February 2023

Predicting Food Delivery Times

Alexandra Whiteford

Springboard Capstone Project

Introduction

Why?

As food delivery services are important to life in the city, and are often unreliable. No one wants food delivered cold or late, so the main goal of this project is to predict delivery times that are accurate in order to keep the customer satisfied.

Audience

Officials at companies such as:

- GrubHub
- DoorDash
- Uber Eats
- Deliveroo

Goal

Predicting accurate pick-up and drop-off times

The ideal estimation for food delivery windows are within 5 minutes or less of the predicted delivery time. This must be done so that the estimated drop-off times are honest and reflect accurate estimations for both the customers and the drivers.

- Early/On-Time delivery: > 5 minutes early
- Exactly On-Time: +/- 0 minutes
- Basically On-Time: < 5 minutes late
- A bit late: < 10 minutes late
- Late: 10+ minutes

Data Source

Kaggle: "Food Delivery Dataset"

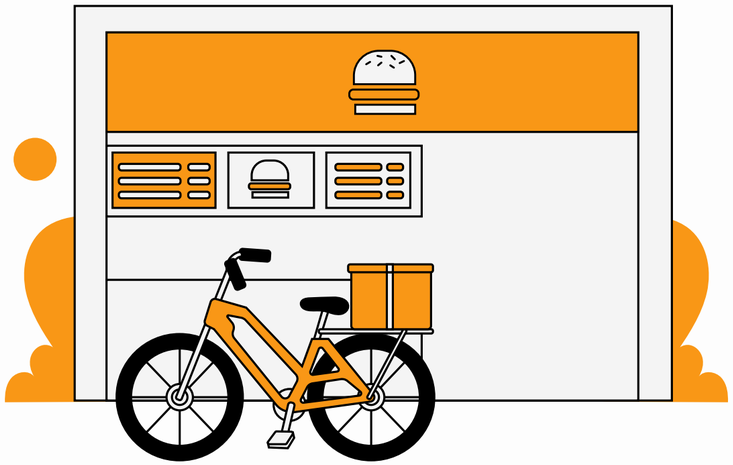
This dataset was downloaded from Kaggle, and includes the following columns:

- Delivery order ID
- Delivery person ID
- Delivery person age
- Delivery person ratings
- Restaurant latitude
- Restaurant longitude
- Delivery location latitude
- Delivery location longitude
- Order date
- Time ordered
- Time order picked up by driver
- Time taken to deliver food
- Weather conditions
- Road traffic density
- Vehicle condition
- Type of order
- Multiple deliveries
- Festival
- City environment

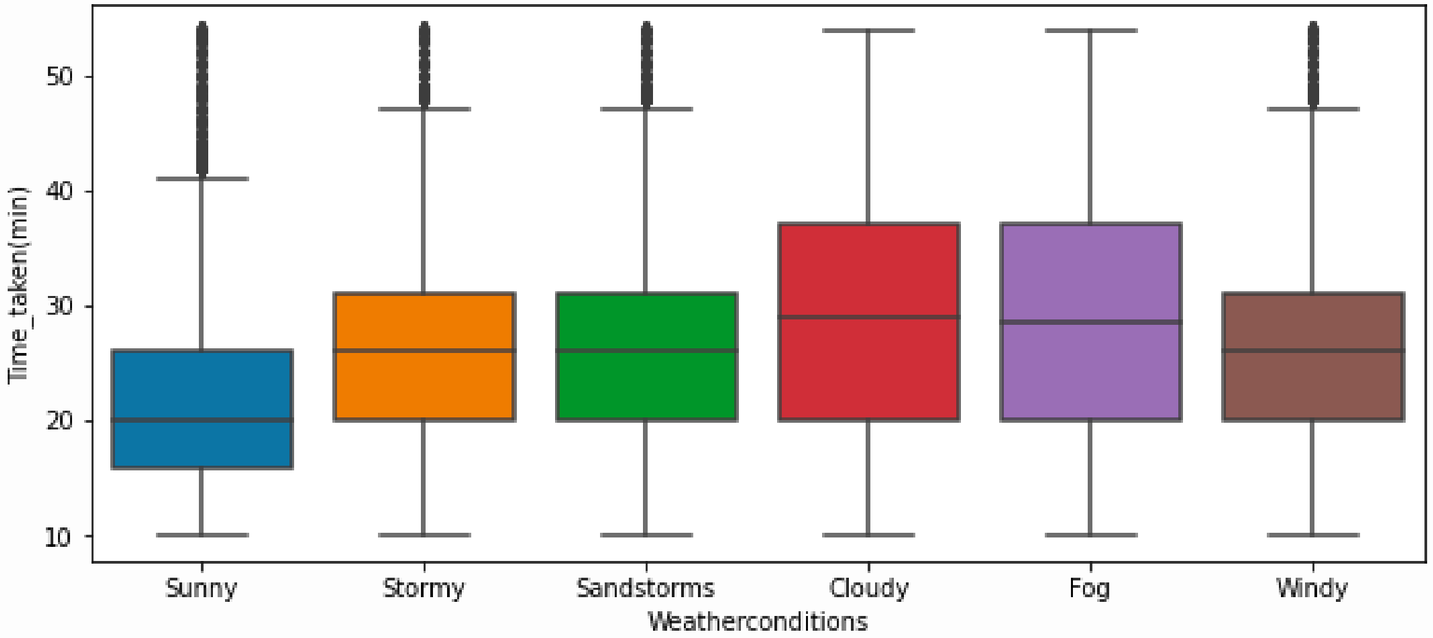
Data Wrangling

Issues encountered during data exploration and cleaning:

- Getting rid of null values
  - Solved using np.nan
- Converting Time_taken(min) column from str value to float value for analysis
  - Solved using astype() within a for loop
- Unknown distance
  - Using geopy to calculate distance in km between four coordinate values
  - Issue: portion of the data is unreliable as the results are difficult to interpret

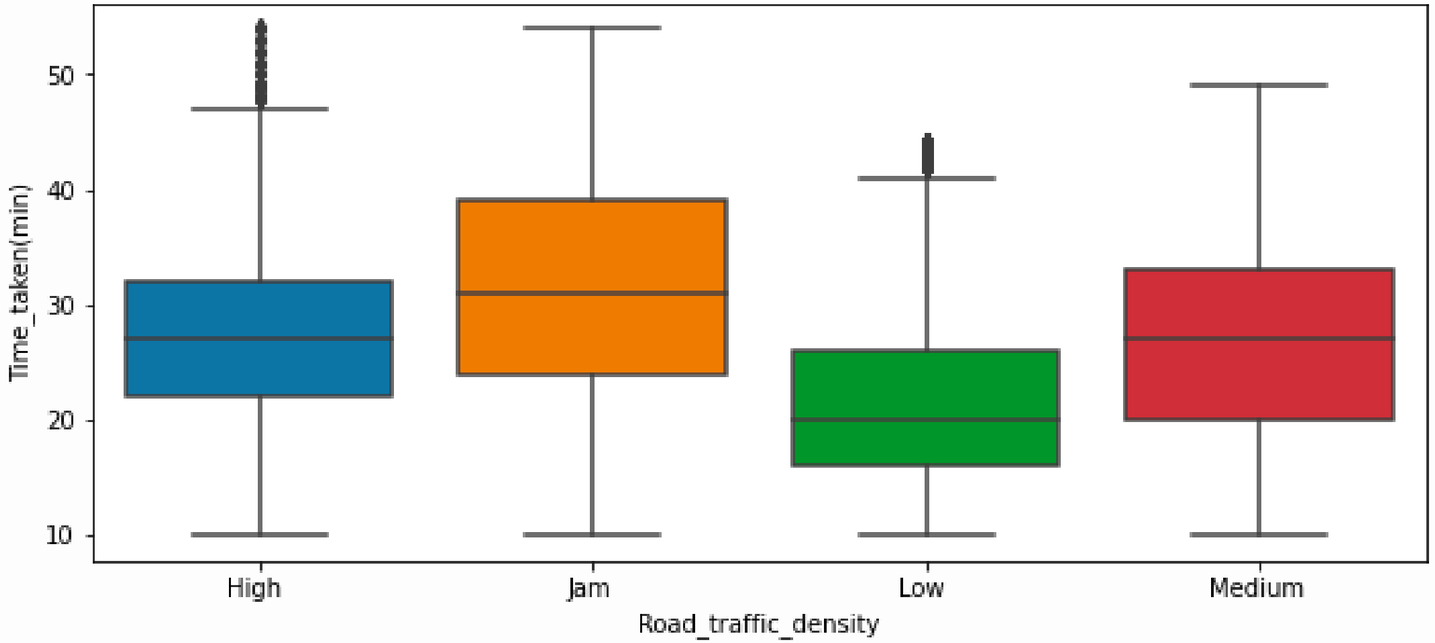


Weather conditions



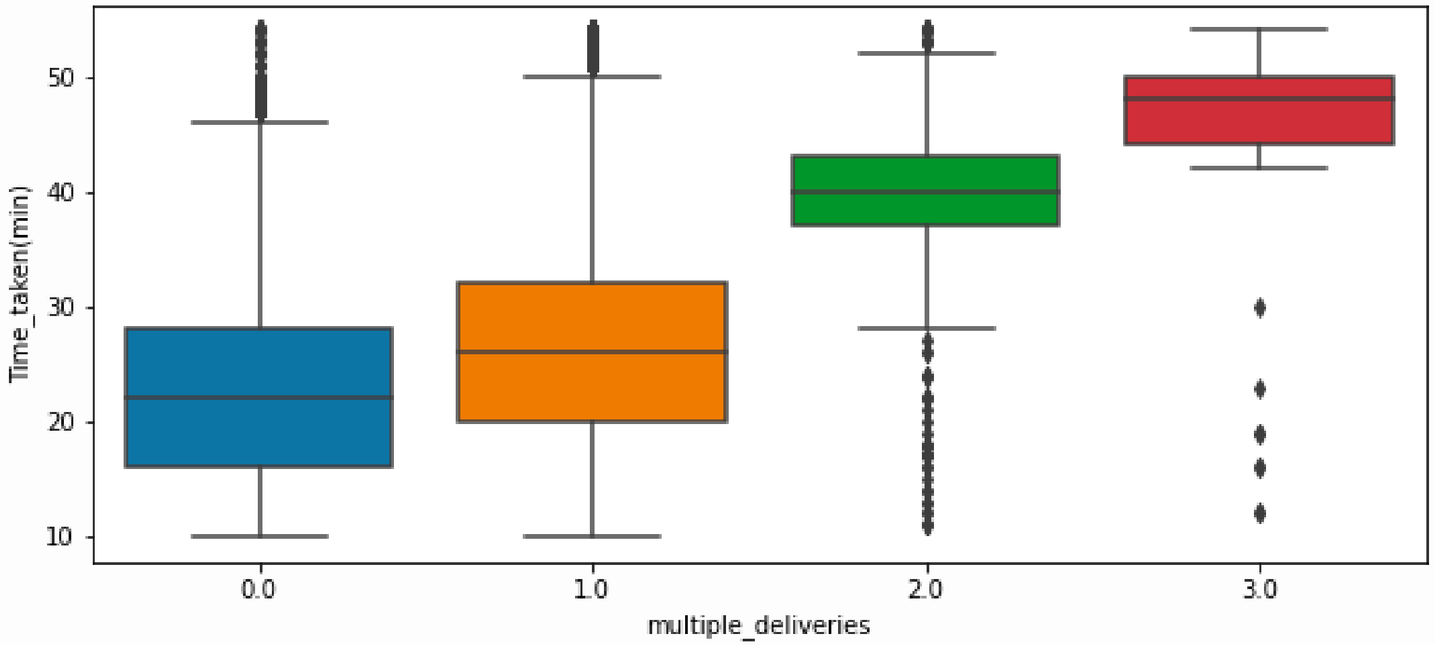
- Fastest delivery time on sunny days
- More variation in delivery time on cloudy or foggy days, including longer wait time on average

Road Traffic Density



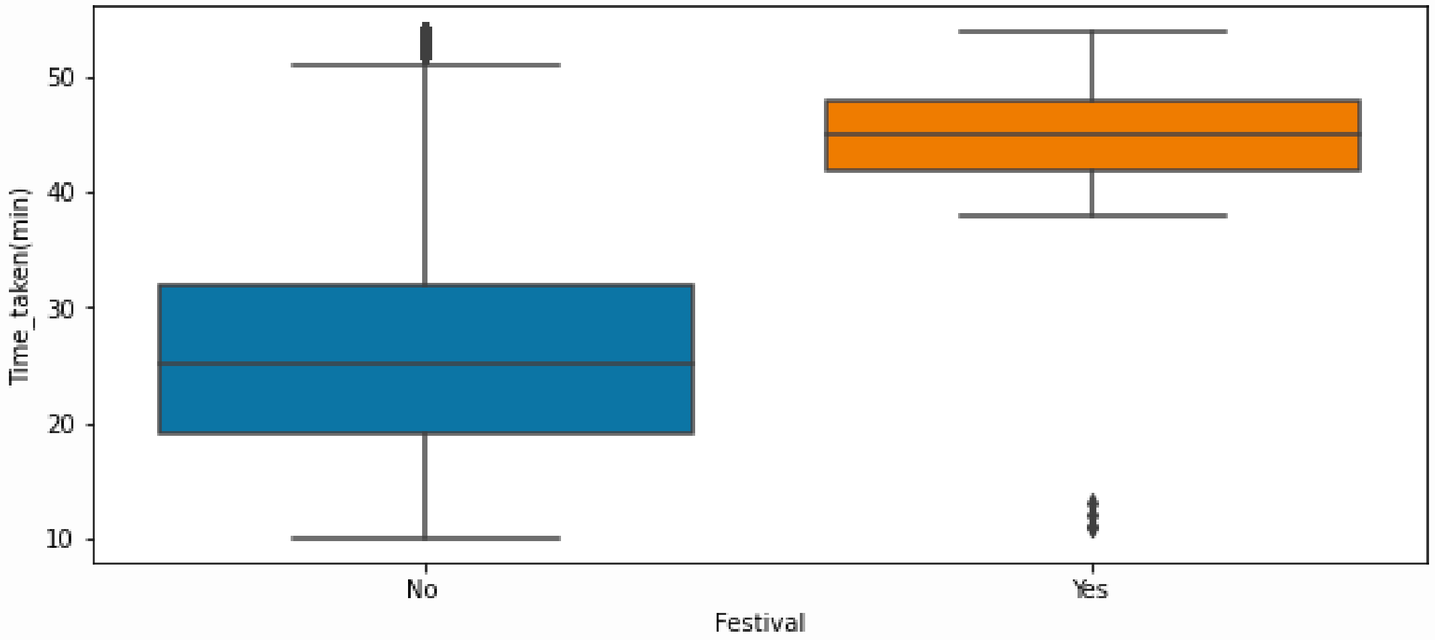
- Fasted delivery time when low traffic
- Slightly slower delivery time with medium traffic
- Slow delivery time with less variation with high traffic
- Traffic jam provides longer delivery times and more variation - hard to predict

Multiple Deliveries



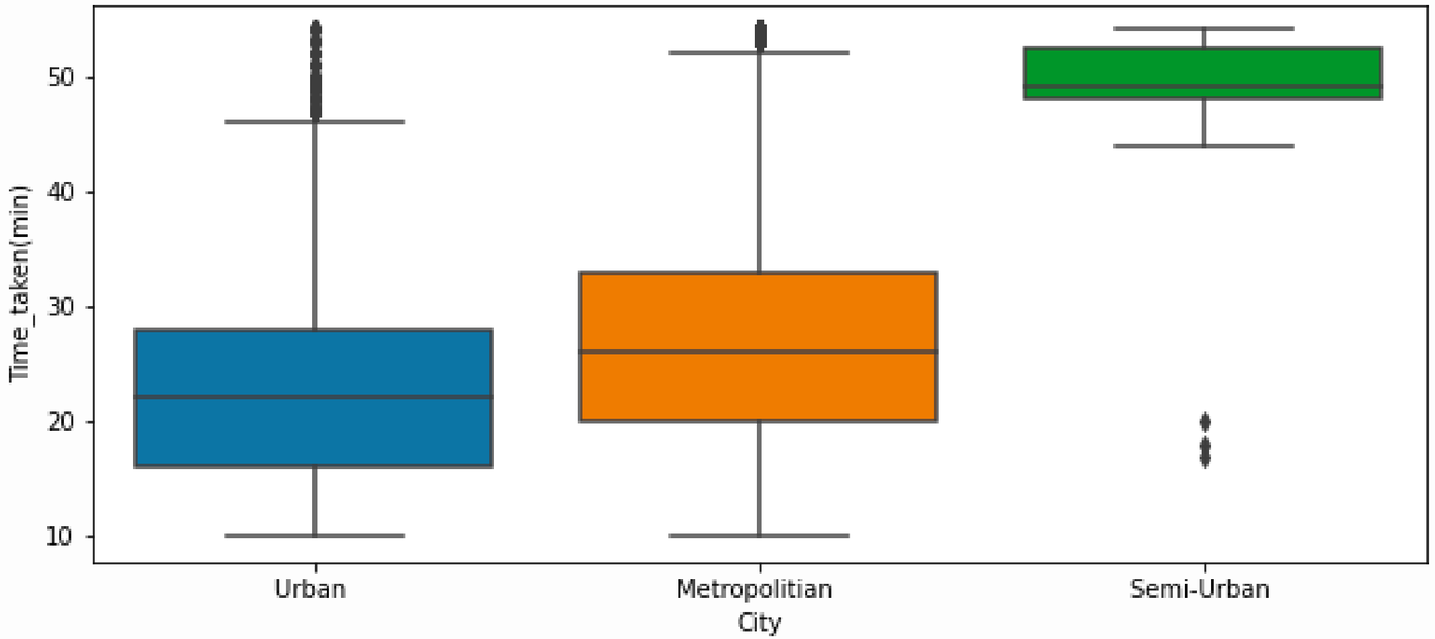
- Longer anticipated delivery time with each additional delivery order
- Lots of variation, but less frequent variation as there become more deliveries

Festival



- Festivals in the urban areas take much longer to reach the delivery address
- Festivals are few, only 2%

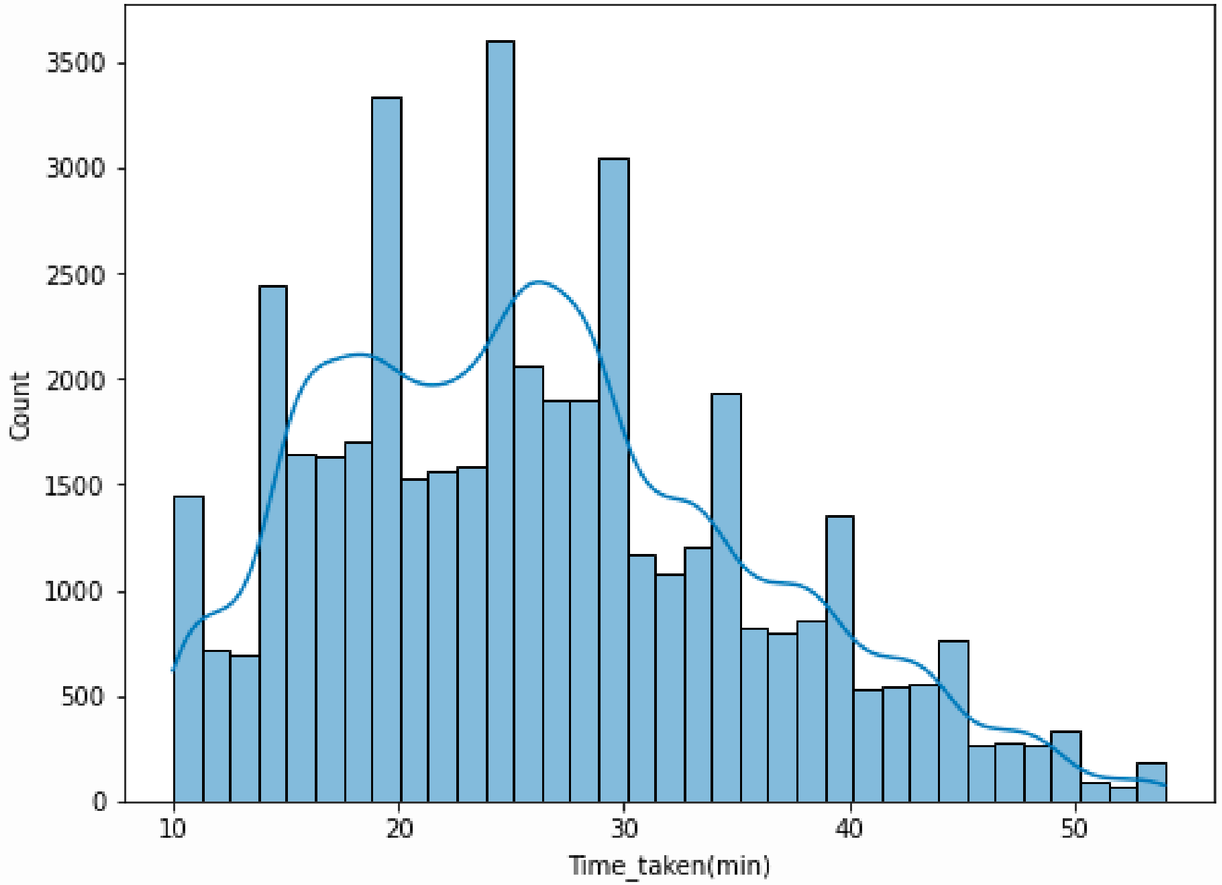
City Type



- Urban and Metropolitan cities are more compact than sub-urban cities
- Restaurants and delivery addresses are geographically closer together
- Semi-Urban delivery addresses are physically further away, which typically takes more time for the driver to leave the city.

Time Taken

- Average overall time taken for deliveries is about 25-27 minutes
- Majority of deliveries take between 15-32 minutes
- High occurrence at intervals of 5, suggesting lazy data collection or rounded values



Modeling comparison table

Out of the 4 models tested, the Linear Regression was the most successful and reliable:

- Lowest RMSE value
- Highest R2 value (although still very low)

Therefore, using the Linear Regression model provides us with confidence that the data is following a significant trend line representing the correlation between the observed v predicted data

	RMSE	R2
LinearRegression	**0.14537212334877098**	**0.021389304824339917**
KNNeighbors	0.15159863732138348	-0.05022433040104393
RandomForestRegressor	0.15028156528536282	-0.04635969968657615
XGBRegressor	0.14904654	-0.06903199681413394

Constraints

- Parking accessibility
- Building security
- Confusing delivery address input
- Delayed supplier

Any of these issues may render the predicted delivery time unreliable, leading to a potentially unsatisfied customer

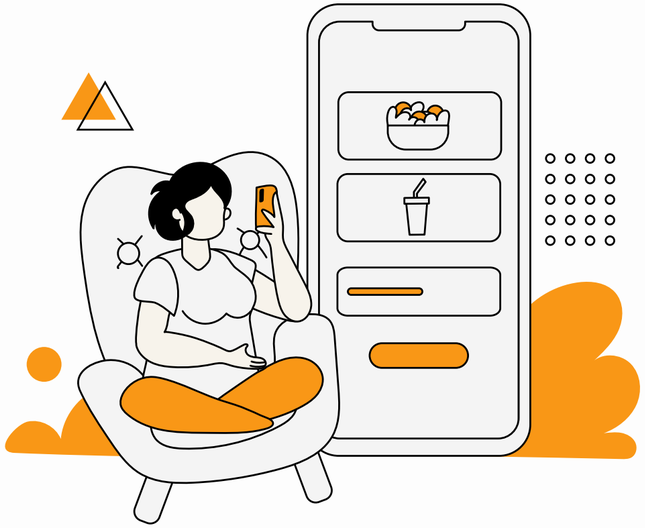
Recommendations

- Ensure the customer verifies the delivery address
- Limit additional deliveries to 2 maximum to keep delivery times shorter
    - If this is not possible due to high volume of orders, ensure that the predicted delivery time portrays the longer wait time accurately
- Verify the temperature-secure delivery bags retain the appropriate temperature

In Summary

In order to keep customers satisfied with their order and the company, ensure that predicted delivery times are as honest

- Consider the implementation of a +/- 5 minute buffer period to account for any minor delays or constraints such as security or parking issues



By ensuring the delivery will be made by a certain time, not only will the customer be satisfied, but the company leaders will be satisfied as well knowing that their deliveries are being made in the most time-efficient manner possible.