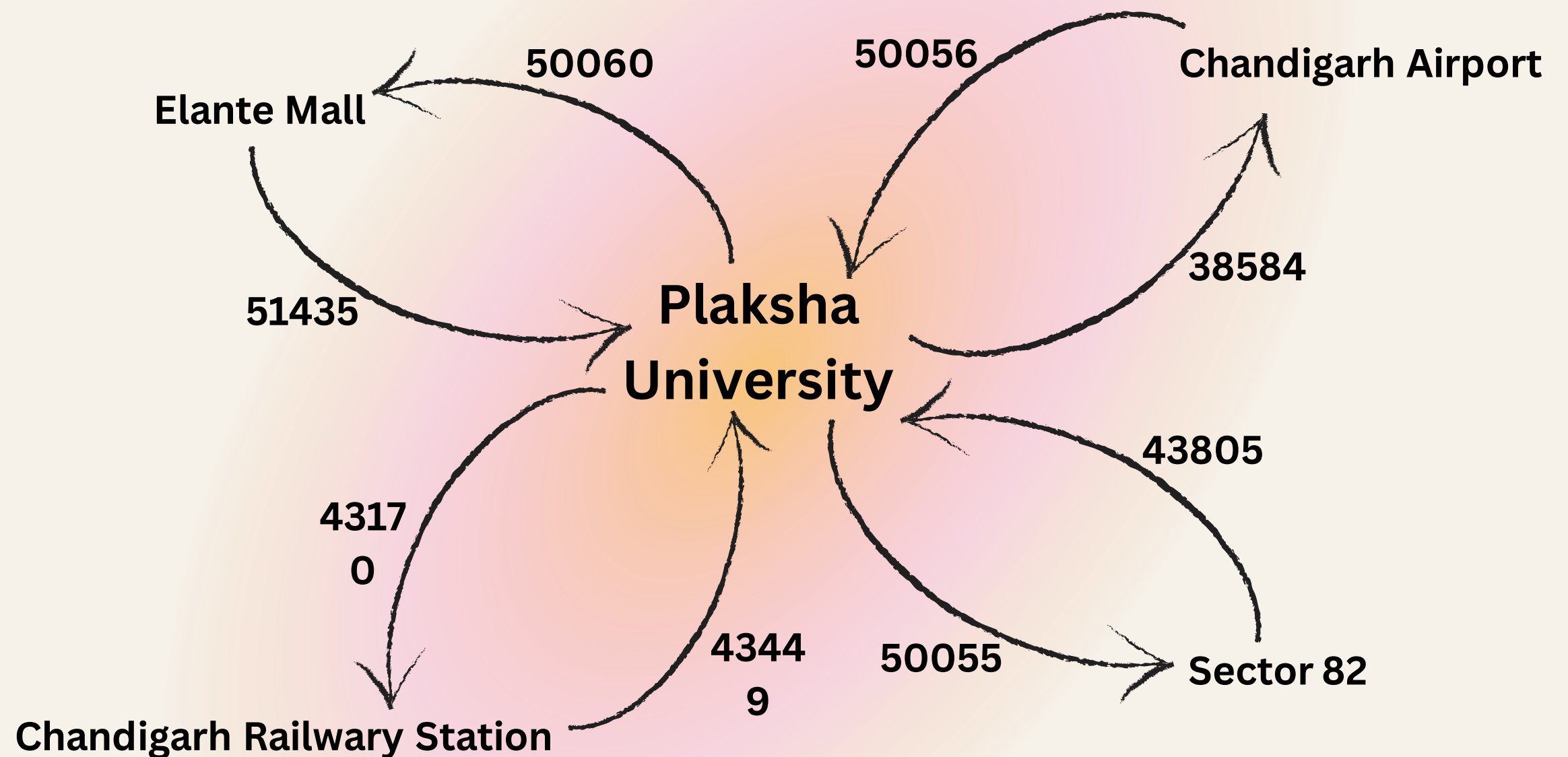


# Let's talk Data

We have collected a total of 3,75,348 datapoints across 7 attributes

## Attributes

Cab_Type	object
Cab_Price	object
Arrival_Time	object
Pick_Up	object
Destination	object
Current Time	object
Current Day	object
dtype:	object



# Data Preprocessing

Raw Data

	Cab_Type	Cab_Price	Arrival_Time	Pick_Up	Destination	Current Time	Current Day
0	Hatchbacks	₹280.96	2 mins away•21:16 drop-off	Elante Mall	Plaksha University	09/09/2023, 20:39:04	Saturday
1	SUV	₹473.86	3 mins away•21:16 drop-off	Elante Mall	Plaksha University	09/09/2023, 20:39:04	Saturday
2	Auto	₹225.62	4 mins away•21:16 drop-off	Elante Mall	Plaksha University	09/09/2023, 20:39:04	Saturday
3	Bike	₹156.81	2 mins away•21:15 drop-off	Elante Mall	Plaksha University	09/09/2023, 20:39:04	Saturday
4	Sedan	₹319.02	7 mins away•21:17 drop-off	Elante Mall	Plaksha University	09/09/2023, 20:39:04	Saturday

Preprocessed Data

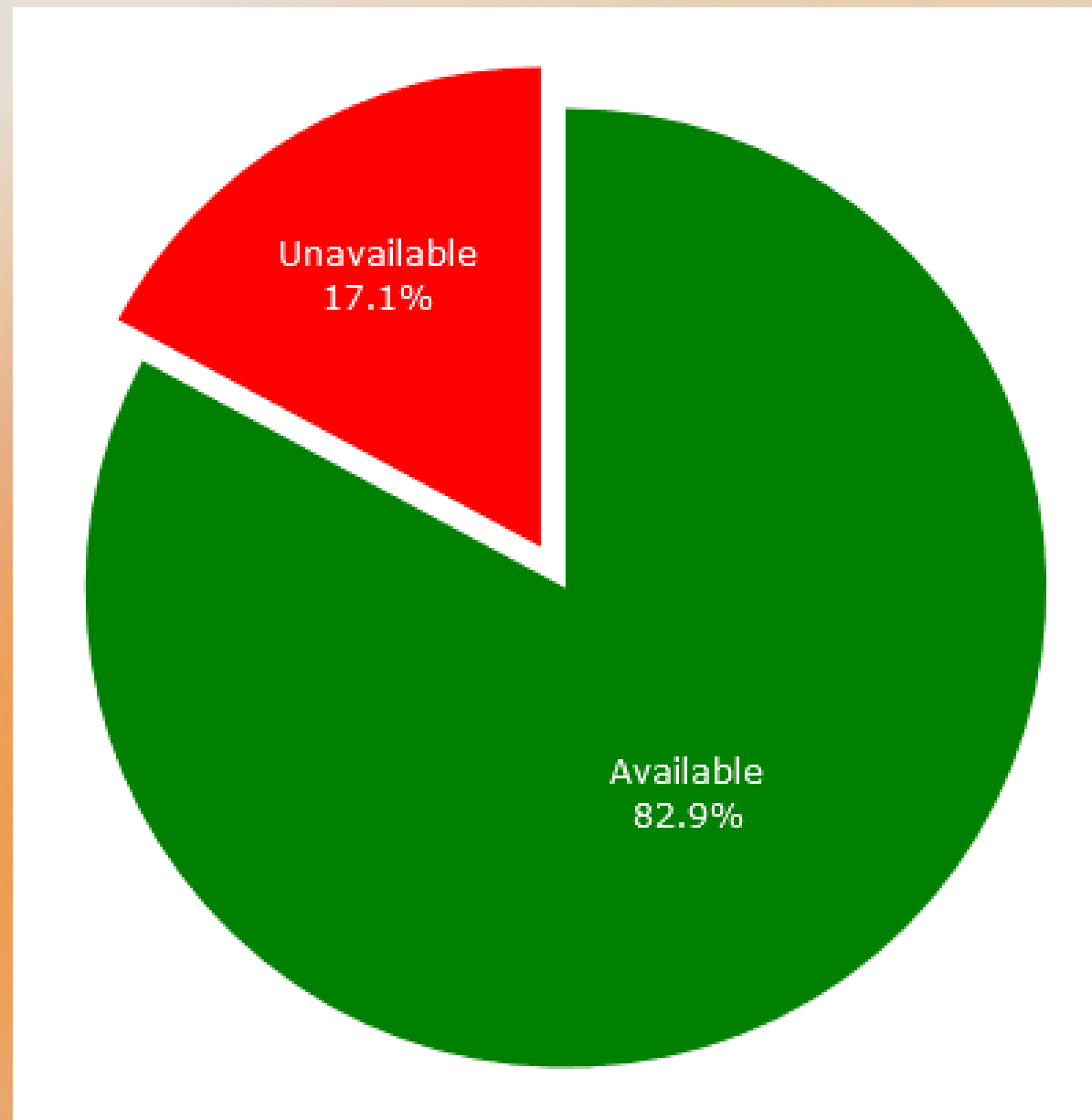
	Cab_Type	Cab_Price	Pick_Up	Destination	Current Day	Cab_Arrival_Time	Cab_Destination_Time	Current_Date	Current_Time	Availability	Route_Time
0	Hatchbacks	280.96	Elante Mall	Plaksha University	Saturday	2 mins away	21:16	09/09/2023	20:39	1	37.0
1	SUV	473.86	Elante Mall	Plaksha University	Saturday	3 mins away	21:16	09/09/2023	20:39	1	37.0
2	Auto	225.62	Elante Mall	Plaksha University	Saturday	4 mins away	21:16	09/09/2023	20:39	1	37.0
3	Bike	156.81	Elante Mall	Plaksha University	Saturday	2 mins away	21:15	09/09/2023	20:39	1	36.0
4	Sedan	319.02	Elante Mall	Plaksha University	Saturday	7 mins away	21:17	09/09/2023	20:39	1	38.0



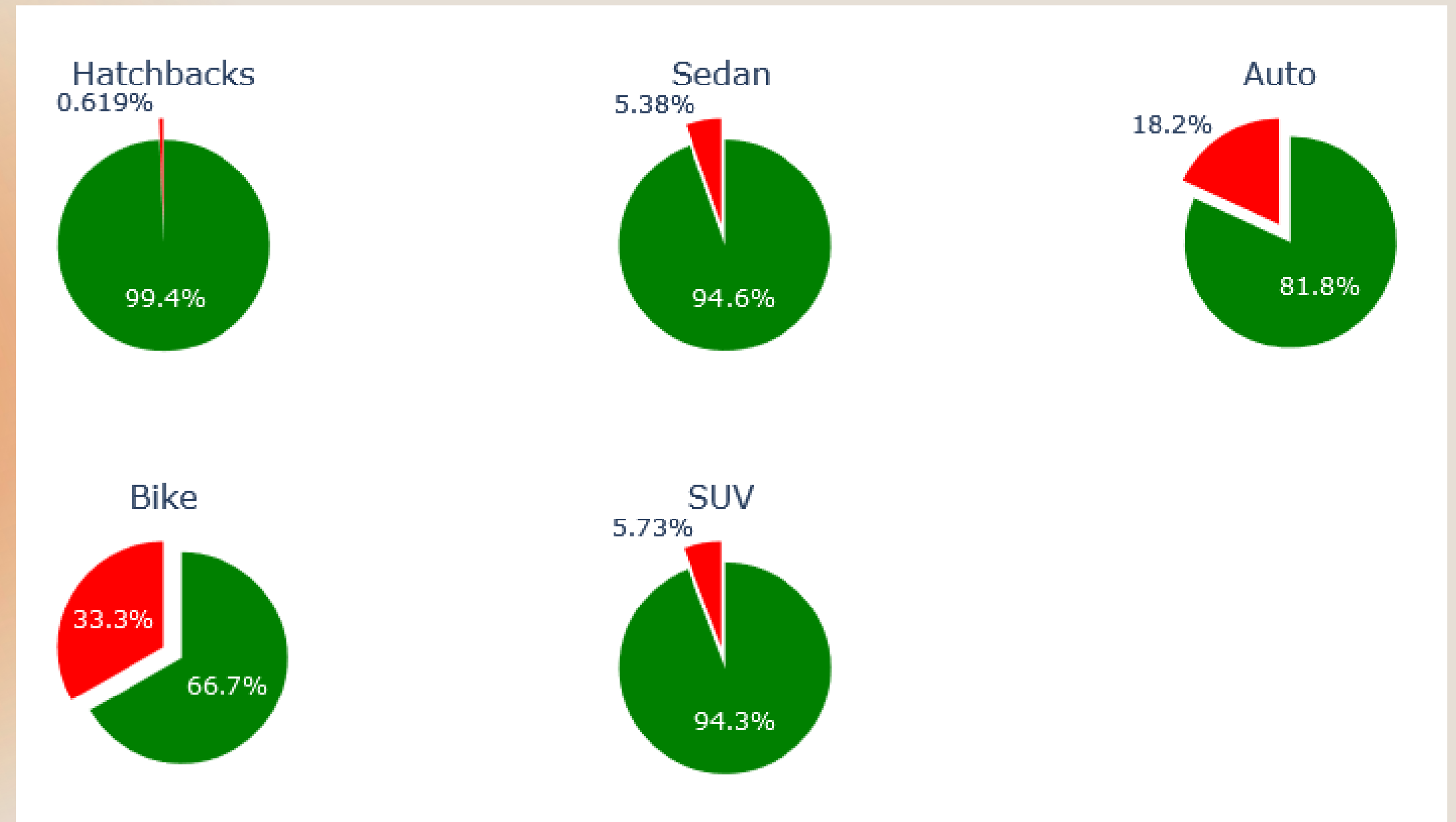
# Cab Availability



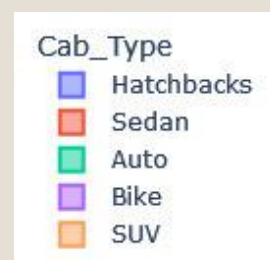
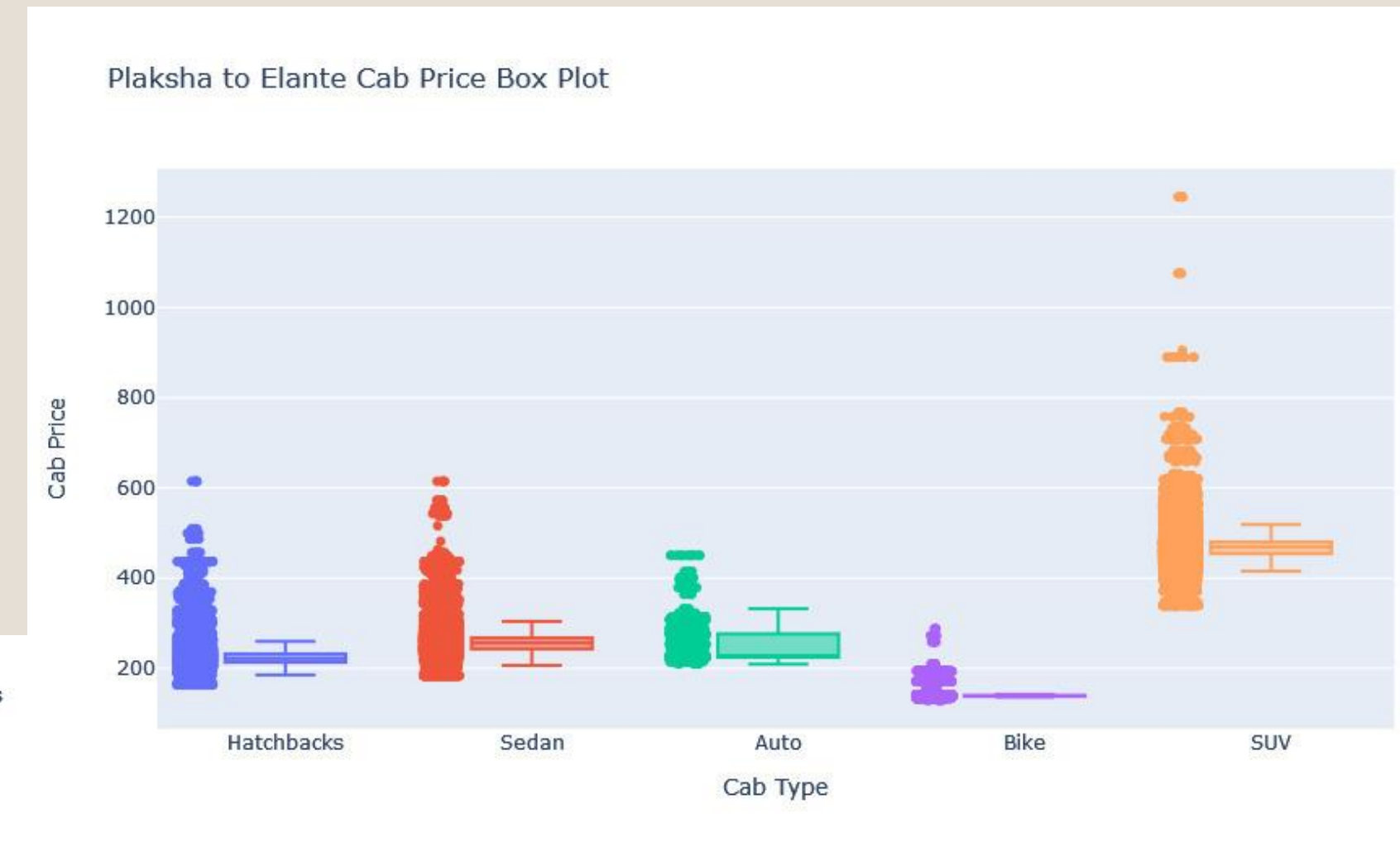
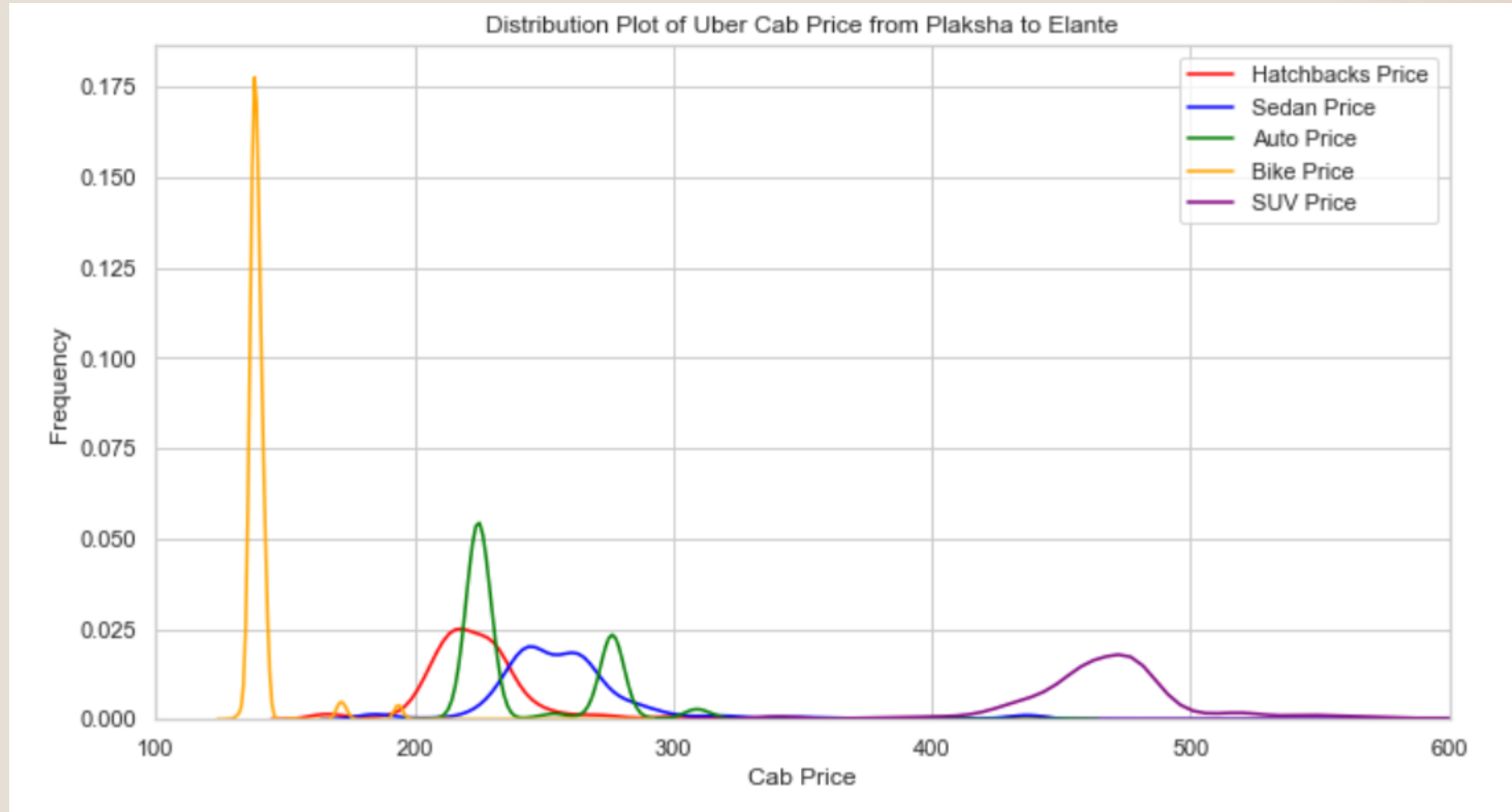
Availability of All Types of Cabs



Availability of Cabs from Plaksha University to Elante Mall



# Cab Price Distribution

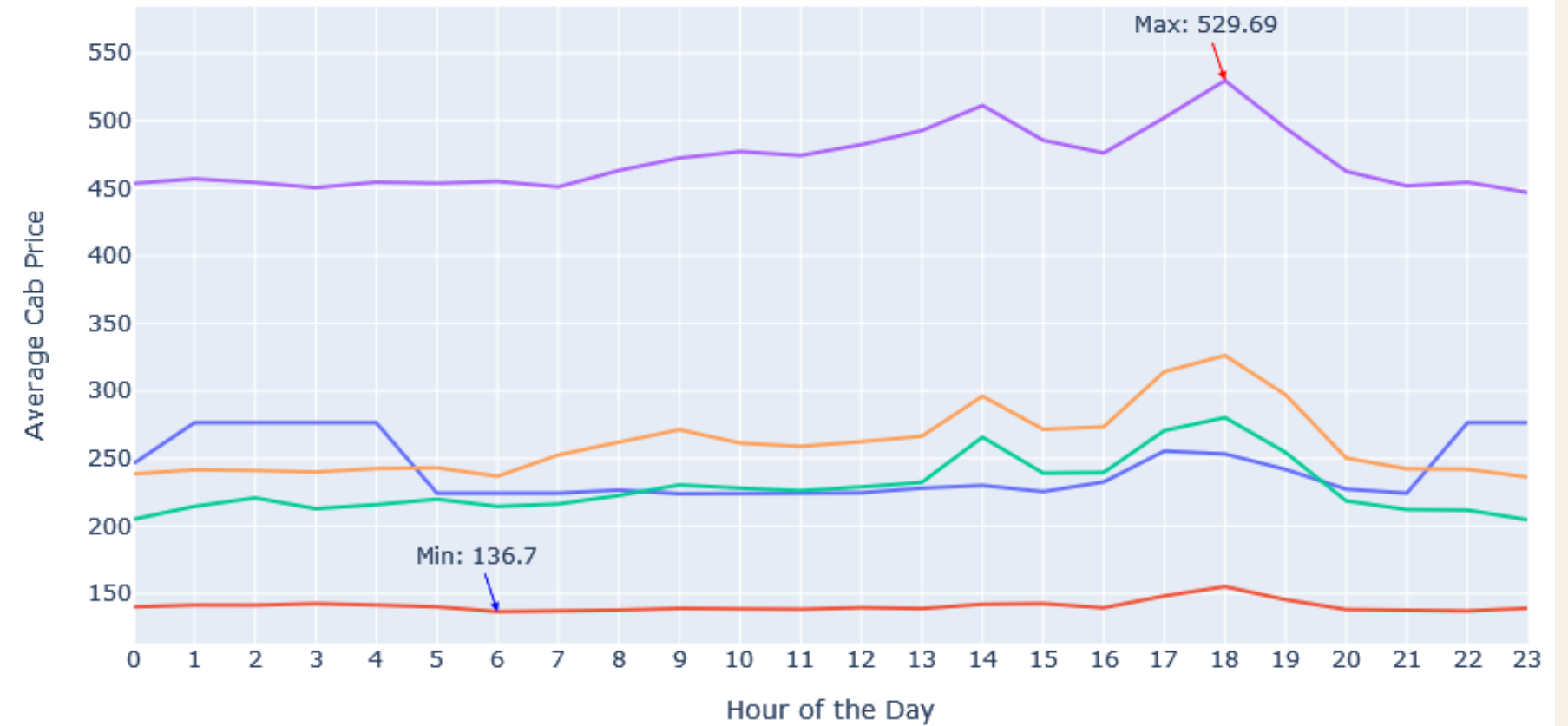


# Hourly Cab Price

Plaksha to Elante Cab Price



Plaksha to Elante Cab Price for each Cab Type



- Cab Type
- Auto
- Bike
- Hatchbacks
- SUV
- Sedan



# ML Methodology

## Cab Price Prediction

### **Linear Regression**

Linear Regression is ideal for predicting cab prices when the connection between factors like distance, time, and traffic conditions and the fare is roughly linear. Its interpretability provides insights into how individual features impact the predicted fare.

### **Random Forest Regressor**

For cab fare prediction with intricate, potentially non-linear relationships among factors, Random Forest Regressor excels at capturing complex dependencies.

### **K-Nearest Neighbors**

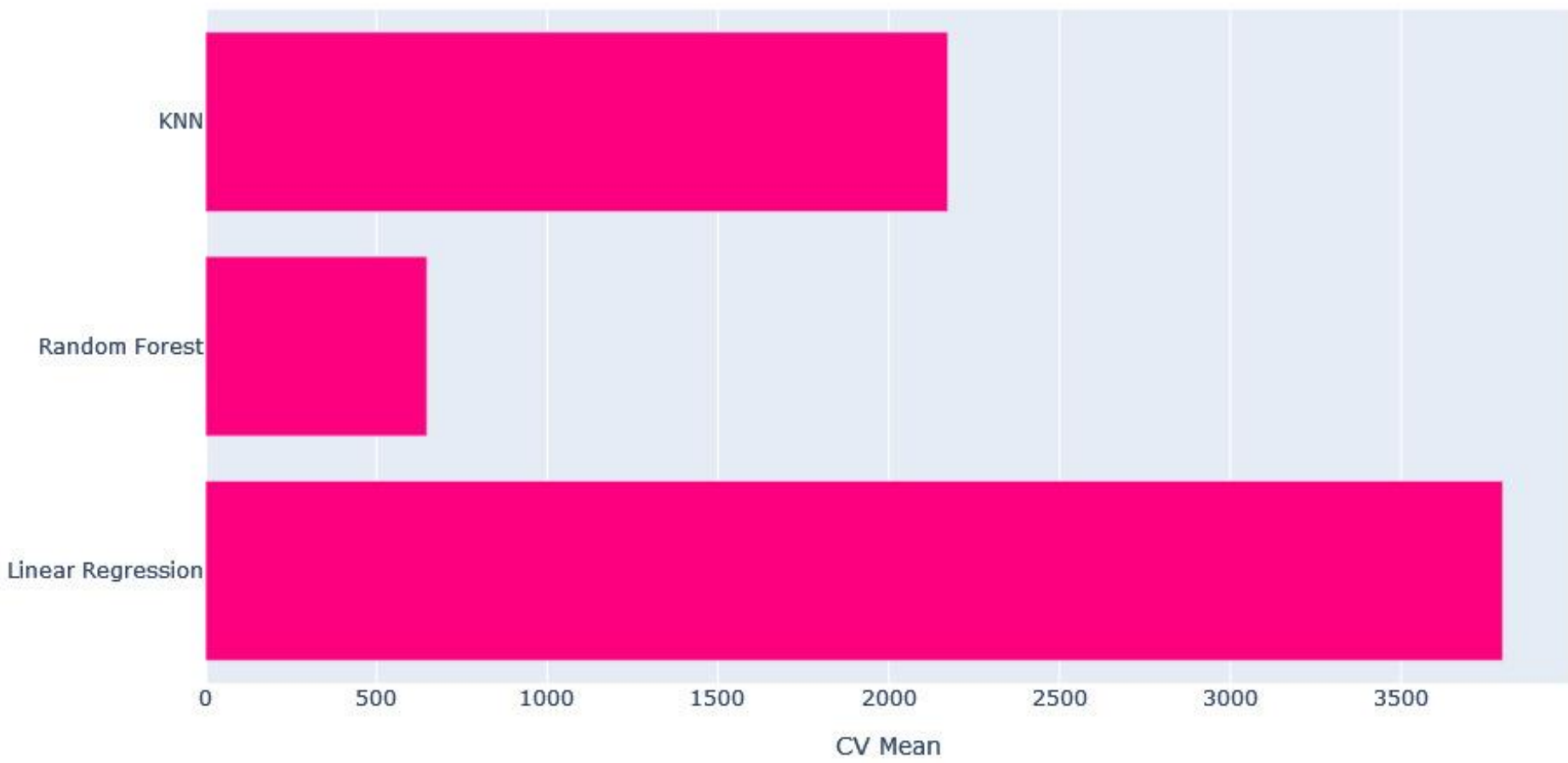
K-Nearest Neighbors is valuable for cab price prediction when there are local pricing patterns, such as neighborhood-specific or time-dependent variations. KNN effectively captures these patterns, adapting to location proximity or temporal factors.

# Performance Metrics

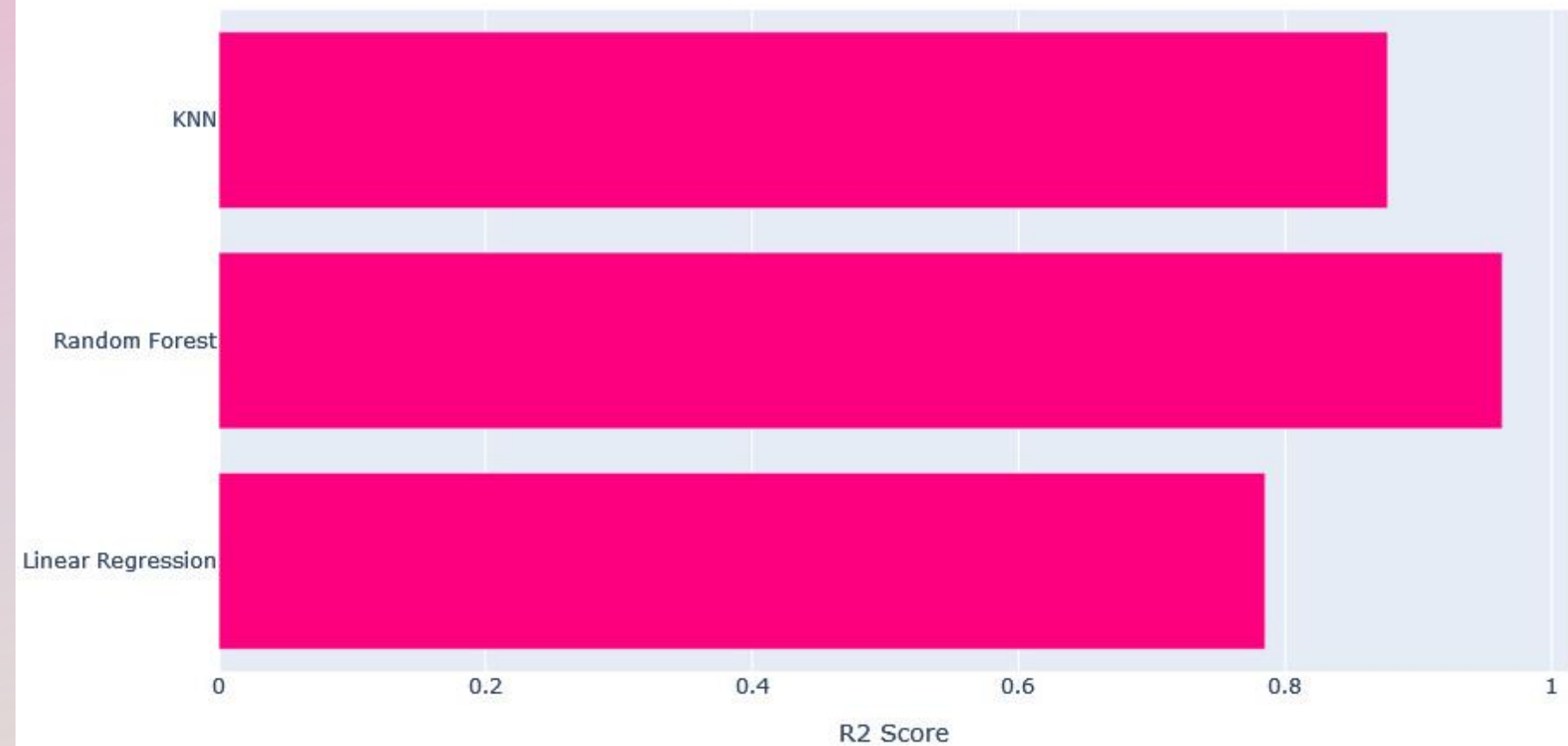
## Cab Price Prediction

	CV Mean	Std	R2 Score
<b>Linear Regression</b>	3797.716902	162.524465	0.784966
<b>Random Forest</b>	656.601318	33.316359	0.962803
<b>KNN</b>	2173.335933	52.001714	0.876900

Average CV Mean Accuracy



Average R2 Score





# ML Methodology

## Route Time Prediction

### **Linear Regression**

Linear Regression is a suitable choice for predicting route time when input features, such as distance, traffic conditions, and time of day, exhibit a linear relationship.

### **Random Forest Regressor**

Similar to cab fare prediction, Random Forest Regressor is advantageous for route time prediction when intricate, non-linear relationships exist between input features and route duration.

### **K-Nearest Neighbors**

K Nearest Neighbors is useful when predicting route time, especially in scenarios where the travel time varies based on local patterns.

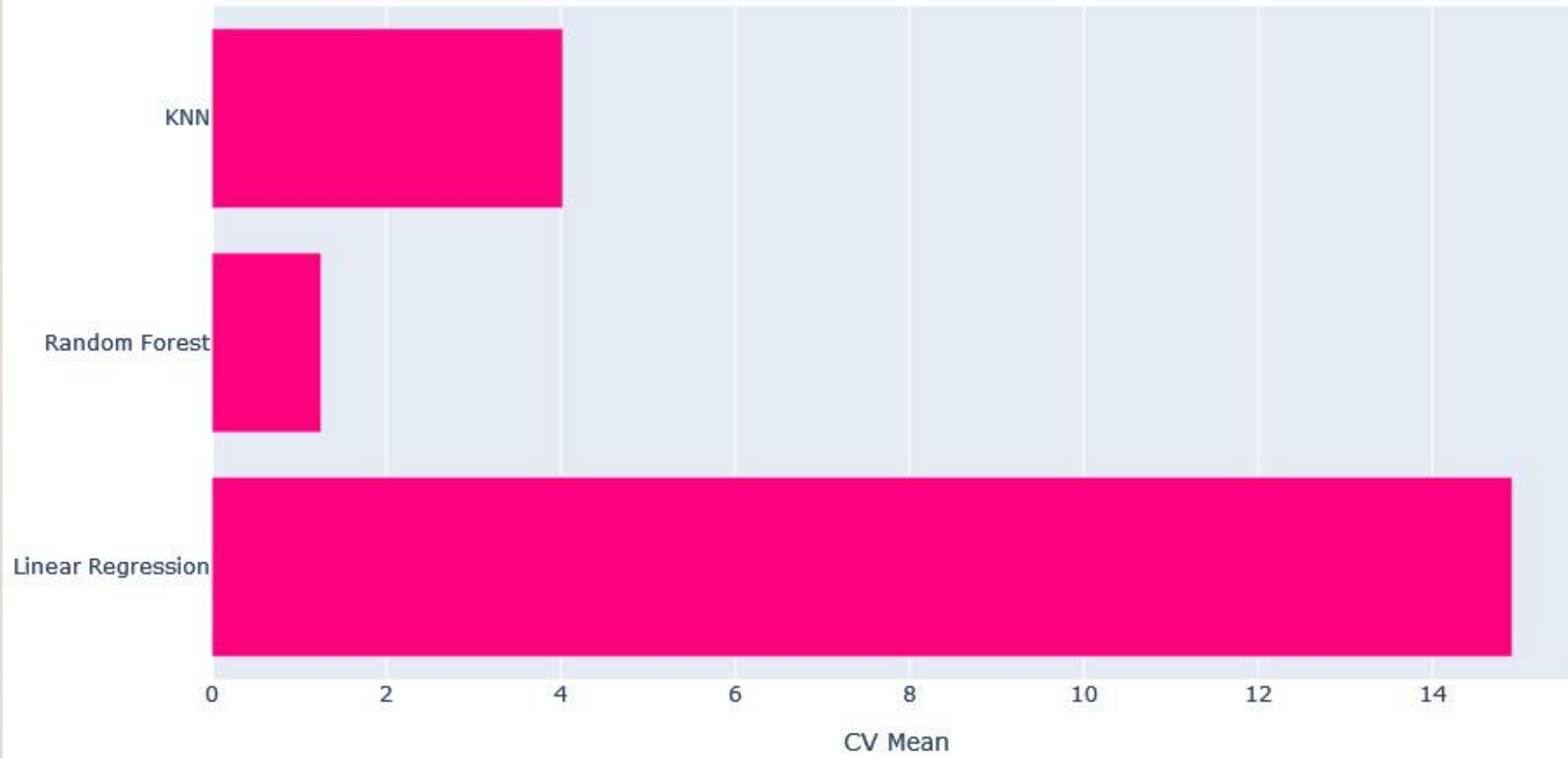


# Performance Metrics

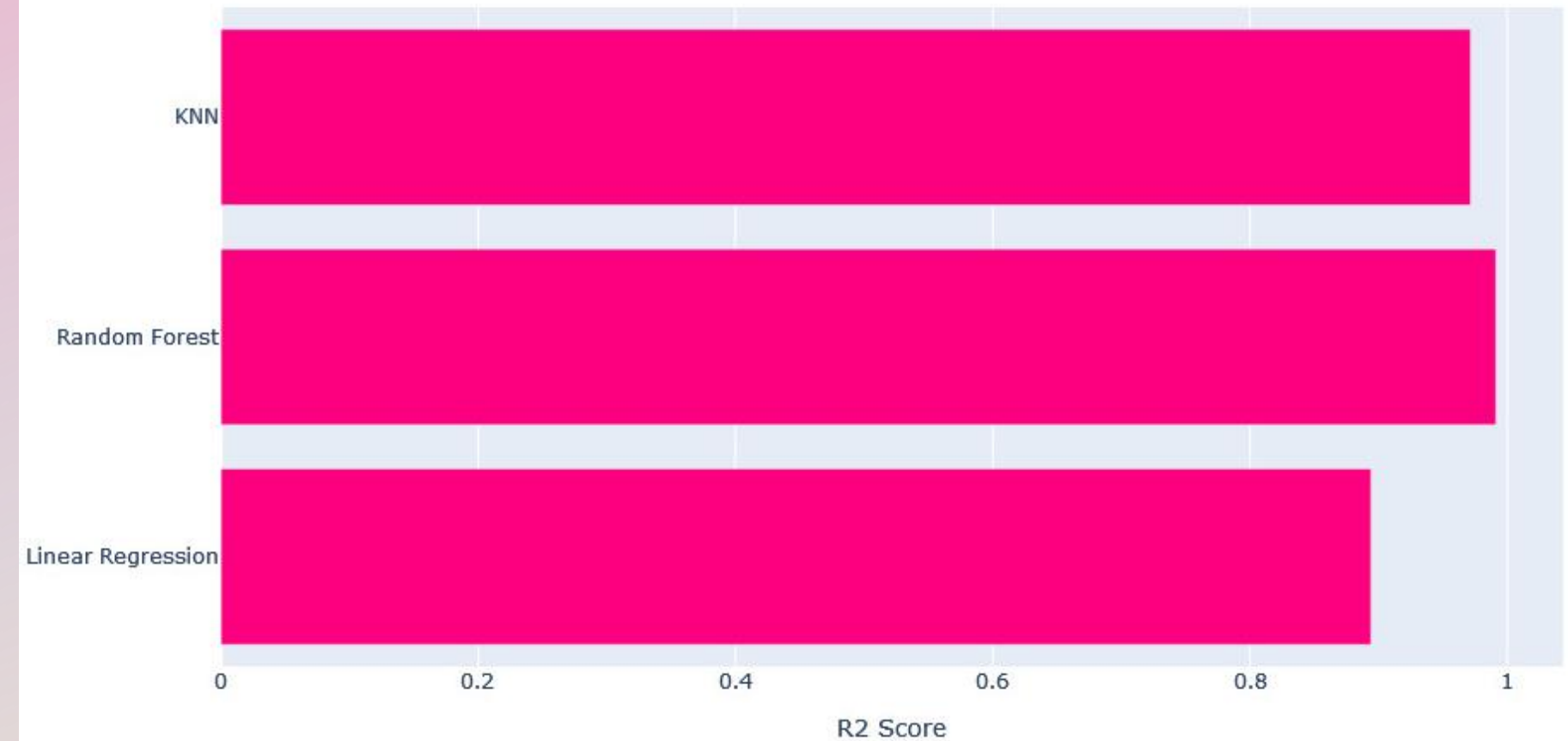
## Route Time Prediction

	CV Mean	Std	R2 Score
<b>Linear Regression</b>	14.913452	0.113069	0.893972
<b>Random Forest</b>	1.251514	0.026682	0.991102
<b>KNN</b>	4.022650	0.113075	0.971400

Average CV Mean Accuracy



Average R2 Score



# ML Modeling Obstacles



## **Persistent Web Scraping**

A laptop was relentlessly employed in the continuous extraction of data from the internet through web scraping



## **Limited Storage Capacity**

We continuously created new Google Sheets after reaching the 250,000-row limit in each, to accommodate our extensive data storage needs



## **Time Intensive Processing**

Training our machine learning model proved to be a time-consuming task, particularly due to the substantial size of our dataset

# Deployment Hurdles

## Deploy

### **Data Volume and Management:**

Handling larger volumes of data effectively, including storage and processing, becomes more complex as the scale grows.

### **Real-Time Data Processing:**

As your service scales, the need for real-time data processing increases, which can be technically demanding.

### **Changing Data Patterns:**

Over time, patterns in cab usage, pricing, and traffic conditions may change, requiring continuous updates to the model to maintain accuracy.



Thank you!  
Got questions?

