Final Project: Customer's Rating Predictive Model on Delivery Operations Business

Presented By

Calista Damara

Outline

01 Project Background

02 Data Preprocessing

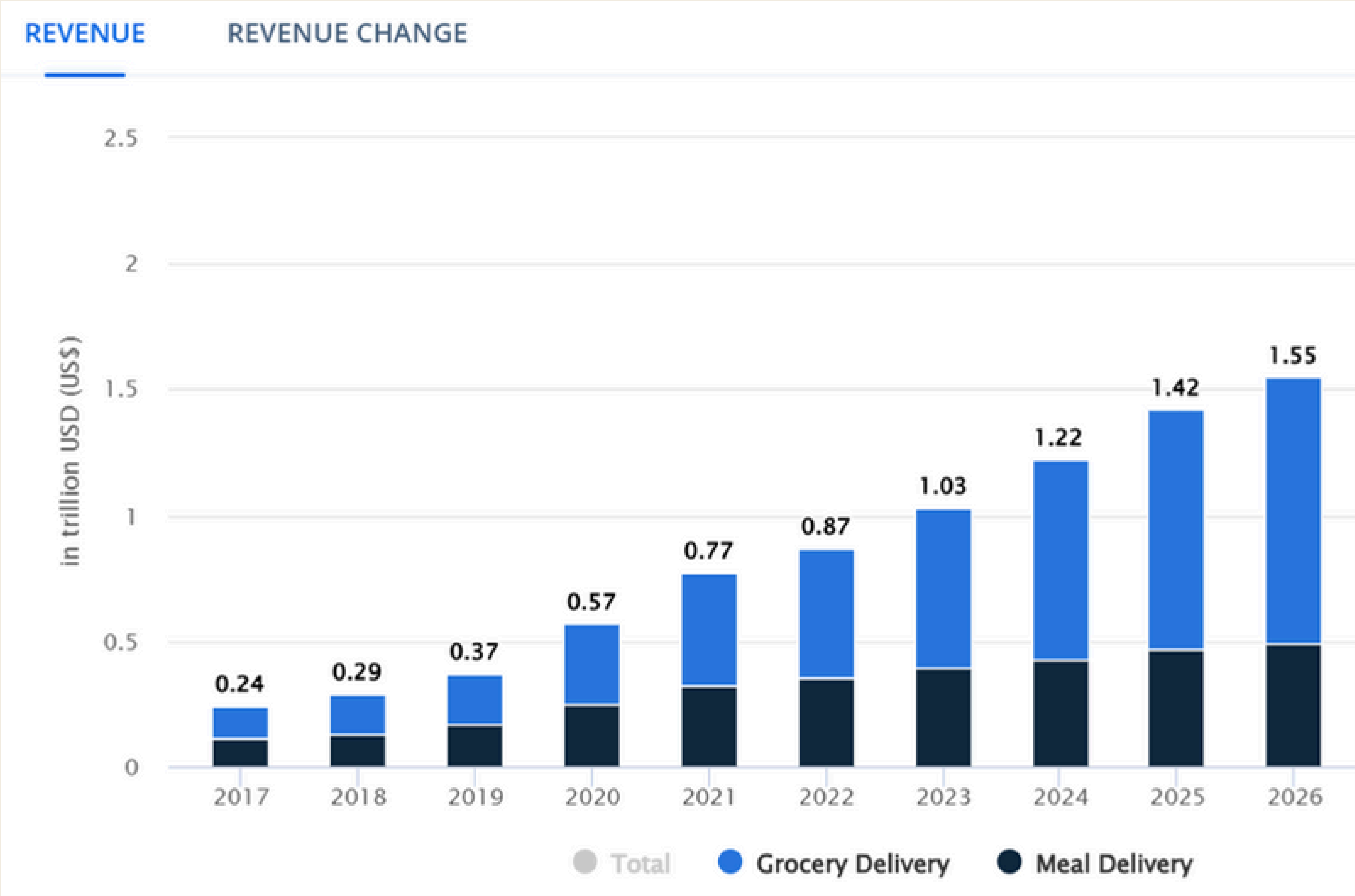
03 Exploratory Data Analysis

04 Machine Learning

05 Conclusion & Recommendation

Project Background

Growth of Delivery Operations Business

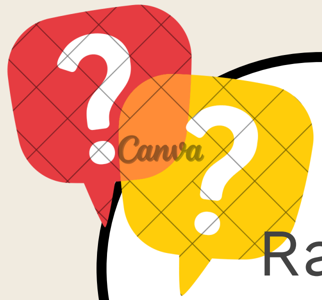


Source: Statista Market Analysis

Economic Growth

Social Behavioural

Digitalization



Problem

Rapid growth in Delivery Operations Business makes this sector has high competition

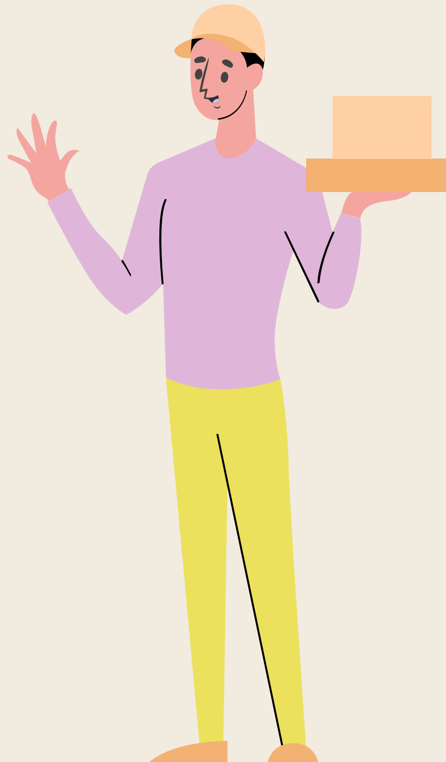
Project Background

Goals

- **Analyzing** Customer Behaviours on Deliv Operation Services and found features that influence Rating Rate
- **Create a predictive model** that can accurately predict Rating Rate

Business Objective

Increase Rating rate 30% within 3 months by create predictive model on Rating Rate and optimizing features that influences it.



Data Information

1

Data Shape

Rows: 45.584
Columns: 20

2

Missing Value

8 columns

	feature	missing_value	percentage
0	Delivery_person_Ratings	1908	4.19
1	Delivery_person_Age	1854	4.07
2	Time_Orderd	1731	3.80
3	City	1200	2.63
4	multiple_deliveries	993	2.18
5	Weather_conditions	616	1.35
6	Road_traffic_density	601	1.32
7	Festival	228	0.50

```
RangeIndex: 45584 entries, 0 to 45583
Data columns (total 20 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   ID                            45584 non-null  object
 1   Delivery_person_ID            45584 non-null  object
 2   Delivery_person_Age           43730 non-null  float64
 3   Delivery_person_Ratings       43676 non-null  float64
 4   Restaurant_latitude           45584 non-null  float64
 5   Restaurant_longitude          45584 non-null  float64
 6   Delivery_location_latitude    45584 non-null  float64
 7   Delivery_location_longitude   45584 non-null  float64
 8   Order_Date                    45584 non-null  object
 9   Time_Orderd                   43853 non-null  object
 10  Time_Order_picked             45584 non-null  object
 11  Weather_conditions            44968 non-null  object
 12  Road_traffic_density          44983 non-null  object
 13  Vehicle_condition             45584 non-null  int64
 14  Type_of_order                 45584 non-null  object
 15  Type_of_vehicle               45584 non-null  object
 16  multiple_deliveries           44591 non-null  float64
 17  Festival                      45356 non-null  object
 18  City                          44384 non-null  object
 19  Time_taken (min)              45584 non-null  int64
dtypes: float64(7), int64(2), object(11)
memory usage: 7.0+ MB
```

3

Unmatch Datatype

3 columns

Pre-processing

Data Preprocessing

Handling Missing Value

	Features	Treatment
Categorical	City (2.63%)	Replace with Mode value
	multiple_deliveries (2.18%)	
	Weather_conditions (1.35%)	
	Road_traffic_density (1.32%)	
	Festival (0.5%)	
Numerical	Delivery_person_Ratings (4.19%)	Replace with Median from each features
	Delivery_person_Age (4.07%)	
	Time_Order (3.80%)	Drop

```
Mode from City: Metropolitian
Mode from multiple_deliveries: 1.0
Mode from Weather_conditions: Fog
Mode from Road_traffic_density: Low
Mode from Festival: No
```

```
Median from Delivery_person_Ratings: 4.7
Median from Delivery_person_Age: 30.0
```

```
['21:55', '14:55', '17:30', '09:20', '19:50', '20:25', '20:
 '20:40', '21:15', '20:20', '22:30', '08:15', '19:30', '12:
 '18:35', '20:35', '23:20', '21:20', '23:35', '22:35', '23:
 '13:35', '21:35', '18:55', '14:15', '0.458333333', '09:45
 '08:40', '0.958333333', '17:25', '19:45', '19:10', '10:55'
 '21:40', '0.791666667', '16:45', '11:30', '15:10', '22:45'
 '22:10', '20:45', '22:50', '17:55', '09:25', '20:15', '22:
 '22:40', '23:50', '15:25', '10:20', '20:55', '10:40', '15:
 '20:10', '12:10', '15:30', '10:35', '21:10', '20:50', '12:
```

--> filter uncorrect format of time

Duplicate Value

No duplicate value detected

Data Type Convert

Features	Data Type
Order_Date	object -> datetime
Time_ordered	object -> datetime
Time_Order_picked	object -> datetime

Feature Engineering

~ Creating new features

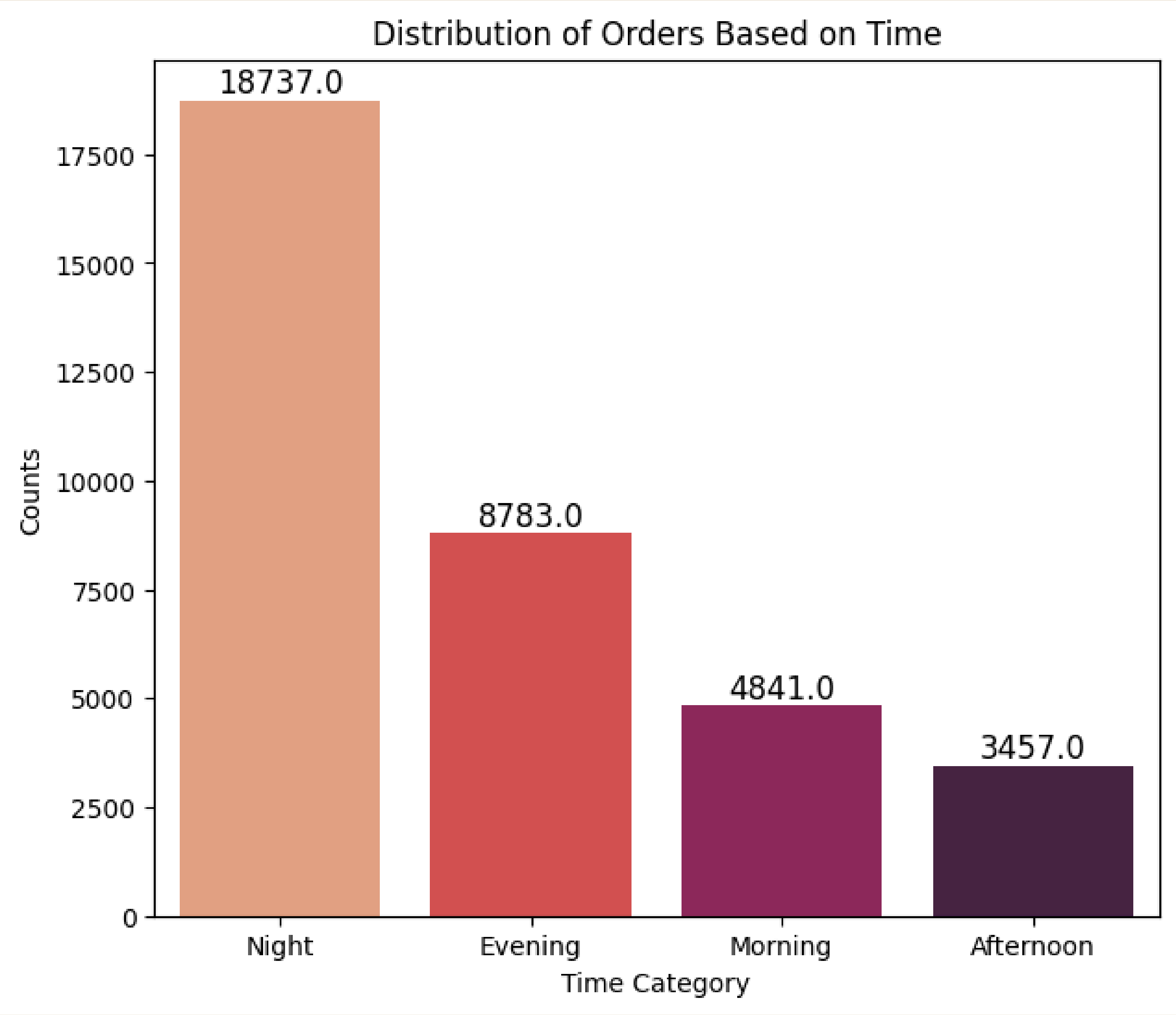
New Features	Description
order_prepare_time	the difference between **Time_Ordered** and **Time_Order_picked**, reflecting the time taken for order preparation
distance	the spatial separation between **Restaurant_coordinates** and **Delivery_coordinates**.
day_of_week	Extracted from **Order_Date**, specifying the day (Sunday/Monday/etc)
time_category	**Time_Order** grouping -> Morning (00.00-10.00), Afternoon(10.01-14.00), Evening (14.01-18.00), Night (18.01-23.59)

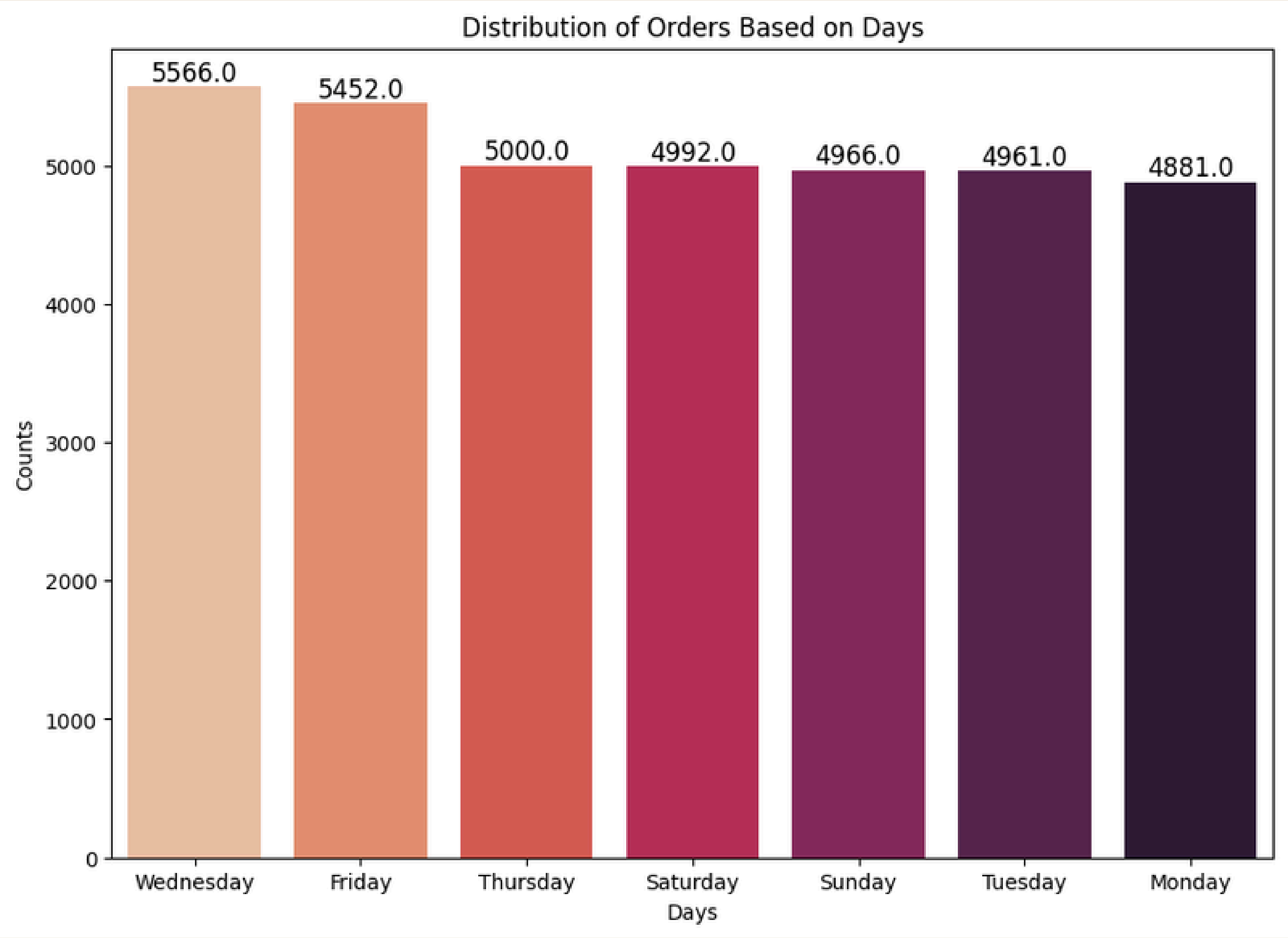
~ Drop unnecessary features

Restaurant_latitude, Restaurant_longitude, Delivery_location_latitude , Delivery_location_longitude, Time_Orderd, Time_Order_picked

Exploratory Data Analysis

How is customer behavior on the time and day features?

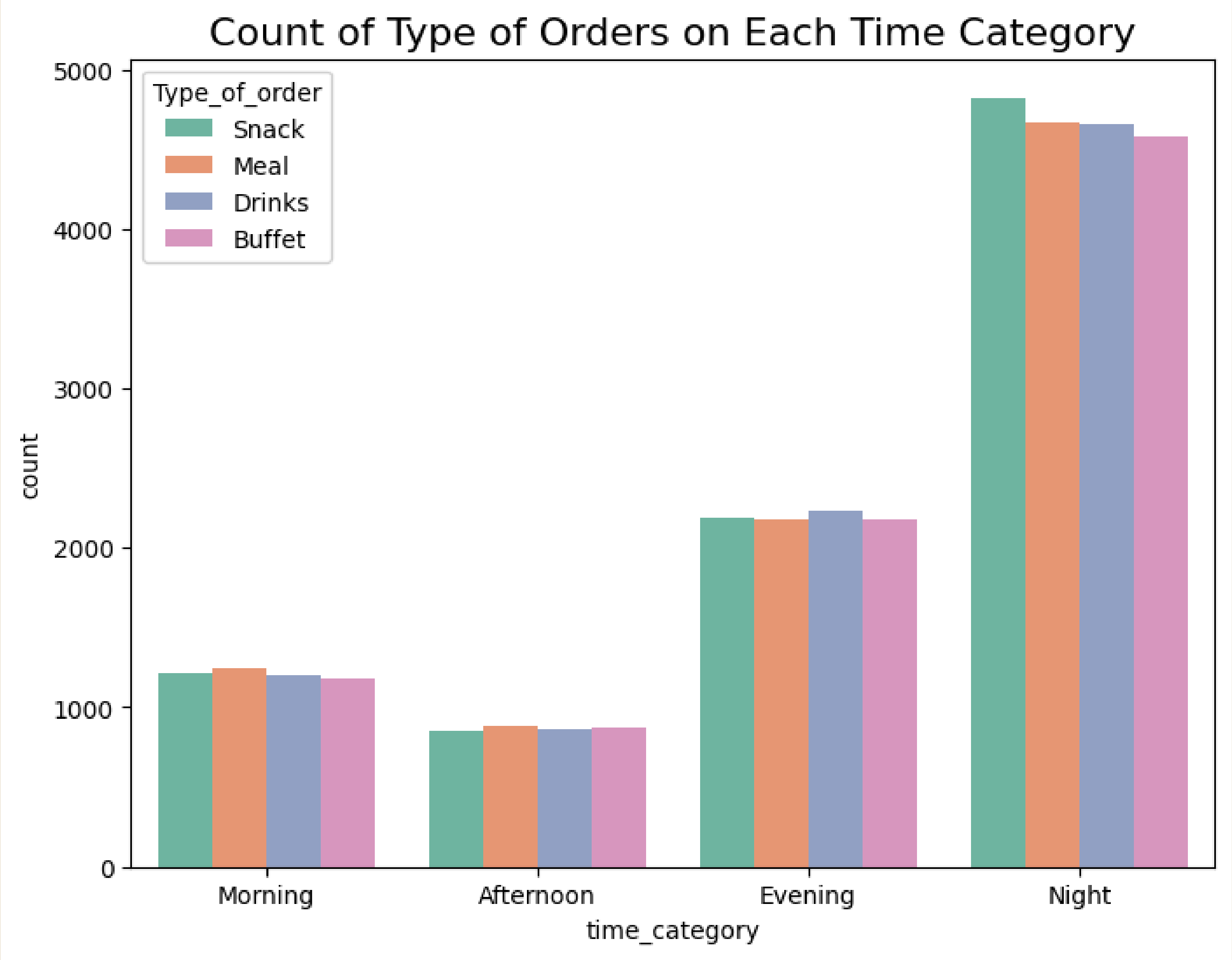




Customers most likely to use Delivery Service on the night (after 18.00 - 23.59) and The day with the most customer's orders is Wednesday.

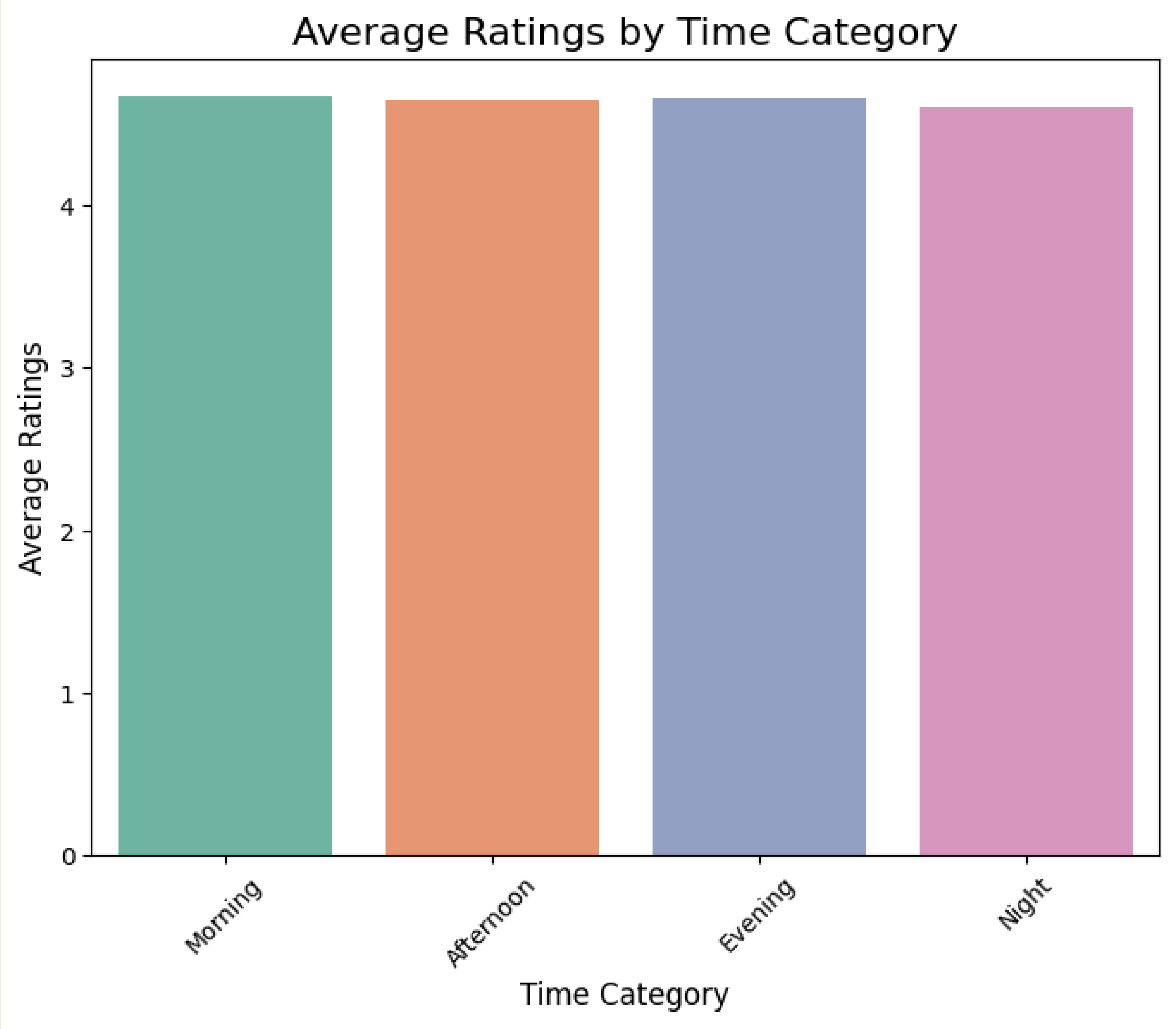
Exploratory Data Analysis

How is customer behavior on type of Order?



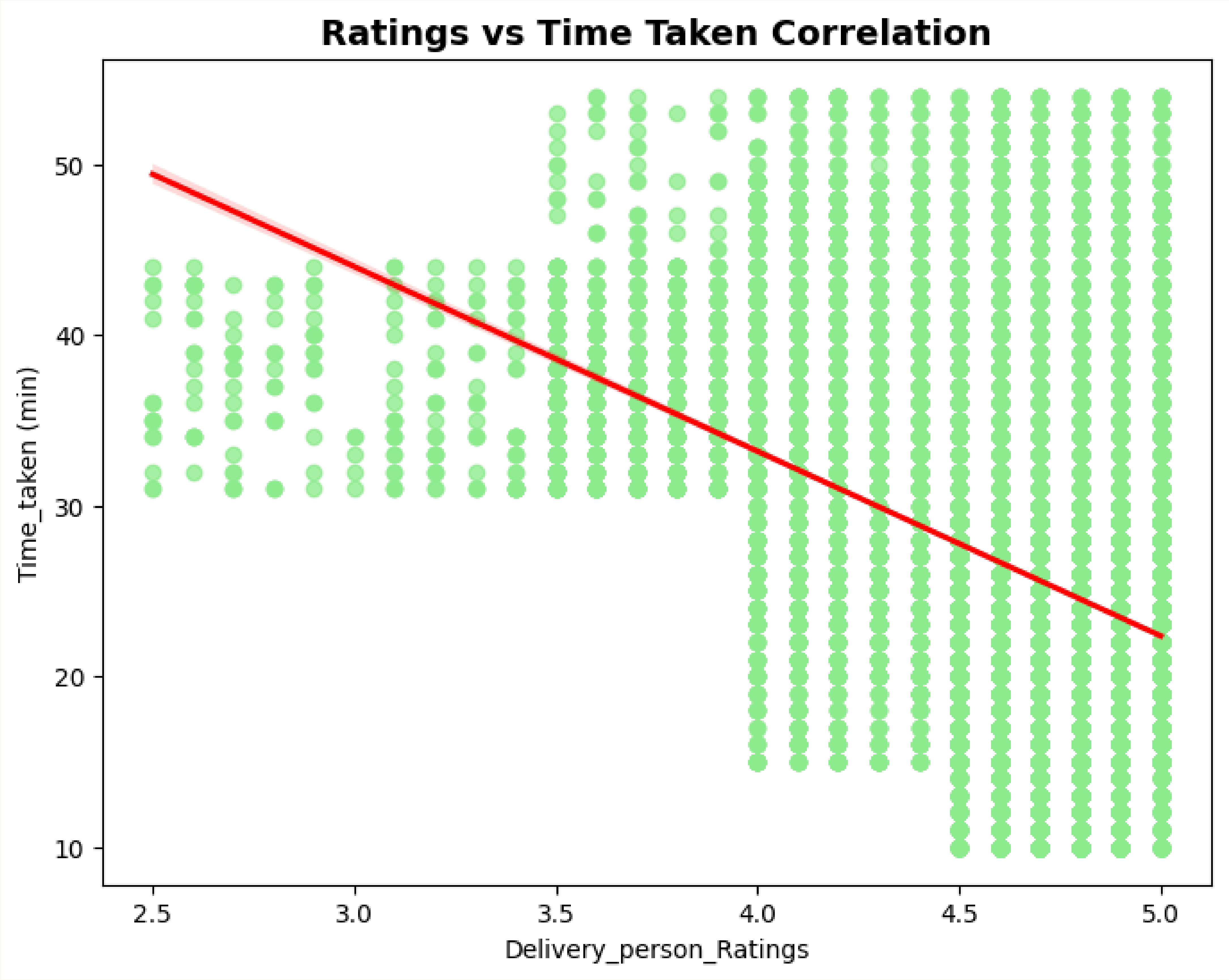
Favorite order in each time:
Morning: Meal
Afternoon: Meal
Evening: Drink
Night: Snack

Does the large number of orders coming in at night affect the rating?



Although most of customers using delivery services on the night, It doesn’t affects the Ratings. Night shift drivers still has good performance.
So, what features that impact to Rating Rate?

Exploratory Data Analysis



One of the features that impact to Rating is Time (Delivery) Taken. It has negative correlation.

It means the longer time can make the lowest ratings.

Feature Engineering

~ Scaling the Numerical Features

'Delivery_person_Age','distance','Time_taken (min)','order_prepare_time (min)

~ Encoding the Categorical Features

Festival

Value	Encode
No	0
Yes	1

Traffic_road_density

Value	Encode
Low	1
Medium	2
High	3
Jam	4

Weather_condi

Value	Encode
Sunny	1
Cloudy	2
Windy	3
Fog	4
Sandstorms	5
Stormy	6

Type of vehicle

Value	Encode
motorcycle	1
scooter	2
electric scooter	3

City

Value	Encode
Metropolitan	1
Urban	2
Semi-Urban	3

time category

Value	Encode
Morning	1
Evening	2
Afternoon	3
Night	4

type of order

Value	Encode
Buffet	1
Drinks	2
Meal	3
Snack	4

day

Value	Encode
Monday	1
Tuesday	2
Wednesday	3
Thursday	4
Friday	5
Saturday	6
Sunday	7

Machine Learning

Split the Data

Train vs Test

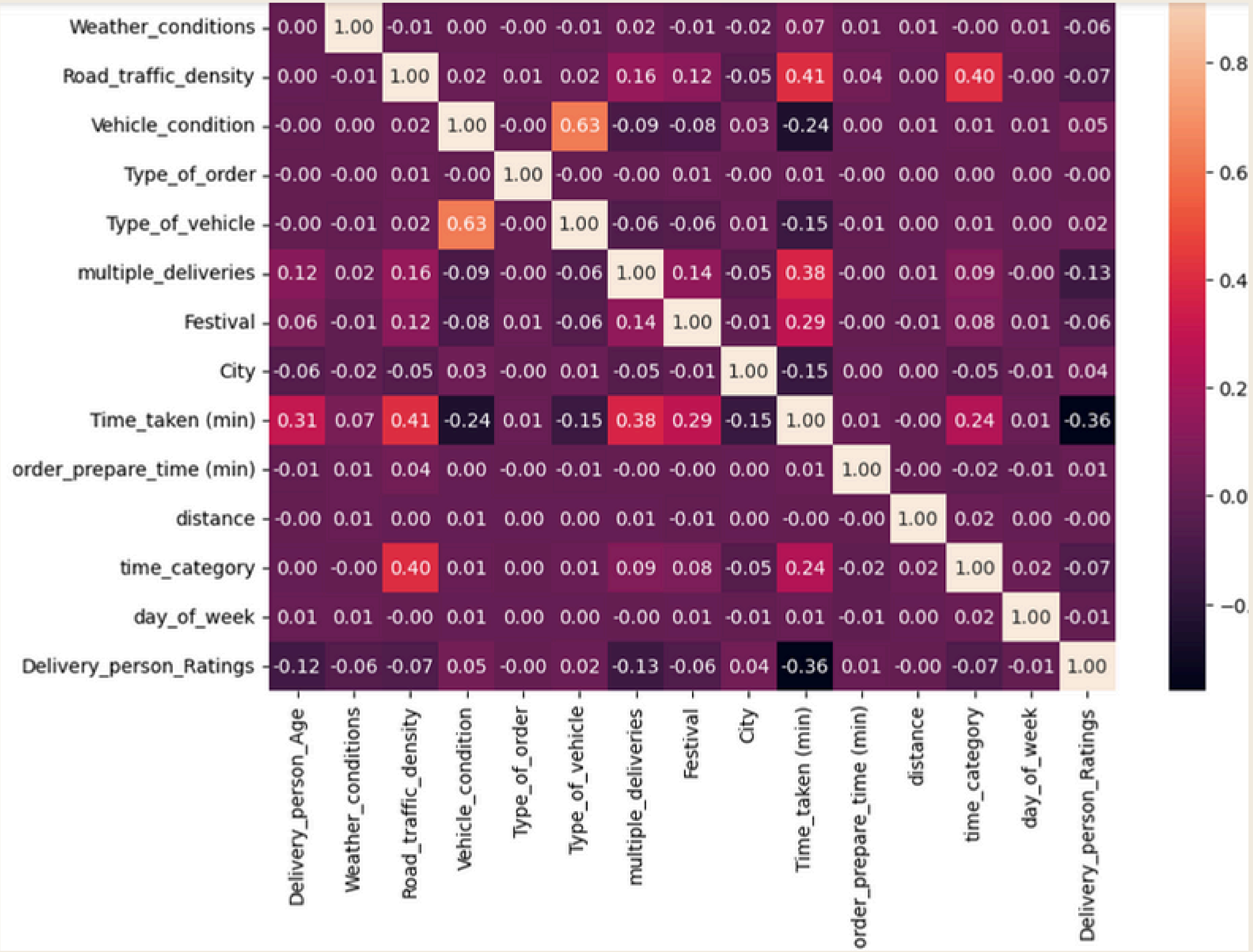
(80 : 20)

VIF Score

	feature	vif_score
1	Delivery_person_Age	1.140712
2	Weather_conditions	1.009250
3	Road_traffic_density	1.416486
4	Vehicle_condition	1.754595
5	Type_of_order	1.000424
6	Type_of_vehicle	1.659652
7	multiple_deliveries	1.167387
8	Festival	1.094867
9	City	1.024379
10	Time_taken (min)	1.787629
11	order_prepare_time (min)	1.003937
12	distance	1.000584
13	time_category	1.211087
14	day_of_week	1.000993

VIF Score < 4.0, Multicollinearity between independent variables is low and model is stable, so there is no need for multicollinearity handling.

Heatmap Correlation



Modeling & Evaluation

Training Evaluation

Models	Metrics		
	RMSE	MAE	MAPE
Ridge Reg	0.2922	0.2170	0.0491
Lasso Reg	0.2936	02161	0.0491
Random Forest Reg	**0.0948**	**0.0682**	**0.015**
Xgboost Reg	0.1663	0.1344	0.0295

Testing Evaluation

Models	Metrics		
	RMSE	MAE	MAPE
Ridge Reg	0.2871	0.2131	0.0482
Lasso Reg	0.2883	0.2124	0.0482
Random Forest Reg	**0.2261**	**0.1773**	**0.039**
Xgboost Reg	0.2213	0.1750	0.038

From 4 models, **RandomForest Regression** is the best model, The model has **smallest error** and experience **increase in performance** when implemented from training data to test data.

Hyperparameter Tuning

	params	mean_test_score	rank_test_score
20	{'max_depth': 5, 'n_estimators': 10}	0.493689	1
21	{'max_depth': 5, 'n_estimators': 20}	0.493640	2
24	{'max_depth': 5, 'n_estimators': 50}	0.493585	3

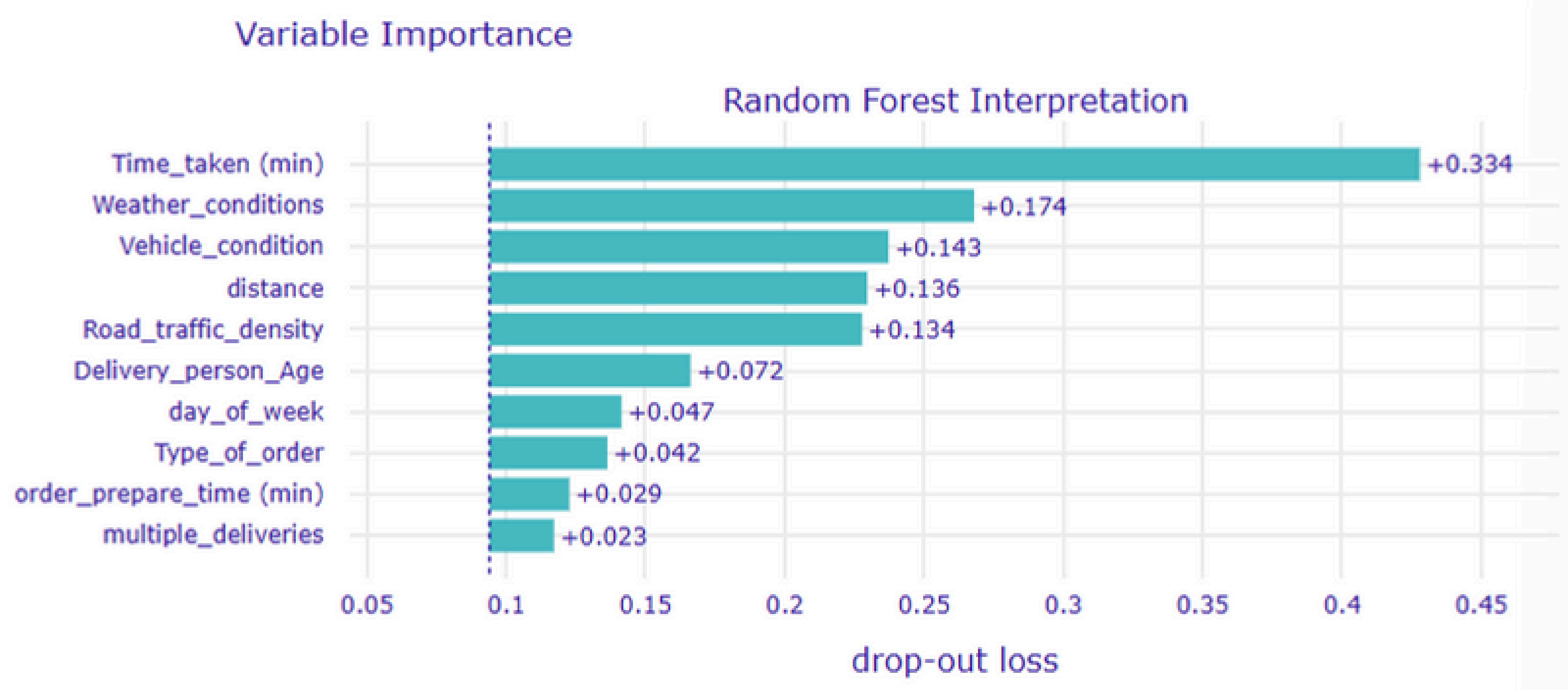
Using cross validation with GridSearchCV
The highest score (0.493689) was achieved by the model with max_depth=5 and n_estimators=10.

Hyperparameter Tuning

	Metrics	Before Tuning	After Tuning
0	RMSE	0.226105	0.224063
1	MAE	0.177353	0.179030
2	MAPE	0.039190	0.039698

After tuning, the model shows better accuration of test data

Importance Features



Top 3 features that influences Rating:

1. Time taken
2. Weather condition
3. Vehicle condition

Conclusion & Recommendation

Optimize Night Operational Service

Customers most likely to use Delivery Service on the **night (after 18.00 - 23.59)**

Recommendations

- Prioritizing **resource allocation** such as drivers, vehicles, and night monitoring systems **to accommodate high demand.**
- Provide **incentives** to night drivers to maintain **motivation and performance**,

Peak Days Strategy

The day with the most customer's orders is **Wednesday**

Recommendation

Special promotions such as discounts or bundling on that day **to increase** customer **satisfaction** and potential **repeat orders**.

Favorite Order

Favorite order in each time:
Morning: Meal; Afternoon: Meal; Evening: Drink; Night: Snack

Recommendation

Morning and Afternoon: Promote meal
Afternoon: Offer discounts on popular drinks.
Evening: Create special promotions for late-night snacks such as discounts on snack orders.

Conclusion & Recommendation

- Predictive Model succesfully build, the best model is **Random Forest Regressor**,
- **Top 3 The Most Importance Features** are:
  1. Time taken
  2. Weather condition
  3. Vehicle condition

Recommendations to Manage Time- taken

- Optimize delivery routing algorithms to ensure more efficient trips.
- Invest in technology like traffic prediction and delivery time estimation to assist drivers.

Recommendations to Manage Weather Condition

Offer customers notification of potential delays due to extreme weather conditions.

Recommendations to Manage Vehicle Conditions

Design regulation of vehicle inspections into operational flows to minimize the risk of delays.

THANK YOU!

Notebook: https://bit.ly/DeliveryOp_Project

calistadmra@gmail.com

Calista Damara

github.com/calistadamara