Github:

<https://github.com/kansonc/LHL_final_project>

Business Problems:

1. A company would like to know a time estimate it may take to have their order delivered once placed. They would like to know:
   1. What are the main factors contributing to longer delivery times, such as:
      1. Type of Vehicles:
         1. A screenshot of a computer

            Description automatically generated
         2. Assumptions:
            1. Scooters and Electric Scooters take the bike lane, not always the car lanes on the road.
            2. Scooters and Electric Scooters can be combined into Scooters Category
            3. Motorcycles must use car lanes on the road.
      2. The impact of weather Conditions, where:
         1. A screenshot of a computer

            Description automatically generated
         2. Assumptions:
            1. 616 Records are NaN

Of these records, all do not have a road traffic density, Time ordered, delivery person age, delivery person Ratings,

* + - * 1. Ordinal Hierarchy from worst to best conditions:

Sandstorm

Stormy

Fog

Windy

Cloudy

Sunny

* + 1. The impact of Road Traffic Density
       1. A group of check boxes with black and white text

          Description automatically generated
       2. 601 Null Values
          1. Null values correspond with Nulls for the Columns:

Time ordered

Other columns but not all are nulls in those columns

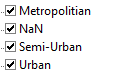
* + - 1. Assumptions:
         1. The Null Values should all be removed unless they can be imputed.
         2. The ordinal Hierarchy from Worst to Best conditions are:

Jam

High

Medium

Low

* + - * 1. It would be worth exploring whether our assumptions on Scooters as they function like a bicycle and can use bike lanes, or cut across sidewalks. Worth it to see if traffic does have an effect on the mode of transportation.
    1. The impact of Vehicle Condition:
       1. This cannot be used as the definition of cannot be found or assumed. It does not detail delays or any specific instances were it might actually cause a delay.
       2. DROP THIS COLUMN
    2. Multiple Deliveries
       1. This may itself be a decision tree branch for the model considering we cannot find the exact route and which was taken by the deliverer. The data is not constant where for the delivery person ID does not match the multiple delivery orders within the timeframe. The Delivery ID also has many different ages per a single ID, so it can be assumed that multiple delivery personnel are assigned a the same ID.
    3. City
       1. 
       2. 1200 records are Null
       3. The city cannot be imputed, but with more mapping of the geolocation coordinates on a map, it could be found out. It could ultimately be a rural part near the outskirts of the local where the delivery location or restaurant has not been assigned
       4. We can try an Ordinal Hierarchy from worst to best as
          1. Metro
          2. Urban
          3. Semi
    4. Festival
       1. A screenshot of a computer screen

          Description automatically generated
       2. 18 Null records, which are all null in the the categories:
          1. City
       3. Only 13 “Yes” records.
       4. DROP THIS COLUMN
    5. Day of the Week
       1. Assumptions are:
          1. More delivery persons are needed for weekends
          2. There are less people out for
    6. Distance Travelled
       1. Assumptions:
          1. The person has taken the most direct route
          2. From the time ordered to the time picked up, a direct route cannot be assumed from the delivery person’s previous delivery location to the new restaurant location.
    7. Time Taken
       1. The Delivery time is from when the food was picked up. A new feature called time from order to pickup can be computed
  1. What are some desirable factors contributing to low ratings by the deliverer

1. Ratings
   1. Difference based on meals ordered against time taken to deliver.
   2. How ratings are affected by more deliveries and how long it takes – Chi Square stats to see if there is a difference between multiple deliveries or not.
   3. Ratings does not imply a tipping amount, so this would be something to ask for in the data and additionally the “service fee”, whether that is relative to the location area or if it is the same across.
2. Column Analysis:
   1. Continuation of the things above
      1. Restaurant Location Coordinates
         1. There are many columns that have 0 coordinates.

Checklist of cleaned:

Time Taken- minutes

City – Ordinal

Type of Vehicle -> Vehicle -> Dummy

Festival -> Dropped

Vehicle Condition ->Dropped

Weather Conditions -> Ordinal