



SC1015: Mini Project

Estimating Food Delivery

A127 - Team 6:
Teo Xuan Ming
Lim Yan Xuan
Pascalis Pandey

Content

1

Introduction

2

Data Cleaning

3

EDA

4

Machine Learning

5

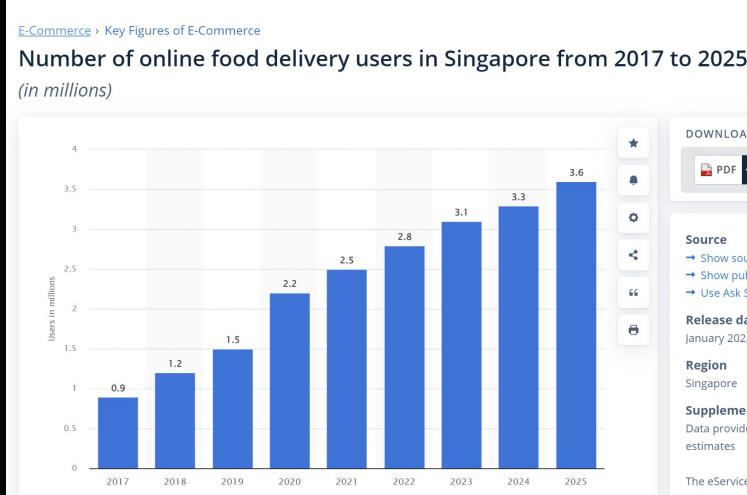
Conclusion



01

Introduction

Introduction



Delivery services see spike in business because of COVID-19



Introduction

Home > India > 48% people unhappy with online food delivery platforms' grievance redressal, survey finds

India

48% people unhappy with online food delivery platforms' grievance redressal, survey finds

According to a survey by LocalCircles, 88% people want govt to implement 'stringent food packaging standards', while 70% say high fees and surcharges remain key concerns.

TARAN DEOL 23 September, 2021 11:46 am IST

South China Morning Post
https://www.scmp.com › ... › Health & Environment

Fivefold surge in complaints against food delivery apps ... ✓

15 Jul 2020 — Most customers seeking Consumer Council's help **unhappy** with issues ranging from wrong orders to poor service quality.

AsiaOne
https://www.asiaone.com › singapore › citizen-singapo...

'Citizen of Singapore': Food delivery rider blurts out after ... ✓

7 Mar 2023 — A man got into an argument with a **food delivery** rider in Toa Payoh where both declared they ... The man appeared to be **unhappy** with where the.

The New York Times
https://www.nytimes.com › food-delivery-apps

Why Restaurants Are Fed Up With Apps ✓

10 Jun 2020 — More people are using **food delivery** apps like Grubhub and Uber Eats to order ... But can this trend last if so many restaurants are **unhappy**?

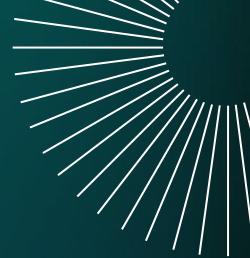
- The Hindu**
https://www.thehindu.com › News › Cities › Bengaluru
- Unhappy with new payment system, food delivery partners ... ✓**
- 17 Apr 2022 — At HSR Layout, a section of the **food delivery** partners logged off from the mobile app and held a protest, saying the new system has ...
- Mothership.SG**
https://mothership.sg › 2020/09 › impatient-grabfood-...
- GrabFood rider told to 'hurry up', replies that he's soaking ... ✓**
- 8 Sept 2020 — One GrabFood **delivery** rider shared his experience with an impatient customer, who was **unhappy** about his **food** being delayed for two hours, ...
- The Guardian**
https://www.theguardian.com › food › feb › i-quit-foo...
- I quit food delivery apps – the absurd convenience was not ... ✓**
- 10 Feb 2021 — Not because I didn't like using it. I loved it. Ordering precisely what you feel like eating, silently and seamlessly, only to have that hot ...
- Moneycontrol**
https://www.moneycontrol.com › current-affairs-trends
- Twitter unhappy with Zomato's 10-minute food delivery ... ✓**
- 22 Mar 2022 — Zomato Instant will be launched in Gurugram next month with four stations, founder Deepinder Goyal announced adding that the company will ...



Dataset:

Determining estimated time of food delivery



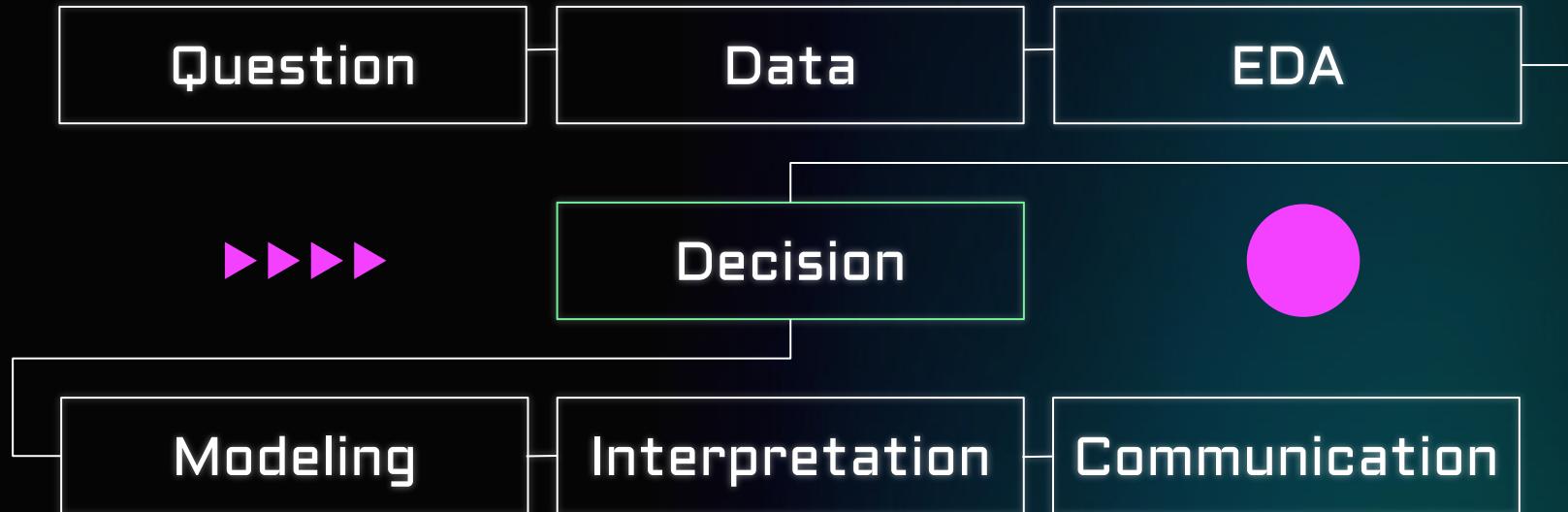


PROBLEM DEFINITION

Can we estimate the delivery time based on the delivery rider's attributes?



DATA PROJECT ARCHITECTURE



02

Data Cleansing

Data Retrieval & Preparation

Data Cleansing

- Typecasting columns
- Fill NaNs
- Outliers correction
- Columns renaming

Retrieval

Kaggle Dataset



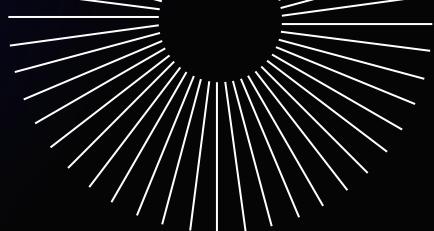
Export

Clean CSV exported
for future use

EDA



Data Cleansing



Typecast

Object -> int64

Stripped string along to obtain numeric value in “Time_taken” column.

Some float columns casted to int64 for seaborn charts.

NaNs

Hours spent

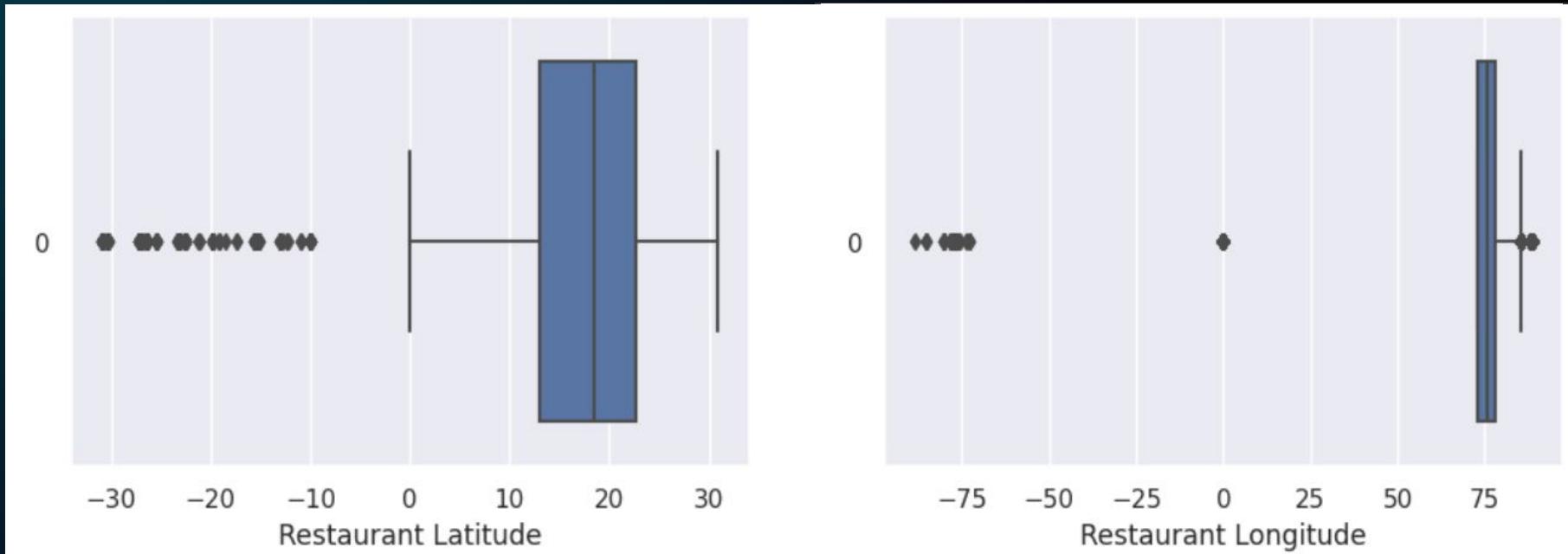
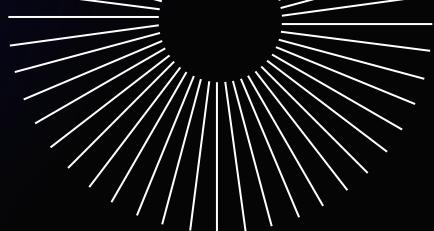
Most categorical columns’ NaN values filled with mode

“Time_taken” NaNs was dropped as it is the response variable

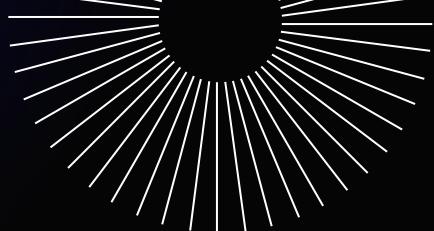
Dropped Lat/Lon 0 values



Data Cleansing



Data Cleansing



Renaming Columns



CamelCasing convention

Stripping of unnecessary
whitespaces in variable values

Exporting

Exported cleaned data to a new CSV
and pushed it to a repo for use in
Google Colab



03

EDA

Exploratory Data Analysis

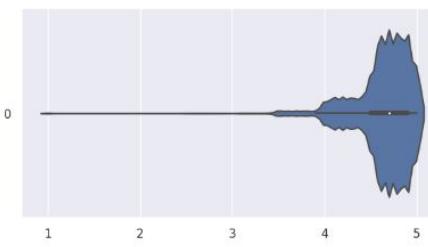
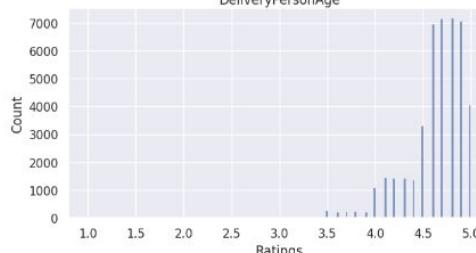
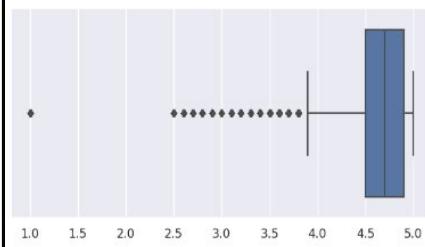
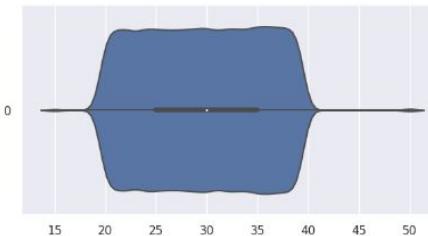
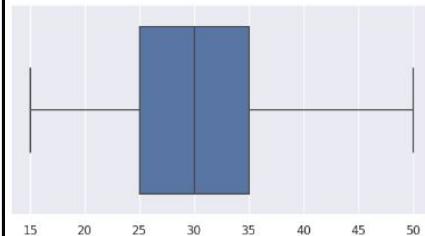
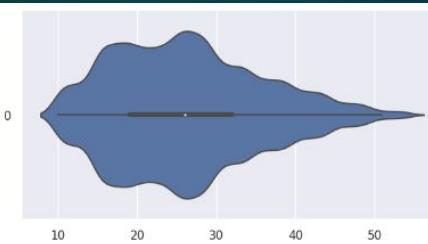
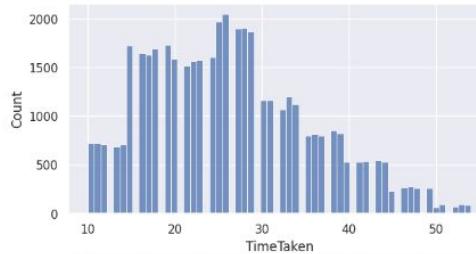
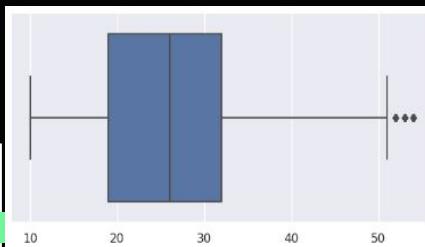


Numeric Columns

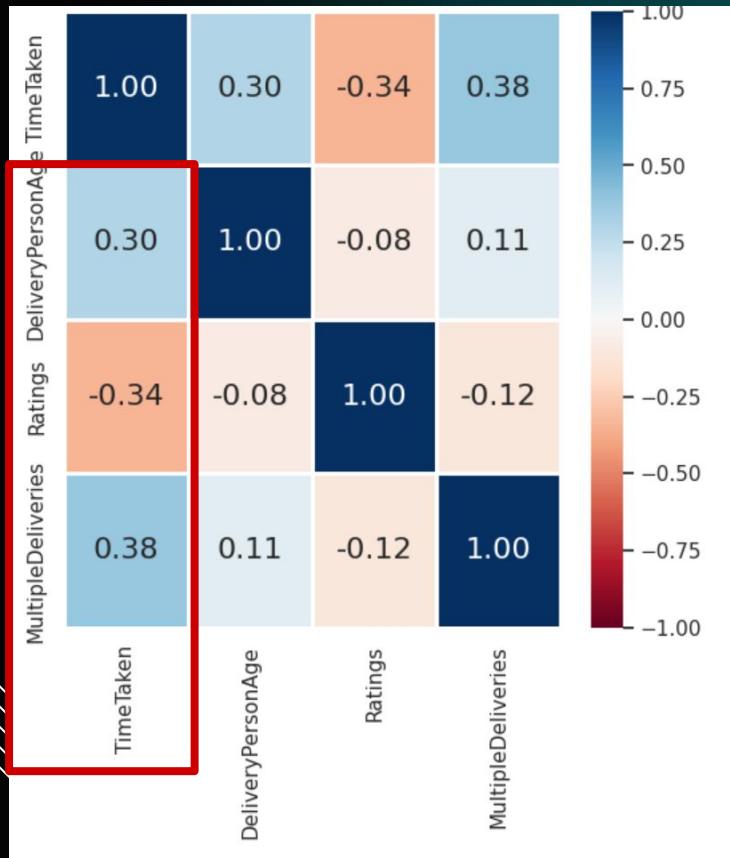
Exploratory Data Analysis



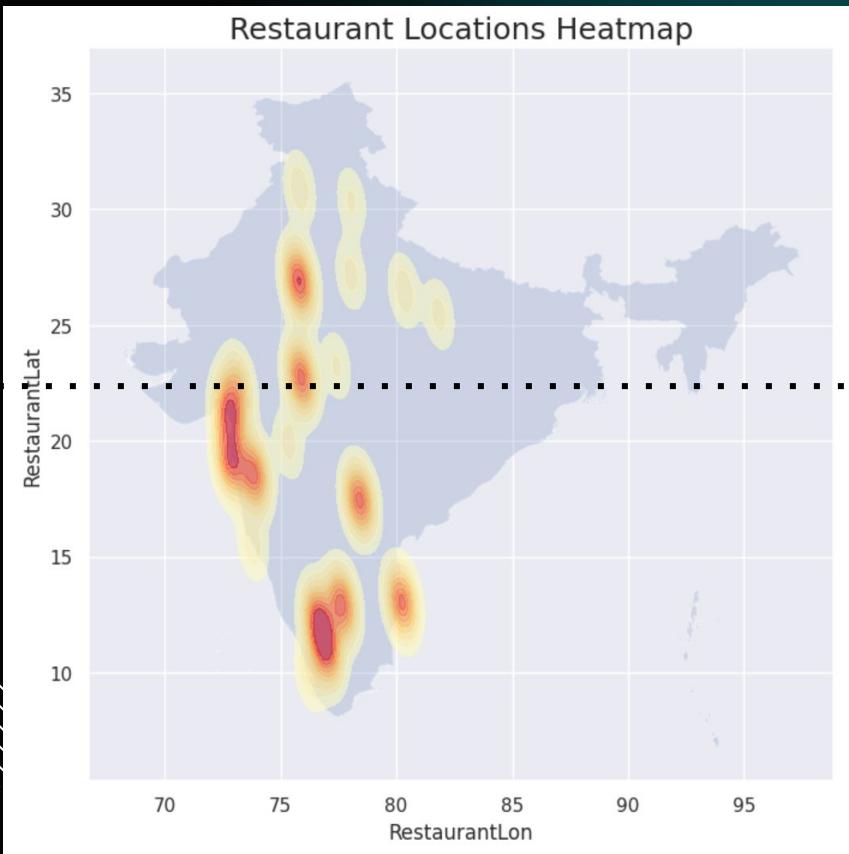
Distribution of Numeric Columns



Correlation of Numeric Columns



Distribution of Location Geographically



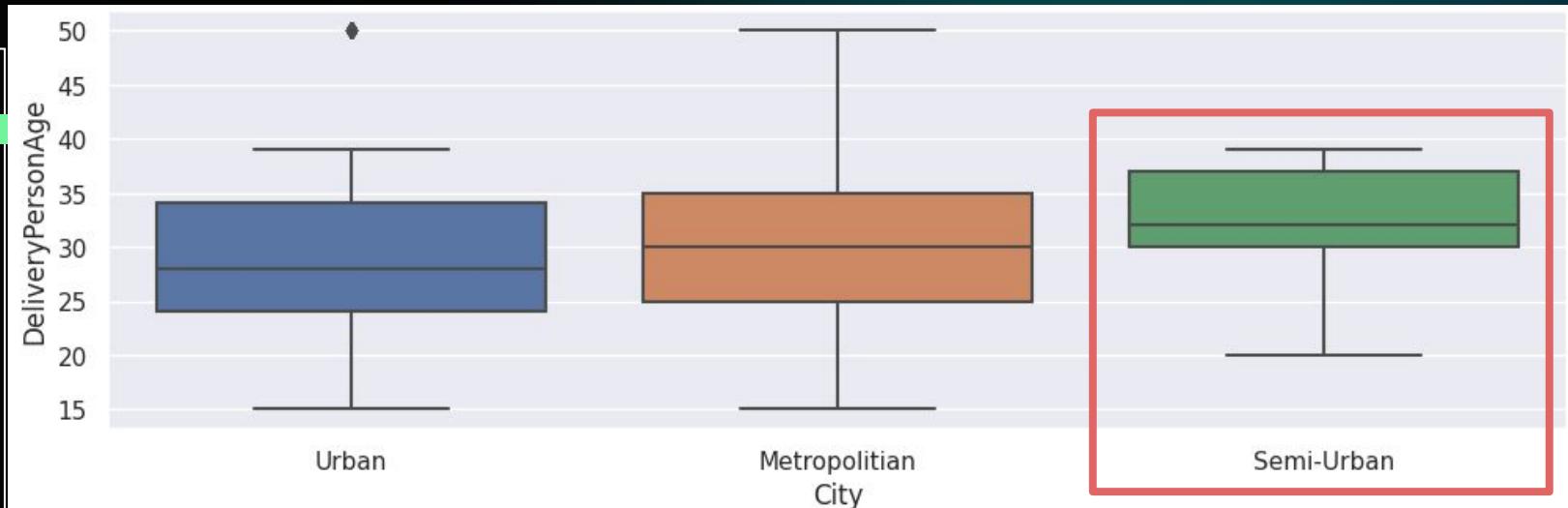
Categorical Columns



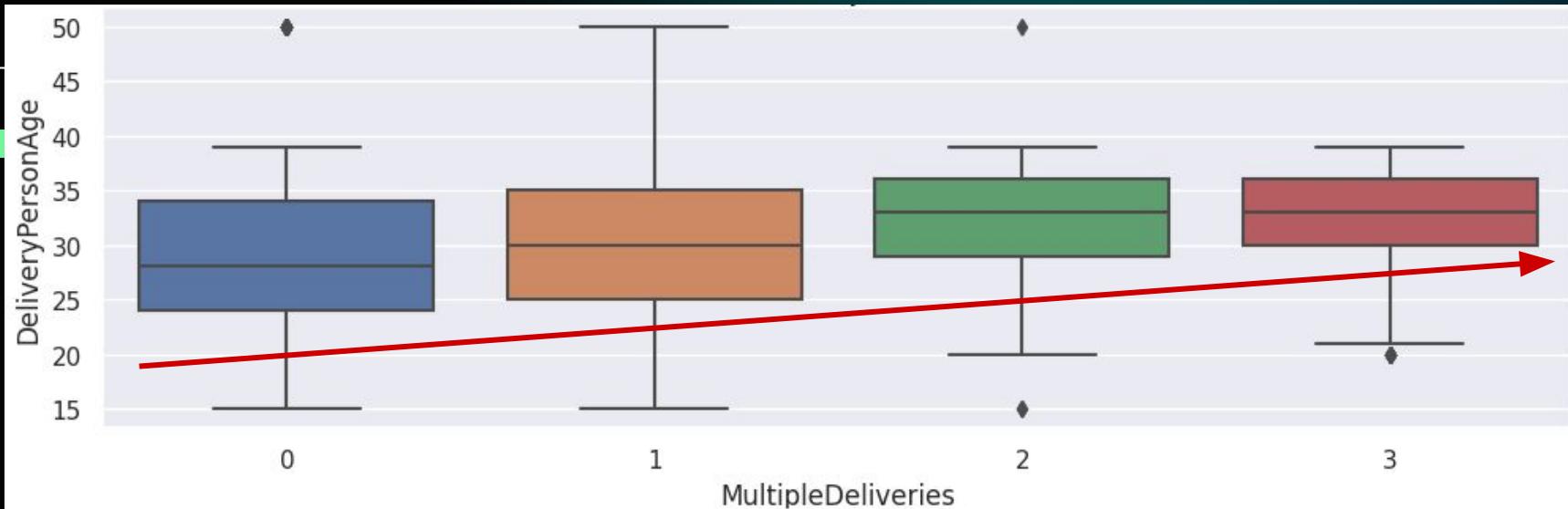
Exploratory Data Analysis



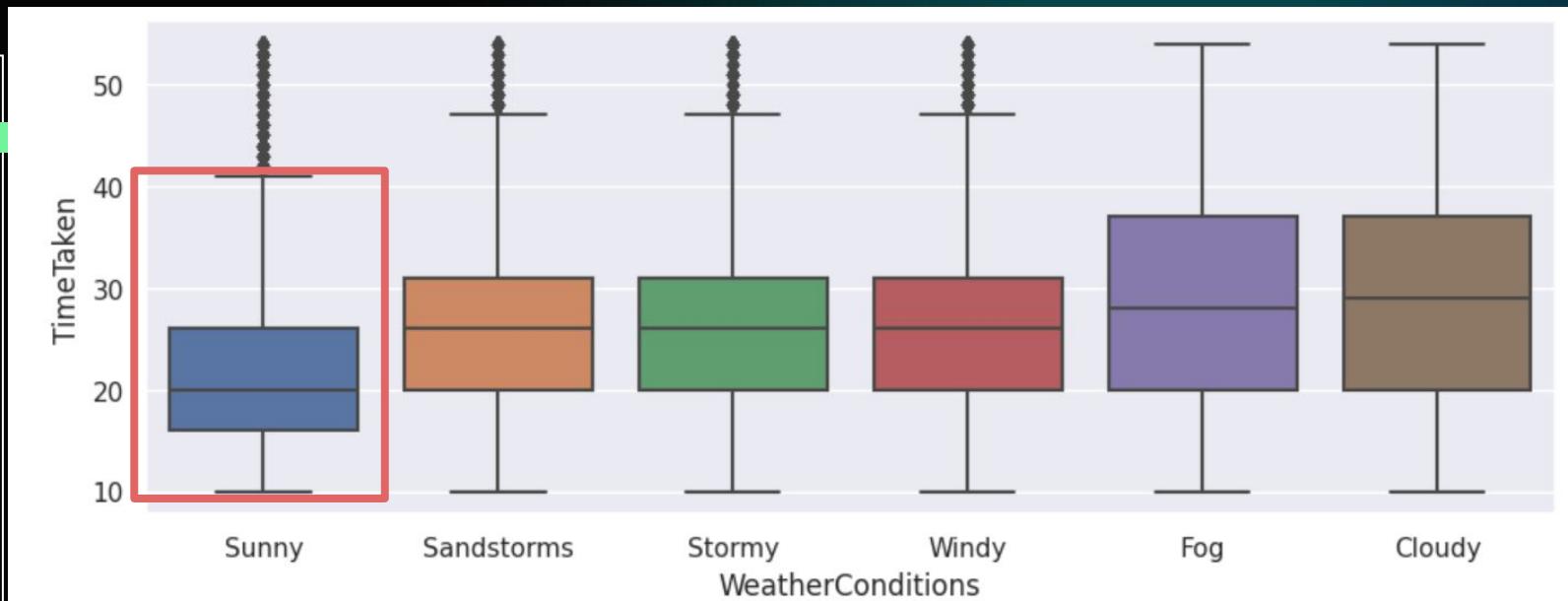
City & Age



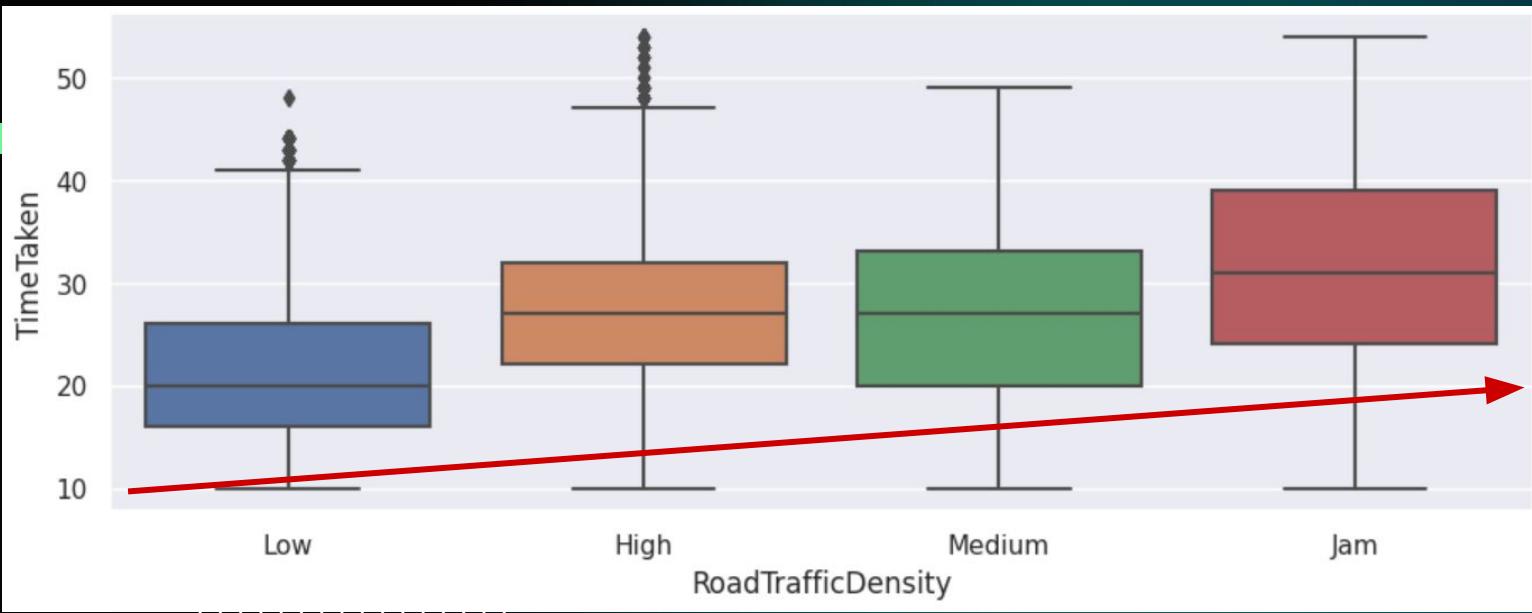
Multiple Deliveries & Age



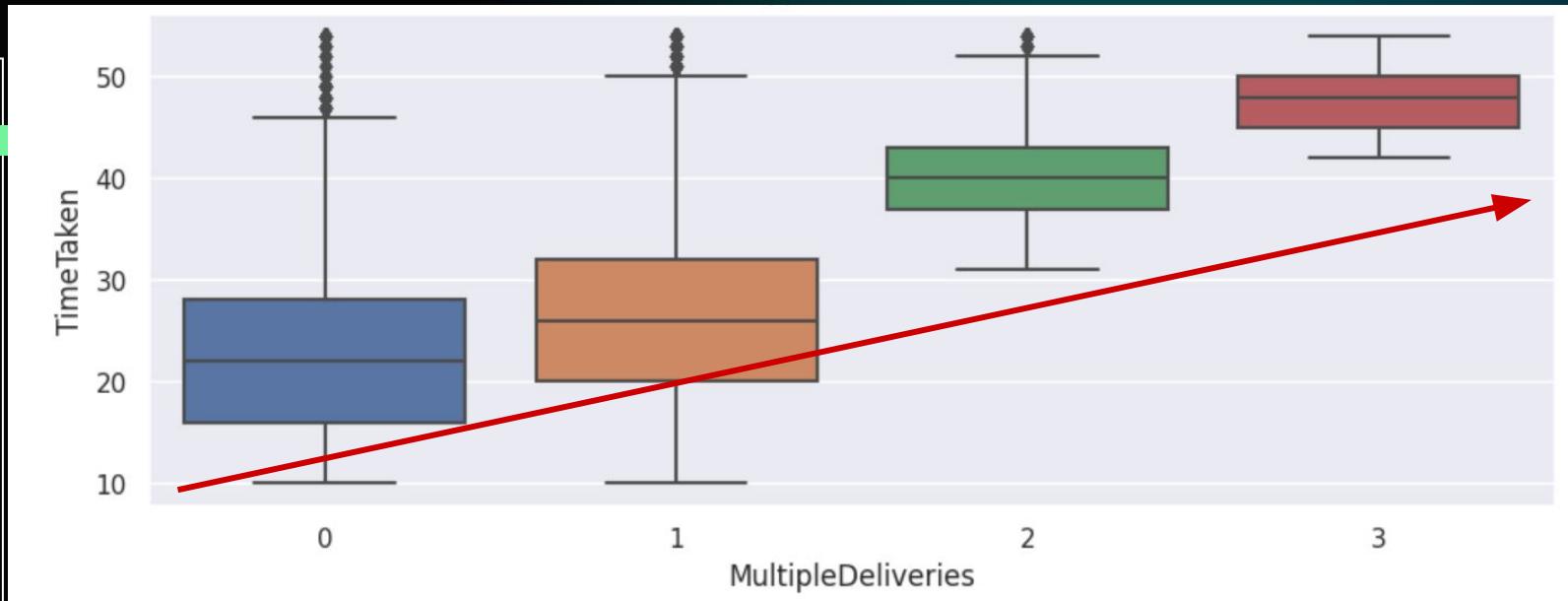
Weather & Delivery Time



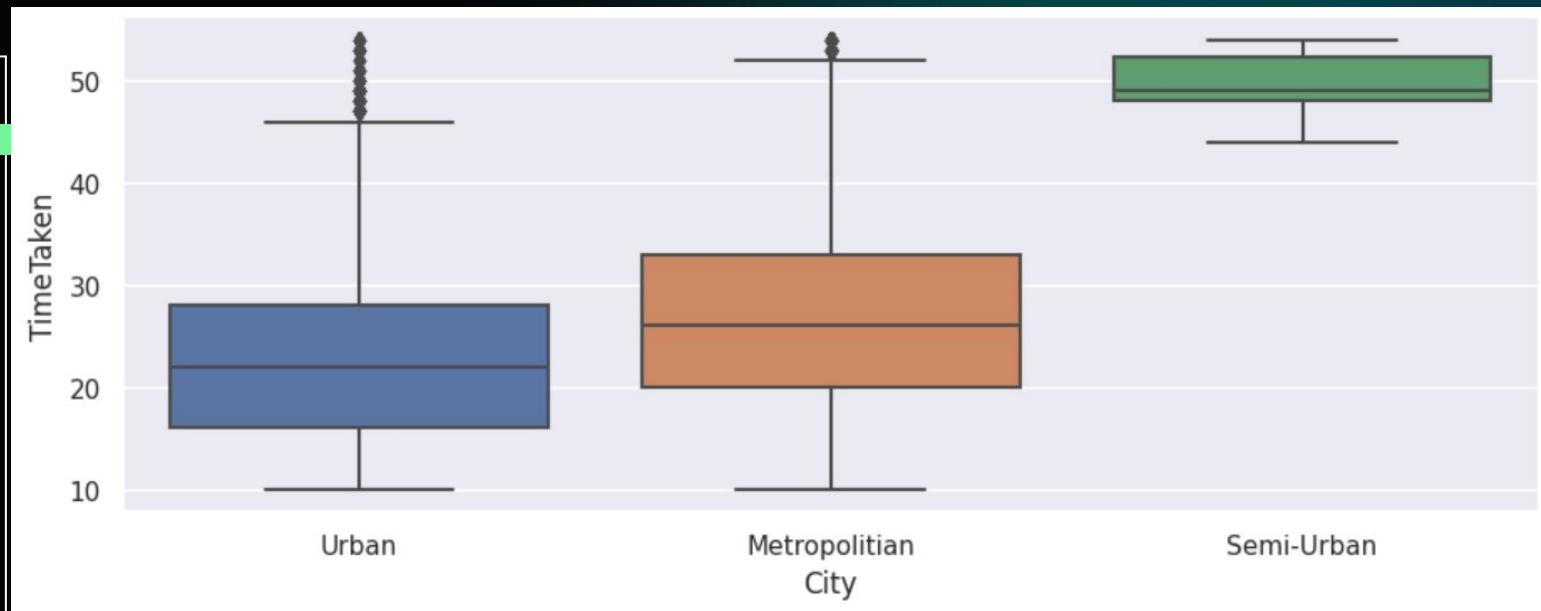
Traffic Density & Delivery Time



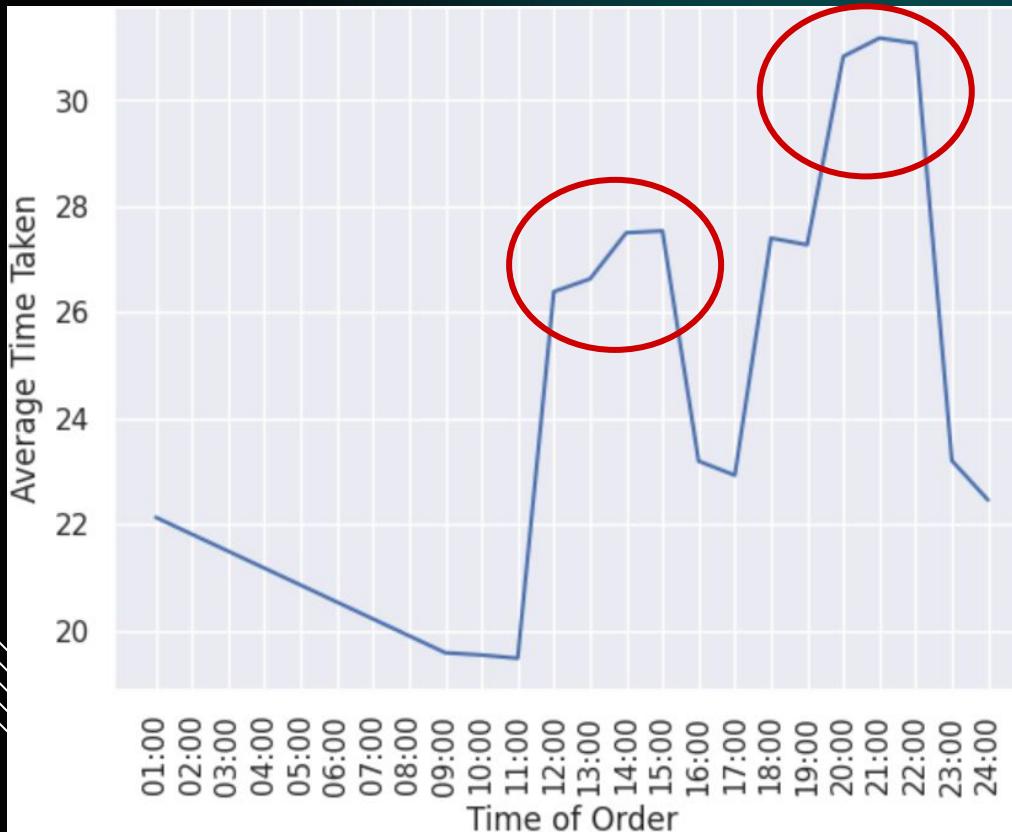
Multiple Deliveries & Delivery Time



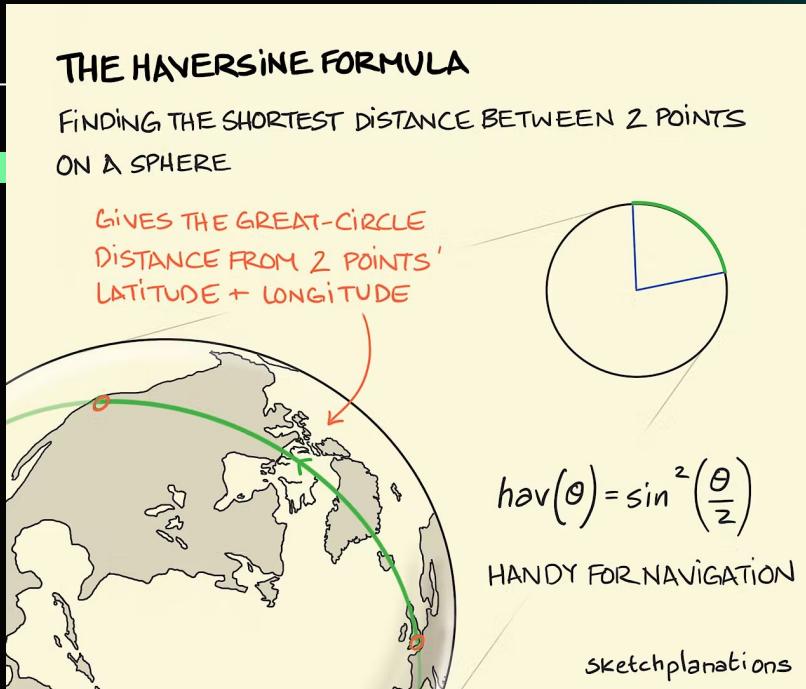
City Terrain & Delivery Time



Average Delivery Time throughout a day



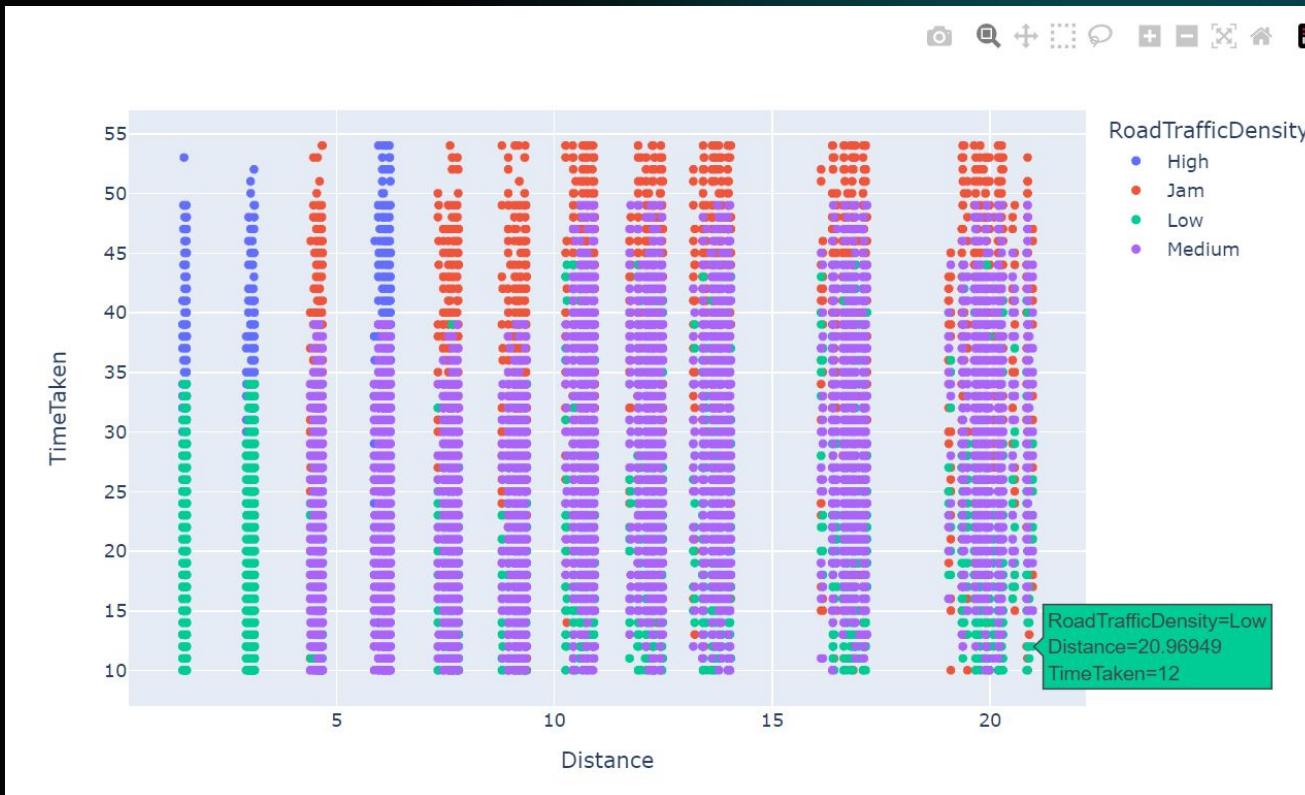
Feature Engineering - Distance



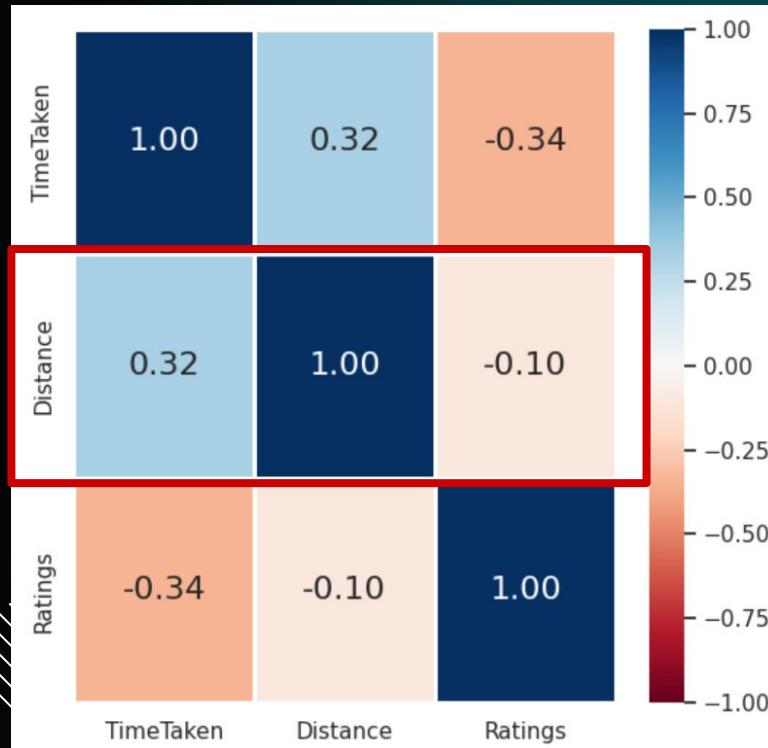
Retrieved from: <https://sketchplanations.com/the-haversine-formula>

```
def haversine(lat1, lon1, lat2, lon2):
    """calculate haversine distance between two co-ordinates"""
    lat1, lon1, lat2, lon2 = map(np.radians, (lat1, lon1, lat2, lon2))
    AVG_EARTH_RADIUS = 6371 # in km
    lat = lat2 - lat1
    lon = lon2 - lon1
    d = np.sin(lat * 0.5) ** 2 + np.cos(lat1) * np.cos(lat2) * np.sin(lon * 0.5) ** 2
    h = 2 * AVG_EARTH_RADIUS * np.arcsin(np.sqrt(d))
    return(h)
```

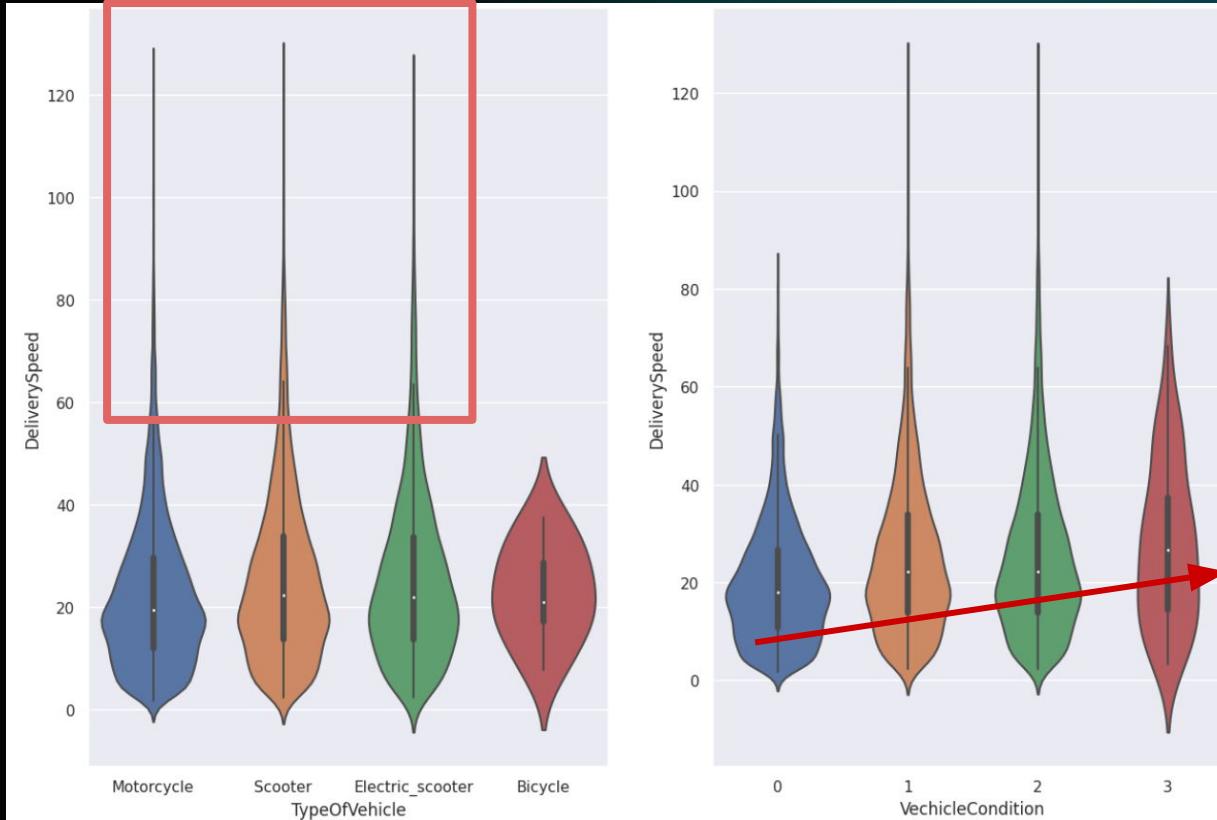
Relationship with Time and Traffic



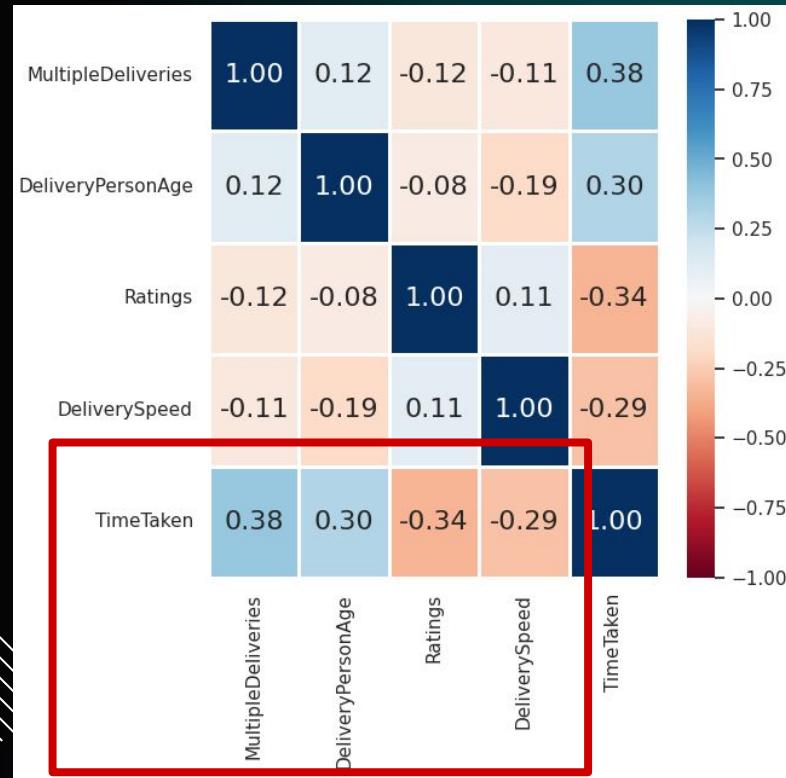
Correlation with Distance



Feature Engineering - Speed



Correlation with Speed

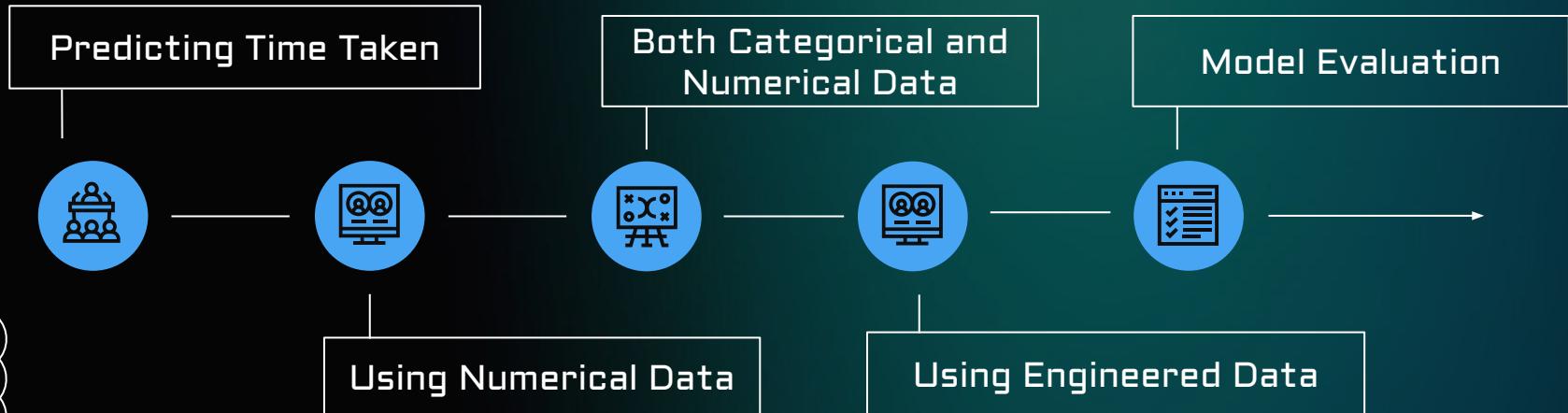


Machine Learning

04

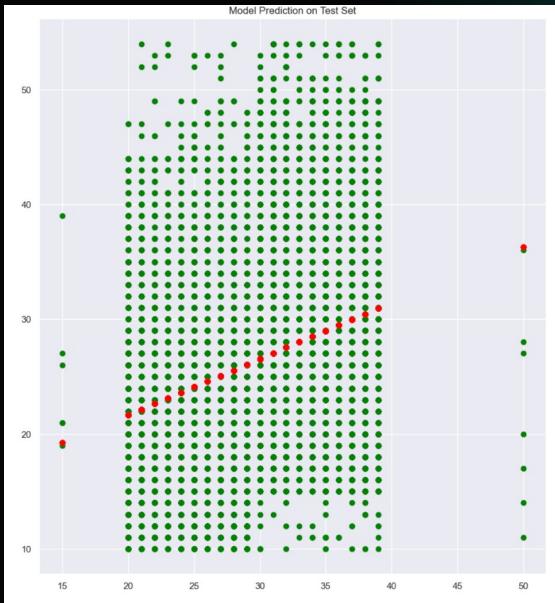


Machine Learning Process



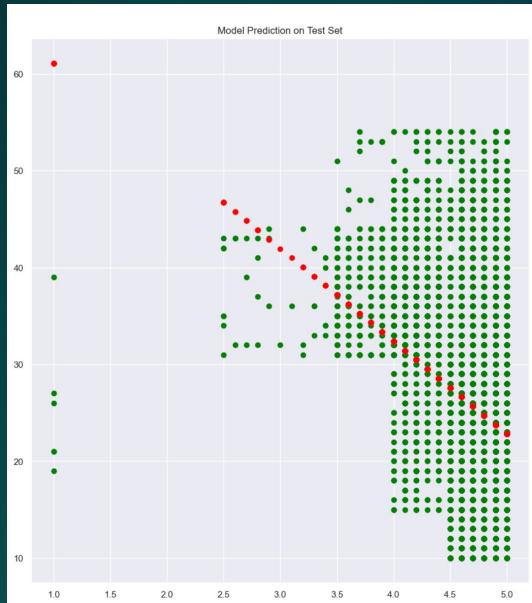
Numerical Data: Univariate LinReg

Using DeliveryPersonAge



	R ² Score	MSE
Train Set	0.116175	77.657425
Test Set	0.116581	77.225152

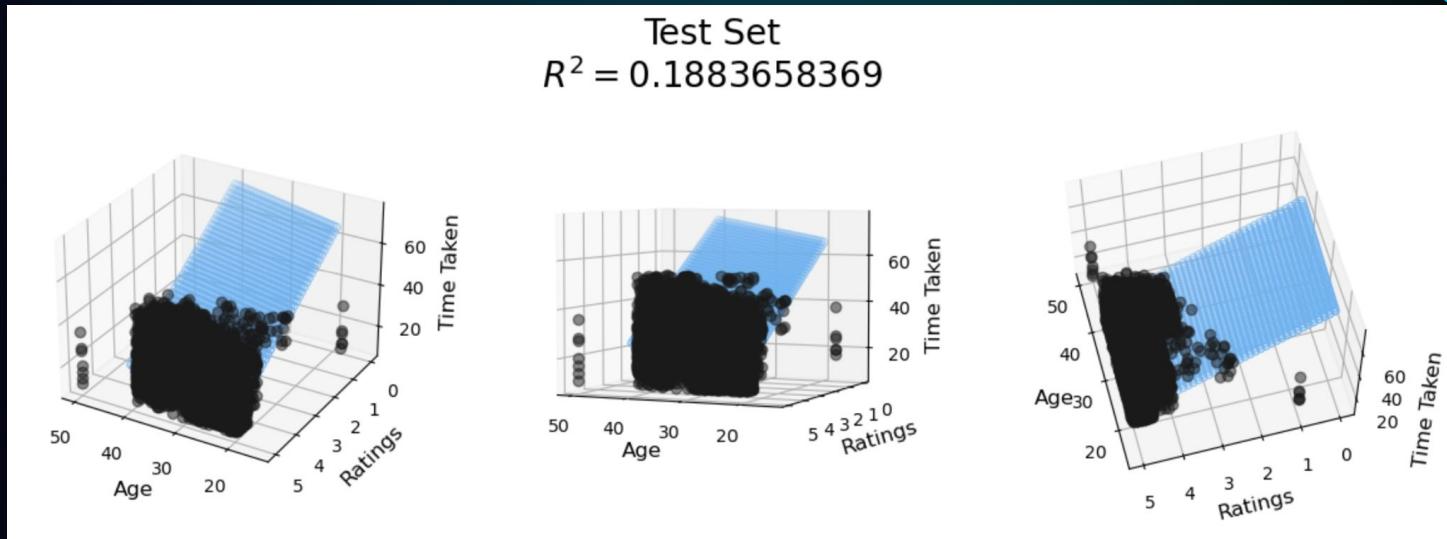
Using Ratings



	R ² Score	MSE
Train Set	0.116175	77.657425
Test Set	0.116581	77.225152



Multivariate LinReg



	R ² Score	MSE
Train Set	0.191209	71.064530
Test Set	0.187739	71.004839





Adding Categorical Data: ANOVA Test & OH Encoding



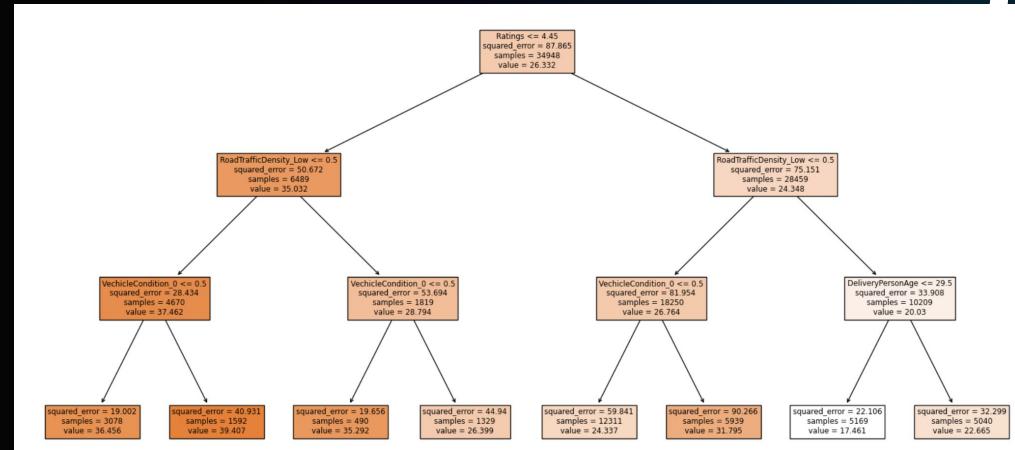
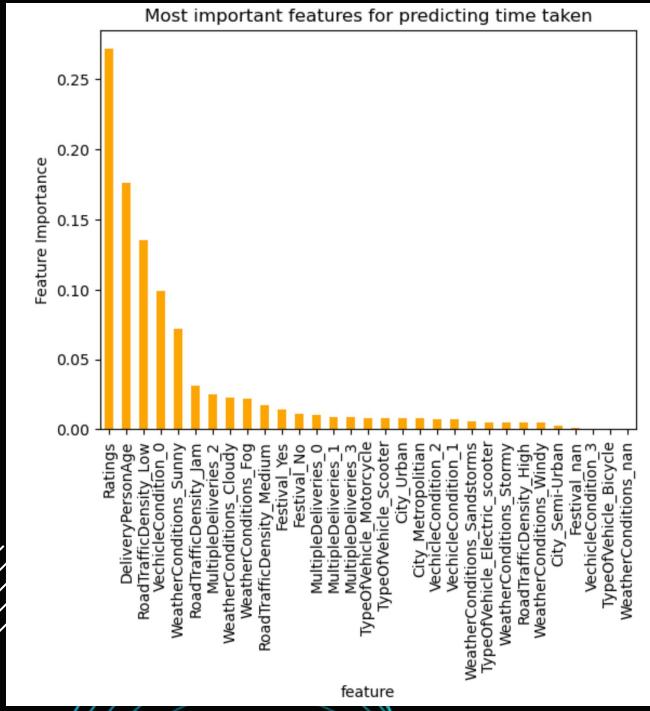
	F-Value	P-Value
WeatherConditions	594.653620	0.000000e+00
RoadTrafficDensity	3264.776322	0.000000e+00
VechicleCondition	1272.417725	0.000000e+00
TypeOfOrder	1.031479	3.773037e-01
TypeOfVehicle	412.443909	3.121529e-264
MultipleDeliveries	3279.864846	0.000000e+00
City	1283.722912	0.000000e+00
Festival	3997.863317	0.000000e+00

	MultipleDeliveries_0	MultipleDeliveries_1	MultipleDeliveries_2
	1.0	0.0	0.0
	0.0	1.0	0.0
	0.0	1.0	0.0
	0.0	1.0	0.0
	0.0	1.0	0.0





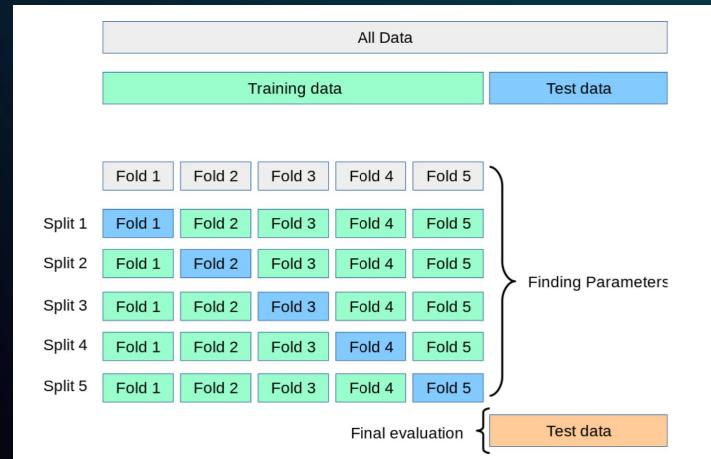
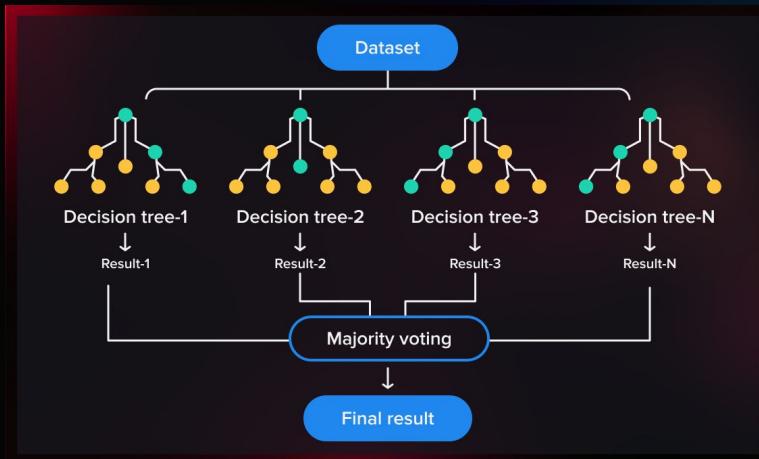
Using Categorical & Numerical: Decision Tree



R² Score **MSE**
Decision Tree 0.427998 50.002259



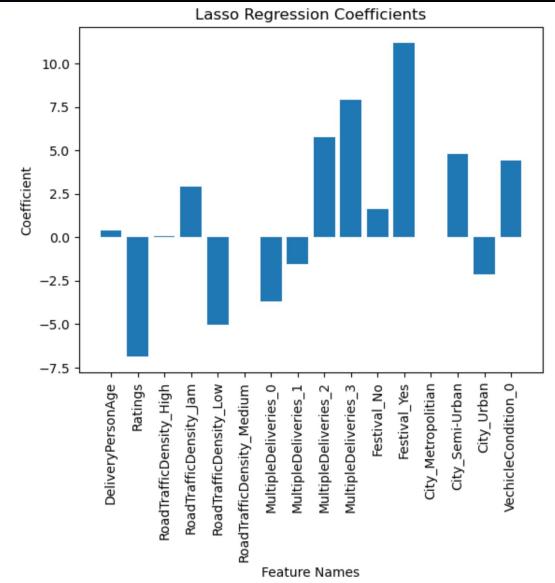
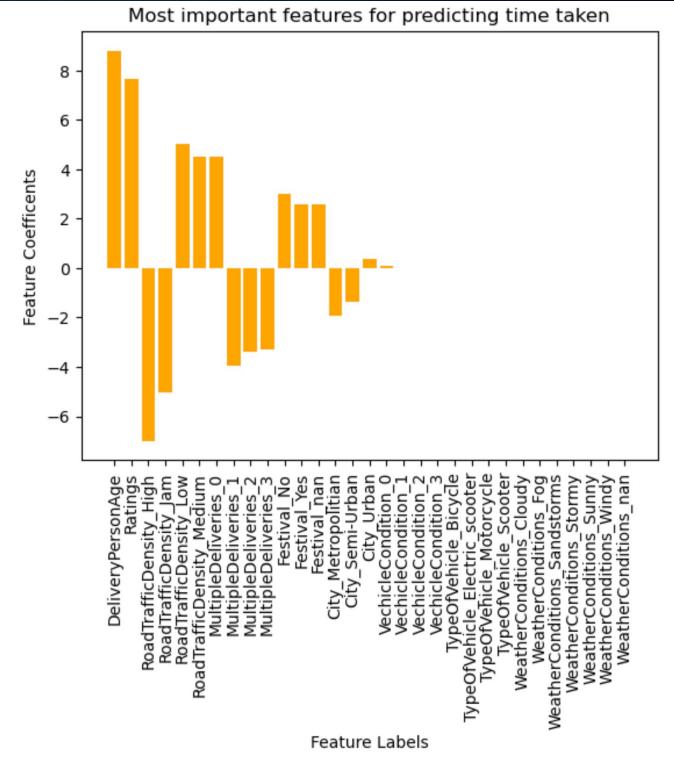
Random Forest



	R ² Score	MSE
Random Forest	0.626945	32.611058

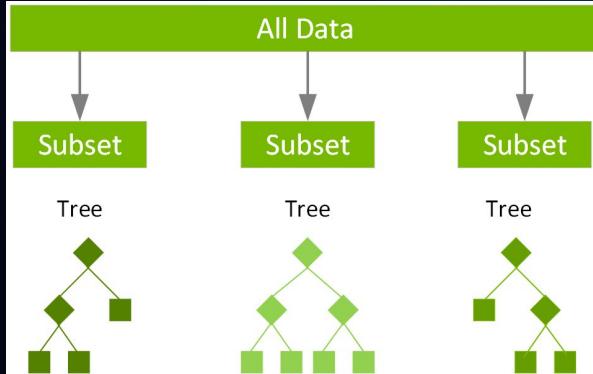
	R ² Score	MSE
Random Forest with CV	0.657132	29.972224

Lasso Regression



	R ² Score	MSE
Lasso Regression	0.531724	40.934936

XGBoost

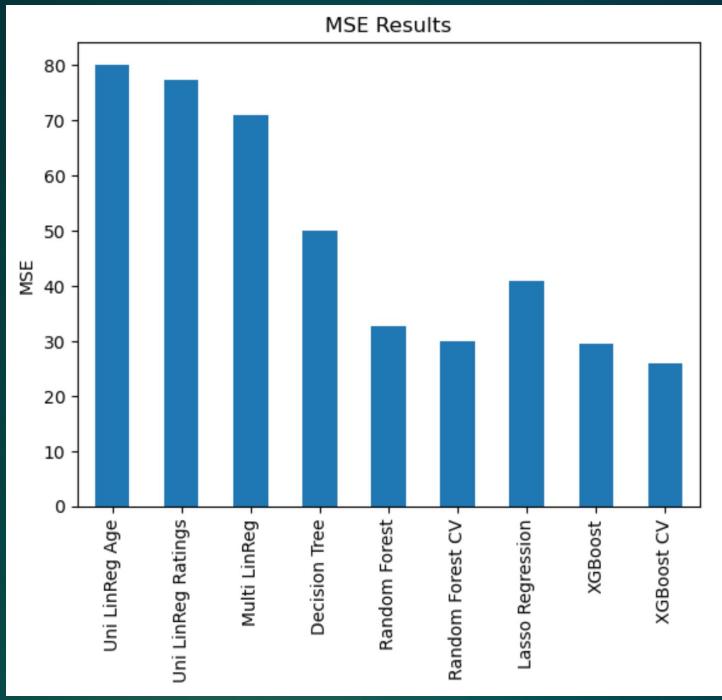
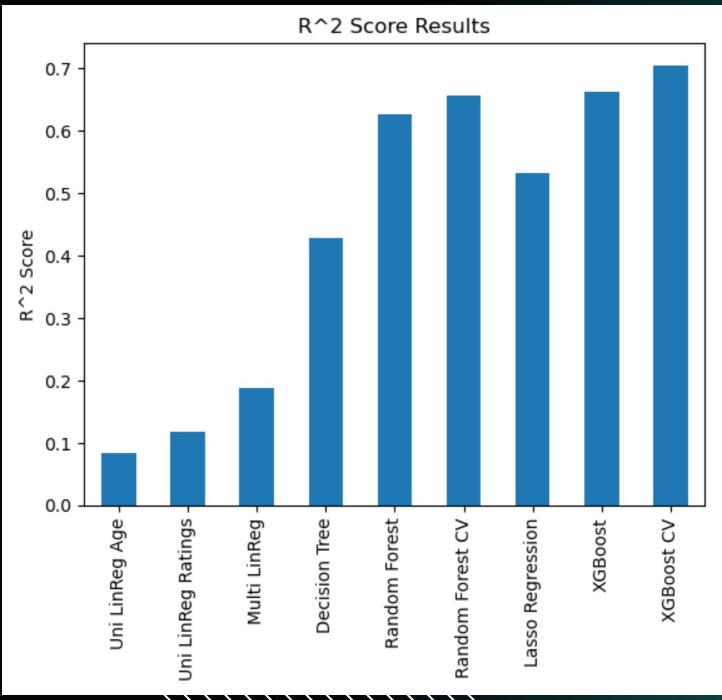


	R ² Score	MSE
Normal XGBoost	0.661872	29.557848

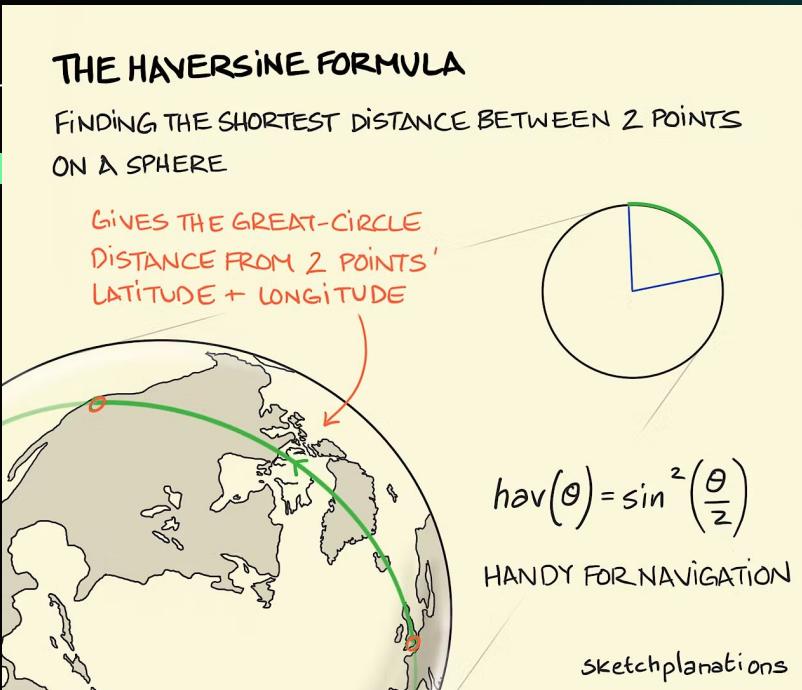
	train-rmse-mean	train-rmse-std	test-rmse-mean	test-rmse-std
0	24.954790	0.013145	24.954997	0.127169
1	22.710379	0.030813	22.711117	0.146320
2	20.690285	0.050261	20.688506	0.157408
3	18.869738	0.049534	18.869297	0.147581
4	17.257230	0.055036	17.254277	0.136107
...
145	4.678109	0.049234	5.011351	0.047971
146	4.676891	0.049189	5.010760	0.048632
147	4.674684	0.049253	5.010112	0.048347
148	4.672793	0.049602	5.009287	0.048028
149	4.671336	0.049648	5.008824	0.047791

	R ² Score	MSE
CV XGBoost	0.705442	25.749111

Results & Evaluation



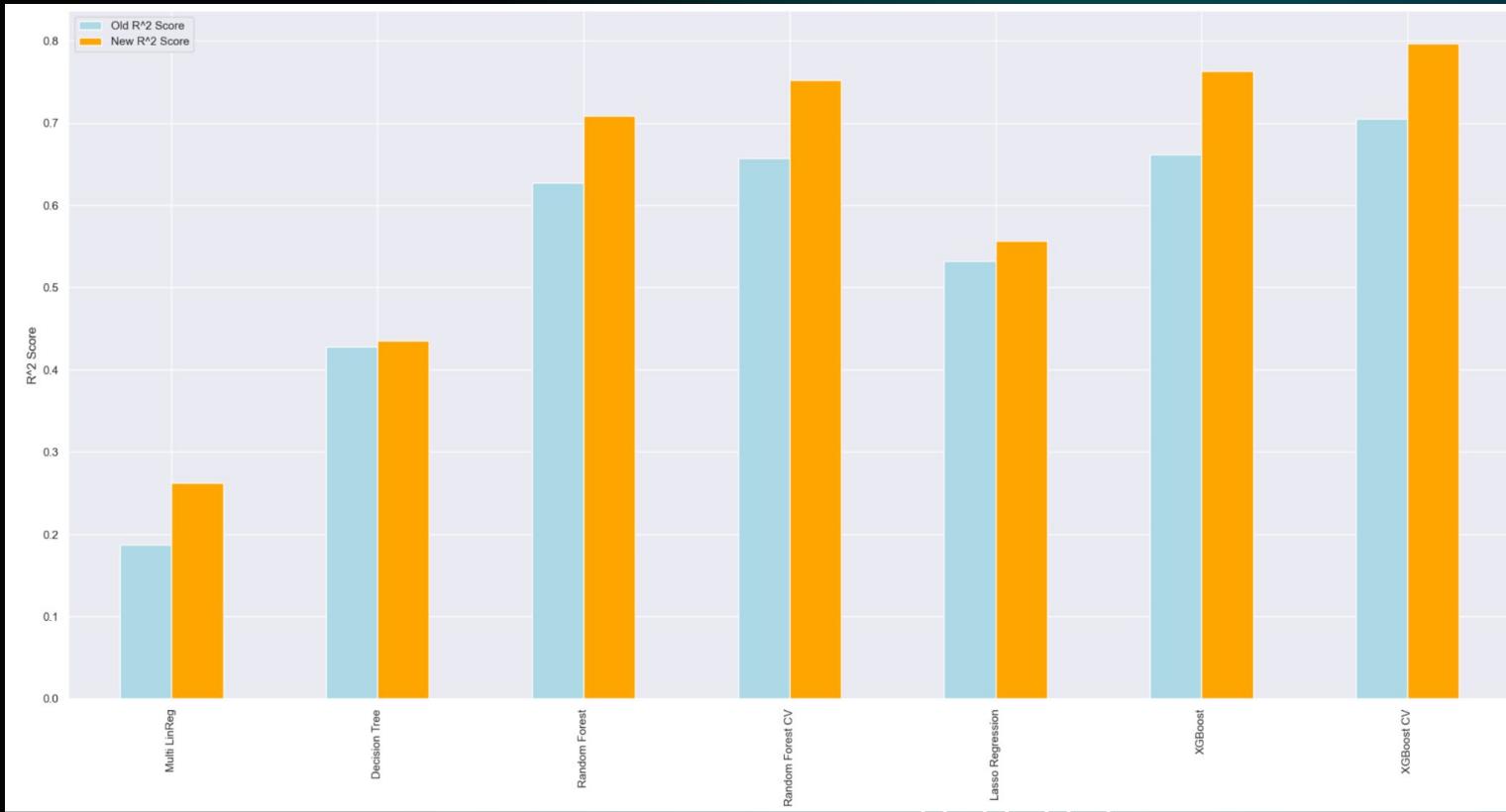
Using Engineered Feature - Distance



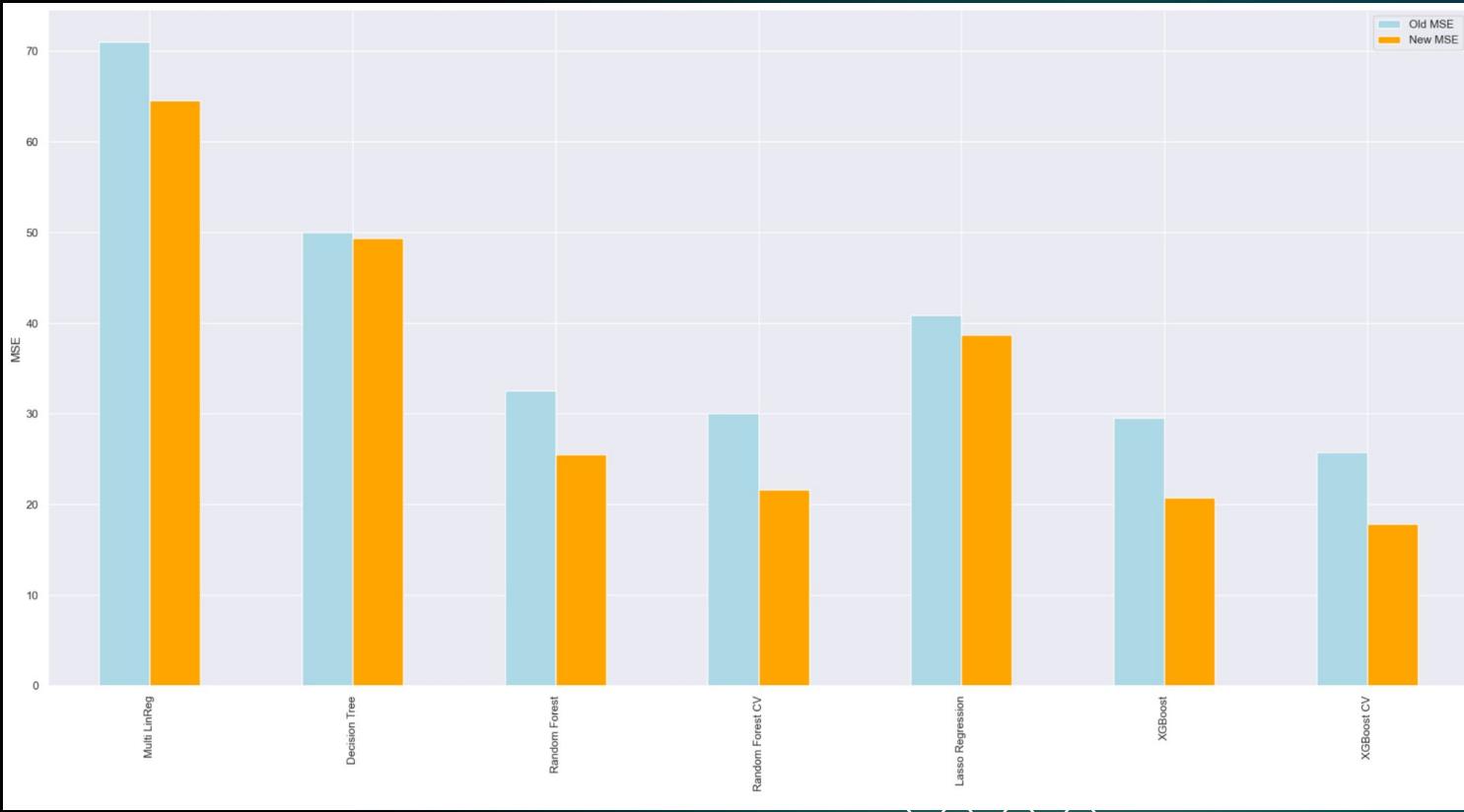
Retrieved from:<https://sketchplanations.com/the-haversine-formula>

```
def haversine(lat1, lon1, lat2, lon2):
    """calculate haversine distance between two co-ordinates"""
    lat1, lon1, lat2, lon2 = map(np.radians, (lat1, lon1, lat2, lon2))
    AVG_EARTH_RADIUS = 6371 # in km
    lat = lat2 - lat1
    lon = lon2 - lon1
    d = np.sin(lat * 0.5) ** 2 + np.cos(lat1) * np.cos(lat2) * np.sin(lon * 0.5) ** 2
    h = 2 * AVG_EARTH_RADIUS * np.arcsin(np.sqrt(d))
    return(h)
```

Results With Engineered Feature



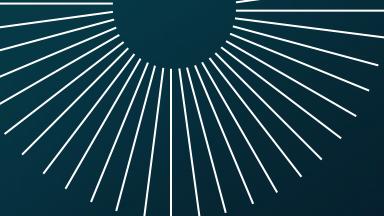
Results With Engineered Feature



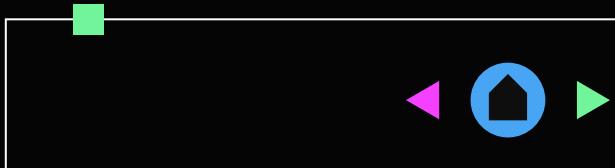
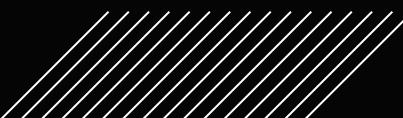


Conclusion

Data-Driven Insights

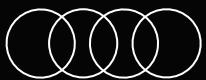


05





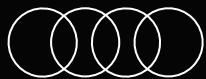
Any there any biases to
Deliverer Ratings?





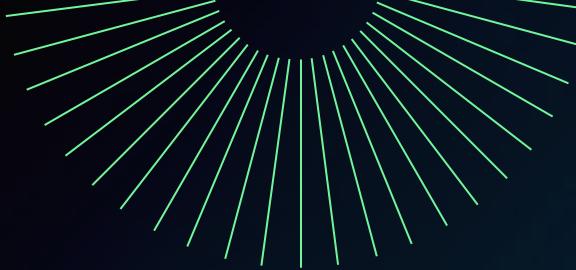
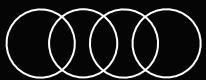
Delivery time &
Multiple Deliveries.

Nothing otherwise.





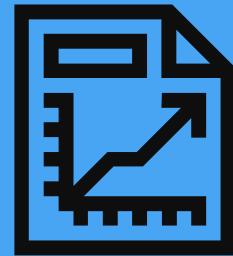
What affects Delivery Time?



What affects Delivery Time

Positively Correlated

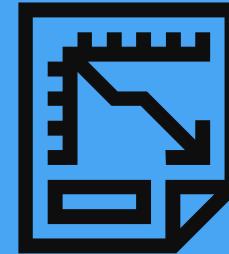
- Distance
- Multiple deliveries
- Age



What affects Delivery Time

Negatively Correlated

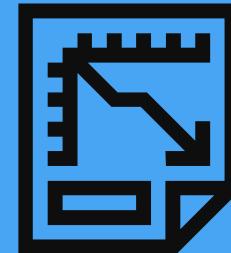
- Ratings
 - Higher Ratings = Lower Delivery Time



What affects Delivery Time

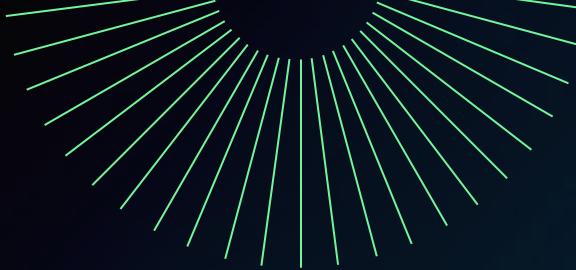
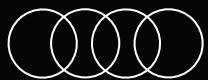
Categorical Variables

- **Traffic density**
 - Low traffic vs Jams
- **City terrain**
 - Urban & Metropolitan vs Semi-urban
- **Weather**
 - Sunny





Machine Learning to estimate **Delivery Time**



Estimate delivery time with XGBoost

	R^2 Score	MSE
CV XGBoost	0.791355	18.238924

79% accuracy

with a **low** error in estimation

Data Driven Insights & Recommendations

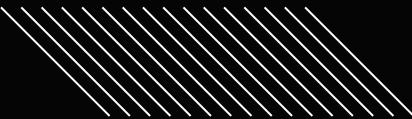
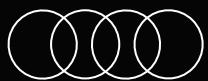
What can be done to reduce delivery time?

Problem	Distance & attending to multiple deliveries leads to long delivery times
Solution	<p>The food delivery algorithm should aim to further minimise delivery distances.</p> <p>The food delivery platform can allow users to “prioritise” their delivery, paying a greater fee ensure the delivery is for 1 only.</p> <p>A user can try to order before the peak lunch & dinner period to avoid longer delivery.</p>

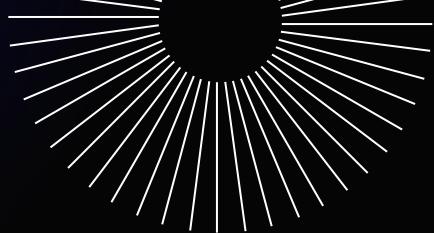


(

THANK
YOU



Outcome of Project



1,200

Planned hours

Despite being red, Mars is actually a cold place. It's full of iron oxide dust

892

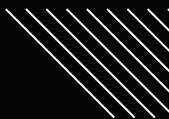
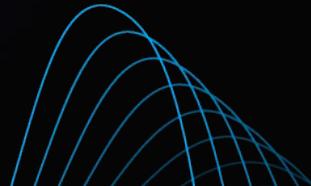
Hours spent

Venus has a beautiful name and is the second planet from the Sun

63

Adjustments

Mercury is the closest planet to the Sun and the smallest one



What we learnt



Calls

Venus is the second planet from the Sun



Foundations

Despite being red, Mars is a cold place



Portals

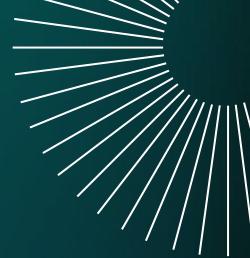
Jupiter is the biggest planet of them all



Fundraising

Saturn is a gas giant and has several rings





- ➊ Are we able to identify any biases in ratings based strictly on the attributes of the delivery rider, omitting delivery timings?

PROBLEM MOTIVATION

