In [3]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import statsmodels.api as sm
sns.set(style="whitegrid")
sns.set(font_scale = 1.1)
```

## Problem Statement

- **Define Problem Statement**
  - Definition of problem (as per given problem statement with additional views)

- **Primary Goal**

  - **Explore capabilities** using this data that helps to widen the gap between **the quality, efficiency, and profitability of their business versus their competitors**.
  - **Predicting delivery time segmented by route , source , destination**
  - **Clean, sanitize and manipulate data**
  - Recognizing **significant features** that will drive more orders.
  - How well those features describe the volume of orders
  - Recognizing **Demand pattern** based on **route type , trips booking time (days of week/month), source , destination , distance, time** etc.
  - How to **drive sales** , across route types and other factors mentioned above
    - Data driven discounting / offers among customer segments
- **Long term benefits** : **Sales growth** , **More market penetration** where (i.e. states, place, cities) there are less volume of orders, **Customer acquisition** , **Balance short and long trips** and **retention**

## Basic Analysis

- **Exploratory Data Analysis** (10 points)
  - Observations on shape of data, data types of all the attributes, conversion of categorical attributes to 'category' (If required), missing value detection, statistical summary.
  - Visual Analysis (distribution plots of all the continuous variable(s), boxplots of all the categorical variables)

In [4]:

```
df = pd.read_csv("delhivery_data.csv")
```

**Data types - structure & characteristics of the dataset**

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144867 entries, 0 to 144866
Data columns (total 24 columns):
 #   Column                          Non-Null Count   Dtype  
---  ------                          --------------   -----  
 0   data                            144867 non-null  object 
 1   trip_creation_time              144867 non-null  object 
 2   route_schedule_uuid             144867 non-null  object 
 3   route_type                      144867 non-null  object 
 4   trip_uuid                       144867 non-null  object 
 5   source_center                   144867 non-null  object 
 6   source_name                     144574 non-null  object 
 7   destination_center              144867 non-null  object 
 8   destination_name                144606 non-null  object 
 9   od_start_time                   144867 non-null  object 
 10  od_end_time                     144867 non-null  object 
 11  start_scan_to_end_scan          144867 non-null  float64
 12  is_cutoff                       144867 non-null  bool   
 13  cutoff_factor                   144867 non-null  int64  
 14  cutoff_timestamp                144867 non-null  object 
 15  actual_distance_to_destination  144867 non-null  float64
 16  actual_time                     144867 non-null  float64
 17  osrm_time                       144867 non-null  float64
 18  osrm_distance                   144867 non-null  float64
 19  factor                          144867 non-null  float64
 20  segment_actual_time             144867 non-null  float64
 21  segment_osrm_time               144867 non-null  float64
 22  segment_osrm_distance           144867 non-null  float64
 23  segment_factor                  144867 non-null  float64
dtypes: bool(1), float64(10), int64(1), object(12)
memory usage: 25.6+ MB
```

### Observations on shape of data

In [6]:
```
df.shape
```

Out[6]:
```
(144867, 24)
```

- **Observation** - Not a small size sample

### Statistical summary

In [7]:
```
df.describe()
```

Out[7]:

| | start_scan_to_end_scan | cutoff_factor | actual_distance_to_destination | actual_time | osrm_time | osrm_c |
|---|---|---|---|---|---|---|
| **count** | 144867.000000 | 144867.000000 | 144867.000000 | 144867.000000 | 144867.000000 | 144867 |
| **mean** | 961.262986 | 232.926567 | 234.073372 | 416.927527 | 213.868272 | 284 |
| **std** | 1037.012769 | 344.755577 | 344.990009 | 598.103621 | 308.011085 | 421 |
| **min** | 20.000000 | 9.000000 | 9.000045 | 9.000000 | 6.000000 | 9 |
| **25%** | 161.000000 | 22.000000 | 23.355874 | 51.000000 | 27.000000 | 29 |
| **50%** | 449.000000 | 66.000000 | 66.126571 | 132.000000 | 64.000000 | 78 |
| **75%** | 1634.000000 | 286.000000 | 286.708875 | 513.000000 | 257.000000 | 343 |
| **max** | 7898.000000 | 1927.000000 | 1927.447705 | 4532.000000 | 1686.000000 | 2326 |

In [8]:
```
df.describe(include=object)
```

Out[8]:

| | data | trip_creation_time | route_schedule_uuid | route_type | trip_uuid | source_center | |
|---|---|---|---|---|---|---|---|
| **count** | 144867 | 144867 | 144867 | 144867 | 144867 | 144867 | |
| **unique** | 2 | 14817 | 1504 | 2 | 14817 | 1508 | |
| **top** | training | 2018-09-28 05:23:15.359220 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153811219535896559 | IND000000ACB | Gurg |
| **freq** | 104858 | 101 | 1812 | 99660 | 101 | 23347 | |

## Converting data type objects to datetime

In [9]:

```
df["trip_creation_time"]= pd.to_datetime(df["trip_creation_time"])
df["od_start_time"]= pd.to_datetime(df["od_start_time"])
df["od_end_time"]= pd.to_datetime(df["od_end_time"])
df["cutoff_timestamp"]= pd.to_datetime(df["cutoff_timestamp"])
```

## Non-Graphical Analysis (Part 1)

In [10]:

```
df.head()
```

Out[10]:

| | data | trip_creation_time | route_schedule_uuid | route_type | trip_uuid | source_center | sour |
|---|---|---|---|---|---|---|---|
| **0** | training | 2018-09-20 02:35:36.476840 | thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3... | Carting | trip-153741093647649320 | IND388121AAA | Anand_VUN |
| **1** | training | 2018-09-20 02:35:36.476840 | thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3... | Carting | trip-153741093647649320 | IND388121AAA | Anand_VUN |
| **2** | training | 2018-09-20 02:35:36.476840 | thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3... | Carting | trip-153741093647649320 | IND388121AAA | Anand_VUN |
| **3** | training | 2018-09-20 02:35:36.476840 | thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3... | Carting | trip-153741093647649320 | IND388121AAA | Anand_VUN |
| **4** | training | 2018-09-20 02:35:36.476840 | thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3... | Carting | trip-153741093647649320 | IND388121AAA | Anand_VUN |

5 rows × 24 columns

In [11]:

```
df["data"].value_counts(normalize=True)*100
```

Out[11]:

```
training    72.382254
test        27.617746
Name: data, dtype: float64
```

In [12]:

```
df["route_type"].value_counts(normalize=True)*100
```

Out[12]:

```
FTL        68.794135
Carting    31.205865
Name: route_type, dtype: float64
```

In [13]:
```
df["source_name"].value_counts(normalize=True)*100
```

Out[13]:
```
Gurgaon_Bilaspur_HB (Haryana)            16.148823
Bangalore_Nelmngla_H (Karnataka)          6.899581
Bhiwandi_Mankoli_HB (Maharashtra)         6.286054
Pune_Tathawde_H (Maharashtra)             2.808942
Hyderabad_Shamshbd_H (Telangana)          2.310236
                                             ...
Shahjhnpur_NavdaCln_D (Uttar Pradesh)     0.000692
Soro_UttarDPP_D (Orissa)                  0.000692
Kayamkulam_Bhrnikvu_D (Kerala)            0.000692
Krishnanagar_AnadiDPP_D (West Bengal)     0.000692
Faridabad_Old (Haryana)                   0.000692
Name: source_name, Length: 1498, dtype: float64
```

In [14]:
```
df["destination_name"].value_counts(normalize=True)*100
```

Out[14]:
```
Gurgaon_Bilaspur_HB (Haryana)        10.505788
Bangalore_Nelmngla_H (Karnataka)      7.620016
Bhiwandi_Mankoli_HB (Maharashtra)     3.797906
Hyderabad_Shamshbd_H (Telangana)      3.555869
Kolkata_Dankuni_HB (West Bengal)      3.382985
                                         ...
Hyd_Trimulgherry_Dc (Telangana)       0.000692
Vijayawada (Andhra Pradesh)           0.000692
Baghpat_Barout_D (Uttar Pradesh)      0.000692
Mumbai_Sanpada_CP (Maharashtra)       0.000692
Basta_Central_DPP_1 (Orissa)          0.000692
Name: destination_name, Length: 1468, dtype: float64
```

In [15]:
```
df["segment_factor"].value_counts(normalize=True)*100
```

Out[15]:
```
 2.000000     4.142420
 1.500000     3.200867
 1.000000     1.636674
 1.666667     1.635983
-1.000000     1.620107
                ...
 1.844444     0.000690
 1.380000     0.000690
 4.103448     0.000690
 2.614458     0.000690
 29.777778    0.000690
Name: segment_factor, Length: 5675, dtype: float64
```

- **Filtering Feature types by data type**

In [16]:
```
continious_features = df.select_dtypes(include=['int64','float64']).columns
continious_features
```

Out[16]:
```
Index(['start_scan_to_end_scan', 'cutoff_factor',
       'actual_distance_to_destination', 'actual_time', 'osrm_time',
       'osrm_distance', 'factor', 'segment_actual_time', 'segment_osrm_time',
       'segment_osrm_distance', 'segment_factor'],
      dtype='object')
```

In [17]:
```
categorical_features = df.select_dtypes(include=['object']).columns
categorical_features
```

Out[17]:
```
Index(['data', 'route_schedule_uuid', 'route_type', 'trip_uuid',
       'source_center', 'source_name', 'destination_center',
       'destination_name'],
      dtype='object')
```

## Missing Value Detection

In [18]:
```python
percent_missing = df.isnull().sum() * 100 / len(df)
missing_value_df = pd.DataFrame({'column_name': df.columns,
                                 'percent_missing': percent_missing})
missing_value_df.sort_values('percent_missing', ascending=False)
```

Out[18]:

| | column_name | percent_missing |
|---|---|---|
| **source_name** | source_name | 0.202254 |
| **destination_name** | destination_name | 0.180165 |
| **data** | data | 0.000000 |
| **cutoff_factor** | cutoff_factor | 0.000000 |
| **segment_osrm_distance** | segment_osrm_distance | 0.000000 |
| **segment_osrm_time** | segment_osrm_time | 0.000000 |
| **segment_actual_time** | segment_actual_time | 0.000000 |
| **factor** | factor | 0.000000 |
| **osrm_distance** | osrm_distance | 0.000000 |
| **osrm_time** | osrm_time | 0.000000 |
| **actual_time** | actual_time | 0.000000 |
| **actual_distance_to_destination** | actual_distance_to_destination | 0.000000 |
| **cutoff_timestamp** | cutoff_timestamp | 0.000000 |
| **is_cutoff** | is_cutoff | 0.000000 |
| **trip_creation_time** | trip_creation_time | 0.000000 |
| **start_scan_to_end_scan** | start_scan_to_end_scan | 0.000000 |
| **od_end_time** | od_end_time | 0.000000 |
| **od_start_time** | od_start_time | 0.000000 |
| **destination_center** | destination_center | 0.000000 |
| **source_center** | source_center | 0.000000 |
| **trip_uuid** | trip_uuid | 0.000000 |
| **route_type** | route_type | 0.000000 |
| **route_schedule_uuid** | route_schedule_uuid | 0.000000 |
| **segment_factor** | segment_factor | 0.000000 |

- **Observations** :
  - **source_name (0.20%) and destination_name (0.18%)** have **missing values** and proportion is very less
  - **No missing values for rest of the features**

## Missing values Treatment

In [19]:
```
df["source_name"].fillna("Unknown", inplace=True)
df["destination_name"].fillna("Unknown", inplace=True)
```

In [20]:
```
percent_missing = df.isnull().sum() * 100 / len(df)
missing_value_df = pd.DataFrame({'column_name': df.columns,
                                 'percent_missing': percent_missing})
missing_value_df.sort_values('percent_missing', ascending=False)
```

Out[20]:

| | column_name | percent_missing |
|---|---|---|
| **data** | data | 0.0 |
| **trip_creation_time** | trip_creation_time | 0.0 |
| **segment_osrm_distance** | segment_osrm_distance | 0.0 |
| **segment_osrm_time** | segment_osrm_time | 0.0 |
| **segment_actual_time** | segment_actual_time | 0.0 |
| **factor** | factor | 0.0 |
| **osrm_distance** | osrm_distance | 0.0 |
| **osrm_time** | osrm_time | 0.0 |
| **actual_time** | actual_time | 0.0 |
| **actual_distance_to_destination** | actual_distance_to_destination | 0.0 |
| **cutoff_timestamp** | cutoff_timestamp | 0.0 |
| **cutoff_factor** | cutoff_factor | 0.0 |
| **is_cutoff** | is_cutoff | 0.0 |
| **start_scan_to_end_scan** | start_scan_to_end_scan | 0.0 |
| **od_end_time** | od_end_time | 0.0 |
| **od_start_time** | od_start_time | 0.0 |
| **destination_name** | destination_name | 0.0 |
| **destination_center** | destination_center | 0.0 |
| **source_name** | source_name | 0.0 |
| **source_center** | source_center | 0.0 |
| **trip_uuid** | trip_uuid | 0.0 |
| **route_type** | route_type | 0.0 |
| **route_schedule_uuid** | route_schedule_uuid | 0.0 |
| **segment_factor** | segment_factor | 0.0 |

## Visual Analysis (Part 1)

- Distribution plots of all the continuous variable(s), boxplots of all the categorical variables)

In [21]:
```
continious_features
```

Out[21]:
```
Index(['start_scan_to_end_scan', 'cutoff_factor',
       'actual_distance_to_destination', 'actual_time', 'osrm_time',
       'osrm_distance', 'factor', 'segment_actual_time', 'segment_osrm_time',
       'segment_osrm_distance', 'segment_factor'],
      dtype='object')
```

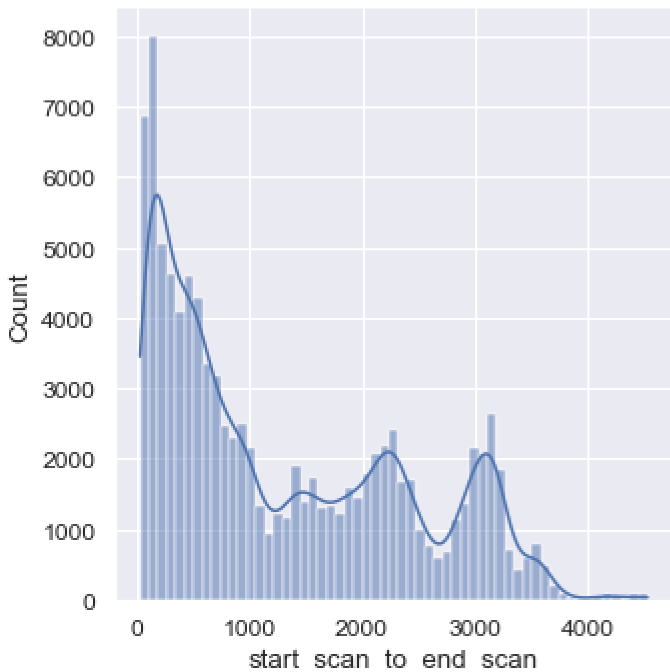
### Seperating data by route type - Full truck load and Carting

In [22]:
```
df_full_truck_load = df[df["route_type"] == 'FTL']
df_carting = df[df["route_type"] == 'Carting']
```

### Distribution plot of route type - Full truck load

In [23]:
```
sns.displot(df_full_truck_load['start_scan_to_end_scan'], kde=True)
```

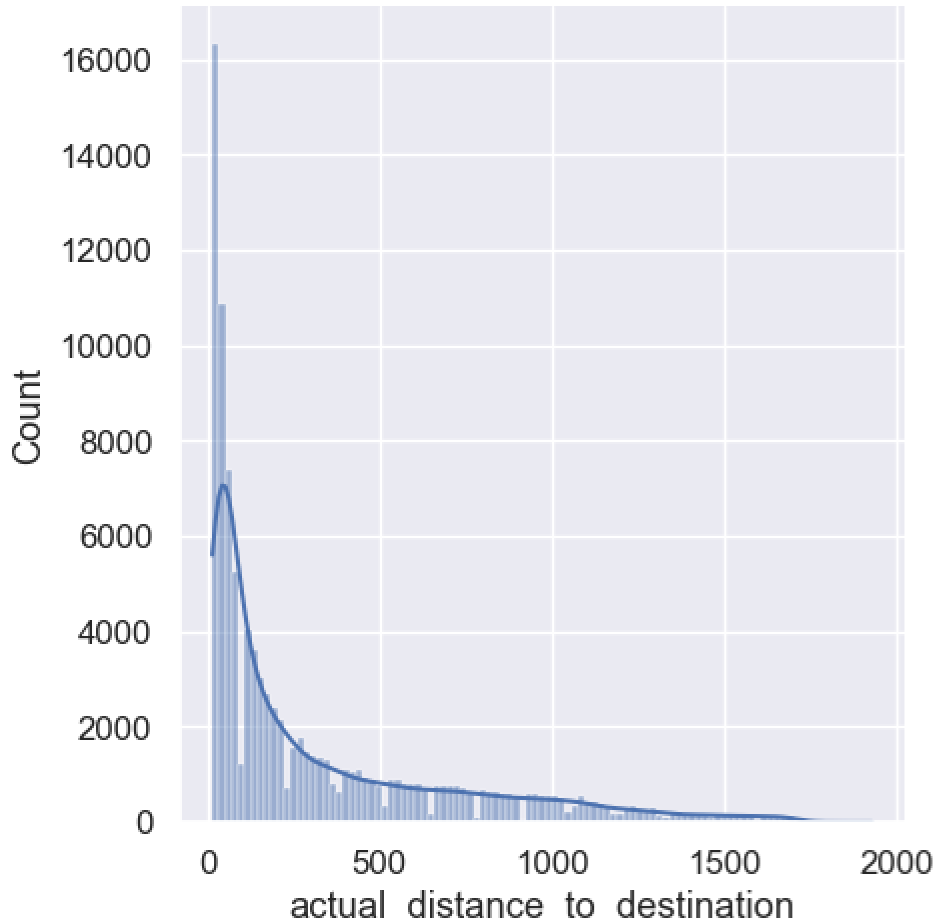
Out[23]:
```
<seaborn.axisgrid.FacetGrid at 0x1a5c5e8d600>
```




- **Observation** :
  - Multi-modal distribution . Need to explore if "start_scan_to_end_scan i.e. Time taken to deliver from source to destination) can be categorized as sub categories such small time ,medium and large time (in consultation with domain experts)

In [24]:
```
sns.displot(df_full_truck_load['actual_distance_to_destination'], kde=True)
```

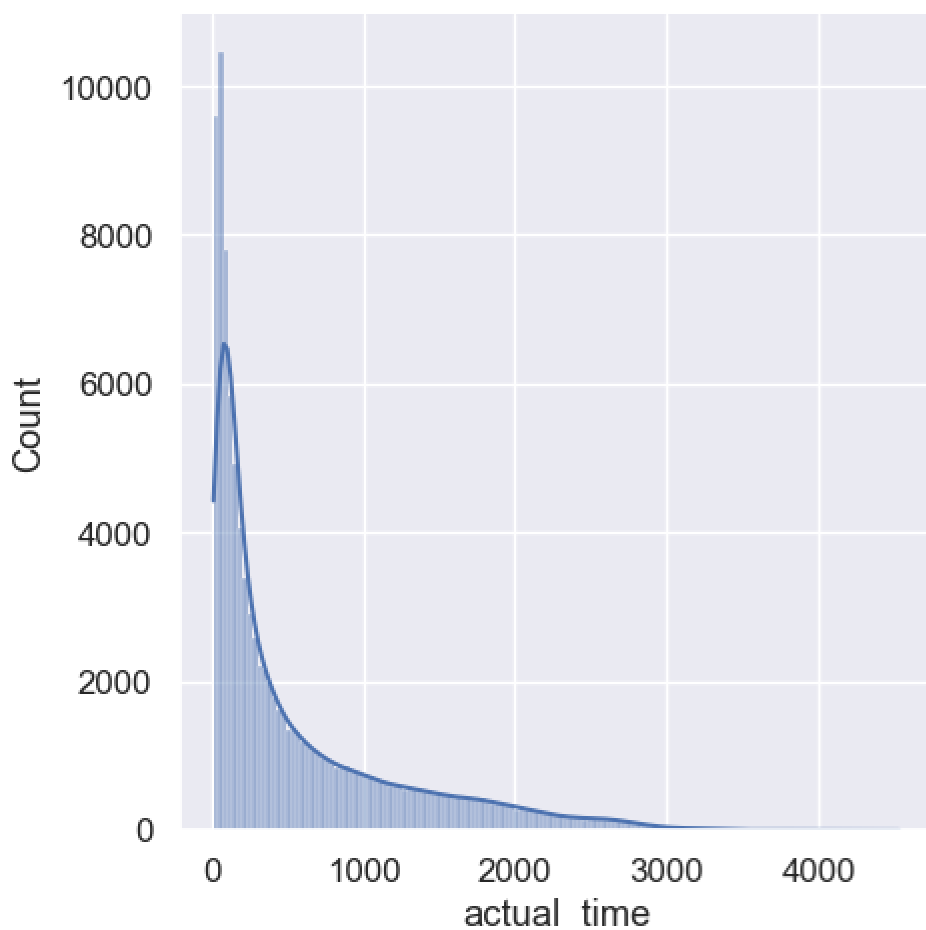
Out[24]:
```
<seaborn.axisgrid.FacetGrid at 0x1a5c5e8da50>
```

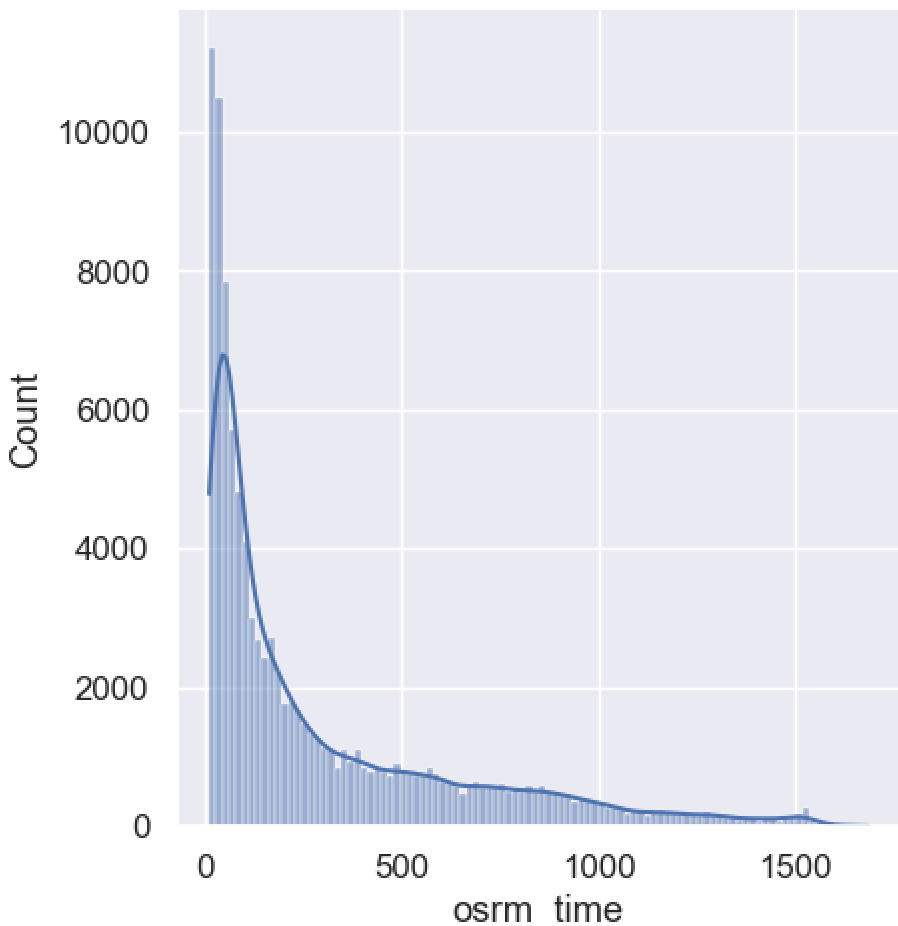



In [25]:

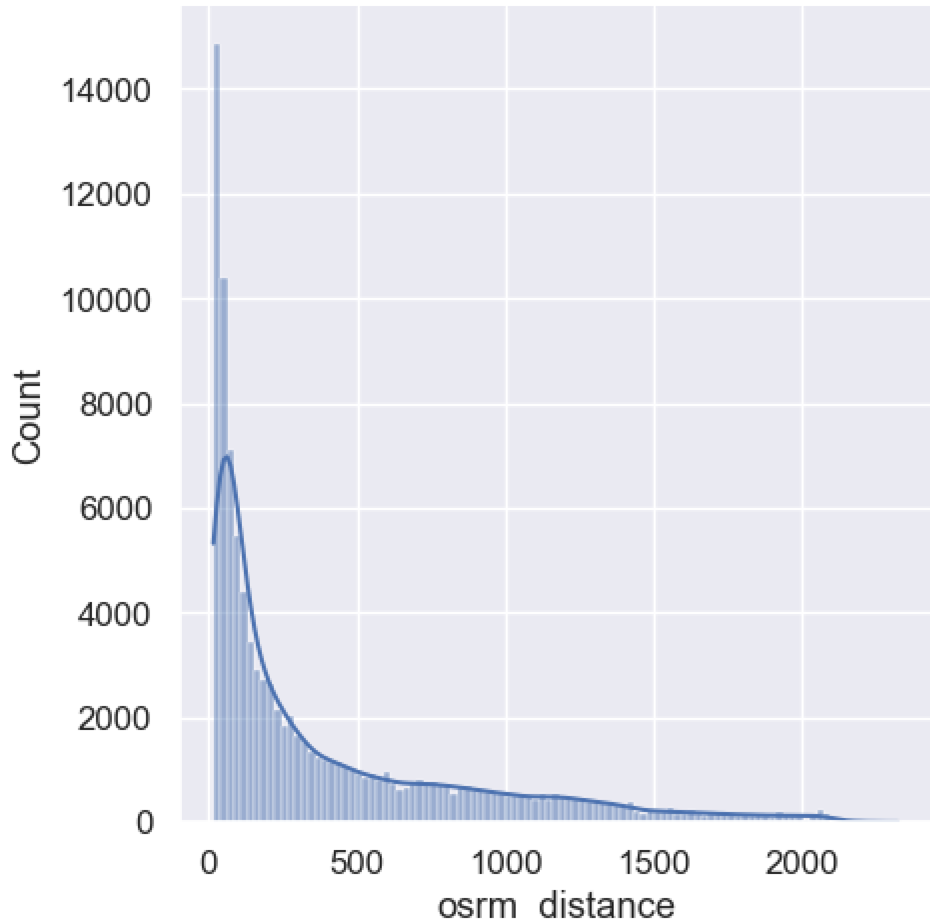
```
sns.displot(df_full_truck_load['actual_time'], kde=True)
```

Out[25]: <seaborn.axisgrid.FacetGrid at 0x1a5ea2d0e50>



In [26]:
```
sns.displot(df_full_truck_load['osrm_time'], kde=True)
```
Out[26]:
```
<seaborn.axisgrid.FacetGrid at 0x1a5ea032c50>
```


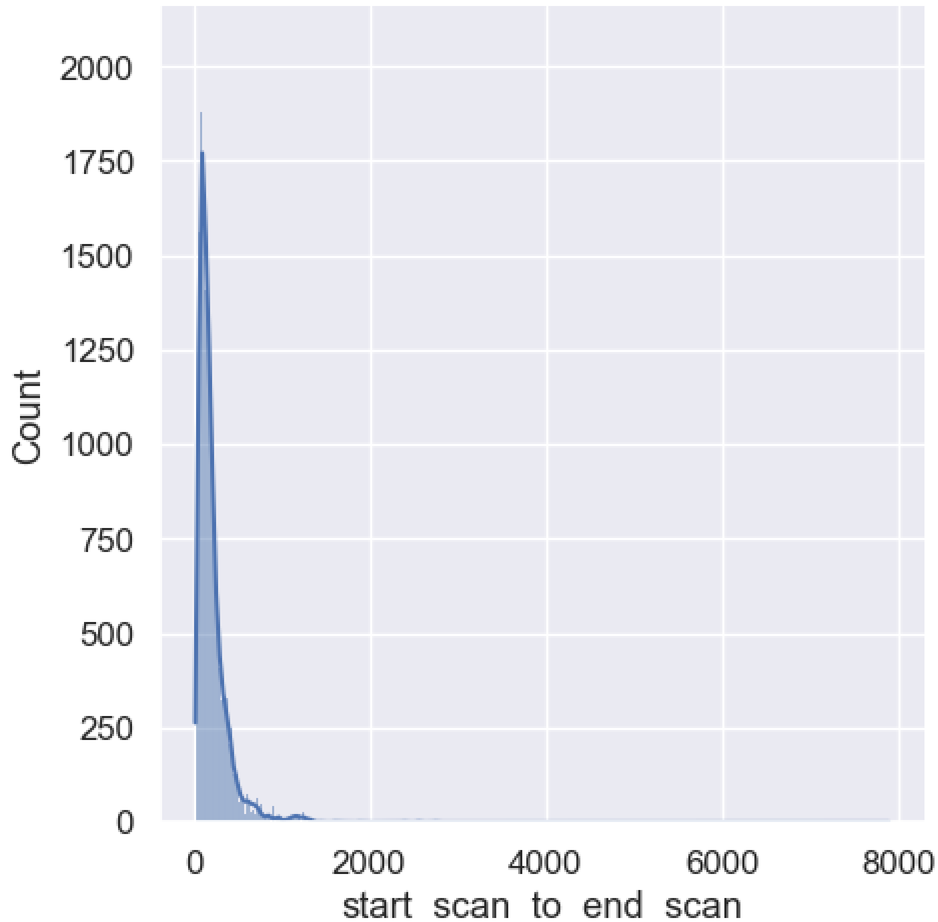

In [27]:
```
sns.displot(df_full_truck_load['osrm_distance'], kde=True)
```
Out[27]:
```
<seaborn.axisgrid.FacetGrid at 0x1a5e8560b20>
```




### Distribution plot of route type - Carting

In [28]:
```
sns.displot(df_carting['start_scan_to_end_scan'], kde=True)
```

Out[28]: <seaborn.axisgrid.FacetGrid at 0x1a5e9c95720>

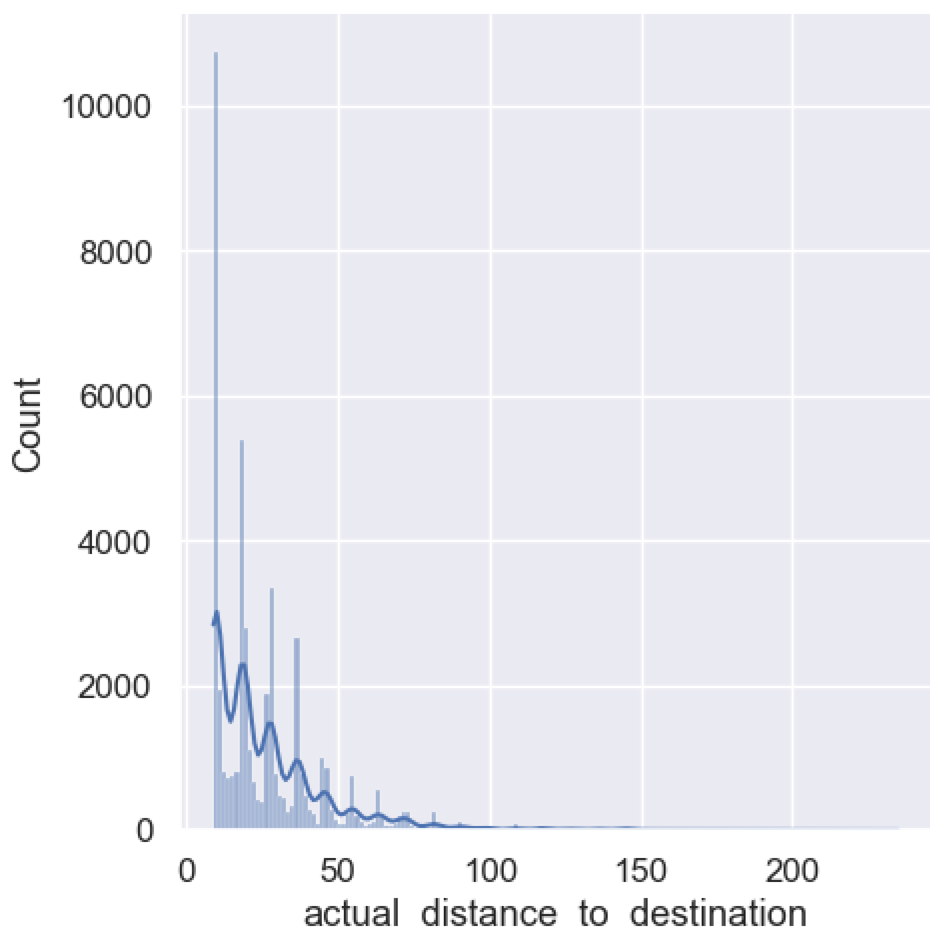



In [29]:

```
sns.displot(df_carting['actual_distance_to_destination'], kde=True)
```

Out[29]:

```
<seaborn.axisgrid.FacetGrid at 0x1a5e82c4700>
```

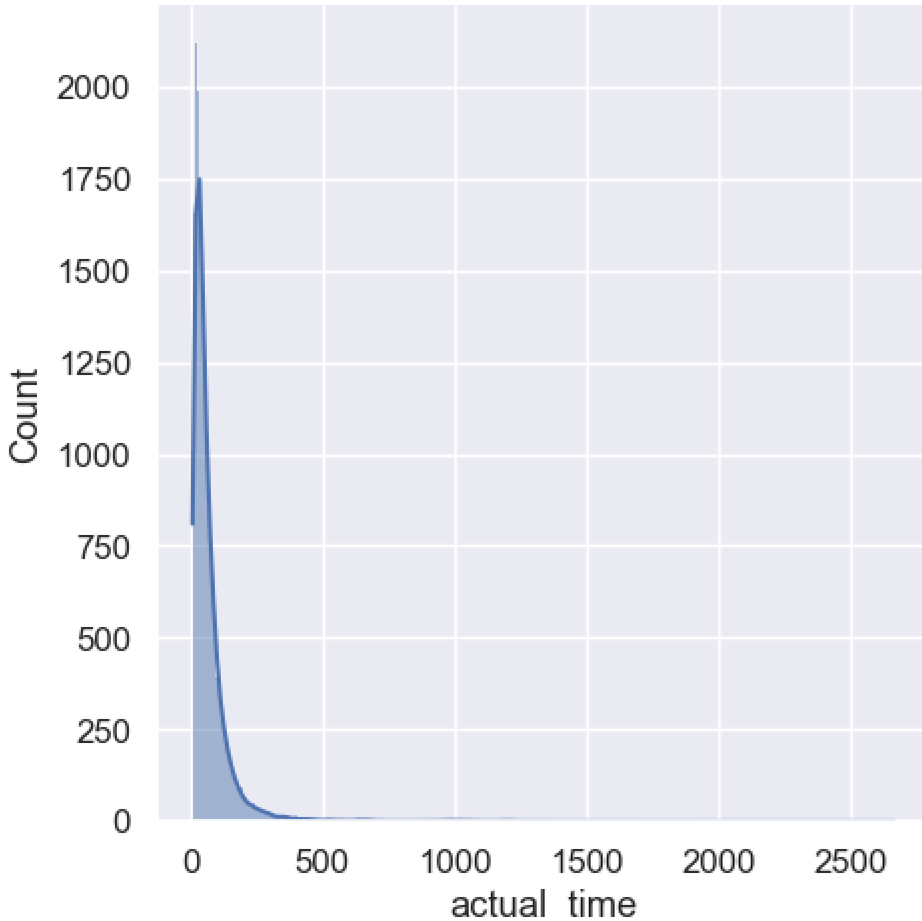


-**Observations**

- Multiple sub category of distances can be created by consulting with domain experts

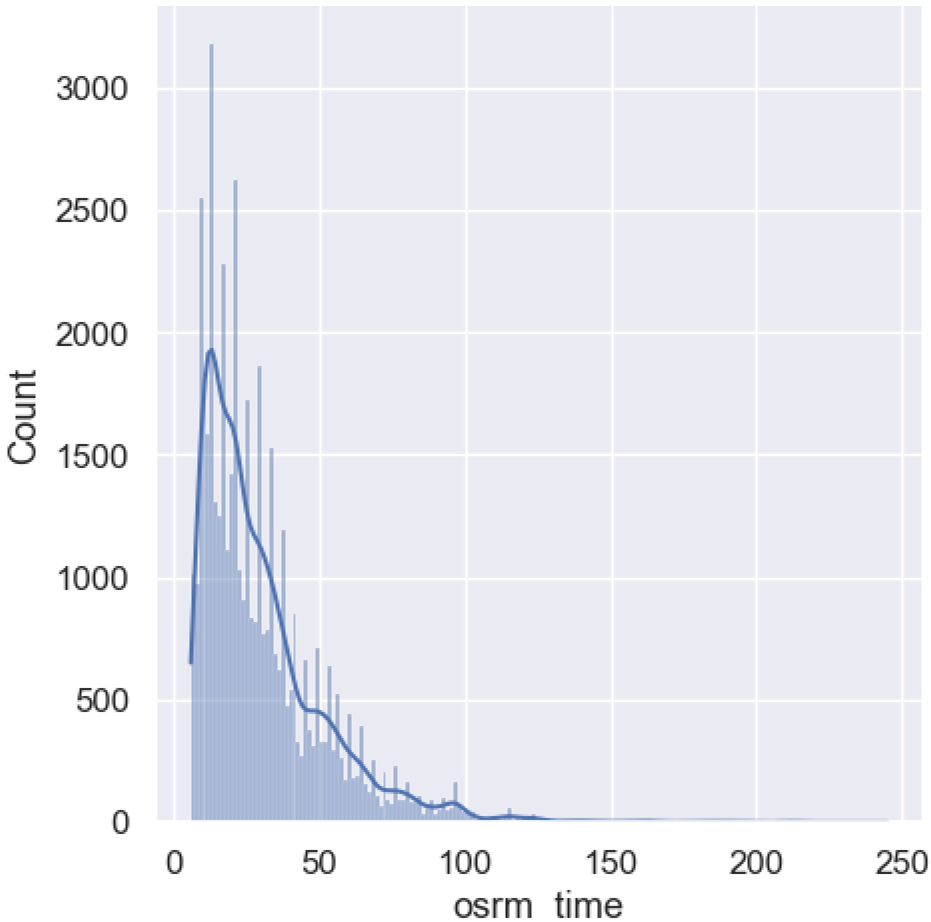
In [30]:
```
sns.displot(df_carting['actual_time'], kde=True)
```

Out[30]: <seaborn.axisgrid.FacetGrid at 0x1a5e816afb0>




In [31]:
```
sns.displot(df_carting['osrm_time'], kde=True)
```

Out[31]: <seaborn.axisgrid.FacetGrid at 0x1a5ebdfd210>

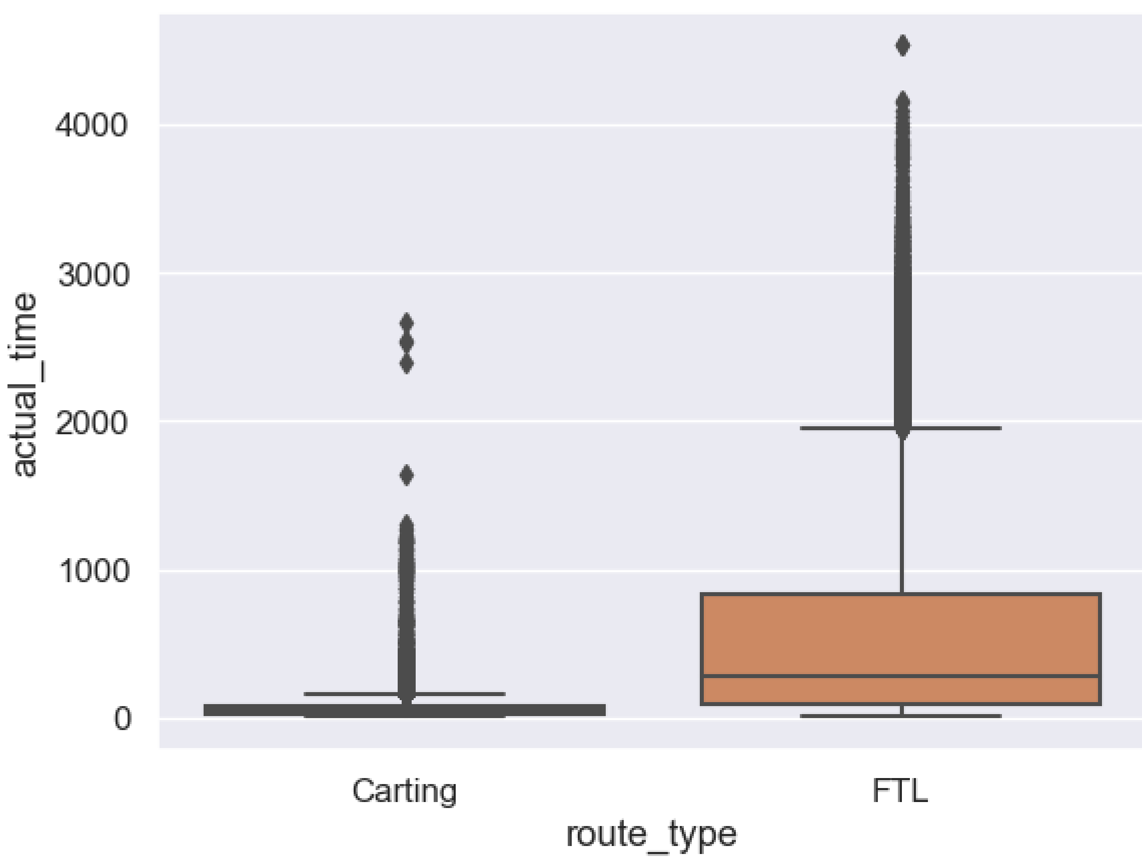



**Boxplot route type vs actual_time**

In [32]:

```
sns.boxplot(x="route_type", y="actual_time", data=df)
```

Out[32]: <AxesSubplot:xlabel='route_type', ylabel='actual_time'>



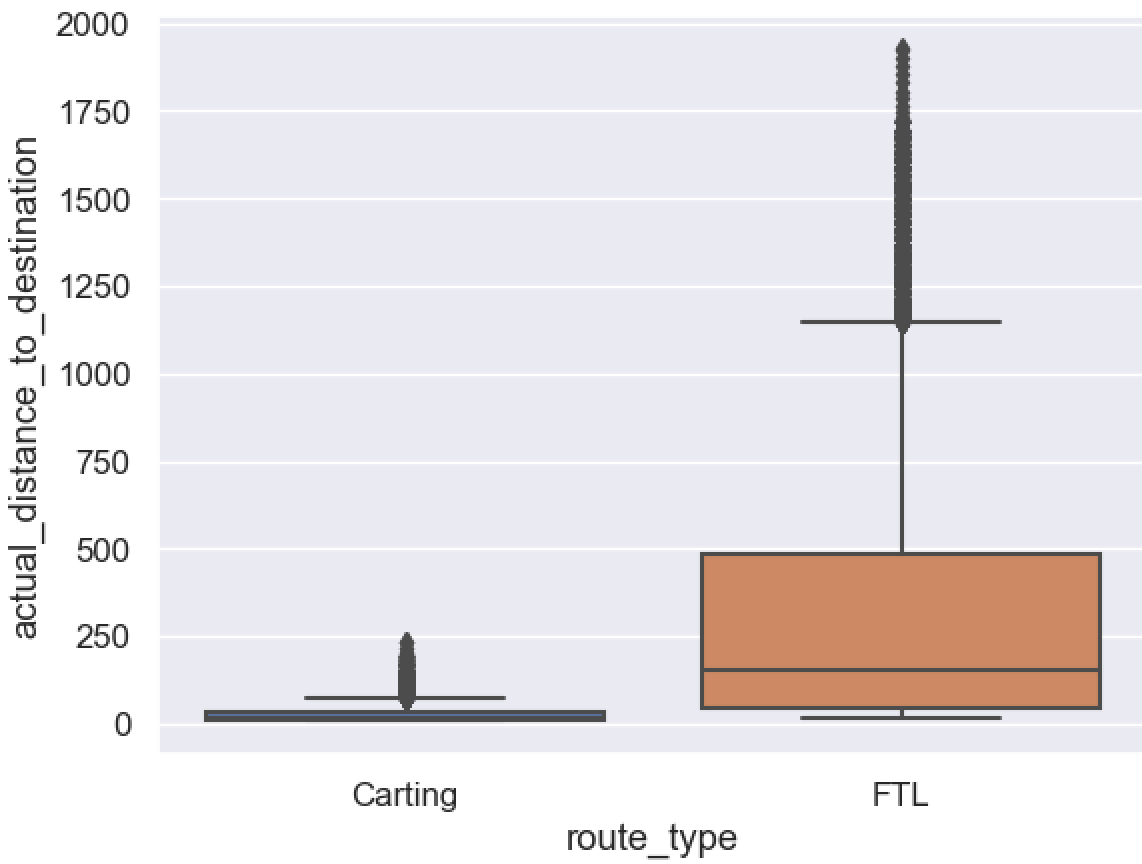
### Boxplot route type vs actual_distance_to_destination

In [33]:

```
sns.boxplot(x="route_type", y="actual_distance_to_destination", data=df)
```

Out[33]:

```
<AxesSubplot:xlabel='route_type', ylabel='actual_distance_to_destination'>
```



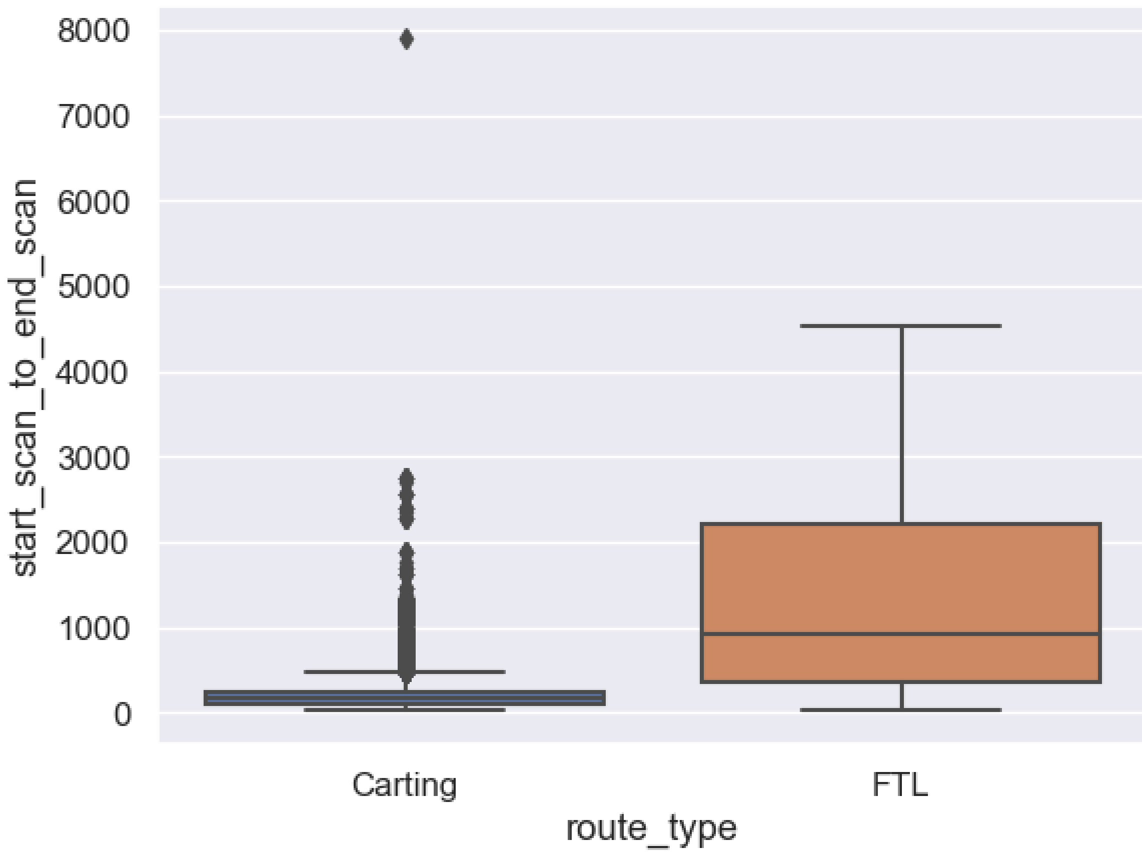

### Boxplot route type vs start_scan_to_end_scan

In [34]:

```
sns.boxplot(x="route_type", y="start_scan_to_end_scan", data=df)
```

Out[34]:

```
<AxesSubplot:xlabel='route_type', ylabel='start_scan_to_end_scan'>
```




- **Insights**
  - Insights based on EDA
    - **Outliers** - Most of the continious features have outliers
    - **Distribution** - **None of the features follow Normal distribution** even for both route category i.e Full truck load and Carting
      - **start_scan_to_end_scan** follow **multi-modal distribution** , very likely it contains multiple sub-category in the distribution
      - **Other continious variables follow Exponential distribution**

## Feature Engineering (Part 1)

- **Trip_creation_time**: Extracting features like month, year and day from trip booking date

In [35]:
```
df["trip_creation_time_year"] = df["trip_creation_time"].dt.year
df["trip_creation_time_month"] = df["trip_creation_time"].dt.month_name()
df["trip_creation_time_weekday"] = df["trip_creation_time"].dt.weekday
df["trip_creation_time_day"] = df["trip_creation_time"].dt.day
```

In [36]:
```
df["od_start_time_year"] = df["od_start_time"].dt.year
df["od_start_time_month"] = df["od_start_time"].dt.month_name()
df["od_start_time_weekday"] = df["od_start_time"].dt.weekday
df["od_start_time_day"] = df["od_start_time"].dt.day
```

In [37]:
```
df["od_end_time_year"] = df["od_end_time"].dt.year
df["od_end_time_month"] = df["od_end_time"].dt.month_name()
df["od_end_time_weekday"] = df["od_end_time"].dt.weekday
df["od_end_time_day"] = df["od_end_time"].dt.day
```

- Feature extraction from **"Source Name"** to **City, Place, Code and State**

In [38]:
```
df[['source_name_city','source_name_place','source_name_code','source_name_state']] = d
```

In [39]:
```
df[['source_name','source_name_city','source_name_place','source_name_code','source_name
```

Out[39]:

| | source_name | source_name_city | source_name_place | source_name_code | source_name_state |
|---|---|---|---|---|---|
| **0** | Anand_VUNagar_DC (Gujarat) | Anand | VUNagar | DC | Gujarat |
| **1** | Anand_VUNagar_DC (Gujarat) | Anand | VUNagar | DC | Gujarat |
| **2** | Anand_VUNagar_DC (Gujarat) | Anand | VUNagar | DC | Gujarat |
| **3** | Anand_VUNagar_DC (Gujarat) | Anand | VUNagar | DC | Gujarat |
| **4** | Anand_VUNagar_DC (Gujarat) | Anand | VUNagar | DC | Gujarat |
| **...** | ... | ... | ... | ... | ... |
| **144862** | Sonipat_Kundli_H (Haryana) | Sonipat | Kundli | H | Haryana |
| **144863** | Sonipat_Kundli_H (Haryana) | Sonipat | Kundli | H | Haryana |
| **144864** | Sonipat_Kundli_H (Haryana) | Sonipat | Kundli | H | Haryana |
| **144865** | Sonipat_Kundli_H (Haryana) | Sonipat | Kundli | H | Haryana |
| **144866** | Sonipat_Kundli_H (Haryana) | Sonipat | Kundli | H | Haryana |

144867 rows × 5 columns

## Data cleanup | source_name_city | Level Bangalore , Bengaluru to Bengaluru

In [40]:
```
# Re-leveling source city from Bangalore to Bengaluru as both signifies same city
df["source_name_city"]= df["source_name_city"].str.replace("Bangalore", "Bengaluru", ca
df["source_name_city"].replace(to_replace=r'(^Del$)', value='Delhi', regex=True,inplace
```

- Feature extraction from **"Destination Name"** to **City, Place, Code and State**

In [41]:
```
df[['destination_name_city','destination_name_place','destination_name_code','destinati
```

In [42]:
```
df[['destination_name_city','destination_name_place','destination_name_code','destinati
```

Out[42]:

| | destination_name_city | destination_name_place | destination_name_code | destination_name_state |
|---|---|---|---|---|
| **0** | Khambhat | MotvdDPP | D | Gujarat |
| **1** | Khambhat | MotvdDPP | D | Gujarat |
| **2** | Khambhat | MotvdDPP | D | Gujarat |
| **3** | Khambhat | MotvdDPP | D | Gujarat |
| **4** | Khambhat | MotvdDPP | D | Gujarat |
| **...** | ... | ... | ... | ... |
| **144862** | Gurgaon | Bilaspur | HB | Haryana |
| **144863** | Gurgaon | Bilaspur | HB | Haryana |
| **144864** | Gurgaon | Bilaspur | HB | Haryana |
| **144865** | Gurgaon | Bilaspur | HB | Haryana |
| **144866** | Gurgaon | Bilaspur | HB | Haryana |

144867 rows × 4 columns

### Data cleanup | destination_name_city | Level Bangalore , Bengaluru to Bengaluru

In [43]:

```
# Re-leveling destination city from Bangalore to Bengaluru as both signifies same city
df["destination_name_city"]= df["destination_name_city"].str.replace("Bangalore", "Benga
```

## Non-Graphical Analysis (Part 2)

In [44]:

```
df_full_truck_load = df[df["route_type"] == 'FTL']
df_carting = df[df["route_type"] == 'Carting']
```

In [45]:

```
# Viewing top proportion of orders based on source and destination states
source_destination_state_ct= pd.crosstab(df_full_truck_load["source_name_state"],df_full
source_destination_state_ct.sort_values('All', ascending=False).iloc[:5]
```

Out[45]:

| destination_name_state | Assam | Bihar | Chandigarh | Chhattisgarh | Delhi | Gujarat | Haryana | Jharkhand | Karnataka |
|---|---|---|---|---|---|---|---|---|---|
| **source_name_state** | | | | | | | | | |
| **All** | 0.59 | 4.05 | 0.1 | 0.25 | 2.79 | 1.88 | 21.45 | 0.89 | 23.26 |
| **Haryana** | 0.00 | 1.78 | 0.0 | 0.00 | 0.19 | 0.95 | 1.24 | 0.00 | 12.27 |
| **Maharashtra** | 0.00 | 0.00 | 0.0 | 0.00 | 2.38 | 0.00 | 7.23 | 0.00 | 3.91 |
| **Karnataka** | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 8.37 | 0.00 | 4.89 |
| **Telangana** | 0.00 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.70 |

In [46]:

```
# Viewing top proportion of orders based on source and destination cities
source_destination_city_ct= pd.crosstab(df_full_truck_load["source_name_city"],df_full_t
source_destination_city_ct.sort_values('All', ascending=False).iloc[:5]
```

Out[46]:

| destination_name_city | Achrol | Agartala | Aizawl | Ajmer | Akola | Almora | Aluva | Ambajogai | Amreli | AnandprShb |
|---|---|---|---|---|---|---|---|---|---|---|
| **source_name_city** | | | | | | | | | | |
| **All** | 0.03 | 0.01 | 0.02 | 0.01 | 1.06 | 0.05 | 0.67 | 0.03 | 0.15 | 0.09 |
| **Gurgaon** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **Bhiwandi** | 0.00 | 0.00 | 0.00 | 0.00 | 1.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **Bengaluru** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.62 | 0.00 | 0.00 | 0.00 |
| **Pune** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

5 rows × 317 columns

In [47]:

```
# Viewing top proportion of orders based on source and destination place
source_destination_place_ct= pd.crosstab(df_full_truck_load["source_name_place"],df_full
source_destination_place_ct.sort_values(by = ['All', 'source_name_place'], ascending =
```

Out[47]:

| destination_name_place | Adargchi | AdrshSt | AgrohDPP | Airport | AkkolRD | Alngjuri | AmbedDPP | AnugrDPP | Aravi |
|---|---|---|---|---|---|---|---|---|---|
| **source_name_place** | | | | | | | | | |
| **All** | 0.72 | 0.0 | 0.05 | 2.79 | 0.05 | 0.06 | 0.01 | 0.02 | 0 |
| **Bilaspur** | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| **Mankoli** | 0.00 | 0.0 | 0.00 | 2.38 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| **Nelmngla** | 0.11 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |
| **Tathawde** | 0.23 | 0.0 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0 |

5 rows × 319 columns

## Visual Analysis (Part 2) - Post Initial Feature engineering

In [515…

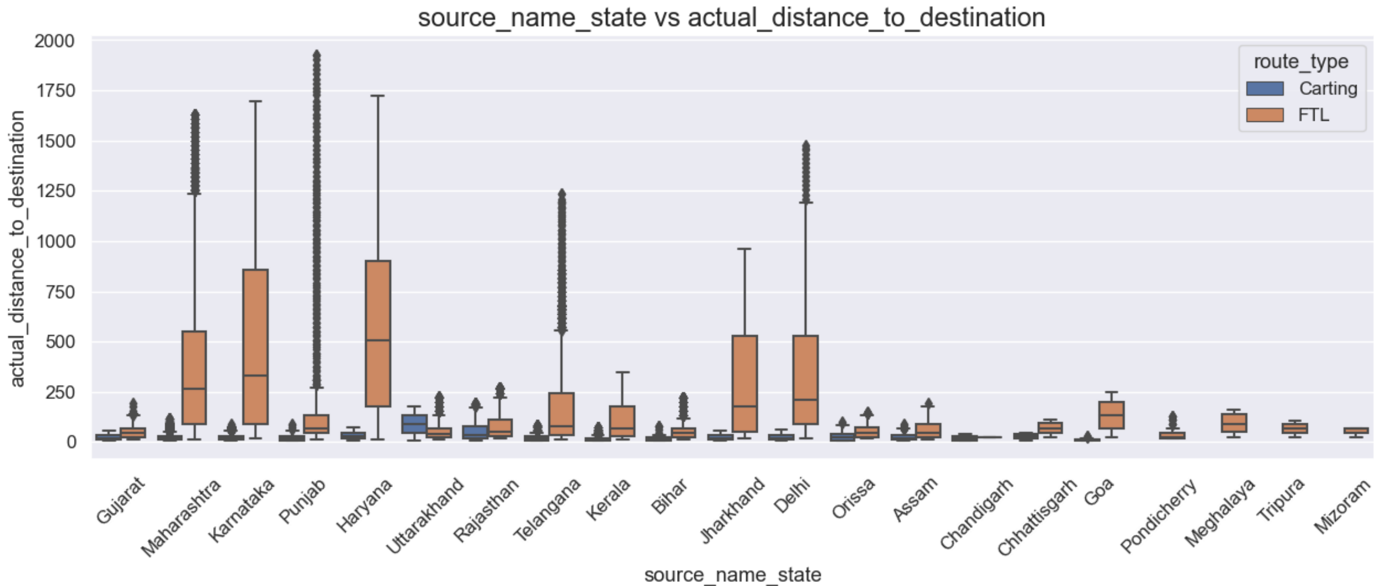
```
# selecting specific categorical features
selected_categorical_features = ['source_name_state','destination_name_state','trip_crea
       'od_start_time_day','od_end_time_weekday', 'od_end_time_day']
```

In [516…

```
# selecting specific numeric/continious features
selected_continious_features = ['start_scan_to_end_scan', 'actual_distance_to_destinatio
       'osrm_distance']
```

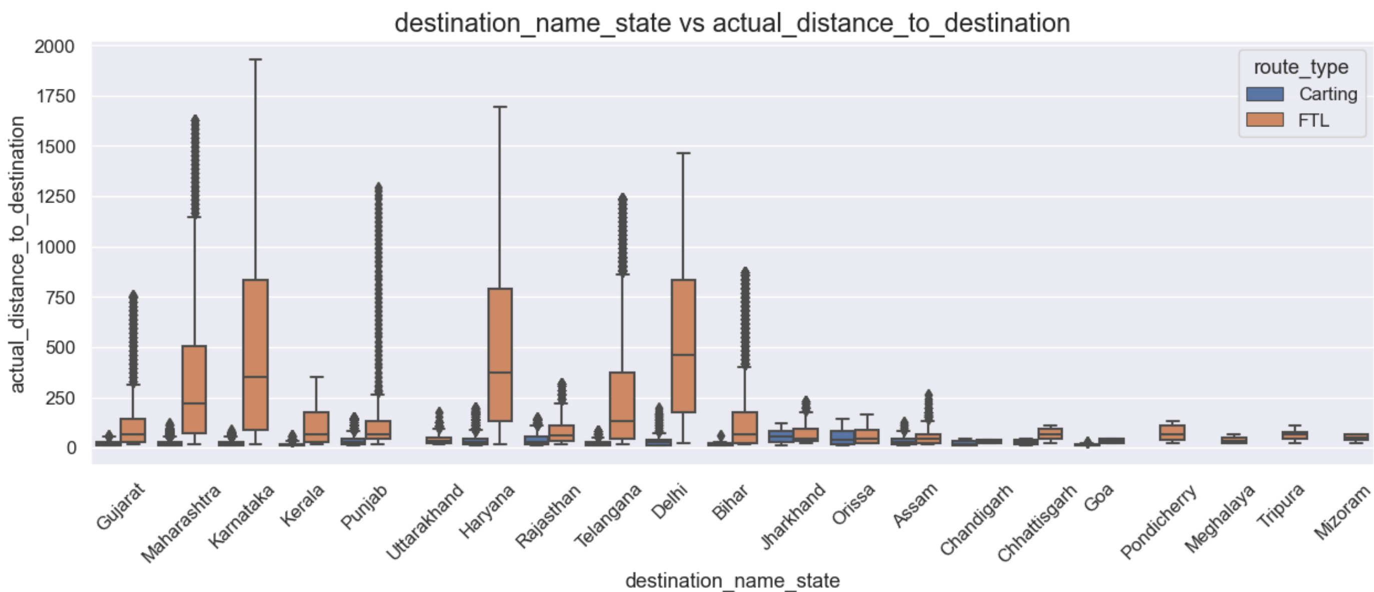
In [201…

```
# plotting box plots and adding title , rotation for better visibility of levels i.e sta
fig, axes = plt.subplots(figsize=(15,5))
plt.title("source_name_state vs actual_distance_to_destination",fontdict ={"fontsize": 1
plt.xticks(rotation = 45)
sns.boxplot(x="source_name_state", y="actual_distance_to_destination", hue="route_type",
plt.show()
```



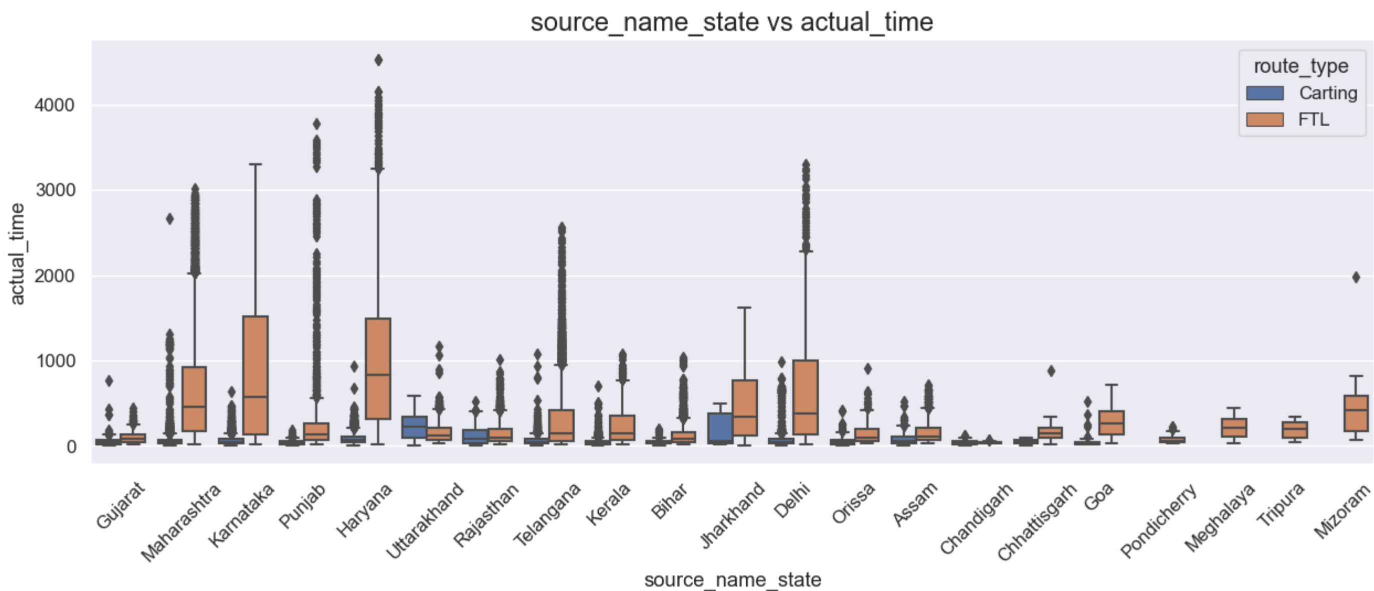

In [202…

```
fig, axes = plt.subplots(figsize=(15,5))
plt.title("destination_name_state vs actual_distance_to_destination",fontdict ={"fontsiz
plt.xticks(rotation = 45)
sns.boxplot(x="destination_name_state", y="actual_distance_to_destination", hue="route_t
plt.show()
```




In [203…

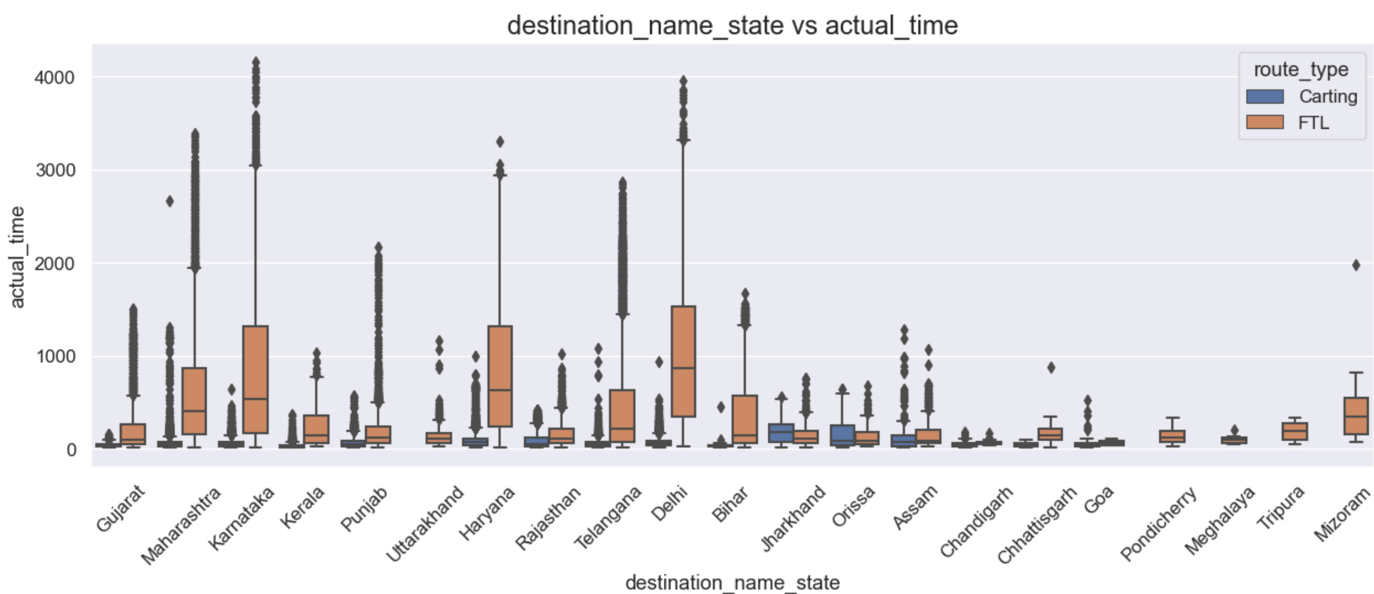
```
fig, axes = plt.subplots(figsize=(15,5))
plt.title("source_name_state vs actual_time",fontdict ={"fontsize": 17})
plt.xticks(rotation = 45)
sns.boxplot(x="source_name_state", y="actual_time", hue="route_type", data=df)
plt.show()
```




- **Observation** :
  - Relatively **more Carting trips originated from "Uttarakhand" , "Rajasthan" , "Jharkhand"** etc.

In [204…

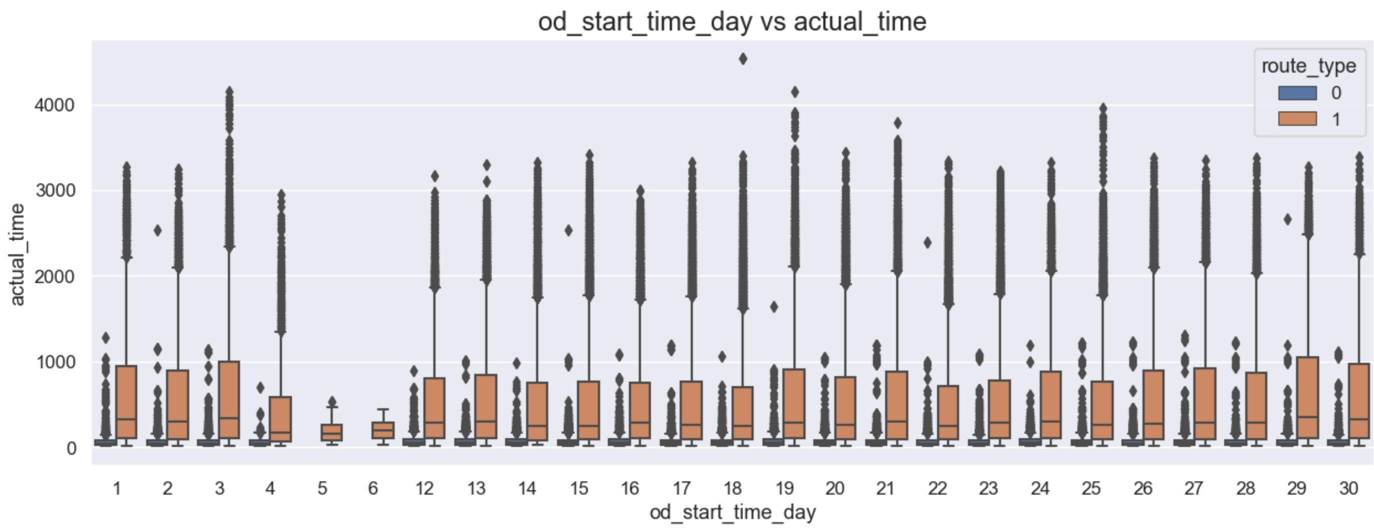
```
fig, axes = plt.subplots(figsize=(15,5))
plt.title("destination_name_state vs actual_time",fontdict ={"fontsize": 17})
plt.xticks(rotation = 45)
sns.boxplot(x="destination_name_state", y="actual_time", hue="route_type", data=df)
plt.show()
```




- **Observation** :
  - Relatively more **Carting trips** booked **for destination "Jharkhand" , "Orissa" and Assam**

In [293…

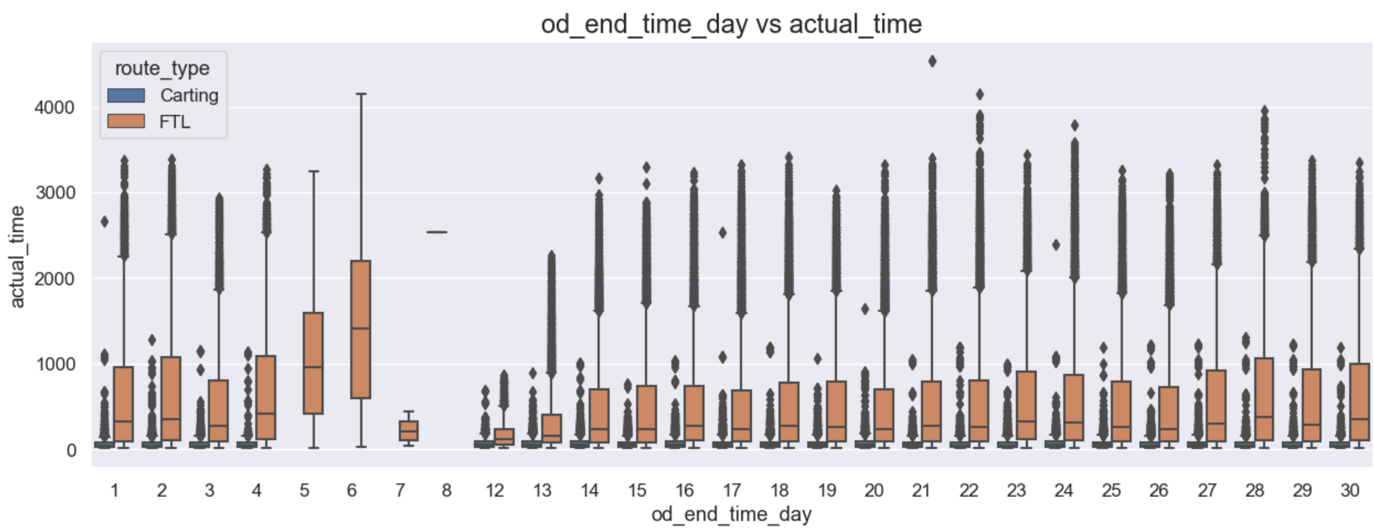
```
fig, axes = plt.subplots(figsize=(15,5))
plt.title("od_start_time_day vs actual_time",fontdict ={"fontsize": 17})
sns.boxplot(x="od_start_time_day", y="actual_time", hue="route_type", data=df)
plt.show()
```




- **Observation** :
  - Relatively less full truck loads trips starting on 5th or 6th day of the month

In [206…

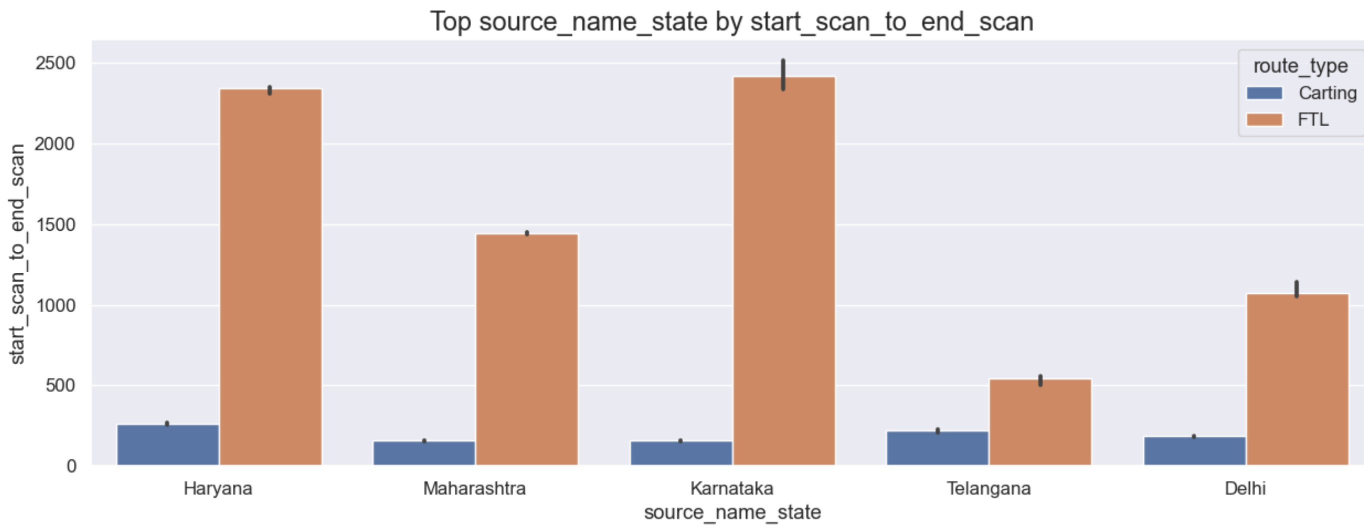
```
fig, axes = plt.subplots(figsize=(15,5))
plt.title("od_end_time_day vs actual_time",fontdict ={"fontsize": 17})
sns.boxplot(x="od_end_time_day", y="actual_time", hue="route_type", data=df)
plt.show()
```




- **Observation** :
  - Relatively **large full truck loads** trips **ending on 5th or 6th day of the month**
  - Relatively **less full truck loads** trips **ending on 7th or 8th day of the month**

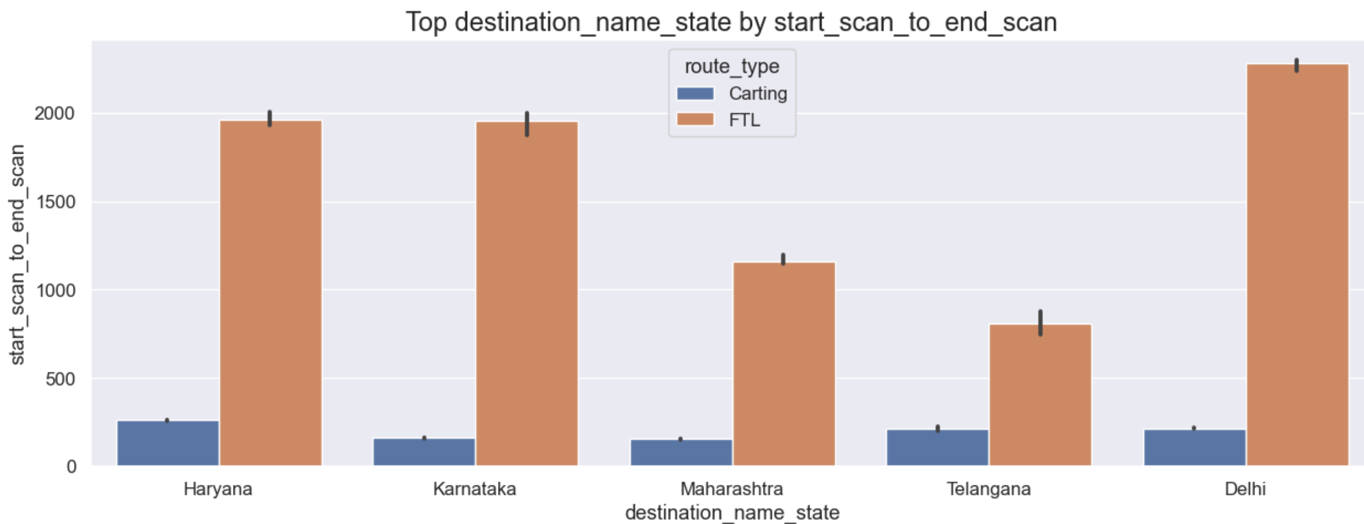
In [207…

```
# Showing top 5 source_names which registers max time of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("Top source_name_state by start_scan_to_end_scan",fontdict ={"fontsize": 17})
sns.barplot(x="source_name_state", y="start_scan_to_end_scan", hue="route_type", data=di
plt.show()
```



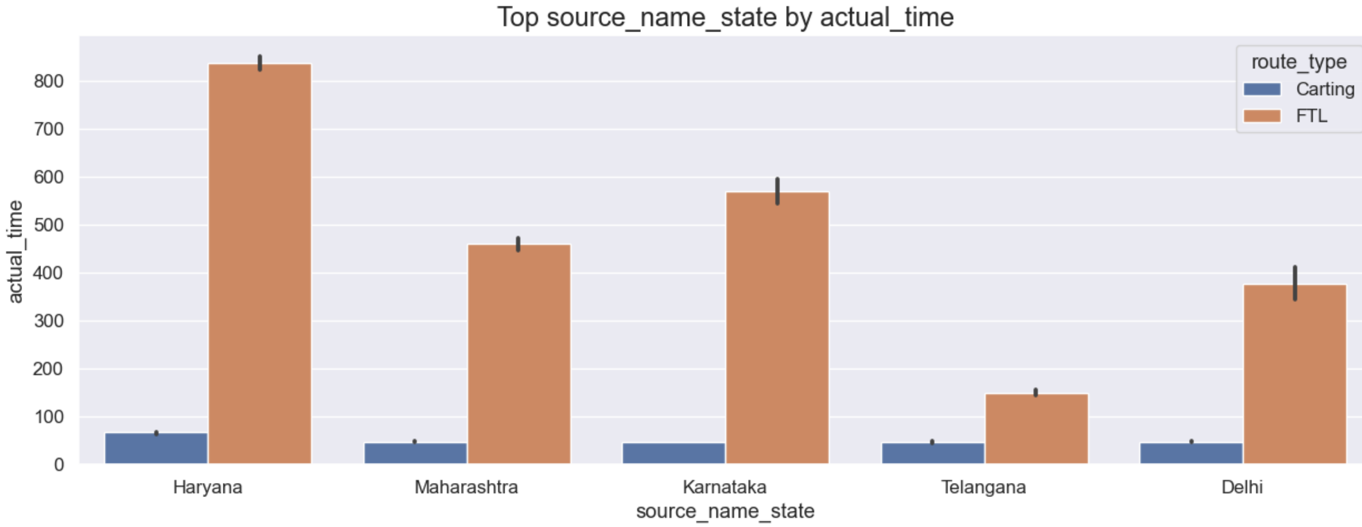

In [208…

```
# Showing top 5 destination_name_states which registers max time of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("Top destination_name_state by start_scan_to_end_scan",fontdict ={"fontsize":
# estimator used as median as there are many outliers in the data
sns.barplot(x="destination_name_state", y="start_scan_to_end_scan", hue="route_type", da
plt.show()
```



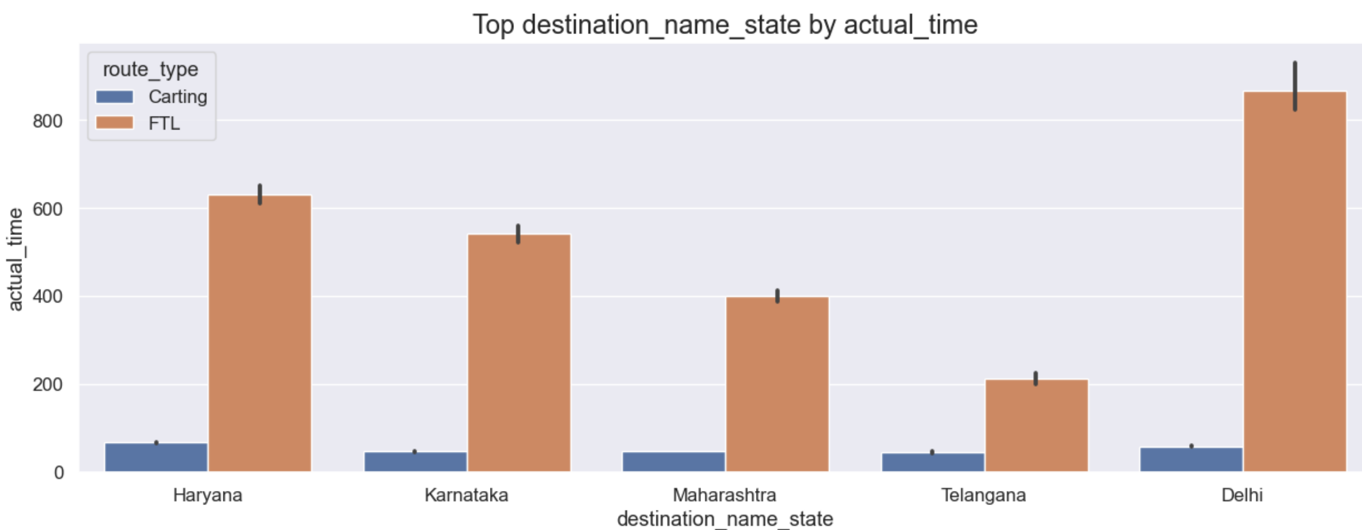

In [209…

```
# Showing top 5 source_name_states which registers max actual time of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("Top source_name_state by actual_time",fontdict ={"fontsize": 17})
sns.barplot(x="source_name_state", y="actual_time", hue="route_type", data=df, order=df
plt.show()
```



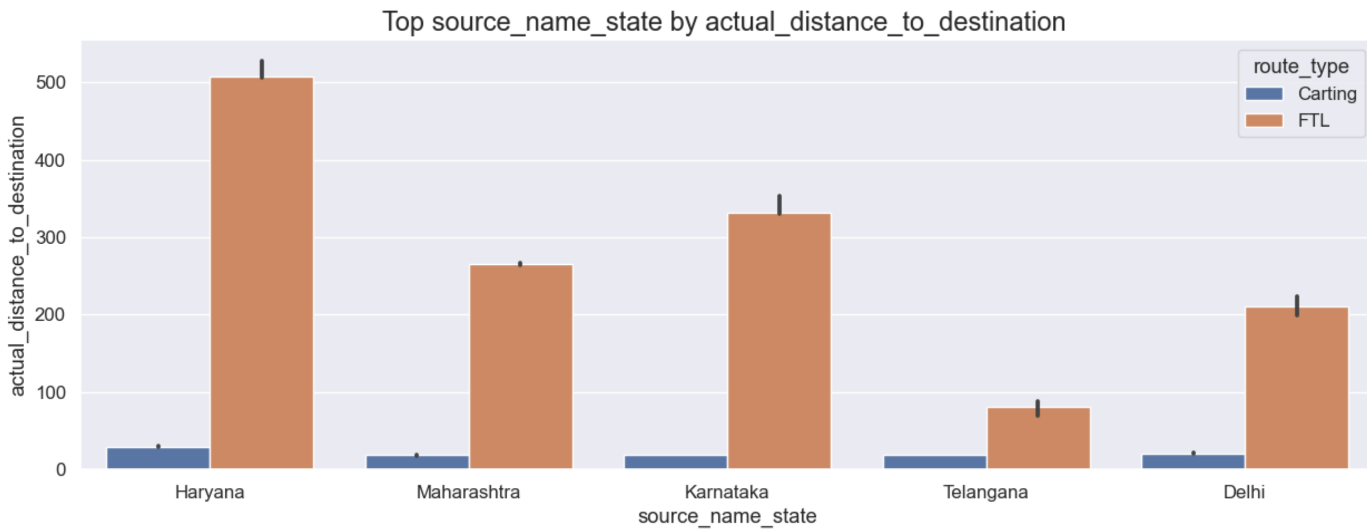

In [210…

```
# Showing top 5 destination_name_states which registers max actual time of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("Top destination_name_state by actual_time",fontdict ={"fontsize": 17})
sns.barplot(x="destination_name_state", y="actual_time", hue="route_type", data=df, orde
plt.show()
```



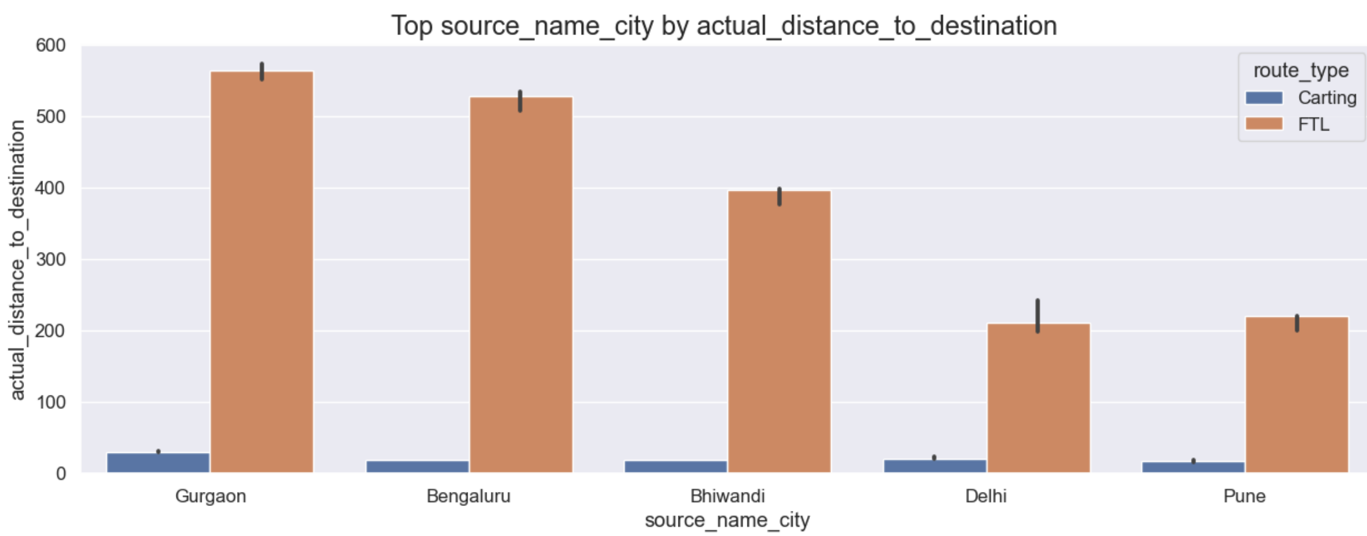

In [211…

```
# Showing top 5 source_name_states which registers max actual distance of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("Top source_name_state by actual_distance_to_destination ",fontdict ={"fontsiz
sns.barplot(x="source_name_state", y="actual_distance_to_destination", hue="route_type",
plt.show()
```



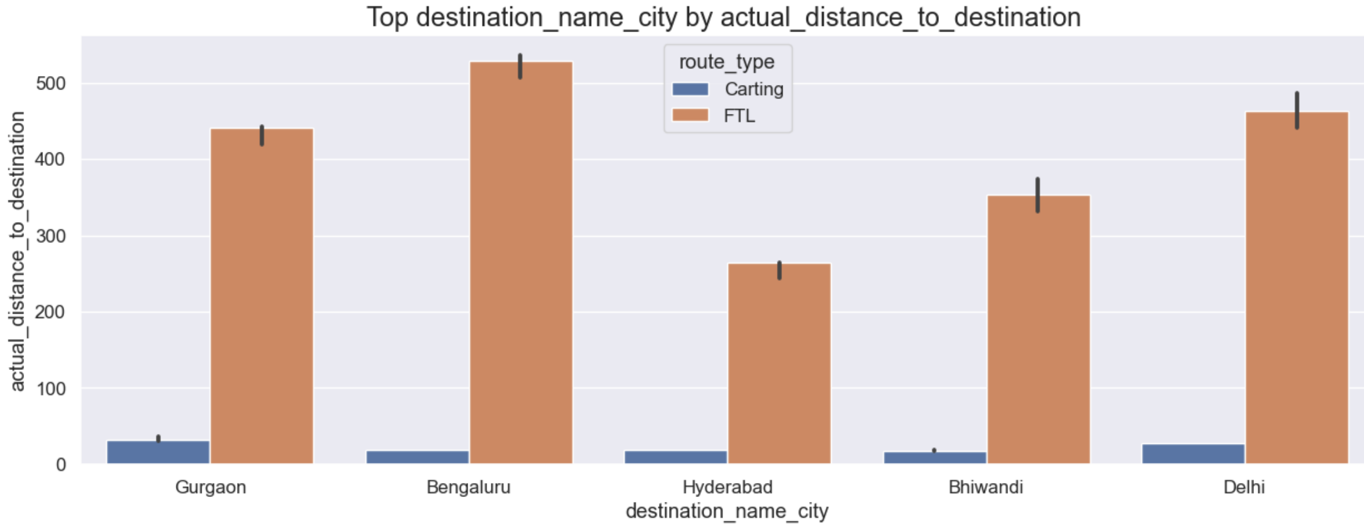

In [212…

```
# Showing top 5 source_name_cities which registers max actual distance of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("Top source_name_city by actual_distance_to_destination ",fontdict ={"fontsize
sns.barplot(x="source_name_city", y="actual_distance_to_destination", hue="route_type",
plt.show()
```



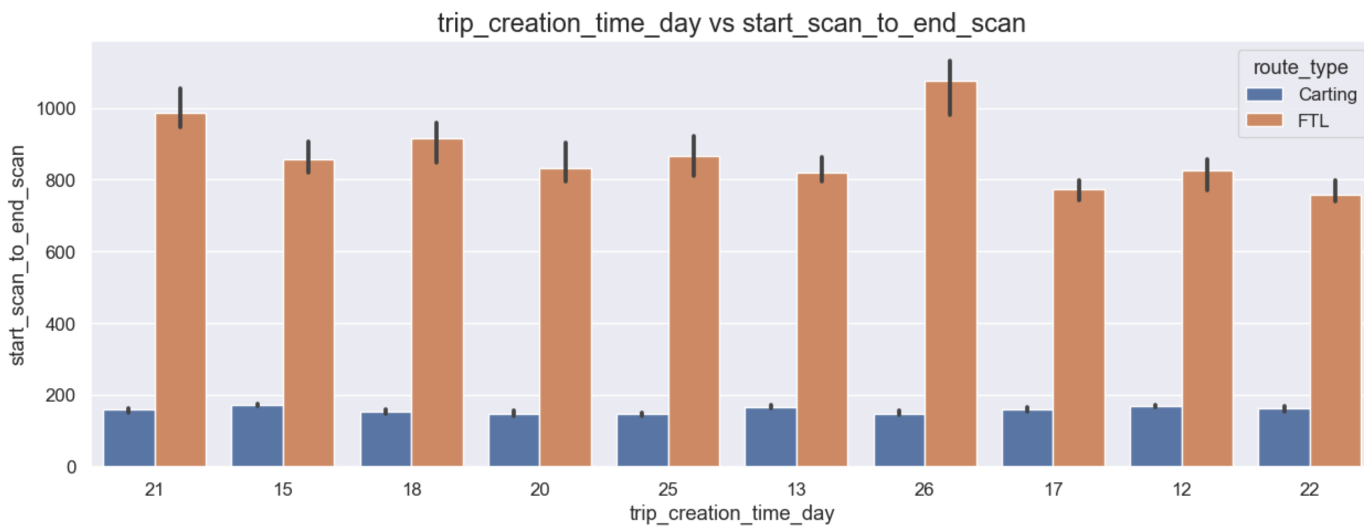

In [213…

```
# Showing top 5 destination_name_cities which registers max actual distance of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("Top destination_name_city by actual_distance_to_destination ",fontdict ={"fon
sns.barplot(x="destination_name_city", y="actual_distance_to_destination", hue="route_ty
plt.show()
```




In [214…

```
# Showing top 10 trip creation days which registers max time of delivery
fig, axes = plt.subplots(figsize=(15,5))
plt.title("trip_creation_time_day vs start_scan_to_end_scan",fontdict ={"fontsize": 17})
sns.barplot(x="trip_creation_time_day", y="start_scan_to_end_scan", hue="route_type", da
plt.show()
```




**Observations**

- **Most Time taken**(i.e. for route type full truck load) to **deliver from source to destination** are from **Haryana , Karnataka , Maharshtra , Delhi , Telangana**
- **Actual time taken to complete the delivery** are high for source/destination **Haryana , Karnataka , Maharshtra , Delhi , Telangana**
- **Most distanced cities** for full truck deliveries are **Gurgaon , Bengaluru , Bhiwandi**
- **Most trips are being booked on 12,13,15,17,18,20,21,22,25,26** days of the month

# Aggregation of fields and Merging of rows

## Aggregation and Merge by trip_uuid

In [48]:

```
# Aggregation by trip_uuid
df_agg = df.groupby("trip_uuid")[["actual_time", "actual_distance_to_destination","osrm_
df_agg.reset_index(inplace=True)
df_agg.rename(columns = {'actual_time':'actual_time_aggregated','actual_distance_to_dest
df_agg.head()
```

Out[48]:

| | trip_uuid | actual_time_aggregated | osrm_time_aggregated | segment_osrm_time_aggregated | segment_a |
|---|---|---|---|---|---|
| 0 | trip-153671041653548748 | 15682.0 | 7787.0 | 1008.0 | |
| 1 | trip-153671042288605164 | 399.0 | 210.0 | 65.0 | |
| 2 | trip-153671043369099517 | 112225.0 | 65768.0 | 1941.0 | |
| 3 | trip-153671046011330457 | 82.0 | 24.0 | 16.0 | |
| 4 | trip-153671052974046625 | 556.0 | 207.0 | 115.0 | |

In [49]:

```
# Selecting specific features from non aggregated data , before merging
df_org = df[["trip_uuid","route_type","trip_creation_time_year","trip_creation_time_mont
df_org.shape
```

Out[49]: (144867, 6)

In [50]:

```
# Removing duplicate datas , keep first occurance of the data
df_org = df_org.drop_duplicates(keep='first')
df_org.shape
```

Out[50]: (14817, 6)

In [51]:

```
# viewing selected rows post removing duplicate rows
df_org.head()
```

Out[51]:

| | trip_uuid | route_type | trip_creation_time_year | trip_creation_time_month | trip_creation_time_weekday |
|---|---|---|---|---|---|
| 0 | trip-153741093647649320 | Carting | 2018 | September | 3 |
| 10 | trip-153768492602129387 | FTL | 2018 | September | 6 |
| 15 | trip-153693976643699843 | Carting | 2018 | September | 4 |
| 17 | trip-153687145942424248 | FTL | 2018 | September | 3 |
| 35 | trip-153825970514894360 | FTL | 2018 | September | 5 |

In [52]:

```
# Removing duplicate records post aggregation
df_agg = df_agg.drop_duplicates(keep='first')
df_agg.shape
```

Out[52]: (14817, 7)

In [53]:

```
# Merging aggregated and raw features
df_agg = pd.merge(df_agg, df_org, how="inner", on="trip_uuid")
df_agg.shape
```

Out[53]: (14817, 12)

In [54]:
```
df_agg.head()
```

Out[54]:

| | trip_uuid | actual_time_aggregated | osrm_time_aggregated | segment_osrm_time_aggregated | segment_a |
|---|---|---|---|---|---|
| 0 | trip-153671041653548748 | 15682.0 | 7787.0 | 1008.0 | |
| 1 | trip-153671042288605164 | 399.0 | 210.0 | 65.0 | |
| 2 | trip-153671043369099517 | 112225.0 | 65768.0 | 1941.0 | |
| 3 | trip-153671046011330457 | 82.0 | 24.0 | 16.0 | |
| 4 | trip-153671052974046625 | 556.0 | 207.0 | 115.0 | |

## Aggregation and Merge by 'trip_uuid','source_name', 'destination_name' - Select first and last

- **Aggregation by 'trip_uuid','source_center', 'destination_center'**

In [57]:
```
# Aggregation by trip_uuid , source_center, destination_center
#df_agg_by_trip_src_dest= df.groupby(['trip_uuid','source_center', 'destination_center']
df_agg_by_trip_src_dest_center= df.groupby(['trip_uuid','source_center', 'destination_ce

# Unfolding grouped data to rows
df_agg_by_trip_src_dest_center.reset_index(inplace=True)
# Renaming grouped columns
df_agg_by_trip_src_dest_center.columns = ['_'.join(col) for col in df_agg_by_trip_src_de
# Removing underscore from column names source_center_ and  destination_center_
df_agg_by_trip_src_dest_center.rename(columns = {'trip_uuid_':'trip_uuid','source_cente
# Merge by trip_uuid
df_agg_by_trip_src_dest_center_merged = pd.merge(df_agg_by_trip_src_dest_center, df_org,
df_agg_by_trip_src_dest_center_merged.head()
```

Out[57]:

| | trip_uuid | source_center | destination_center | actual_time_first | actual_time_last | osrm_time_first | osrr |
|---|---|---|---|---|---|---|---|
| 0 | trip-153671041653548748 | IND209304AAA | IND000000ACB | 50.0 | 732.0 | 33.0 | |
| 1 | trip-153671041653548748 | IND462022AAA | IND209304AAA | 43.0 | 830.0 | 39.0 | |
| 2 | trip-153671042288605164 | IND561203AAB | IND562101AAA | 18.0 | 47.0 | 10.0 | |
| 3 | trip-153671042288605164 | IND572101AAA | IND561203AAB | 14.0 | 96.0 | 8.0 | |
| 4 | trip-153671043369099517 | IND000000ACB | IND160002AAC | 36.0 | 611.0 | 19.0 | |

- **Aggregation by 'trip_uuid','source_name', 'destination_name'**

In [524…
```
# Aggregation by trip_uuid , source_name, destination_name
df_agg_by_trip_src_dest= df.groupby(['trip_uuid','source_name', 'destination_name'])[["a
```

```
# Unfolding grouped data to rows
df_agg_by_trip_src_dest.reset_index(inplace=True)
# Renaming grouped columns
df_agg_by_trip_src_dest.columns = ['_'.join(col) for col in df_agg_by_trip_src_dest.colu
# Removing underscore from column names source_center_ and  destination_center_
df_agg_by_trip_src_dest.rename(columns = {'trip_uuid_':'trip_uuid','source_name_':'sourc
df_agg_by_trip_src_dest.head()
```

Out[524]:

| | trip_uuid | source_name | destination_name | actual_time_first | actual_time_last | osrm_ti |
|---|---|---|---|---|---|---|
| **0** | trip-153671041653548748 | Bhopal_Trnsport_H (Madhya Pradesh) | Kanpur_Central_H_6 (Uttar Pradesh) | 43.0 | 830.0 | |
| **1** | trip-153671041653548748 | Kanpur_Central_H_6 (Uttar Pradesh) | Gurgaon_Bilaspur_HB (Haryana) | 50.0 | 732.0 | |
| **2** | trip-153671042288605164 | Doddablpur_ChikaDPP_D (Karnataka) | Chikblapur_ShntiSgr_D (Karnataka) | 18.0 | 47.0 | |
| **3** | trip-153671042288605164 | Tumkur_Veersagr_I (Karnataka) | Doddablpur_ChikaDPP_D (Karnataka) | 14.0 | 96.0 | |
| **4** | trip-153671043369099517 | Bangalore_Nelmngla_H (Karnataka) | Gurgaon_Bilaspur_HB (Haryana) | 60.0 | 2736.0 | |

In [525…

```
df_agg_by_trip_src_dest.shape
```

Out[525]:

```
(26368, 15)
```

- **Merge by 'trip_uuid'**

In [526…

```
df_agg_by_trip_src_dest_merged = pd.merge(df_agg_by_trip_src_dest, df_org, how="inner",
df_agg_by_trip_src_dest_merged.shape
```

Out[526]:

```
(26368, 20)
```

In [527…

```
df_agg_by_trip_src_dest_merged.head()
```

Out[527]:

| | trip_uuid | source_name | destination_name | actual_time_first | actual_time_last | osrm_ti |
|---|---|---|---|---|---|---|
| **0** | trip-153671041653548748 | Bhopal_Trnsport_H (Madhya Pradesh) | Kanpur_Central_H_6 (Uttar Pradesh) | 43.0 | 830.0 | |
| **1** | trip-153671041653548748 | Kanpur_Central_H_6 (Uttar Pradesh) | Gurgaon_Bilaspur_HB (Haryana) | 50.0 | 732.0 | |
| **2** | trip-153671042288605164 | Doddablpur_ChikaDPP_D (Karnataka) | Chikblapur_ShntiSgr_D (Karnataka) | 18.0 | 47.0 | |
| **3** | trip-153671042288605164 | Tumkur_Veersagr_I (Karnataka) | Doddablpur_ChikaDPP_D (Karnataka) | 14.0 | 96.0 | |
| **4** | trip-153671043369099517 | Bangalore_Nelmngla_H (Karnataka) | Gurgaon_Bilaspur_HB (Haryana) | 60.0 | 2736.0 | |

In [528…

```
df_agg_by_trip_src_dest_merged.head()
```

Out[528]:

| | trip_uuid | source_name | destination_name | actual_time_first | actual_time_last | osrm_ti |
|---|---|---|---|---|---|---|
| 0 | trip-153671041653548748 | Bhopal_Trnsport_H (Madhya Pradesh) | Kanpur_Central_H_6 (Uttar Pradesh) | 43.0 | 830.0 | |
| 1 | trip-153671041653548748 | Kanpur_Central_H_6 (Uttar Pradesh) | Gurgaon_Bilaspur_HB (Haryana) | 50.0 | 732.0 | |
| 2 | trip-153671042288605164 | Doddablpur_ChikaDPP_D (Karnataka) | Chikblapur_ShntiSgr_D (Karnataka) | 18.0 | 47.0 | |
| 3 | trip-153671042288605164 | Tumkur_Veersagr_I (Karnataka) | Doddablpur_ChikaDPP_D (Karnataka) | 14.0 | 96.0 | |
| 4 | trip-153671043369099517 | Bangalore_Nelmngla_H (Karnataka) | Gurgaon_Bilaspur_HB (Haryana) | 60.0 | 2736.0 | |

In [529…

```
# Applying feature engineering part 1 on aggregated data
# Extraction of city , place , code , state
df_agg_by_trip_src_dest_merged[['source_name_city','source_name_place','source_name_code
# Data cleanup
df_agg_by_trip_src_dest_merged["source_name_city"]= df_agg_by_trip_src_dest_merged["sour
df_agg_by_trip_src_dest_merged["source_name_city"].replace(to_replace=r'(^Del$)', value=

# Above steps on destination name column
df_agg_by_trip_src_dest_merged[['destination_name_city','destination_name_place','destin
df_agg_by_trip_src_dest_merged["destination_name_city"]= df_agg_by_trip_src_dest_merged
```

In [530…

```
df_agg_by_trip_src_dest_merged.head()
```

Out[530]:

| | trip_uuid | source_name | destination_name | actual_time_first | actual_time_last | osrm_ti |
|---|---|---|---|---|---|---|
| 0 | trip-153671041653548748 | Bhopal_Trnsport_H (Madhya Pradesh) | Kanpur_Central_H_6 (Uttar Pradesh) | 43.0 | 830.0 | |
| 1 | trip-153671041653548748 | Kanpur_Central_H_6 (Uttar Pradesh) | Gurgaon_Bilaspur_HB (Haryana) | 50.0 | 732.0 | |
| 2 | trip-153671042288605164 | Doddablpur_ChikaDPP_D (Karnataka) | Chikblapur_ShntiSgr_D (Karnataka) | 18.0 | 47.0 | |
| 3 | trip-153671042288605164 | Tumkur_Veersagr_I (Karnataka) | Doddablpur_ChikaDPP_D (Karnataka) | 14.0 | 96.0 | |
| 4 | trip-153671043369099517 | Bangalore_Nelmngla_H (Karnataka) | Gurgaon_Bilaspur_HB (Haryana) | 60.0 | 2736.0 | |

5 rows × 28 columns

# In-depth analysis and Feature engineering (Part 2)

- Calculate the **time taken between od_start_time and od_end_time** and keep it as a feature. Drop the original columns, if required

In [227…

```
df['time_taken_between_od_start_and_od_end_time'] = df['od_end_time'] - df['od_start_tir
df['time_taken_between_od_start_and_od_end_time']= df['time_taken_between_od_start_and_c
df['time_taken_between_od_start_and_od_end_time'] = df['time_taken_between_od_start_and_
```

In [228…

```
df[['time_taken_between_od_start_and_od_end_time', 'od_end_time', 'od_start_time']]
```

Out[228]:

| | time_taken_between_od_start_and_od_end_time | od_end_time | od_start_time |
|---|---|---|---|
| **0** | 1.437 | 2018-09-20 04:47:45.236797 | 2018-09-20 03:21:32.418600 |
| **1** | 1.437 | 2018-09-20 04:47:45.236797 | 2018-09-20 03:21:32.418600 |
| **2** | 1.437 | 2018-09-20 04:47:45.236797 | 2018-09-20 03:21:32.418600 |
| **3** | 1.437 | 2018-09-20 04:47:45.236797 | 2018-09-20 03:21:32.418600 |
| **4** | 1.437 | 2018-09-20 04:47:45.236797 | 2018-09-20 03:21:32.418600 |
| **...** | ... | ... | ... |
| **144862** | 7.128 | 2018-09-20 23:32:09.618069 | 2018-09-20 16:24:28.436231 |
| **144863** | 7.128 | 2018-09-20 23:32:09.618069 | 2018-09-20 16:24:28.436231 |
| **144864** | 7.128 | 2018-09-20 23:32:09.618069 | 2018-09-20 16:24:28.436231 |
| **144865** | 7.128 | 2018-09-20 23:32:09.618069 | 2018-09-20 16:24:28.436231 |
| **144866** | 7.128 | 2018-09-20 23:32:09.618069 | 2018-09-20 16:24:28.436231 |

144867 rows × 3 columns

- **Aggregated data (by trip_uuid, source and destination)**
  - Calculate the **time difference between first and last actual_time**

In [58]:

```
df_agg_by_trip_src_dest_center_merged['actual_time_diff'] = df_agg_by_trip_src_dest_cent
```

- Calculate the **time difference between first and last osrm_time**

In [60]:

```
df_agg_by_trip_src_dest_center_merged['osrm_time_diff'] = df_agg_by_trip_src_dest_cente
```

## Compare the difference between "time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan". Do hypothesis testing/ Visual analysis to check.

In [229…

```
df['start_scan_to_end_scan_in_hrs']= df['start_scan_to_end_scan'] / 60
df['start_scan_to_end_scan_in_hrs']= df['start_scan_to_end_scan_in_hrs'].round(3)
```

In [230…

```
df[["time_taken_between_od_start_and_od_end_time","start_scan_to_end_scan_in_hrs"]]
```

Out[230]:

| | time_taken_between_od_start_and_od_end_time | start_scan_to_end_scan_in_hrs |
|---|---|---|
| **0** | 1.437 | 1.433 |
| **1** | 1.437 | 1.433 |
| **2** | 1.437 | 1.433 |
| **3** | 1.437 | 1.433 |
| **4** | 1.437 | 1.433 |
| **...** | ... | ... |
| **144862** | 7.128 | 7.117 |
| **144863** | 7.128 | 7.117 |
| **144864** | 7.128 | 7.117 |
| **144865** | 7.128 | 7.117 |
| **144866** | 7.128 | 7.117 |

144867 rows × 2 columns

- **Sample T-Test** to check if difference between "time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan"
  - **Define H0 and Ha**
    - **Null hypothesis (H0)** : **Group means** of "time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan" are **equal**
    - **Alternate hypothesis (Ha)** : **Group means** of "time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan" are **NOT equal**
  - **Define experiment and "sensible" (i.e. distribution of test under H0) test statistics**
    - **Two sample T-test (Independent)** .
      - Note : Could have used z-test as well because sample size is more than 30. However , T-test will be turnout to Z-Test as sample size is large (i.e. more than 10k)
    - Independent T-Test : Two diffrent random variable (i.e. "time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan") being tested
  - **Decide One sided / two-sided tail test**
    - Two-sided as in Ha the measure is "not equal", we're neither checking greater nor lesser
  - **Define alfa (significance level)**
    - Let's assume significance level(alpha value) as 5%
  - **Calculate p-value**

In [231…

```
stats.ttest_ind(df["time_taken_between_od_start_and_od_end_time"],df["start_scan_to_end_
```

Out[231]:

```
Ttest_indResult(statistic=0.12870351906903685, pvalue=0.8975923400909249)
```

- **T-Test Analysis**
  - **Conclusion**
    - Failed to Reject Null hupothesis as p value is 0.8975923400909249 i.e. greater than alpha value .05

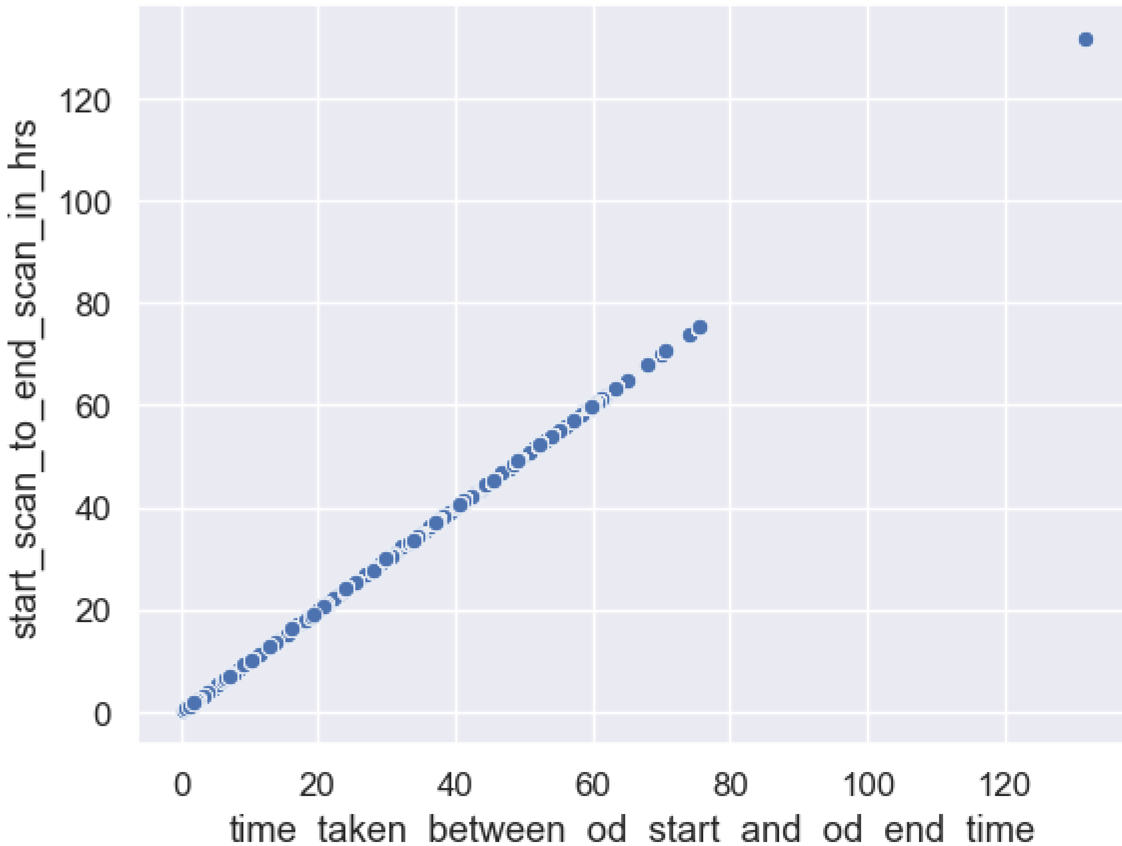
- Can't reject that "time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan" are equal

In [232…

```
sns.scatterplot(x="time_taken_between_od_start_and_od_end_time", y="start_scan_to_end_sc
```

Out[232]:

```
<AxesSubplot:xlabel='time_taken_between_od_start_and_od_end_time', ylabel='start_scan_to
_end_scan_in_hrs'>
```




- Overall **conclusion on the difference between "time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan"**
  - Based on both statistical and visual analysis both features **"time_taken_between_od_start_and_od_end_time" and "start_scan_to_end_scan"** are same
  - We **can drop new feature "time_taken_between_od_start_and_od_end_time"**

## Hypothesis testing/ visual analysis between "actual_time_aggregated" and "osrm_time_aggregated" value (aggregated values are the values you'll get after merging the rows on the basis of trip_uuid)

- **Sample T-Test** to check if there is a difference between "actual_time_aggregated" and "osrm_time_aggregated"
  - **Define H0 and Ha**
    - **Null hypothesis (H0)** : **Group means** of "actual_time_aggregated" and "osrm_time_aggregated" are **equal**
    - **Alternate hypothesis (Ha)** : **Group means** of "actual_time_aggregated" and "osrm_time_aggregated" are **NOT equal**
  - **Define experiment and "sensible" (i.e. distribution of test under H0) test statistics**
    - **Two sample T-test (Independent)** .

- Note : Could have used z-test as well because sample size is more than 30. However , T-test will be turnout to Z-Test as sample size is large (i.e. more than 10k)
- Independent T-Test : Two diffrent random variable (i.e. "actual_time_aggregated" and "osrm_time_aggregated") being tested

- **Decide One sided / two-sided tail test**
  - Two-sided as in Ha the measure is "not equal", we're neither checking greater nor lesser
- **Define alfa (significance level)**
  - Let's assume significance level(alpha value) as 5%
- **Calculate p-value**

In [233…

```
stats.ttest_ind(df_agg["actual_time_aggregated"],df_agg["osrm_time_aggregated"])
```

Out[233]:

```
Ttest_indResult(statistic=14.073444960610715, pvalue=7.714905383019579e-45)
```

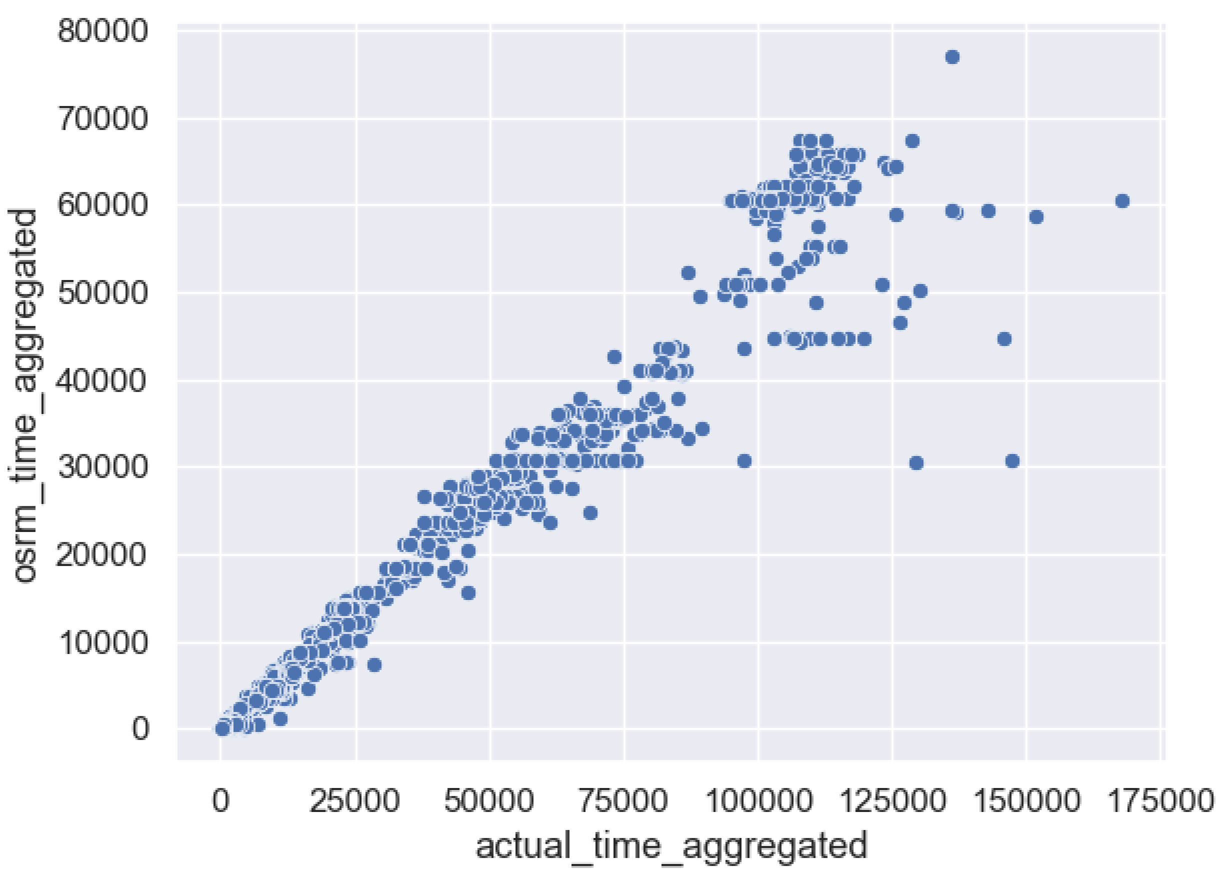
- **T-Test Analysis**
  - **Conclusion**
    - Reject Null hypothesis as p value is 7.714905383019579e-45 i.e. less than alpha value .05
    - **"actual_time_aggregated" and "osrm_time_aggregated" are two different features**
    - **Need to reatin both features i.e. "actual_time_aggregated" and "osrm_time_aggregated"**

In [234…

```
sns.scatterplot(x="actual_time_aggregated", y="osrm_time_aggregated", data=df_agg)
```

Out[234]:

```
<AxesSubplot:xlabel='actual_time_aggregated', ylabel='osrm_time_aggregated'>
```




## Compare feature **actual_time vs segment_actual_time**

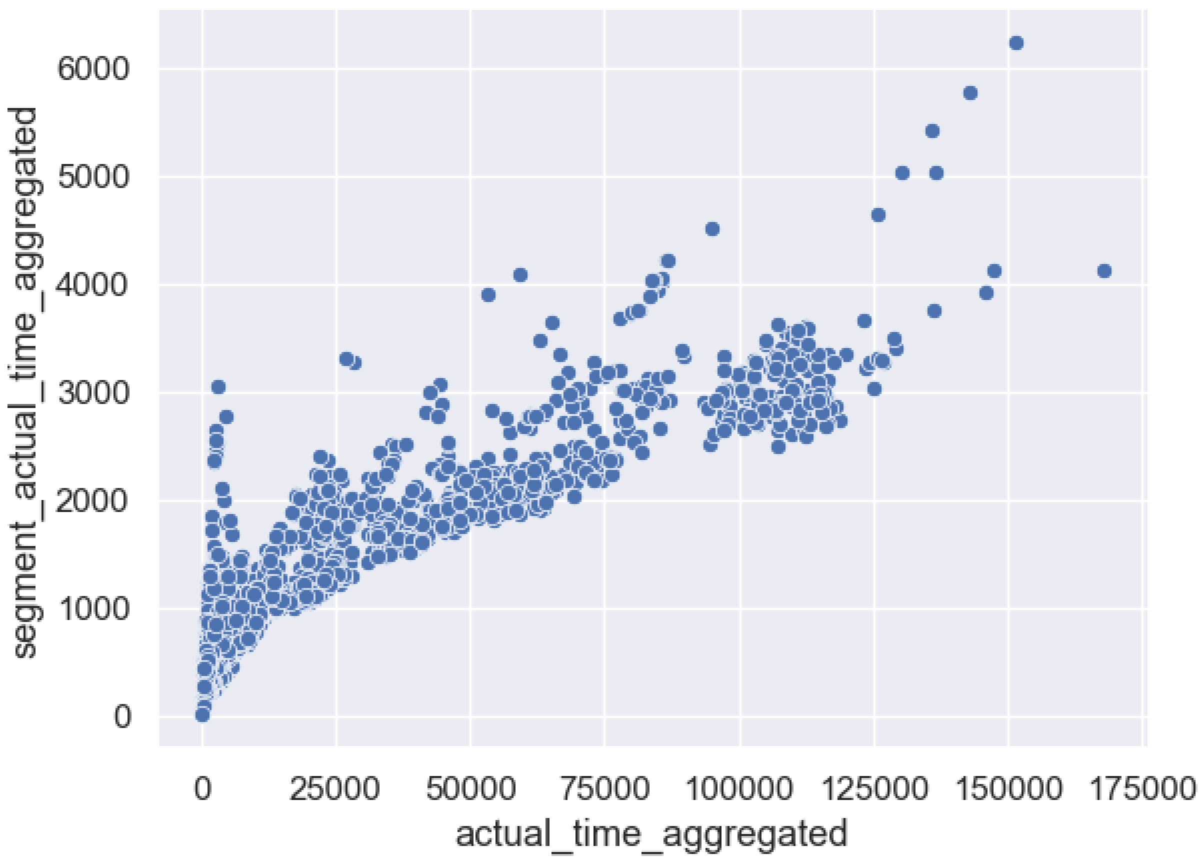
In [235…

```
stats.ttest_ind(df_agg["actual_time_aggregated"],df_agg["segment_actual_time_aggregated'
```

Out[235]: Ttest_indResult(statistic=29.75724632324628, pvalue=9.305532733717133e-192)

In [236…
```
sns.scatterplot(x="actual_time_aggregated", y="segment_actual_time_aggregated", data=df_
```

Out[236]: <AxesSubplot:xlabel='actual_time_aggregated', ylabel='segment_actual_time_aggregated'>




- **T-Test Analysis**
  - **Conclusion**
    - Reject Null hupothesis as p value is 9.305532733717133e-192 i.e. less than alpha value .05
    - **"actual_time_aggregated" and "segment_actual_time_aggregated" are two different features**
    - **Need to reatin both features i.e. "actual_time_aggregated" and "segment_actual_time_aggregated"**

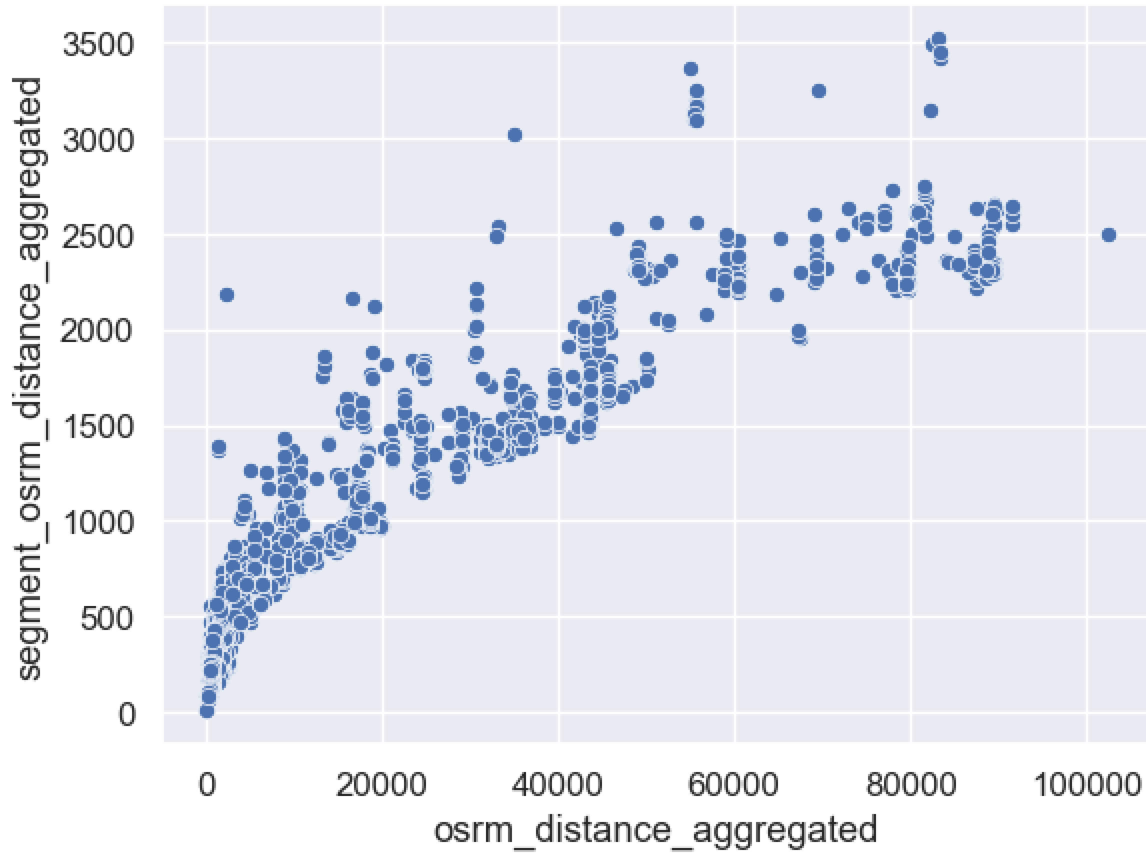
## Compare feature **osrm_distance vs segment_osrm_distance**

In [237…
```
stats.ttest_ind(df_agg["osrm_distance_aggregated"],df_agg["segment_osrm_distance_aggrega
```

Out[237]: Ttest_indResult(statistic=28.952997899197353, pvalue=8.78329034932333e-182)

In [238…
```
sns.scatterplot(x="osrm_distance_aggregated", y="segment_osrm_distance_aggregated", data
```

Out[238]: <AxesSubplot:xlabel='osrm_distance_aggregated', ylabel='segment_osrm_distance_aggregate
d'>




- **T-Test Analysis**
  - **Conclusion**
    - Reject Null hupothesis as p value is 8.78329034932333e-182 i.e. less than alpha value .05
    - **"osrm_distance_aggregated" and "segment_osrm_distance_aggregated" are two different features**
    - **Need to reatin both features i.e. "osrm_distance_aggregated" and "segment_osrm_distance_aggregated"**

## Compare feature **osrm_time vs segment_osrm_time**

In [239…

```
stats.ttest_ind(df_agg["osrm_time_aggregated"],df_agg["segment_osrm_time_aggregated"])
```

Out[239]:

```
Ttest_indResult(statistic=29.19742674380395, pvalue=8.695112641096768e-185)
```

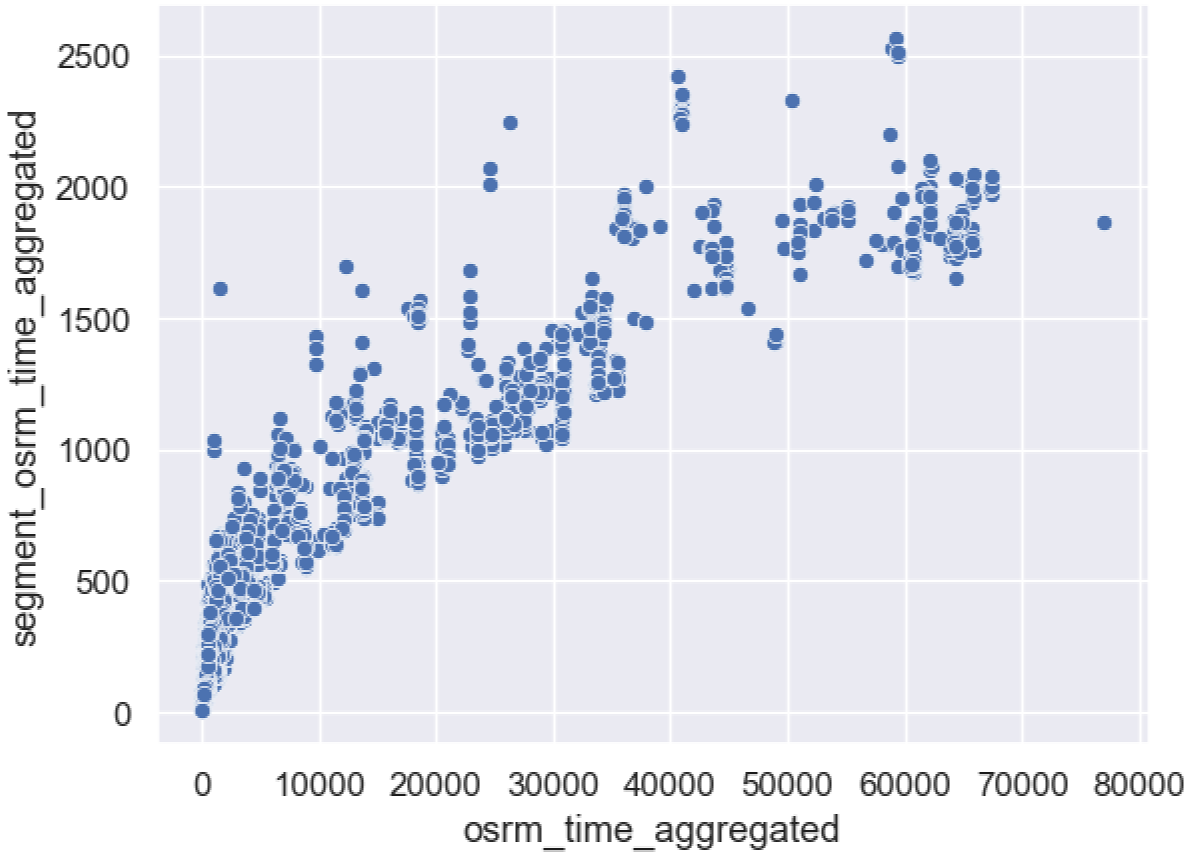
- **T-Test Analysis**
  - **Conclusion**
    - Reject Null hupothesis as p value is 8.695112641096768e-185 i.e. less than alpha value .05
    - **"osrm_time_aggregated" and "segment_osrm_time_aggregated" are two different features**
    - **Need to reatin both features i.e. "osrm_time_aggregated" and "segment_osrm_time_aggregated"**

In [240…

```
sns.scatterplot(x="osrm_time_aggregated", y="segment_osrm_time_aggregated", data=df_agg)
```

Out[240]:

```
<AxesSubplot:xlabel='osrm_time_aggregated', ylabel='segment_osrm_time_aggregated'>
```




- **Visual Analysis**
  - **Conclusion**
    - Follows close linear relationship to some time range
    - Beyond that range , **"actual_time_aggregated" and "osrm_time_aggregated" varies differently**

## Compare feature **osrm_time_diff vs actual_time_diff**

- **Hypothesis**
  - Null hypothesis - Mean of **osrm_time_diff time and actual_time_diff have no diffrence**
  - Alternate hypothesis - Mean of **osrm_time_diff time and actual_time_diff have diffrence**

In [63]:

```
stats.ttest_ind(df_agg_by_trip_src_dest_center_merged["osrm_time_diff"],df_agg_by_trip_s
```

Out[63]:

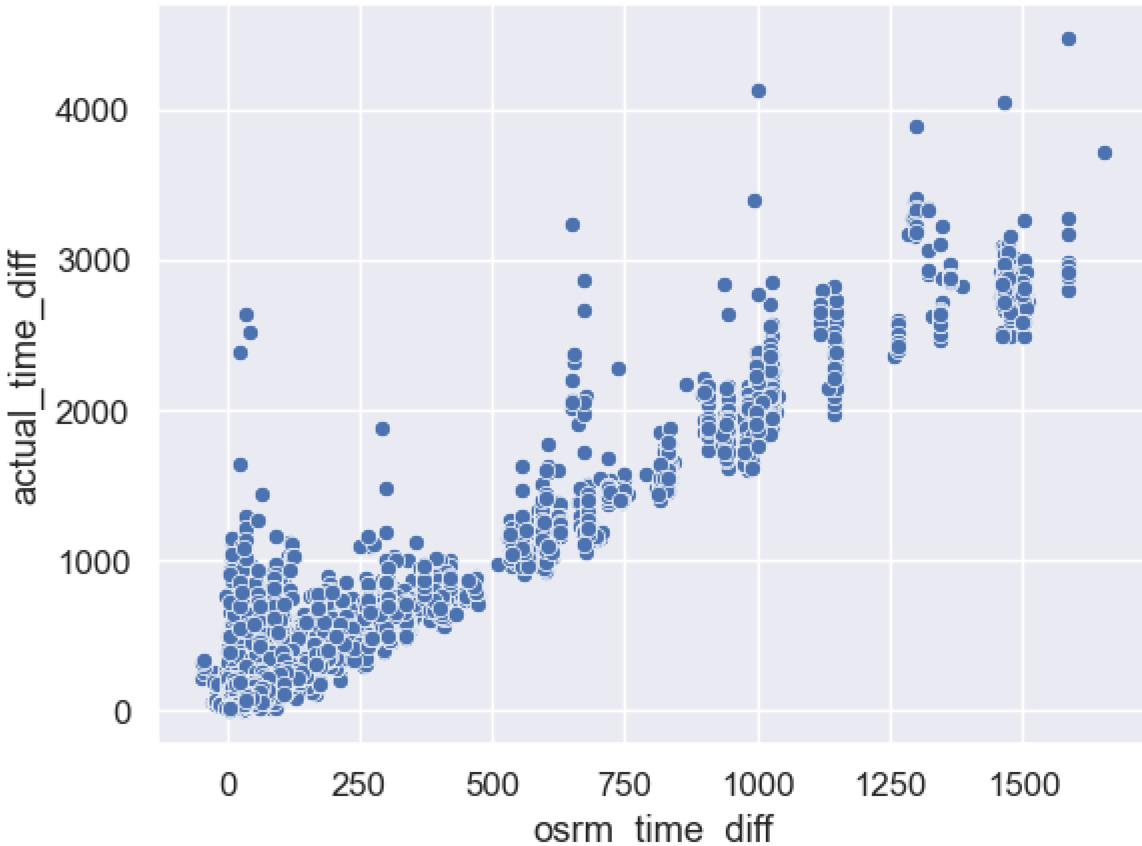
```
Ttest_indResult(statistic=-33.067911482916315, pvalue=2.3296445482665114e-237)
```

In [64]:

```
sns.scatterplot(x="osrm_time_diff", y="actual_time_diff", data=df_agg_by_trip_src_dest_c
```

Out[64]:

```
<AxesSubplot:xlabel='osrm_time_diff', ylabel='actual_time_diff'>
```




- **Visual Analysis**
  - **Conclusion**
    - Follows random linear relationship to some time range
    - Beyond that range , **"osrm_time_diff" and "actual_time_diff" varies differently**
    - **More accurate predictions required for better estimate of time**
    - Need to observe orders which are following accurate relationship and take more such orders

## Comparison & Visualization of time and distance fields

In [241…

```
distance_feature_list = [ "start_scan_to_end_scan","actual_distance_to_destination","osr
time_feature_list= ["actual_time","osrm_time","segment_actual_time","segment_osrm_time"]
```

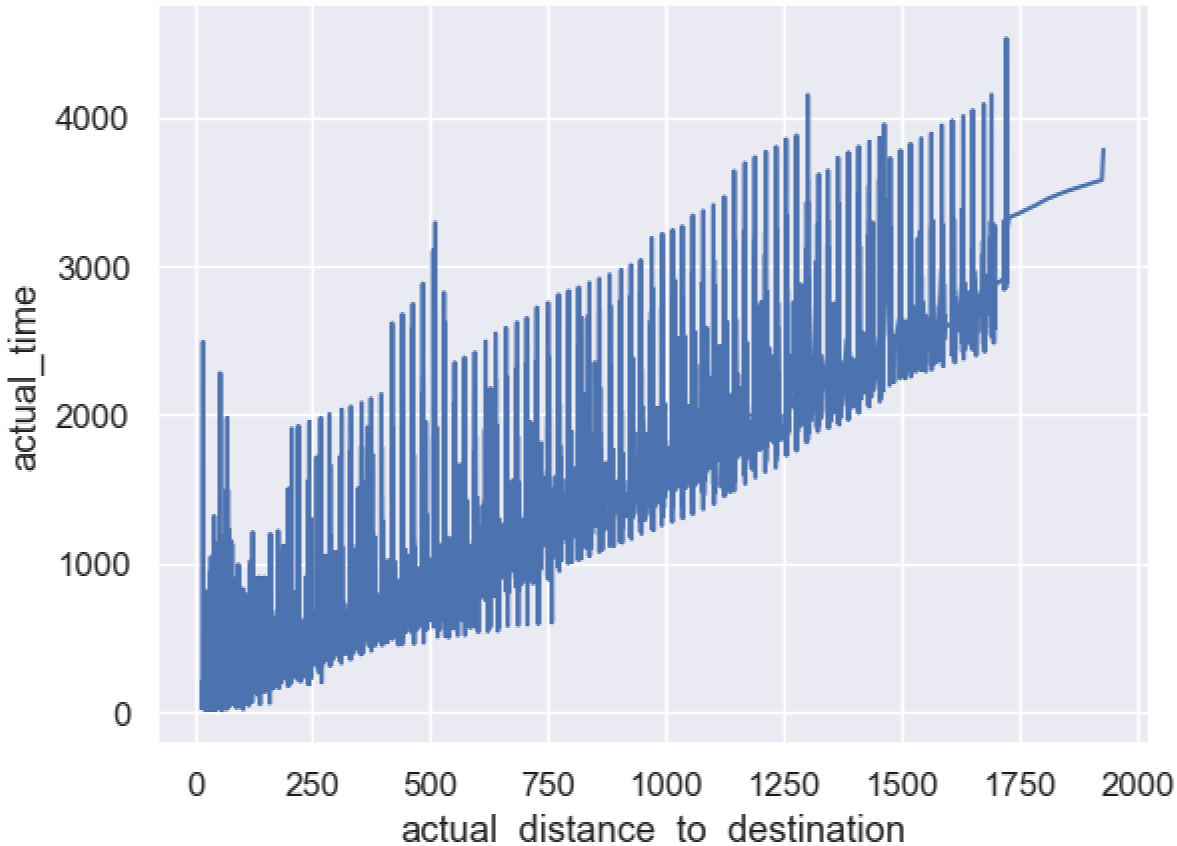
### Full truck load

In [242…

```
sns.lineplot(x="actual_distance_to_destination", y="actual_time",data = df_full_truck_lo
```

Out[242]:

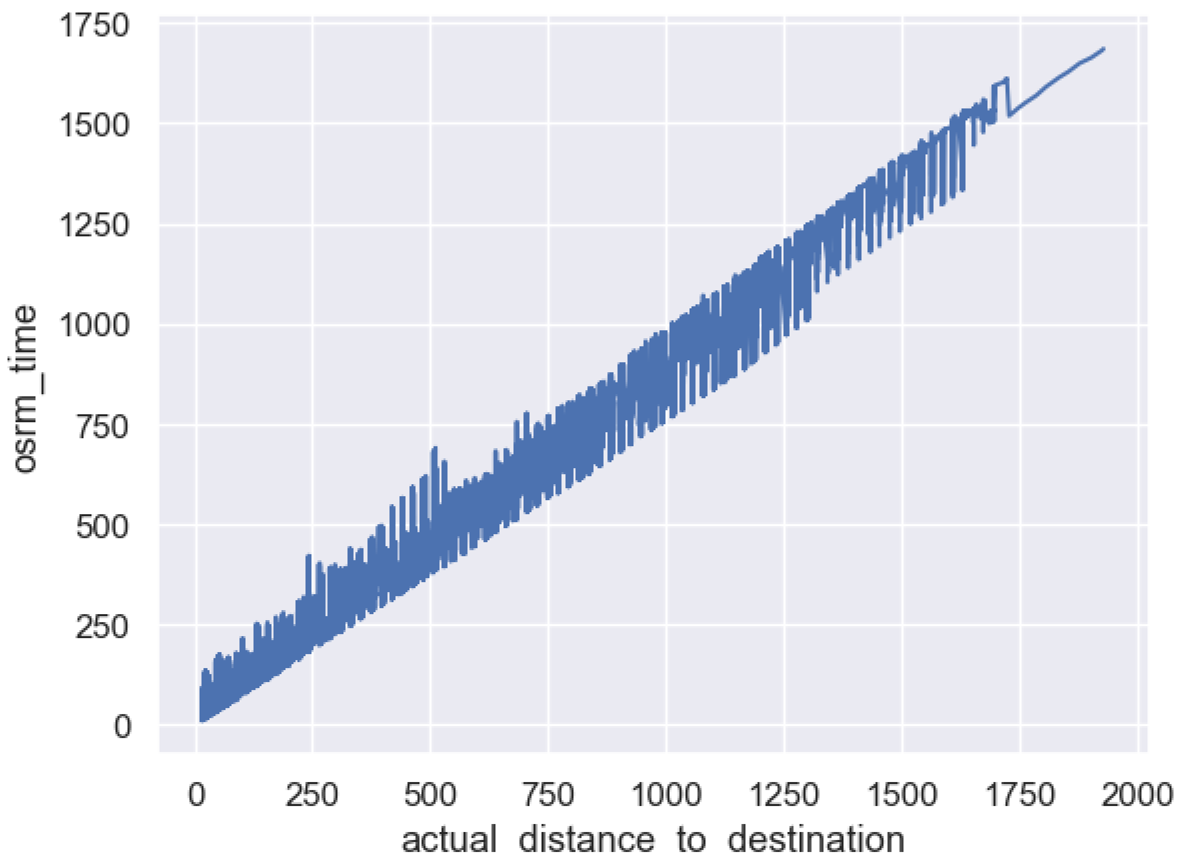
```
<AxesSubplot:xlabel='actual_distance_to_destination', ylabel='actual_time'>
```




In [243…

```
sns.lineplot(x="actual_distance_to_destination", y="osrm_time",data = df_full_truck_load
```

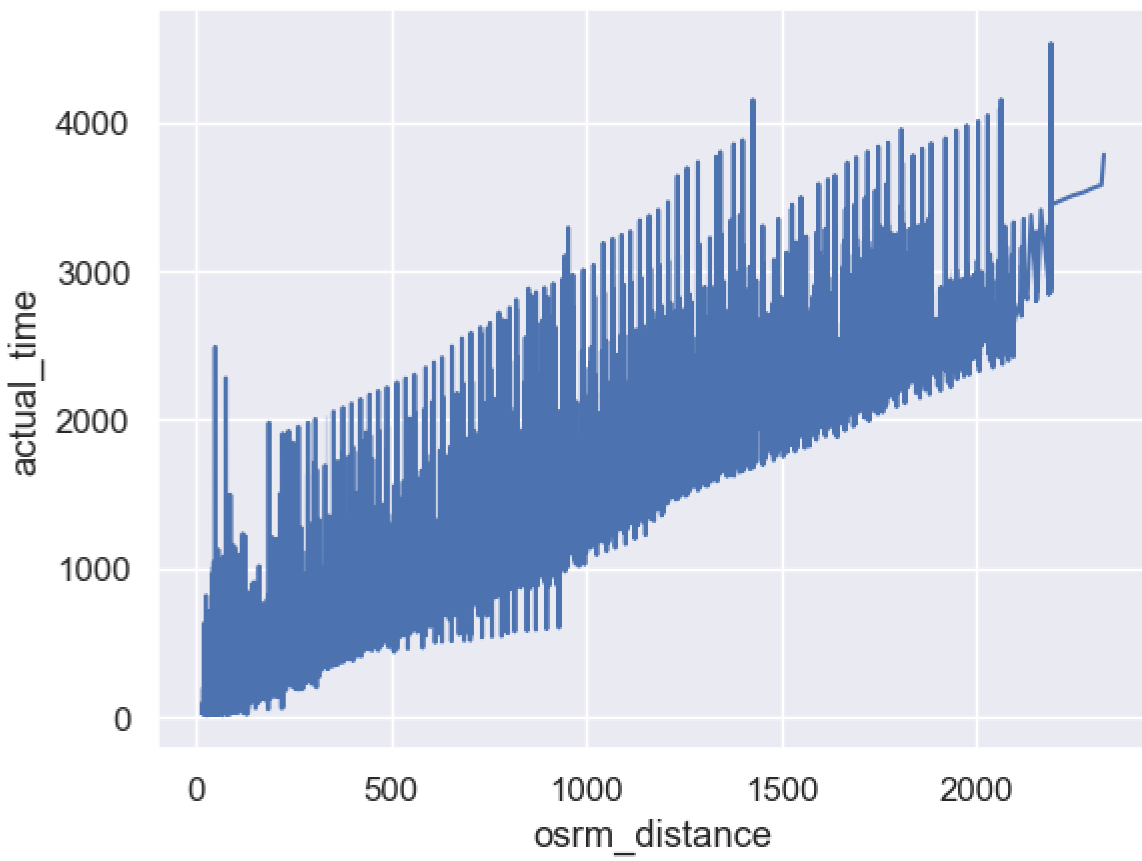
Out[243]: <AxesSubplot:xlabel='actual_distance_to_destination', ylabel='osrm_time'>




In [244…

```
sns.lineplot(x="osrm_distance", y="actual_time",data = df_full_truck_load)
```

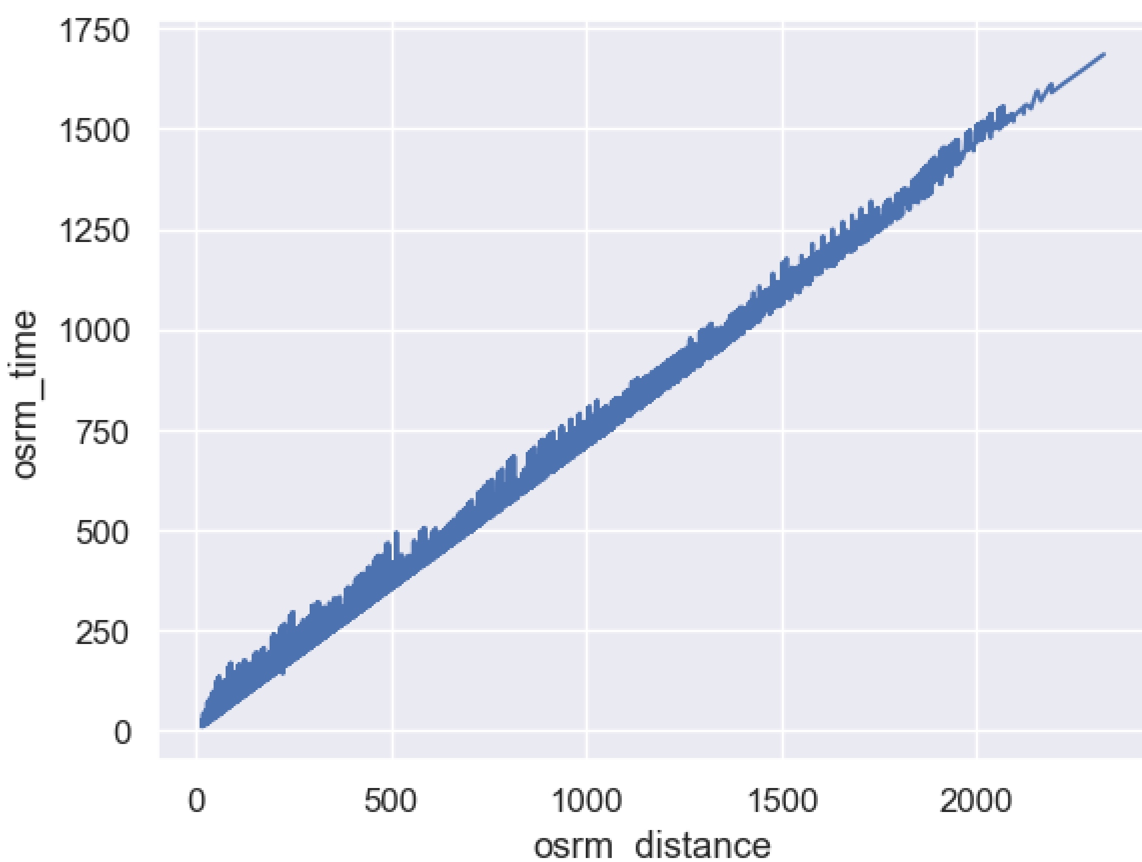
Out[244]: <AxesSubplot:xlabel='osrm_distance', ylabel='actual_time'>




In [245…

```
sns.lineplot(x="osrm_distance", y="osrm_time",data = df_full_truck_load)
```

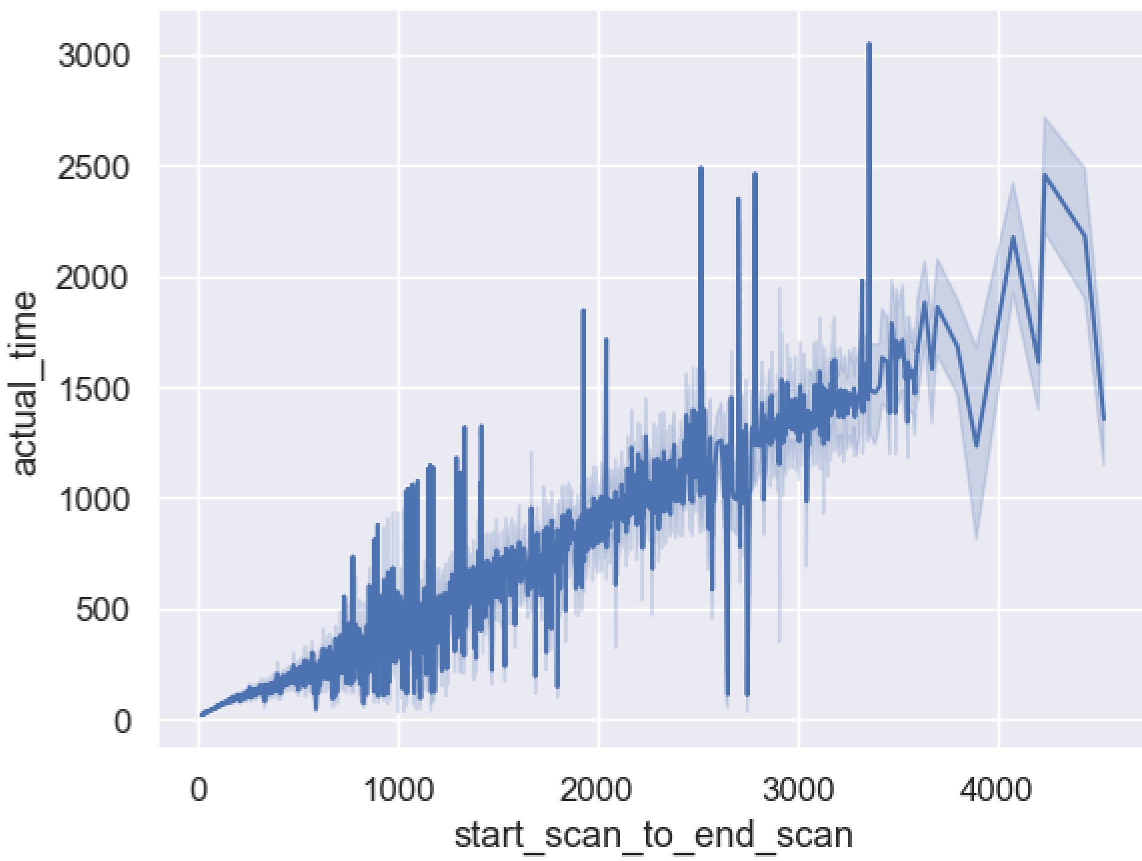
Out[245]: <AxesSubplot:xlabel='osrm_distance', ylabel='osrm_time'>




In [246…

```
sns.lineplot(x="start_scan_to_end_scan", y="actual_time",data = df_full_truck_load)
```

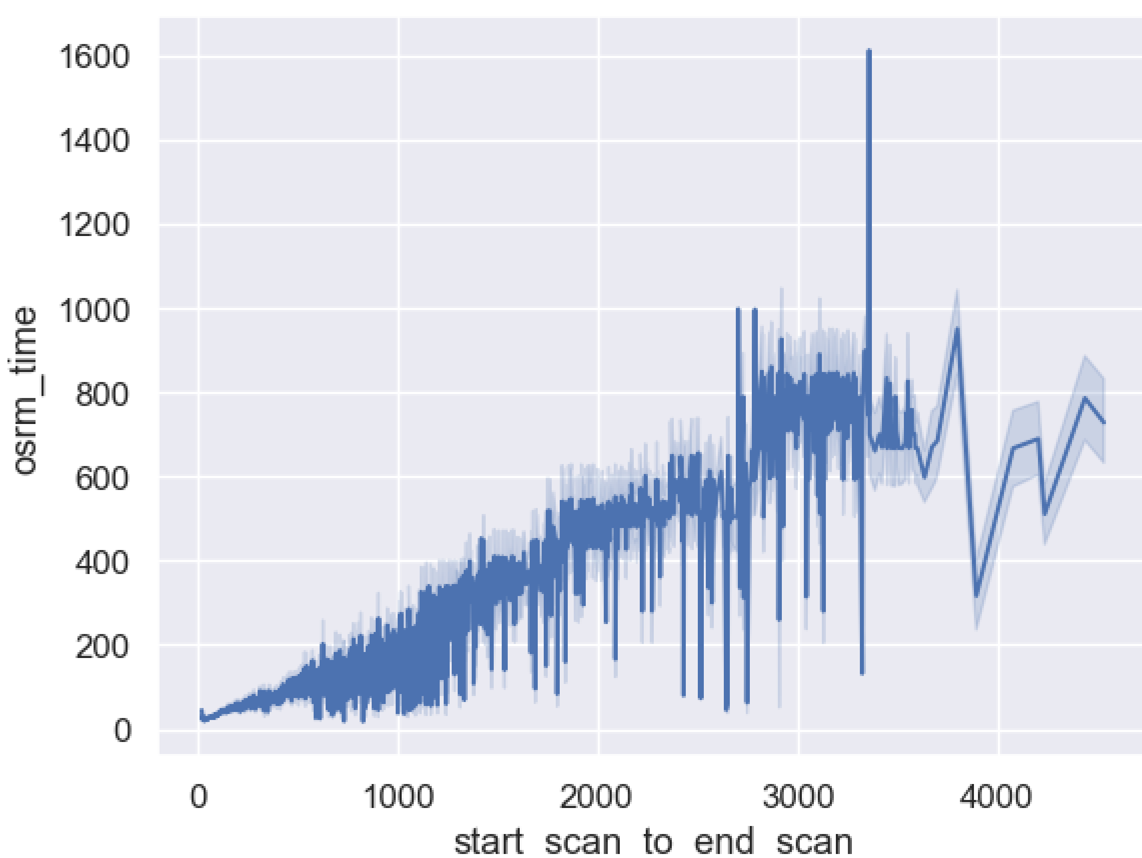
Out[246]: <AxesSubplot:xlabel='start_scan_to_end_scan', ylabel='actual_time'>




In [247…

```
sns.lineplot(x="start_scan_to_end_scan", y="osrm_time",data = df_full_truck_load)
```

Out[247]: <AxesSubplot:xlabel='start_scan_to_end_scan', ylabel='osrm_time'>

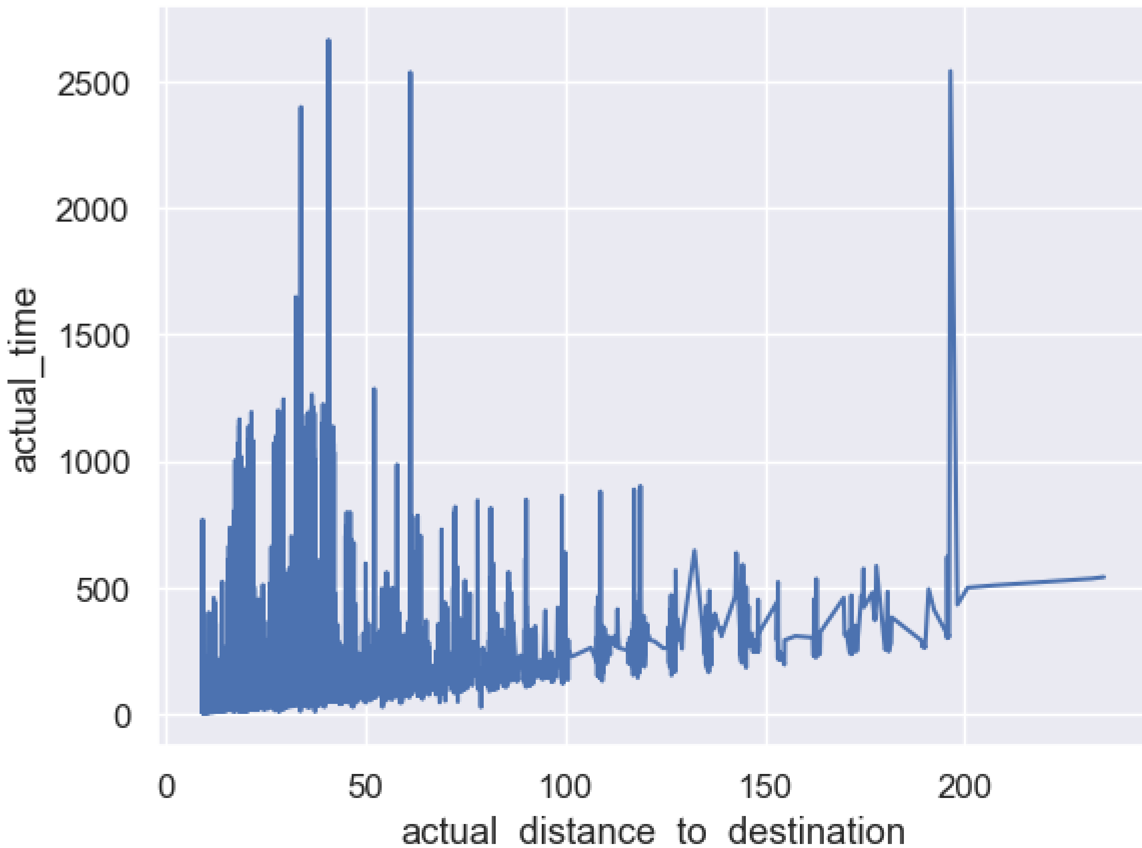


## Carting

In [248…

```
sns.lineplot(x="actual_distance_to_destination", y="actual_time",data = df_carting)
```

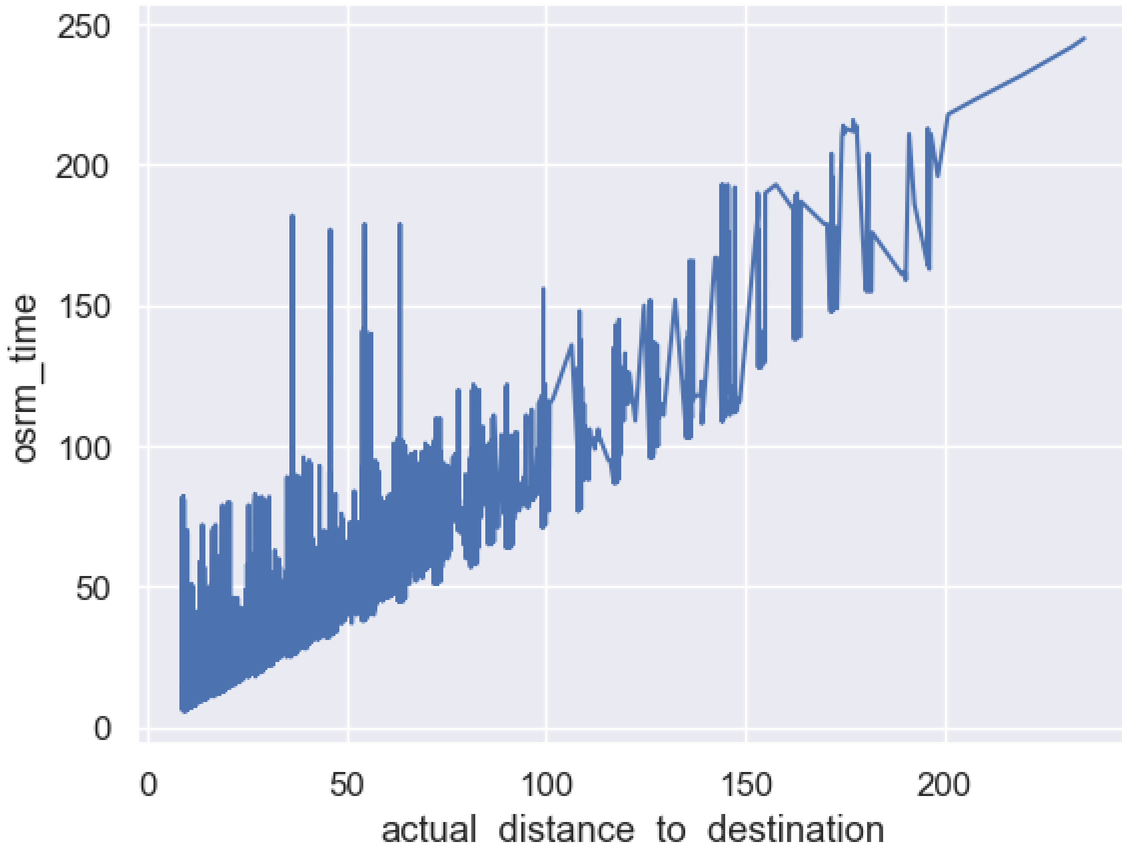
Out[248]: <AxesSubplot:xlabel='actual_distance_to_destination', ylabel='actual_time'>




In [249…

```
sns.lineplot(x="actual_distance_to_destination", y="osrm_time",data = df_carting)
```

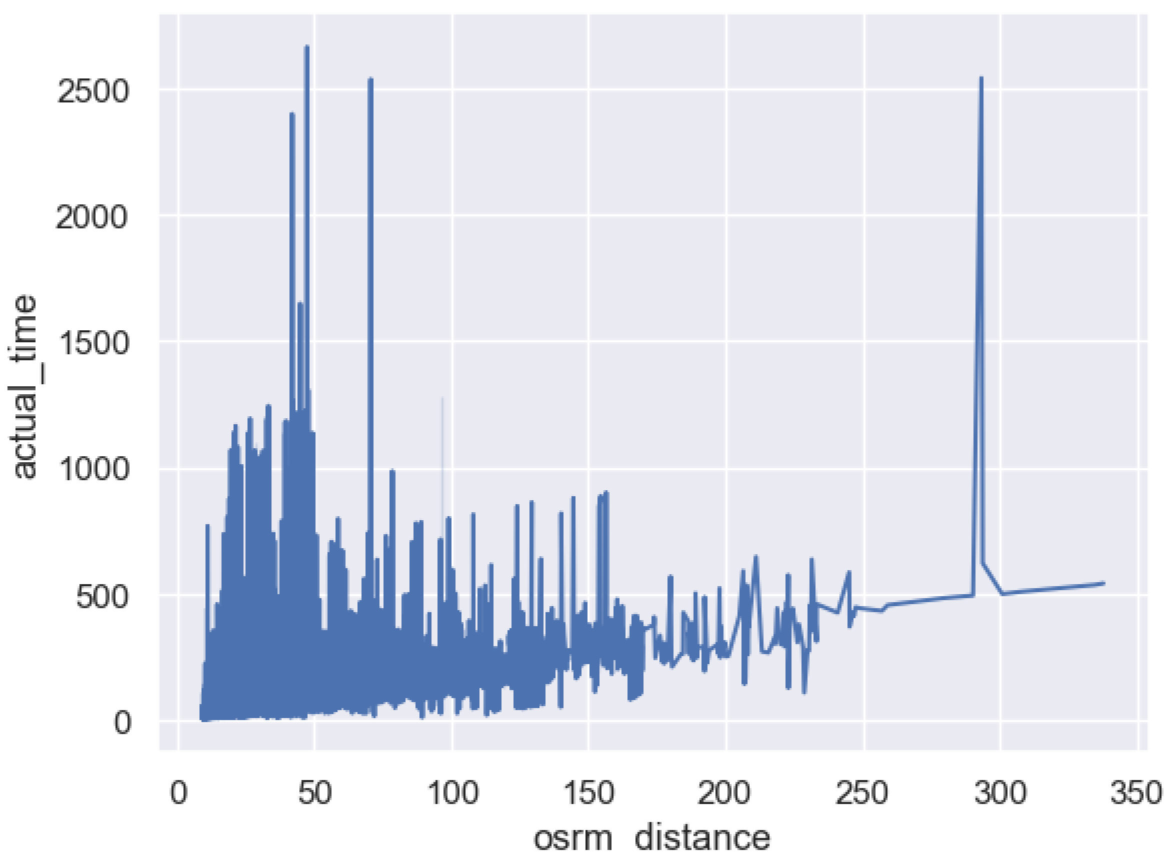
Out[249]: <AxesSubplot:xlabel='actual_distance_to_destination', ylabel='osrm_time'>




In [250…

```
sns.lineplot(x="osrm_distance", y="actual_time",data = df_carting)
```

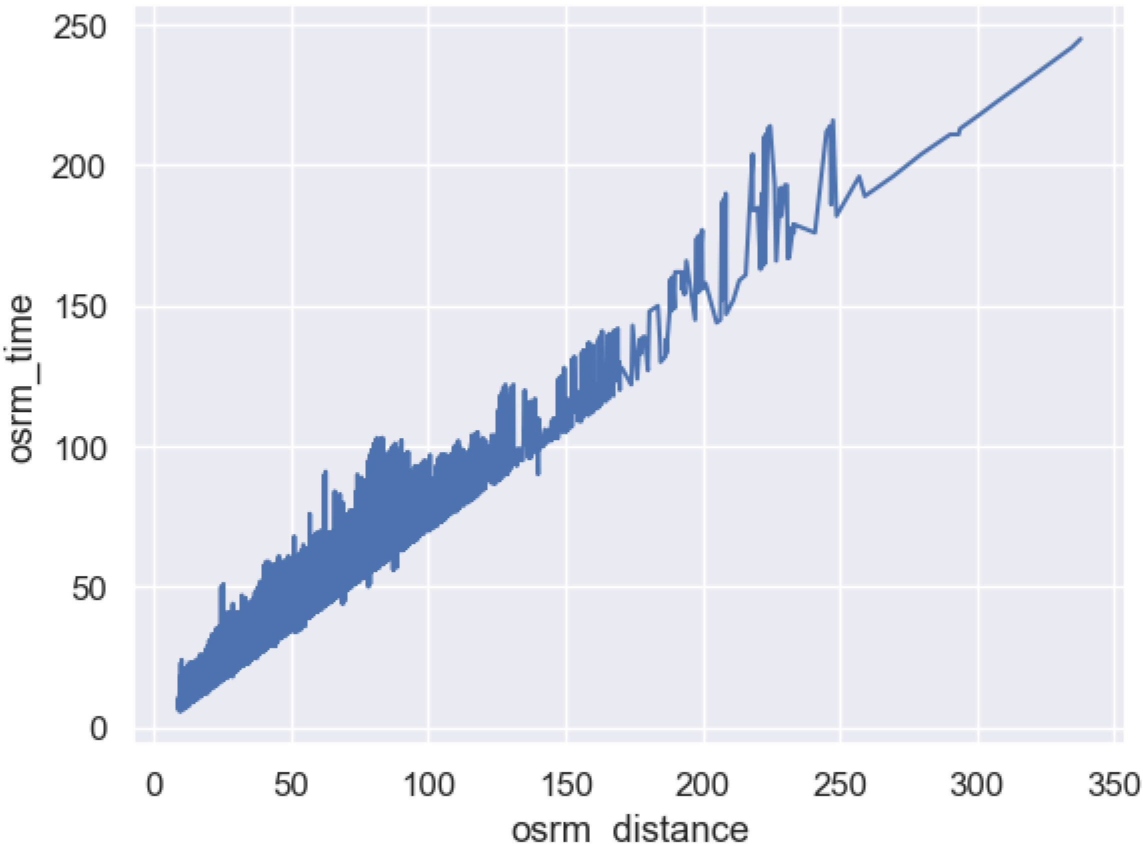
Out[250]: <AxesSubplot:xlabel='osrm_distance', ylabel='actual_time'>




In [251…

```
sns.lineplot(x="osrm_distance", y="osrm_time",data = df_carting)
```

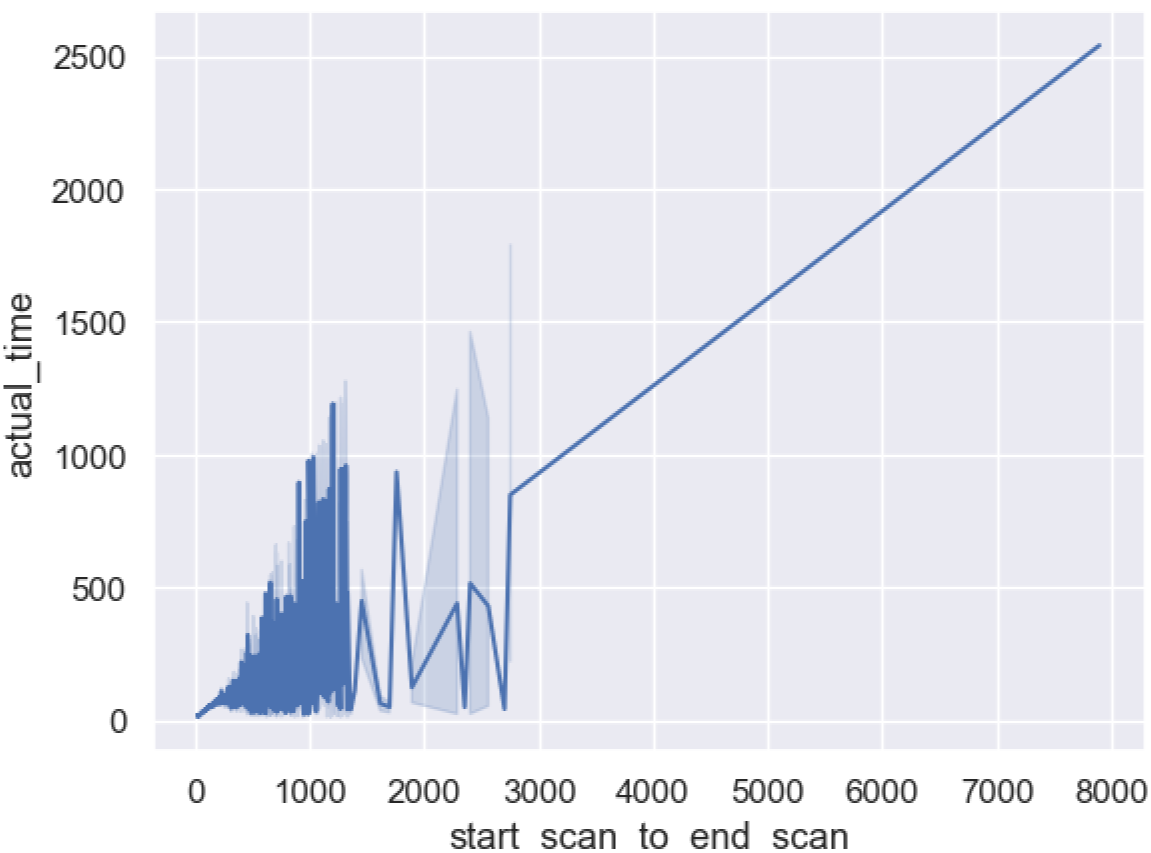
Out[251]: <AxesSubplot:xlabel='osrm_distance', ylabel='osrm_time'>




In [252…

```
sns.lineplot(x="start_scan_to_end_scan", y="actual_time",data = df_carting)
```

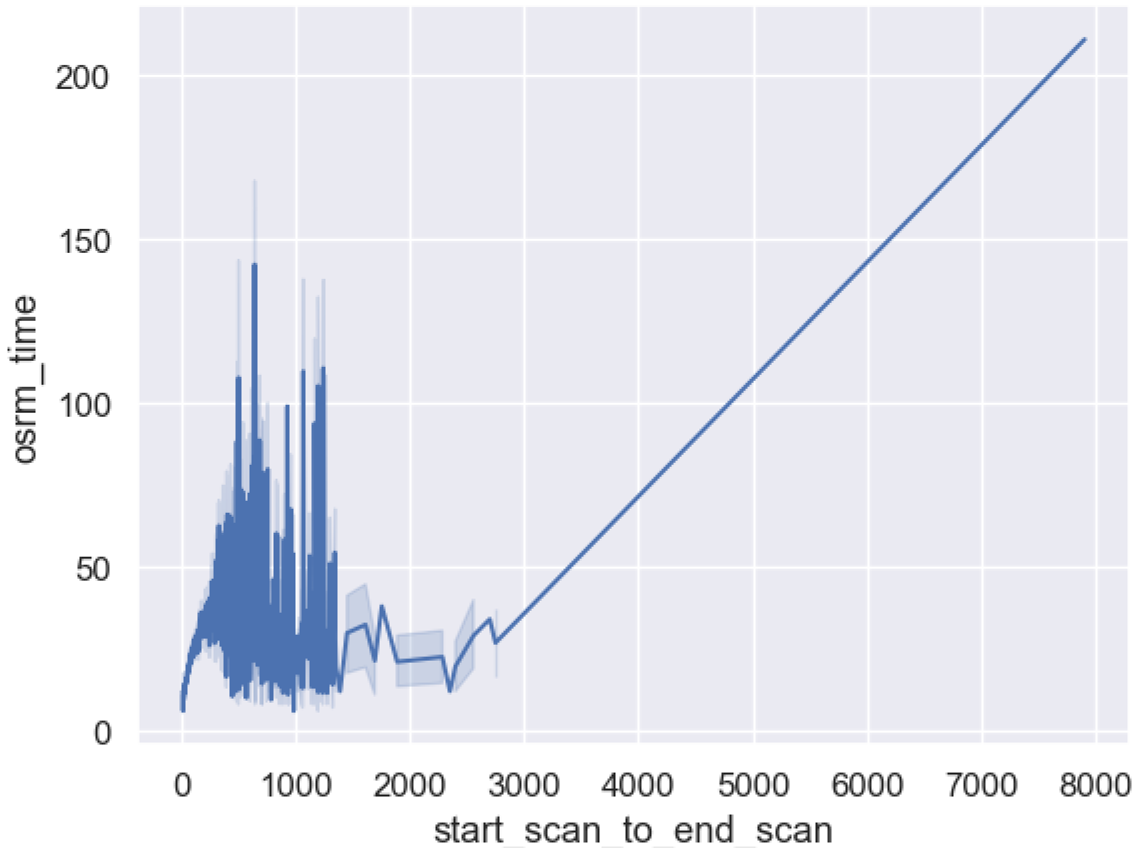
Out[252]: <AxesSubplot:xlabel='start_scan_to_end_scan', ylabel='actual_time'>




In [253…

```
sns.lineplot(x="start_scan_to_end_scan", y="osrm_time",data = df_carting)
```

Out[253]: <AxesSubplot:xlabel='start_scan_to_end_scan', ylabel='osrm_time'>




- **Conclusion**
  - **Approximate Linear relationship observed** between following **distance and time features**
    - "actual_distance_to_destination" vs "osrm_time"
    - "osrm_distance" vs "osrm_time"
    - "start_scan_to_end_scan" vs "actual_time"
    - "osrm_distance" vs "osrm_time"

## Outlier Detection and treatment

- Find outliers in the numerical variables and check it using visual analysis

In [254…

```
continious_features = df.select_dtypes(include=['int64','float64']).columns
continious_features
```

Out[254]:

```
Index(['start_scan_to_end_scan', 'cutoff_factor',
       'actual_distance_to_destination', 'actual_time', 'osrm_time',
       'osrm_distance', 'factor', 'segment_actual_time', 'segment_osrm_time',
       'segment_osrm_distance', 'segment_factor', 'trip_creation_time_year',
       'trip_creation_time_weekday', 'trip_creation_time_day',
       'od_start_time_year', 'od_start_time_weekday', 'od_start_time_day',
       'od_end_time_year', 'od_end_time_weekday', 'od_end_time_day',
       'time_taken_between_od_start_and_od_end_time',
       'start_scan_to_end_scan_in_hrs'],
      dtype='object')
```

In [255…

```
def find_outliers_IQR(column_name):
    print("Outliers by feature name --> ",column_name)
    # calculating quartiles
    Q1=df[column_name].quantile(0.25)
```

```
    Q3=df[column_name].quantile(0.75)
    # calculating inter quartile range
    IQR=Q3-Q1
    lower = Q1 - 1.5*IQR
    upper = Q3 + 1.5*IQR
    # filtering data that is outside inter quartile range
    outliers = df[((df[column_name]<lower) | (df[column_name]>upper))]

    return outliers
```

In [256…

```
for feature_name in continious_features:
    print(find_outliers_IQR(feature_name))
```

```
Outliers by feature name -->  start_scan_to_end_scan
            data         trip_creation_time  \
32950   training 2018-09-13 01:28:45.326644
32951   training 2018-09-13 01:28:45.326644
32952   training 2018-09-13 01:28:45.326644
32953   training 2018-09-13 01:28:45.326644
32954   training 2018-09-13 01:28:45.326644
...          ...                        ...
79524   training 2018-09-19 13:44:58.665210
79525   training 2018-09-19 13:44:58.665210
79526   training 2018-09-19 13:44:58.665210
79527   training 2018-09-19 13:44:58.665210
123196      test 2018-10-01 23:35:54.432745

                                      route_schedule_uuid route_type  \
32950   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32951   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32952   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32953   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32954   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
...                                                   ...        ...
79524   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79525   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79526   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79527   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
123196  thanos::sroute:4316e05f-b4cc-4ea7-b801-62a93ae...    Carting

                      trip_uuid source_center  \
32950   trip-153680212532637033  IND712311AAA
32951   trip-153680212532637033  IND712311AAA
32952   trip-153680212532637033  IND712311AAA
32953   trip-153680212532637033  IND712311AAA
32954   trip-153680212532637033  IND712311AAA
...                         ...           ...
79524   trip-153736469866480991  IND000000ACB
79525   trip-153736469866480991  IND000000ACB
79526   trip-153736469866480991  IND000000ACB
79527   trip-153736469866480991  IND000000ACB
123196  trip-153843695443252828  IND764071AAB

                                source_name destination_center  \
32950      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32951      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32952      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32953      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32954      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
...                                     ...                ...
79524         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79525         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79526         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79527         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
123196  Pappadahandi_Central_DPP_2 (Orissa)       IND530012AAA

                                    destination_name              od_start_time  \
32950                         Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32951                         Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32952                         Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32953                         Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32954                         Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
...                                              ...                        ...
79524               Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79525               Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79526               Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79527               Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
123196  Visakhapatnam_Gajuwaka_IP (Andhra Pradesh) 2018-10-02 15:21:51.236205
```

```
       ... source_name_city source_name_place source_name_code  \
32950  ...              NaN               NaN              NaN
32951  ...              NaN               NaN              NaN
32952  ...              NaN               NaN              NaN
32953  ...              NaN               NaN              NaN
32954  ...              NaN               NaN              NaN
...    ...              ...               ...              ...
79524  ...          Gurgaon          Bilaspur               HB
79525  ...          Gurgaon          Bilaspur               HB
79526  ...          Gurgaon          Bilaspur               HB
79527  ...          Gurgaon          Bilaspur               HB
123196 ...              NaN               NaN              NaN

       source_name_state destination_name_city destination_name_place  \
32950                NaN                   NaN                    NaN
32951                NaN                   NaN                    NaN
32952                NaN                   NaN                    NaN
32953                NaN                   NaN                    NaN
32954                NaN                   NaN                    NaN
...                  ...                   ...                    ...
79524            Haryana                   NaN                    NaN
79525            Haryana                   NaN                    NaN
79526            Haryana                   NaN                    NaN
79527            Haryana                   NaN                    NaN
123196               NaN                   NaN                    NaN

       destination_name_code destination_name_state  \
32950                    NaN                    NaN
32951                    NaN                    NaN
32952                    NaN                    NaN
32953                    NaN                    NaN
32954                    NaN                    NaN
...                      ...                    ...
79524                    NaN                    NaN
79525                    NaN                    NaN
79526                    NaN                    NaN
79527                    NaN                    NaN
123196                   NaN                    NaN

       time_taken_between_od_start_and_od_end_time  \
32950                                       64.959
32951                                       64.959
32952                                       64.959
32953                                       64.959
32954                                       64.959
...                                            ...
79524                                       70.658
79525                                       70.658
79526                                       70.658
79527                                       70.658
123196                                     131.643

       start_scan_to_end_scan_in_hrs
32950                         64.950
32951                         64.950
32952                         64.950
32953                         64.950
32954                         64.950
...                              ...
79524                         70.650
79525                         70.650
79526                         70.650
79527                         70.650
123196                       131.633

[373 rows x 46 columns]
```

```
Outliers by feature name -->  cutoff_factor
            data        trip_creation_time  \
402     training 2018-09-25 15:06:59.975279
403     training 2018-09-25 15:06:59.975279
404     training 2018-09-25 15:06:59.975279
405     training 2018-09-25 15:06:59.975279
406     training 2018-09-25 15:06:59.975279
...          ...                        ...
144796      test 2018-10-01 18:17:37.047270
144797      test 2018-10-01 18:17:37.047270
144798      test 2018-10-01 18:17:37.047270
144799      test 2018-10-01 18:17:37.047270
144800      test 2018-10-01 18:17:37.047270

                                      route_schedule_uuid route_type  \
402     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
403     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
404     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
405     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
406     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
...                                                   ...        ...
144796  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144797  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144798  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144799  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144800  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL

                      trip_uuid source_center  \
402     trip-153788801997503817  IND825409AAA
403     trip-153788801997503817  IND825409AAA
404     trip-153788801997503817  IND825409AAA
405     trip-153788801997503817  IND825409AAA
406     trip-153788801997503817  IND825409AAA
...                         ...           ...
144796  trip-153841785704702048  IND000000ACB
144797  trip-153841785704702048  IND000000ACB
144798  trip-153841785704702048  IND000000ACB
144799  trip-153841785704702048  IND000000ACB
144800  trip-153841785704702048  IND000000ACB

                              source_name destination_center  \
402     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
403     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
404     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
405     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
406     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
...                                   ...                ...
144796      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144797      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144798      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144799      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144800      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA

                       destination_name              od_start_time  ...  \
402         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
403         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
404         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
405         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
406         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
...                                   ...                        ...  ...
144796  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144797  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144798  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144799  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144800  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
```

```
       source_name_city source_name_place source_name_code  \
402          JhumriTlya          RadhaCpx                D
403          JhumriTlya          RadhaCpx                D
404          JhumriTlya          RadhaCpx                D
405          JhumriTlya          RadhaCpx                D
406          JhumriTlya          RadhaCpx                D
...                 ...               ...              ...
144796          Gurgaon          Bilaspur               HB
144797          Gurgaon          Bilaspur               HB
144798          Gurgaon          Bilaspur               HB
144799          Gurgaon          Bilaspur               HB
144800          Gurgaon          Bilaspur               HB

        source_name_state destination_name_city destination_name_place  \
402             Jharkhand               Gurgaon               Bilaspur
403             Jharkhand               Gurgaon               Bilaspur
404             Jharkhand               Gurgaon               Bilaspur
405             Jharkhand               Gurgaon               Bilaspur
406             Jharkhand               Gurgaon               Bilaspur
...                   ...                   ...                    ...
144796            Haryana             Bengaluru               Nelmngla
144797            Haryana             Bengaluru               Nelmngla
144798            Haryana             Bengaluru               Nelmngla
144799            Haryana             Bengaluru               Nelmngla
144800            Haryana             Bengaluru               Nelmngla

        destination_name_code destination_name_state  \
402                        HB                Haryana
403                        HB                Haryana
404                        HB                Haryana
405                        HB                Haryana
406                        HB                Haryana
...                       ...                    ...
144796                      H              Karnataka
144797                      H              Karnataka
144798                      H              Karnataka
144799                      H              Karnataka
144800                      H              Karnataka

        time_taken_between_od_start_and_od_end_time  \
402                                          25.852
403                                          25.852
404                                          25.852
405                                          25.852
406                                          25.852
...                                             ...
144796                                       49.162
144797                                       49.162
144798                                       49.162
144799                                       49.162
144800                                       49.162

        start_scan_to_end_scan_in_hrs
402                             25.85
403                             25.85
404                             25.85
405                             25.85
406                             25.85
...                               ...
144796                          49.15
144797                          49.15
144798                          49.15
144799                          49.15
144800                          49.15

[17246 rows x 46 columns]
```

```
Outliers by feature name -->  actual_distance_to_destination
            data        trip_creation_time  \
401     training 2018-09-25 15:06:59.975279
402     training 2018-09-25 15:06:59.975279
403     training 2018-09-25 15:06:59.975279
404     training 2018-09-25 15:06:59.975279
405     training 2018-09-25 15:06:59.975279
...          ...                       ...
144796      test 2018-10-01 18:17:37.047270
144797      test 2018-10-01 18:17:37.047270
144798      test 2018-10-01 18:17:37.047270
144799      test 2018-10-01 18:17:37.047270
144800      test 2018-10-01 18:17:37.047270

                                       route_schedule_uuid route_type  \
401     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
402     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
403     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
404     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
405     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
...                                                   ...        ...
144796  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144797  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144798  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144799  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144800  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL

                      trip_uuid source_center  \
401     trip-153788801997503817  IND825409AAA
402     trip-153788801997503817  IND825409AAA
403     trip-153788801997503817  IND825409AAA
404     trip-153788801997503817  IND825409AAA
405     trip-153788801997503817  IND825409AAA
...                         ...           ...
144796  trip-153841785704702048  IND000000ACB
144797  trip-153841785704702048  IND000000ACB
144798  trip-153841785704702048  IND000000ACB
144799  trip-153841785704702048  IND000000ACB
144800  trip-153841785704702048  IND000000ACB

                               source_name destination_center  \
401     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
402     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
403     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
404     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
405     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
...                                    ...                ...
144796      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144797      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144798      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144799      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144800      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA

                          destination_name              od_start_time  ...  \
401        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
402        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
403        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
404        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
405        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
...                                    ...                        ...  ...
144796  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144797  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144798  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144799  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144800  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
```

```
       source_name_city  source_name_place  source_name_code  \
401          JhumriTlya           RadhaCpx                 D
402          JhumriTlya           RadhaCpx                 D
403          JhumriTlya           RadhaCpx                 D
404          JhumriTlya           RadhaCpx                 D
405          JhumriTlya           RadhaCpx                 D
...                 ...                ...               ...
144796          Gurgaon           Bilaspur                HB
144797          Gurgaon           Bilaspur                HB
144798          Gurgaon           Bilaspur                HB
144799          Gurgaon           Bilaspur                HB
144800          Gurgaon           Bilaspur                HB

        source_name_state destination_name_city  destination_name_place  \
401             Jharkhand               Gurgaon                Bilaspur
402             Jharkhand               Gurgaon                Bilaspur
403             Jharkhand               Gurgaon                Bilaspur
404             Jharkhand               Gurgaon                Bilaspur
405             Jharkhand               Gurgaon                Bilaspur
...                   ...                   ...                     ...
144796            Haryana             Bengaluru                Nelmngla
144797            Haryana             Bengaluru                Nelmngla
144798            Haryana             Bengaluru                Nelmngla
144799            Haryana             Bengaluru                Nelmngla
144800            Haryana             Bengaluru                Nelmngla

        destination_name_code  destination_name_state  \
401                        HB                 Haryana
402                        HB                 Haryana
403                        HB                 Haryana
404                        HB                 Haryana
405                        HB                 Haryana
...                       ...                     ...
144796                      H               Karnataka
144797                      H               Karnataka
144798                      H               Karnataka
144799                      H               Karnataka
144800                      H               Karnataka

        time_taken_between_od_start_and_od_end_time  \
401                                          25.852
402                                          25.852
403                                          25.852
404                                          25.852
405                                          25.852
...                                             ...
144796                                       49.162
144797                                       49.162
144798                                       49.162
144799                                       49.162
144800                                       49.162

        start_scan_to_end_scan_in_hrs
401                             25.85
402                             25.85
403                             25.85
404                             25.85
405                             25.85
...                               ...
144796                          49.15
144797                          49.15
144798                          49.15
144799                          49.15
144800                          49.15

[17992 rows x 46 columns]
```

```
Outliers by feature name -->  actual_time
            data        trip_creation_time  \
406     training 2018-09-25 15:06:59.975279
407     training 2018-09-25 15:06:59.975279
408     training 2018-09-25 15:06:59.975279
409     training 2018-09-25 15:06:59.975279
410     training 2018-09-25 15:06:59.975279
...          ...                       ...
144796      test 2018-10-01 18:17:37.047270
144797      test 2018-10-01 18:17:37.047270
144798      test 2018-10-01 18:17:37.047270
144799      test 2018-10-01 18:17:37.047270
144800      test 2018-10-01 18:17:37.047270

                                       route_schedule_uuid route_type  \
406     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
407     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
408     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
409     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
410     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
...                                                   ...        ...
144796  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144797  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144798  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144799  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144800  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL

                      trip_uuid source_center  \
406     trip-153788801997503817  IND825409AAA
407     trip-153788801997503817  IND825409AAA
408     trip-153788801997503817  IND825409AAA
409     trip-153788801997503817  IND825409AAA
410     trip-153788801997503817  IND825409AAA
...                         ...           ...
144796  trip-153841785704702048  IND000000ACB
144797  trip-153841785704702048  IND000000ACB
144798  trip-153841785704702048  IND000000ACB
144799  trip-153841785704702048  IND000000ACB
144800  trip-153841785704702048  IND000000ACB

                              source_name destination_center  \
406     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
407     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
408     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
409     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
410     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
...                                   ...                ...
144796      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144797      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144798      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144799      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144800      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA

                          destination_name              od_start_time  ...  \
406         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
407         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
408         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
409         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
410         Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
...                                   ...                        ...  ...
144796  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144797  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144798  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144799  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144800  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
```

```
       source_name_city  source_name_place  source_name_code  \
406          JhumriTlya           RadhaCpx                 D
407          JhumriTlya           RadhaCpx                 D
408          JhumriTlya           RadhaCpx                 D
409          JhumriTlya           RadhaCpx                 D
410          JhumriTlya           RadhaCpx                 D
...                 ...                ...               ...
144796          Gurgaon           Bilaspur                HB
144797          Gurgaon           Bilaspur                HB
144798          Gurgaon           Bilaspur                HB
144799          Gurgaon           Bilaspur                HB
144800          Gurgaon           Bilaspur                HB

        source_name_state destination_name_city  destination_name_place  \
406             Jharkhand               Gurgaon                Bilaspur
407             Jharkhand               Gurgaon                Bilaspur
408             Jharkhand               Gurgaon                Bilaspur
409             Jharkhand               Gurgaon                Bilaspur
410             Jharkhand               Gurgaon                Bilaspur
...                   ...                   ...                     ...
144796            Haryana             Bengaluru                Nelmngla
144797            Haryana             Bengaluru                Nelmngla
144798            Haryana             Bengaluru                Nelmngla
144799            Haryana             Bengaluru                Nelmngla
144800            Haryana             Bengaluru                Nelmngla

        destination_name_code  destination_name_state  \
406                        HB                 Haryana
407                        HB                 Haryana
408                        HB                 Haryana
409                        HB                 Haryana
410                        HB                 Haryana
...                       ...                     ...
144796                      H               Karnataka
144797                      H               Karnataka
144798                      H               Karnataka
144799                      H               Karnataka
144800                      H               Karnataka

        time_taken_between_od_start_and_od_end_time  \
406                                          25.852
407                                          25.852
408                                          25.852
409                                          25.852
410                                          25.852
...                                             ...
144796                                       49.162
144797                                       49.162
144798                                       49.162
144799                                       49.162
144800                                       49.162

        start_scan_to_end_scan_in_hrs
406                             25.85
407                             25.85
408                             25.85
409                             25.85
410                             25.85
...                               ...
144796                          49.15
144797                          49.15
144798                          49.15
144799                          49.15
144800                          49.15

[16633 rows x 46 columns]
```

```
Outliers by feature name -->  osrm_time
            data         trip_creation_time  \
404     training 2018-09-25 15:06:59.975279
405     training 2018-09-25 15:06:59.975279
406     training 2018-09-25 15:06:59.975279
407     training 2018-09-25 15:06:59.975279
408     training 2018-09-25 15:06:59.975279
...          ...                        ...
144796      test 2018-10-01 18:17:37.047270
144797      test 2018-10-01 18:17:37.047270
144798      test 2018-10-01 18:17:37.047270
144799      test 2018-10-01 18:17:37.047270
144800      test 2018-10-01 18:17:37.047270

                                      route_schedule_uuid route_type  \
404     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
405     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
406     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
407     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
408     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
...                                                   ...        ...
144796  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144797  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144798  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144799  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144800  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL

                      trip_uuid source_center  \
404     trip-153788801997503817  IND825409AAA
405     trip-153788801997503817  IND825409AAA
406     trip-153788801997503817  IND825409AAA
407     trip-153788801997503817  IND825409AAA
408     trip-153788801997503817  IND825409AAA
...                         ...           ...
144796  trip-153841785704702048  IND000000ACB
144797  trip-153841785704702048  IND000000ACB
144798  trip-153841785704702048  IND000000ACB
144799  trip-153841785704702048  IND000000ACB
144800  trip-153841785704702048  IND000000ACB

                               source_name destination_center  \
404     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
405     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
406     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
407     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
408     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
...                                   ...                ...
144796      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144797      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144798      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144799      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144800      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA

                        destination_name              od_start_time  ...  \
404        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
405        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
406        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
407        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
408        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
...                                  ...                        ...  ...
144796  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144797  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144798  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144799  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144800  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
```

```
       source_name_city  source_name_place  source_name_code  \
404          JhumriTlya           RadhaCpx                 D
405          JhumriTlya           RadhaCpx                 D
406          JhumriTlya           RadhaCpx                 D
407          JhumriTlya           RadhaCpx                 D
408          JhumriTlya           RadhaCpx                 D
...                 ...                ...               ...
144796          Gurgaon           Bilaspur                HB
144797          Gurgaon           Bilaspur                HB
144798          Gurgaon           Bilaspur                HB
144799          Gurgaon           Bilaspur                HB
144800          Gurgaon           Bilaspur                HB

        source_name_state destination_name_city  destination_name_place  \
404             Jharkhand               Gurgaon                Bilaspur
405             Jharkhand               Gurgaon                Bilaspur
406             Jharkhand               Gurgaon                Bilaspur
407             Jharkhand               Gurgaon                Bilaspur
408             Jharkhand               Gurgaon                Bilaspur
...                   ...                   ...                     ...
144796            Haryana             Bengaluru                Nelmngla
144797            Haryana             Bengaluru                Nelmngla
144798            Haryana             Bengaluru                Nelmngla
144799            Haryana             Bengaluru                Nelmngla
144800            Haryana             Bengaluru                Nelmngla

        destination_name_code  destination_name_state  \
404                        HB                 Haryana
405                        HB                 Haryana
406                        HB                 Haryana
407                        HB                 Haryana
408                        HB                 Haryana
...                       ...                     ...
144796                      H               Karnataka
144797                      H               Karnataka
144798                      H               Karnataka
144799                      H               Karnataka
144800                      H               Karnataka

        time_taken_between_od_start_and_od_end_time  \
404                                          25.852
405                                          25.852
406                                          25.852
407                                          25.852
408                                          25.852
...                                             ...
144796                                       49.162
144797                                       49.162
144798                                       49.162
144799                                       49.162
144800                                       49.162

        start_scan_to_end_scan_in_hrs
404                             25.85
405                             25.85
406                             25.85
407                             25.85
408                             25.85
...                               ...
144796                          49.15
144797                          49.15
144798                          49.15
144799                          49.15
144800                          49.15

[17603 rows x 46 columns]
```

```
Outliers by feature name -->  osrm_distance
            data         trip_creation_time  \
404     training 2018-09-25 15:06:59.975279
405     training 2018-09-25 15:06:59.975279
406     training 2018-09-25 15:06:59.975279
407     training 2018-09-25 15:06:59.975279
408     training 2018-09-25 15:06:59.975279
...          ...                        ...
144796      test 2018-10-01 18:17:37.047270
144797      test 2018-10-01 18:17:37.047270
144798      test 2018-10-01 18:17:37.047270
144799      test 2018-10-01 18:17:37.047270
144800      test 2018-10-01 18:17:37.047270

                                      route_schedule_uuid route_type  \
404     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
405     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
406     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
407     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
408     thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6...        FTL
...                                                   ...        ...
144796  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144797  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144798  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144799  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144800  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL

                      trip_uuid source_center  \
404     trip-153788801997503817  IND825409AAA
405     trip-153788801997503817  IND825409AAA
406     trip-153788801997503817  IND825409AAA
407     trip-153788801997503817  IND825409AAA
408     trip-153788801997503817  IND825409AAA
...                         ...           ...
144796  trip-153841785704702048  IND000000ACB
144797  trip-153841785704702048  IND000000ACB
144798  trip-153841785704702048  IND000000ACB
144799  trip-153841785704702048  IND000000ACB
144800  trip-153841785704702048  IND000000ACB

                              source_name destination_center  \
404     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
405     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
406     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
407     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
408     JhumriTlya_RadhaCpx_D (Jharkhand)       IND000000ACB
...                                   ...                ...
144796      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144797      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144798      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144799      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144800      Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA

                        destination_name              od_start_time  ...  \
404        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
405        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
406        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
407        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
408        Gurgaon_Bilaspur_HB (Haryana) 2018-09-26 03:15:43.970231  ...
...                                  ...                        ...  ...
144796  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144797  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144798  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144799  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144800  Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
```

```
       source_name_city  source_name_place  source_name_code  \
404          JhumriTlya           RadhaCpx                 D   
405          JhumriTlya           RadhaCpx                 D   
406          JhumriTlya           RadhaCpx                 D   
407          JhumriTlya           RadhaCpx                 D   
408          JhumriTlya           RadhaCpx                 D   
...                 ...                ...               ...   
144796          Gurgaon           Bilaspur                HB   
144797          Gurgaon           Bilaspur                HB   
144798          Gurgaon           Bilaspur                HB   
144799          Gurgaon           Bilaspur                HB   
144800          Gurgaon           Bilaspur                HB   

        source_name_state destination_name_city  destination_name_place  \
404             Jharkhand               Gurgaon                Bilaspur   
405             Jharkhand               Gurgaon                Bilaspur   
406             Jharkhand               Gurgaon                Bilaspur   
407             Jharkhand               Gurgaon                Bilaspur   
408             Jharkhand               Gurgaon                Bilaspur   
...                   ...                   ...                     ...   
144796            Haryana             Bengaluru                Nelmngla   
144797            Haryana             Bengaluru                Nelmngla   
144798            Haryana             Bengaluru                Nelmngla   
144799            Haryana             Bengaluru                Nelmngla   
144800            Haryana             Bengaluru                Nelmngla   

        destination_name_code  destination_name_state  \
404                        HB                 Haryana   
405                        HB                 Haryana   
406                        HB                 Haryana   
407                        HB                 Haryana   
408                        HB                 Haryana   
...                       ...                     ...   
144796                      H               Karnataka   
144797                      H               Karnataka   
144798                      H               Karnataka   
144799                      H               Karnataka   
144800                      H               Karnataka   

        time_taken_between_od_start_and_od_end_time  \
404                                          25.852   
405                                          25.852   
406                                          25.852   
407                                          25.852   
408                                          25.852   
...                                             ...   
144796                                       49.162   
144797                                       49.162   
144798                                       49.162   
144799                                       49.162   
144800                                       49.162   

        start_scan_to_end_scan_in_hrs  
404                             25.85  
405                             25.85  
406                             25.85  
407                             25.85  
408                             25.85  
...                               ...  
144796                          49.15  
144797                          49.15  
144798                          49.15  
144799                          49.15  
144800                          49.15  

[17816 rows x 46 columns]
```

```
Outliers by feature name -->  factor
            data         trip_creation_time  \
15      training 2018-09-14 15:42:46.437249
16      training 2018-09-14 15:42:46.437249
76          test 2018-09-27 14:16:14.819357
77          test 2018-09-27 14:16:14.819357
78          test 2018-09-27 14:16:14.819357
...          ...                        ...
144634  training 2018-09-18 00:34:51.206487
144658  training 2018-09-12 00:14:49.629525
144848  training 2018-09-14 18:45:34.164734
144854  training 2018-09-17 11:35:28.838714
144866  training 2018-09-20 16:24:28.436231

                                     route_schedule_uuid route_type  \
15      thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d...    Carting
16      thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d...    Carting
76      thanos::sroute:1283977c-889a-4e96-b632-5ba1a69...    Carting
77      thanos::sroute:1283977c-889a-4e96-b632-5ba1a69...    Carting
78      thanos::sroute:1283977c-889a-4e96-b632-5ba1a69...    Carting
...                                                   ...        ...
144634  thanos::sroute:387e7ab9-2237-48b1-af49-2508ce2...        FTL
144658  thanos::sroute:b62ab3ed-c60b-47d2-8c91-fe62135...    Carting
144848  thanos::sroute:40b6dc9c-faa1-4753-8bc8-ac8c3e0...    Carting
144854  thanos::sroute:d8f74492-4484-412a-887a-61c8e6b...    Carting
144866  thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...    Carting

                      trip_uuid source_center  \
15      trip-153693976643699843  IND400011AAA
16      trip-153693976643699843  IND400011AAA
76      trip-153805777481903807  IND600056AAB
77      trip-153805777481903807  IND600056AAB
78      trip-153805777481903807  IND600056AAB
...                         ...           ...
144634  trip-153723089120625505  IND151001AAA
144658  trip-153671128962918389  IND302014AAB
144848  trip-153695073416451616  IND400102AAB
144854  trip-153718412883843340  IND600056AAB
144866  trip-153746066843555182  IND131028AAB

                            source_name destination_center  \
15          LowerParel_CP (Maharashtra)       IND400072AAD
16          LowerParel_CP (Maharashtra)       IND400072AAD
76       MAA_Poonamallee_HB (Tamil Nadu)       IND600032AAB
77       MAA_Poonamallee_HB (Tamil Nadu)       IND600032AAB
78       MAA_Poonamallee_HB (Tamil Nadu)       IND600032AAB
...                                 ...                ...
144634              Bhatinda_DPC (Punjab)       IND151302AAA
144658    Jaipur_Central_I_7 (Rajasthan)       IND302026AAA
144848  Mumbai_Jogeshwri_L (Maharashtra)       IND421302AAG
144854   MAA_Poonamallee_HB (Tamil Nadu)       IND600032AAB
144866        Sonipat_Kundli_H (Haryana)       IND000000ACB

                       destination_name              od_start_time  ...  \
15       Mumbai_Chndivli_PC (Maharashtra) 2018-09-14 15:42:46.437249  ...
16       Mumbai_Chndivli_PC (Maharashtra) 2018-09-14 15:42:46.437249  ...
76               Chennai_Hub (Tamil Nadu) 2018-09-27 14:16:14.819357  ...
77               Chennai_Hub (Tamil Nadu) 2018-09-27 14:16:14.819357  ...
78               Chennai_Hub (Tamil Nadu) 2018-09-27 14:16:14.819357  ...
...                                   ...                        ...  ...
144634    TalwandiSabo_Wardno3_D (Punjab) 2018-09-18 00:34:51.206487  ...
144658     Jaipur_Bhankrot_DC (Rajasthan) 2018-09-12 00:14:49.629525  ...
144848  Bhiwandi_Mankoli_HB (Maharashtra) 2018-09-14 18:45:34.164734  ...
144854           Chennai_Hub (Tamil Nadu) 2018-09-17 11:35:28.838714  ...
144866      Gurgaon_Bilaspur_HB (Haryana) 2018-09-20 16:24:28.436231  ...
```

```
       source_name_city source_name_place source_name_code  \
15                  NaN               NaN              NaN   
16                  NaN               NaN              NaN   
76                  NaN               NaN              NaN   
77                  NaN               NaN              NaN   
78                  NaN               NaN              NaN   
...                 ...               ...              ...   
144634              NaN               NaN              NaN   
144658              NaN               NaN              NaN   
144848           Mumbai         Jogeshwri                L   
144854              NaN               NaN              NaN   
144866          Sonipat            Kundli                H   

        source_name_state destination_name_city destination_name_place  \
15                    NaN                Mumbai               Chndivli   
16                    NaN                Mumbai               Chndivli   
76                    NaN                   NaN                    NaN   
77                    NaN                   NaN                    NaN   
78                    NaN                   NaN                    NaN   
...                   ...                   ...                    ...   
144634                NaN                   NaN                    NaN   
144658                NaN                Jaipur               Bhankrot   
144848        Maharashtra              Bhiwandi                Mankoli   
144854                NaN                   NaN                    NaN   
144866            Haryana               Gurgaon               Bilaspur   

        destination_name_code  destination_name_state  \
15                         PC             Maharashtra   
16                         PC             Maharashtra   
76                        NaN                     NaN   
77                        NaN                     NaN   
78                        NaN                     NaN   
...                       ...                     ...   
144634                    NaN                     NaN   
144658                     DC               Rajasthan   
144848                     HB             Maharashtra   
144854                    NaN                     NaN   
144866                     HB                 Haryana   

        time_taken_between_od_start_and_od_end_time  \
15                                            1.816   
16                                            1.816   
76                                            2.999   
77                                            2.999   
78                                            2.999   
...                                             ...   
144634                                        1.639   
144658                                        1.843   
144848                                        6.565   
144854                                        1.948   
144866                                        7.128   

        start_scan_to_end_scan_in_hrs  
15                              1.800  
16                              1.800  
76                              2.983  
77                              2.983  
78                              2.983  
...                               ...  
144634                          1.633  
144658                          1.833  
144848                          6.550  
144854                          1.933  
144866                          7.117  

[11429 rows x 46 columns]
```

```
Outliers by feature name -->  segment_actual_time
            data         trip_creation_time  \
21      training 2018-09-13 20:44:19.424489
34      training 2018-09-13 20:44:19.424489
72          test 2018-10-01 16:00:45.719099
73          test 2018-10-01 16:00:45.719099
106     training 2018-09-25 08:53:04.377810
...          ...                        ...
144790      test 2018-10-01 18:17:37.047270
144819  training 2018-09-26 14:05:52.096792
144848  training 2018-09-14 18:45:34.164734
144853  training 2018-09-22 11:30:41.399439
144866  training 2018-09-20 16:24:28.436231

                                      route_schedule_uuid route_type  \
21      thanos::sroute:76951383-1608-44e4-a284-46d92e8...        FTL
34      thanos::sroute:76951383-1608-44e4-a284-46d92e8...        FTL
72      thanos::sroute:182fdad4-dc74-4c7d-ada4-0a76d4c...        FTL
73      thanos::sroute:182fdad4-dc74-4c7d-ada4-0a76d4c...        FTL
106     thanos::sroute:4460a38d-ab9b-484e-bd4e-f4201d0...        FTL
...                                                   ...        ...
144790  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144819  thanos::sroute:f7de4133-6bd9-4367-a7f7-ab190b6...        FTL
144848  thanos::sroute:40b6dc9c-faa1-4753-8bc8-ac8c3e0...    Carting
144853  thanos::sroute:d81088e2-9ccd-43e9-9260-3e85633...        FTL
144866  thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...    Carting

                      trip_uuid source_center  \
21      trip-153687145942424248  IND560099AAB
34      trip-153687145942424248  IND560099AAB
72      trip-153840964571880594  IND131028AAB
73      trip-153840964571880594  IND247667AAB
106     trip-153786558437756691  IND306401AAB
...                         ...           ...
144790  trip-153841785704702048  IND000000ACB
144819  trip-153797075209653066  IND413401AAA
144848  trip-153695073416451616  IND400102AAB
144853  trip-153761584139918815  IND421302AAG
144866  trip-153746066843555182  IND131028AAB

                              source_name destination_center  \
21      Bengaluru_Bomsndra_HB (Karnataka)       IND683511AAA
34      Bengaluru_Bomsndra_HB (Karnataka)       IND683511AAA
72             Sonipat_Kundli_H (Haryana)       IND247667AAB
73       Roorkee_IOTCEncl_L (Uttarakhand)       IND249407AAA
106           Pali_Nayagaon_I (Rajasthan)       IND342005AAD
...                                   ...                ...
144790     Gurgaon_Bilaspur_HB (Haryana)        IND562132AAA
144819    Barshi_Kurduwdi_D (Maharashtra)       IND411033AAA
144848   Mumbai_Jogeshwri_L (Maharashtra)       IND421302AAG
144853  Bhiwandi_Mankoli_HB (Maharashtra)       IND411033AAA
144866         Sonipat_Kundli_H (Haryana)       IND000000ACB

                         destination_name              od_start_time  ...  \
21                Aluva_Peedika_H (Kerala) 2018-09-13 23:59:56.061158  ...
34                Aluva_Peedika_H (Kerala) 2018-09-13 23:59:56.061158  ...
72       Roorkee_IOTCEncl_L (Uttarakhand) 2018-10-01 16:00:45.719099  ...
73                 Haridwar (Uttarakhand) 2018-10-01 23:55:44.909377  ...
106           Jodhpur_Basni_I (Rajasthan) 2018-09-25 21:31:00.741991  ...
...                                   ...                        ...  ...
144790   Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144819      Pune_Tathawde_H (Maharashtra) 2018-09-27 02:21:39.762995  ...
144848  Bhiwandi_Mankoli_HB (Maharashtra) 2018-09-14 18:45:34.164734  ...
144853      Pune_Tathawde_H (Maharashtra) 2018-09-22 11:30:41.399439  ...
144866     Gurgaon_Bilaspur_HB (Haryana) 2018-09-20 16:24:28.436231  ...
```

```
       source_name_city  source_name_place  source_name_code  \
21            Bengaluru           Bomsndra                HB
34            Bengaluru           Bomsndra                HB
72              Sonipat             Kundli                 H
73              Roorkee           IOTCEncl                 L
106                Pali           Nayagaon                 I
...                 ...                ...               ...
144790          Gurgaon           Bilaspur                HB
144819           Barshi           Kurduwdi                 D
144848           Mumbai          Jogeshwri                 L
144853         Bhiwandi            Mankoli                HB
144866          Sonipat             Kundli                 H

        source_name_state destination_name_city  destination_name_place  \
21              Karnataka                 Aluva                 Peedika
34              Karnataka                 Aluva                 Peedika
72                Haryana               Roorkee                IOTCEncl
73            Uttarakhand                   NaN                     NaN
106             Rajasthan               Jodhpur                   Basni
...                   ...                   ...                     ...
144790            Haryana             Bengaluru                Nelmngla
144819        Maharashtra                  Pune                Tathawde
144848        Maharashtra              Bhiwandi                 Mankoli
144853        Maharashtra                  Pune                Tathawde
144866            Haryana               Gurgaon                Bilaspur

        destination_name_code  destination_name_state  \
21                          H                  Kerala
34                          H                  Kerala
72                          L             Uttarakhand
73                        NaN                     NaN
106                         I               Rajasthan
...                       ...                     ...
144790                      H               Karnataka
144819                      H             Maharashtra
144848                     HB             Maharashtra
144853                      H             Maharashtra
144866                     HB                 Haryana

        time_taken_between_od_start_and_od_end_time  \
21                                           13.934
34                                           13.934
72                                            7.719
73                                            1.871
106                                           2.849
...                                             ...
144790                                       49.162
144819                                        6.224
144848                                        6.565
144853                                       10.240
144866                                        7.128

        start_scan_to_end_scan_in_hrs
21                             13.933
34                             13.933
72                              7.717
73                              1.867
106                             2.833
...                               ...
144790                         49.150
144819                          6.217
144848                          6.550
144853                         10.233
144866                          7.117

[9298 rows x 46 columns]
```

```
Outliers by feature name -->  segment_osrm_time
            data         trip_creation_time  \
34      training 2018-09-13 20:44:19.424489
38          test 2018-09-29 22:21:45.149226
157     training 2018-09-15 23:58:16.827101
158     training 2018-09-15 23:58:16.827101
214     training 2018-09-17 00:14:20.789064
...          ...                        ...
144802  training 2018-09-26 14:05:52.096792
144829  training 2018-09-26 19:50:29.657378
144837  training 2018-09-26 19:50:29.657378
144843  training 2018-09-26 19:50:29.657378
144845  training 2018-09-26 19:50:29.657378

                                     route_schedule_uuid route_type  \
34      thanos::sroute:76951383-1608-44e4-a284-46d92e8...        FTL
38      thanos::sroute:0904e75c-b3ac-4278-96cf-802835a...        FTL
157     thanos::sroute:fb308c0f-ea3a-48ef-a6c3-4776341...        FTL
158     thanos::sroute:fb308c0f-ea3a-48ef-a6c3-4776341...        FTL
214     thanos::sroute:d0cd2cb2-ce42-4103-b999-f8899e9...        FTL
...                                                   ...        ...
144802  thanos::sroute:f7de4133-6bd9-4367-a7f7-ab190b6...        FTL
144829  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL
144837  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL
144843  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL
144845  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL

                      trip_uuid source_center  \
34      trip-153687145942424248  IND560099AAB
38      trip-153825970514894360  IND141109AAA
157     trip-153705589682687518  IND206001AAA
158     trip-153705589682687518  IND206001AAA
214     trip-153714326609873773  IND431517AAB
...                         ...           ...
144802  trip-153797075209653066  IND411033AAA
144829  trip-153799142965708367  IND454001AAA
144837  trip-153799142965708367  IND457226AAA
144843  trip-153799142965708367  IND457226AAA
144845  trip-153799142965708367  IND457226AAA

                               source_name destination_center  \
34       Bengaluru_Bomsndra_HB (Karnataka)       IND683511AAA
38                      Raikot_DC (Punjab)       IND000000ACA
157      Etawah_MhraChng_D (Uttar Pradesh)       IND000000ACB
158      Etawah_MhraChng_D (Uttar Pradesh)       IND000000ACB
214     Ambajogai_BnsllNgr_D (Maharashtra)       IND411033AAA
...                                    ...                ...
144802        Pune_Tathawde_H (Maharashtra)       IND413002AAA
144829    Dhar_Trimurti_D (Madhya Pradesh)       IND457001AAA
144837   Jaora_RtlamNka_D (Madhya Pradesh)       IND382430AAB
144843   Jaora_RtlamNka_D (Madhya Pradesh)       IND382430AAB
144845   Jaora_RtlamNka_D (Madhya Pradesh)       IND382430AAB

                          destination_name              od_start_time  ...  \
34                Aluva_Peedika_H (Kerala) 2018-09-13 23:59:56.061158  ...
38           Ludhiana_MilrGanj_HB (Punjab) 2018-09-30 05:12:34.503322  ...
157          Gurgaon_Bilaspur_HB (Haryana) 2018-09-17 02:46:57.274441  ...
158          Gurgaon_Bilaspur_HB (Haryana) 2018-09-17 02:46:57.274441  ...
214          Pune_Tathawde_H (Maharashtra) 2018-09-17 05:21:32.158856  ...
...                                    ...                        ...  ...
144802    Solapur_Central_I_2 (Maharashtra) 2018-09-26 14:05:52.096792  ...
144829  Ratlam_Khjurwli_DC (Madhya Pradesh) 2018-09-27 02:48:14.315366  ...
144837         Ahmedabad_East_H_1 (Gujarat) 2018-09-27 06:55:50.265761  ...
144843         Ahmedabad_East_H_1 (Gujarat) 2018-09-27 06:55:50.265761  ...
144845         Ahmedabad_East_H_1 (Gujarat) 2018-09-27 06:55:50.265761  ...
```

```
       source_name_city  source_name_place  source_name_code  \
34            Bengaluru           Bomsndra                HB   
38                  NaN                NaN               NaN   
157                 NaN                NaN               NaN   
158                 NaN                NaN               NaN   
214           Ambajogai           BnsllNgr                 D   
...                 ...                ...               ...   
144802             Pune           Tathawde                 H   
144829              NaN                NaN               NaN   
144837              NaN                NaN               NaN   
144843              NaN                NaN               NaN   
144845              NaN                NaN               NaN   

        source_name_state destination_name_city  destination_name_place  \
34              Karnataka                 Aluva                 Peedika   
38                    NaN              Ludhiana                MilrGanj   
157                   NaN               Gurgaon                Bilaspur   
158                   NaN               Gurgaon                Bilaspur   
214           Maharashtra                  Pune                Tathawde   
...                   ...                   ...                     ...   
144802        Maharashtra                   NaN                     NaN   
144829                NaN                   NaN                     NaN   
144837                NaN                   NaN                     NaN   
144843                NaN                   NaN                     NaN   
144845                NaN                   NaN                     NaN   

        destination_name_code  destination_name_state  \
34                          H                  Kerala   
38                         HB                  Punjab   
157                        HB                 Haryana   
158                        HB                 Haryana   
214                         H             Maharashtra   
...                       ...                     ...   
144802                    NaN                     NaN   
144829                    NaN                     NaN   
144837                    NaN                     NaN   
144843                    NaN                     NaN   
144845                    NaN                     NaN   

        time_taken_between_od_start_and_od_end_time  \
34                                           13.934   
38                                            1.536   
157                                           8.350   
158                                           8.350   
214                                          10.954   
...                                             ...   
144802                                        9.260   
144829                                        2.303   
144837                                       10.379   
144843                                       10.379   
144845                                       10.379   

        start_scan_to_end_scan_in_hrs  
34                             13.933  
38                              1.533  
157                             8.350  
158                             8.350  
214                            10.950  
...                               ...  
144802                          9.250  
144829                          2.300  
144837                         10.367  
144843                         10.367  
144845                         10.367  

[6378 rows x 46 columns]
```

```
Outliers by feature name -->  segment_osrm_distance
            data          trip_creation_time  \
34      training 2018-09-13 20:44:19.424489
157     training 2018-09-15 23:58:16.827101
158     training 2018-09-15 23:58:16.827101
214     training 2018-09-17 00:14:20.789064
316     training 2018-09-24 02:57:00.372087
...          ...                         ...
144774      test 2018-10-01 18:17:37.047270
144802  training 2018-09-26 14:05:52.096792
144829  training 2018-09-26 19:50:29.657378
144837  training 2018-09-26 19:50:29.657378
144845  training 2018-09-26 19:50:29.657378

                                     route_schedule_uuid route_type  \
34      thanos::sroute:76951383-1608-44e4-a284-46d92e8...        FTL
157     thanos::sroute:fb308c0f-ea3a-48ef-a6c3-4776341...        FTL
158     thanos::sroute:fb308c0f-ea3a-48ef-a6c3-4776341...        FTL
214     thanos::sroute:d0cd2cb2-ce42-4103-b999-f8899e9...        FTL
316     thanos::sroute:8f136f2a-7552-4c91-acfa-ff555d1...        FTL
...                                                   ...        ...
144774  thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...        FTL
144802  thanos::sroute:f7de4133-6bd9-4367-a7f7-ab190b6...        FTL
144829  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL
144837  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL
144845  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL

                      trip_uuid source_center  \
34      trip-153687145942424248  IND560099AAB
157     trip-153705589682687518  IND206001AAA
158     trip-153705589682687518  IND206001AAA
214     trip-153714326609873773  IND431517AAB
316     trip-153775782037183132  IND413002AAA
...                         ...           ...
144774  trip-153841785704702048  IND000000ACB
144802  trip-153797075209653066  IND411033AAA
144829  trip-153799142965708367  IND454001AAA
144837  trip-153799142965708367  IND457226AAA
144845  trip-153799142965708367  IND457226AAA

                                source_name destination_center  \
34       Bengaluru_Bomsndra_HB (Karnataka)       IND683511AAA
157      Etawah_MhraChng_D (Uttar Pradesh)       IND000000ACB
158      Etawah_MhraChng_D (Uttar Pradesh)       IND000000ACB
214     Ambajogai_BnsllNgr_D (Maharashtra)       IND411033AAA
316      Solapur_Central_I_2 (Maharashtra)       IND501359AAE
...                                     ...                ...
144774       Gurgaon_Bilaspur_HB (Haryana)       IND562132AAA
144802       Pune_Tathawde_H (Maharashtra)       IND413002AAA
144829    Dhar_Trimurti_D (Madhya Pradesh)       IND457001AAA
144837   Jaora_RtlamNka_D (Madhya Pradesh)       IND382430AAB
144845   Jaora_RtlamNka_D (Madhya Pradesh)       IND382430AAB

                           destination_name              od_start_time  ...  \
34                Aluva_Peedika_H (Kerala) 2018-09-13 23:59:56.061158  ...
157           Gurgaon_Bilaspur_HB (Haryana) 2018-09-17 02:46:57.274441  ...
158           Gurgaon_Bilaspur_HB (Haryana) 2018-09-17 02:46:57.274441  ...
214           Pune_Tathawde_H (Maharashtra) 2018-09-17 05:21:32.158856  ...
316        Hyderabad_Shamshbd_H (Telangana) 2018-09-24 12:46:04.801166  ...
...                                     ...                        ...  ...
144774     Bangalore_Nelmngla_H (Karnataka) 2018-10-02 09:02:19.284969  ...
144802    Solapur_Central_I_2 (Maharashtra) 2018-09-26 14:05:52.096792  ...
144829  Ratlam_Khjurwli_DC (Madhya Pradesh) 2018-09-27 02:48:14.315366  ...
144837          Ahmedabad_East_H_1 (Gujarat) 2018-09-27 06:55:50.265761  ...
144845          Ahmedabad_East_H_1 (Gujarat) 2018-09-27 06:55:50.265761  ...
```

```
       source_name_city  source_name_place  source_name_code  \
34            Bengaluru           Bomsndra                HB   
157                 NaN                NaN               NaN   
158                 NaN                NaN               NaN   
214           Ambajogai           BnsllNgr                 D   
316                 NaN                NaN               NaN   
...                 ...                ...               ...   
144774          Gurgaon           Bilaspur                HB   
144802             Pune           Tathawde                 H   
144829              NaN                NaN               NaN   
144837              NaN                NaN               NaN   
144845              NaN                NaN               NaN   

        source_name_state destination_name_city  destination_name_place  \
34              Karnataka                 Aluva                 Peedika   
157                   NaN               Gurgaon                Bilaspur   
158                   NaN               Gurgaon                Bilaspur   
214           Maharashtra                  Pune                Tathawde   
316                   NaN             Hyderabad                Shamshbd   
...                   ...                   ...                     ...   
144774            Haryana             Bengaluru                Nelmngla   
144802        Maharashtra                   NaN                     NaN   
144829                NaN                   NaN                     NaN   
144837                NaN                   NaN                     NaN   
144845                NaN                   NaN                     NaN   

        destination_name_code  destination_name_state  \
34                          H                  Kerala   
157                        HB                 Haryana   
158                        HB                 Haryana   
214                         H             Maharashtra   
316                         H               Telangana   
...                       ...                     ...   
144774                      H               Karnataka   
144802                    NaN                     NaN   
144829                    NaN                     NaN   
144837                    NaN                     NaN   
144845                    NaN                     NaN   

        time_taken_between_od_start_and_od_end_time  \
34                                           13.934   
157                                           8.350   
158                                           8.350   
214                                          10.954   
316                                          10.822   
...                                             ...   
144774                                       49.162   
144802                                        9.260   
144829                                        2.303   
144837                                       10.379   
144845                                       10.379   

        start_scan_to_end_scan_in_hrs  
34                             13.933  
157                             8.350  
158                             8.350  
214                            10.950  
316                            10.817  
...                               ...  
144774                         49.150  
144802                          9.250  
144829                          2.300  
144837                         10.367  
144845                         10.367  

[4315 rows x 46 columns]
```

```
Outliers by feature name -->  segment_factor
            data         trip_creation_time  \
6       training 2018-09-20 02:35:36.476840
9       training 2018-09-20 02:35:36.476840
15      training 2018-09-14 15:42:46.437249
47      training 2018-09-12 01:33:48.711350
54      training 2018-09-12 01:33:48.711350
...          ...                        ...
144846  training 2018-09-26 19:50:29.657378
144848  training 2018-09-14 18:45:34.164734
144852  training 2018-09-22 11:30:41.399439
144853  training 2018-09-22 11:30:41.399439
144866  training 2018-09-20 16:24:28.436231

                                      route_schedule_uuid route_type  \
6       thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...    Carting
9       thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...    Carting
15      thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d...    Carting
47      thanos::sroute:5f7d8d49-ae14-430e-9333-37361e1...    Carting
54      thanos::sroute:5f7d8d49-ae14-430e-9333-37361e1...    Carting
...                                                   ...        ...
144846  thanos::sroute:f6d1ba62-76a2-4dba-83ec-3ac0803...        FTL
144848  thanos::sroute:40b6dc9c-faa1-4753-8bc8-ac8c3e0...    Carting
144852  thanos::sroute:d81088e2-9ccd-43e9-9260-3e85633...        FTL
144853  thanos::sroute:d81088e2-9ccd-43e9-9260-3e85633...        FTL
144866  thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...    Carting

                      trip_uuid source_center  \
6       trip-153741093647649320  IND388620AAB
9       trip-153741093647649320  IND388620AAB
15      trip-153693976643699843  IND400011AAA
47      trip-153671602871109556  IND362001AAA
54      trip-153671602871109556  IND362720AAA
...                         ...           ...
144846  trip-153799142965708367  IND457226AAA
144848  trip-153695073416451616  IND400102AAB
144852  trip-153761584139918815  IND421302AAG
144853  trip-153761584139918815  IND421302AAG
144866  trip-153746066843555182  IND131028AAB

                                source_name destination_center  \
6          Khambhat_MotvdDPP_D (Gujarat)       IND388320AAA
9          Khambhat_MotvdDPP_D (Gujarat)       IND388320AAA
15           LowerParel_CP (Maharashtra)       IND400072AAD
47                Junagadh_DPC (Gujarat)       IND362220AAA
54          Kodinar_NCplxDPP_D (Gujarat)       IND362560AAA
...                                  ...                ...
144846  Jaora_RtlamNka_D (Madhya Pradesh)       IND382430AAB
144848   Mumbai_Jogeshwri_L (Maharashtra)       IND421302AAG
144852  Bhiwandi_Mankoli_HB (Maharashtra)       IND411033AAA
144853  Bhiwandi_Mankoli_HB (Maharashtra)       IND411033AAA
144866         Sonipat_Kundli_H (Haryana)       IND000000ACB

                          destination_name              od_start_time  ...  \
6              Anand_Vaghasi_IP (Gujarat) 2018-09-20 04:47:45.236797  ...
9              Anand_Vaghasi_IP (Gujarat) 2018-09-20 04:47:45.236797  ...
15       Mumbai_Chndivli_PC (Maharashtra) 2018-09-14 15:42:46.437249  ...
47          Junagadh_keshod_DC (Gujarat) 2018-09-12 01:33:48.711350  ...
54             Una_Mamlatdr_DC (Gujarat) 2018-09-12 06:12:09.579013  ...
...                                    ...                        ...  ...
144846      Ahmedabad_East_H_1 (Gujarat) 2018-09-27 06:55:50.265761  ...
144848  Bhiwandi_Mankoli_HB (Maharashtra) 2018-09-14 18:45:34.164734  ...
144852      Pune_Tathawde_H (Maharashtra) 2018-09-22 11:30:41.399439  ...
144853      Pune_Tathawde_H (Maharashtra) 2018-09-22 11:30:41.399439  ...
144866     Gurgaon_Bilaspur_HB (Haryana) 2018-09-20 16:24:28.436231  ...
```

```
       source_name_city source_name_place source_name_code  \
6              Khambhat          MotvdDPP                D
9              Khambhat          MotvdDPP                D
15                  NaN               NaN              NaN
47                  NaN               NaN              NaN
54              Kodinar          NCplxDPP                D
...                 ...               ...              ...
144846              NaN               NaN              NaN
144848           Mumbai         Jogeshwri                L
144852         Bhiwandi           Mankoli               HB
144853         Bhiwandi           Mankoli               HB
144866          Sonipat            Kundli                H

        source_name_state destination_name_city destination_name_place  \
6                 Gujarat                 Anand                Vaghasi
9                 Gujarat                 Anand                Vaghasi
15                    NaN                Mumbai               Chndivli
47                    NaN              Junagadh                 keshod
54                Gujarat                   Una               Mamlatdr
...                   ...                   ...                    ...
144846                NaN                   NaN                    NaN
144848        Maharashtra              Bhiwandi                Mankoli
144852        Maharashtra                  Pune               Tathawde
144853        Maharashtra                  Pune               Tathawde
144866            Haryana               Gurgaon               Bilaspur

        destination_name_code  destination_name_state  \
6                          IP                 Gujarat
9                          IP                 Gujarat
15                         PC             Maharashtra
47                         DC                 Gujarat
54                         DC                 Gujarat
...                       ...                     ...
144846                    NaN                     NaN
144848                     HB             Maharashtra
144852                      H             Maharashtra
144853                      H             Maharashtra
144866                     HB                 Haryana

        time_taken_between_od_start_and_od_end_time  \
6                                             1.820
9                                             1.820
15                                            1.816
47                                            1.143
54                                            1.121
...                                             ...
144846                                       10.379
144848                                        6.565
144852                                       10.240
144853                                       10.240
144866                                        7.128

        start_scan_to_end_scan_in_hrs
6                               1.817
9                               1.817
15                              1.800
47                              1.133
54                              1.117
...                               ...
144846                         10.367
144848                          6.550
144852                         10.233
144853                         10.233
144866                          7.117

[13976 rows x 46 columns]
```

```
Outliers by feature name -->  trip_creation_time_year
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  trip_creation_time_weekday
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  trip_creation_time_day
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  od_start_time_year
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  od_start_time_weekday
Empty DataFrame
```

```
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  od_start_time_day
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  od_end_time_year
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  od_end_time_weekday
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  od_end_time_day
Empty DataFrame
Columns: [data, trip_creation_time, route_schedule_uuid, route_type, trip_uuid, source_c
enter, source_name, destination_center, destination_name, od_start_time, od_end_time, st
```

```
art_scan_to_end_scan, is_cutoff, cutoff_factor, cutoff_timestamp, actual_distance_to_des
tination, actual_time, osrm_time, osrm_distance, factor, segment_actual_time, segment_os
rm_time, segment_osrm_distance, segment_factor, trip_creation_time_year, trip_creation_t
ime_month, trip_creation_time_weekday, trip_creation_time_day, od_start_time_year, od_st
art_time_month, od_start_time_weekday, od_start_time_day, od_end_time_year, od_end_time_
month, od_end_time_weekday, od_end_time_day, source_name_city, source_name_place, source
_name_code, source_name_state, destination_name_city, destination_name_place, destinatio
n_name_code, destination_name_state, time_taken_between_od_start_and_od_end_time, start_
scan_to_end_scan_in_hrs]
Index: []

[0 rows x 46 columns]
Outliers by feature name -->  time_taken_between_od_start_and_od_end_time
            data         trip_creation_time  \
32950   training 2018-09-13 01:28:45.326644
32951   training 2018-09-13 01:28:45.326644
32952   training 2018-09-13 01:28:45.326644
32953   training 2018-09-13 01:28:45.326644
32954   training 2018-09-13 01:28:45.326644
...          ...                        ...
79524   training 2018-09-19 13:44:58.665210
79525   training 2018-09-19 13:44:58.665210
79526   training 2018-09-19 13:44:58.665210
79527   training 2018-09-19 13:44:58.665210
123196      test 2018-10-01 23:35:54.432745

                                      route_schedule_uuid route_type  \
32950   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32951   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32952   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32953   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32954   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
...                                                   ...        ...
79524   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79525   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79526   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79527   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
123196  thanos::sroute:4316e05f-b4cc-4ea7-b801-62a93ae...    Carting

                      trip_uuid source_center  \
32950   trip-153680212532637033  IND712311AAA
32951   trip-153680212532637033  IND712311AAA
32952   trip-153680212532637033  IND712311AAA
32953   trip-153680212532637033  IND712311AAA
32954   trip-153680212532637033  IND712311AAA
...                         ...           ...
79524   trip-153736469866480991  IND000000ACB
79525   trip-153736469866480991  IND000000ACB
79526   trip-153736469866480991  IND000000ACB
79527   trip-153736469866480991  IND000000ACB
123196  trip-153843695443252828  IND764071AAB

                                source_name destination_center  \
32950      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32951      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32952      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32953      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32954      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
...                                     ...                ...
79524         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79525         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79526         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79527         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
123196  Pappadahandi_Central_DPP_2 (Orissa)       IND530012AAA

                                  destination_name              od_start_time  \
```

```
32950                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32951                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32952                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32953                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32954                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
...                                       ...                        ...
79524           Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79525           Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79526           Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79527           Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
123196  Visakhapatnam_Gajuwaka_IP (Andhra Pradesh) 2018-10-02 15:21:51.236205

        ... source_name_city  source_name_place  source_name_code  \
32950   ...              NaN                NaN               NaN
32951   ...              NaN                NaN               NaN
32952   ...              NaN                NaN               NaN
32953   ...              NaN                NaN               NaN
32954   ...              NaN                NaN               NaN
...     ...              ...                ...               ...
79524   ...          Gurgaon           Bilaspur                HB
79525   ...          Gurgaon           Bilaspur                HB
79526   ...          Gurgaon           Bilaspur                HB
79527   ...          Gurgaon           Bilaspur                HB
123196  ...              NaN                NaN               NaN

        source_name_state destination_name_city  destination_name_place  \
32950                 NaN                   NaN                     NaN
32951                 NaN                   NaN                     NaN
32952                 NaN                   NaN                     NaN
32953                 NaN                   NaN                     NaN
32954                 NaN                   NaN                     NaN
...                   ...                   ...                     ...
79524             Haryana                   NaN                     NaN
79525             Haryana                   NaN                     NaN
79526             Haryana                   NaN                     NaN
79527             Haryana                   NaN                     NaN
123196                NaN                   NaN                     NaN

        destination_name_code  destination_name_state  \
32950                     NaN                     NaN
32951                     NaN                     NaN
32952                     NaN                     NaN
32953                     NaN                     NaN
32954                     NaN                     NaN
...                       ...                     ...
79524                     NaN                     NaN
79525                     NaN                     NaN
79526                     NaN                     NaN
79527                     NaN                     NaN
123196                    NaN                     NaN

        time_taken_between_od_start_and_od_end_time  \
32950                                        64.959
32951                                        64.959
32952                                        64.959
32953                                        64.959
32954                                        64.959
...                                             ...
79524                                        70.658
79525                                        70.658
79526                                        70.658
79527                                        70.658
123196                                      131.643

        start_scan_to_end_scan_in_hrs
32950                          64.950
```

```
32951                           64.950
32952                           64.950
32953                           64.950
32954                           64.950
...                                ...
79524                           70.650
79525                           70.650
79526                           70.650
79527                           70.650
123196                         131.633

[373 rows x 46 columns]
Outliers by feature name -->  start_scan_to_end_scan_in_hrs
            data        trip_creation_time  \
32950   training 2018-09-13 01:28:45.326644
32951   training 2018-09-13 01:28:45.326644
32952   training 2018-09-13 01:28:45.326644
32953   training 2018-09-13 01:28:45.326644
32954   training 2018-09-13 01:28:45.326644
...          ...                       ...
79524   training 2018-09-19 13:44:58.665210
79525   training 2018-09-19 13:44:58.665210
79526   training 2018-09-19 13:44:58.665210
79527   training 2018-09-19 13:44:58.665210
123196      test 2018-10-01 23:35:54.432745

                                      route_schedule_uuid route_type  \
32950   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32951   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32952   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32953   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
32954   thanos::sroute:6b87651c-fdf4-432f-bf80-0e394f3...        FTL
...                                                   ...        ...
79524   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79525   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79526   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
79527   thanos::sroute:bc7dbb1d-9379-4674-b8d3-f9c3b96...        FTL
123196  thanos::sroute:4316e05f-b4cc-4ea7-b801-62a93ae...    Carting

                     trip_uuid source_center  \
32950   trip-153680212532637033  IND712311AAA
32951   trip-153680212532637033  IND712311AAA
32952   trip-153680212532637033  IND712311AAA
32953   trip-153680212532637033  IND712311AAA
32954   trip-153680212532637033  IND712311AAA
...                         ...           ...
79524   trip-153736469866480991  IND000000ACB
79525   trip-153736469866480991  IND000000ACB
79526   trip-153736469866480991  IND000000ACB
79527   trip-153736469866480991  IND000000ACB
123196  trip-153843695443252828  IND764071AAB

                                source_name destination_center  \
32950      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32951      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32952      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32953      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
32954      Kolkata_Dankuni_HB (West Bengal)       IND781018AAB
...                                     ...                ...
79524         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79525         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79526         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
79527         Gurgaon_Bilaspur_HB (Haryana)       IND712311AAA
123196  Pappadahandi_Central_DPP_2 (Orissa)       IND530012AAA

                                 destination_name              od_start_time  \
```

```
32950                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32951                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32952                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32953                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
32954                    Guwahati_Hub (Assam) 2018-09-13 01:28:45.326644
...                                       ...                        ...
79524            Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79525            Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79526            Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
79527            Kolkata_Dankuni_HB (West Bengal) 2018-09-19 13:44:58.665210
123196  Visakhapatnam_Gajuwaka_IP (Andhra Pradesh) 2018-10-02 15:21:51.236205

        ... source_name_city  source_name_place  source_name_code  \
32950   ...              NaN                NaN               NaN
32951   ...              NaN                NaN               NaN
32952   ...              NaN                NaN               NaN
32953   ...              NaN                NaN               NaN
32954   ...              NaN                NaN               NaN
...     ...              ...                ...               ...
79524   ...          Gurgaon           Bilaspur                HB
79525   ...          Gurgaon           Bilaspur                HB
79526   ...          Gurgaon           Bilaspur                HB
79527   ...          Gurgaon           Bilaspur                HB
123196  ...              NaN                NaN               NaN

        source_name_state destination_name_city  destination_name_place  \
32950                 NaN                   NaN                     NaN
32951                 NaN                   NaN                     NaN
32952                 NaN                   NaN                     NaN
32953                 NaN                   NaN                     NaN
32954                 NaN                   NaN                     NaN
...                   ...                   ...                     ...
79524             Haryana                   NaN                     NaN
79525             Haryana                   NaN                     NaN
79526             Haryana                   NaN                     NaN
79527             Haryana                   NaN                     NaN
123196                NaN                   NaN                     NaN

        destination_name_code  destination_name_state  \
32950                     NaN                     NaN
32951                     NaN                     NaN
32952                     NaN                     NaN
32953                     NaN                     NaN
32954                     NaN                     NaN
...                       ...                     ...
79524                     NaN                     NaN
79525                     NaN                     NaN
79526                     NaN                     NaN
79527                     NaN                     NaN
123196                    NaN                     NaN

        time_taken_between_od_start_and_od_end_time  \
32950                                        64.959
32951                                        64.959
32952                                        64.959
32953                                        64.959
32954                                        64.959
...                                             ...
79524                                        70.658
79525                                        70.658
79526                                        70.658
79527                                        70.658
123196                                      131.643

        start_scan_to_end_scan_in_hrs
32950                          64.950
```

```
32951                          64.950
32952                          64.950
32953                          64.950
32954                          64.950
...                               ...
79524                          70.650
79525                          70.650
79526                          70.650
79527                          70.650
123196                        131.633

[373 rows x 46 columns]
```

In [257…

```
continious_features
```

Out[257]:

```
Index(['start_scan_to_end_scan', 'cutoff_factor',
       'actual_distance_to_destination', 'actual_time', 'osrm_time',
       'osrm_distance', 'factor', 'segment_actual_time', 'segment_osrm_time',
       'segment_osrm_distance', 'segment_factor', 'trip_creation_time_year',
       'trip_creation_time_weekday', 'trip_creation_time_day',
       'od_start_time_year', 'od_start_time_weekday', 'od_start_time_day',
       'od_end_time_year', 'od_end_time_weekday', 'od_end_time_day',
       'time_taken_between_od_start_and_od_end_time',
       'start_scan_to_end_scan_in_hrs'],
      dtype='object')
```

In [258…

```
find_outliers_IQR("actual_time")
```

```
Outliers by feature name -->  actual_time
```

Out[258]:

| | data | trip_creation_time | route_schedule_uuid | route_type | trip_uuid | source_center | |
|---|---|---|---|---|---|---|---|
| **406** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **407** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **408** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **409** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **410** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **...** | ... | ... | ... | ... | ... | ... | |
| **144796** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144797** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144798** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144799** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144800** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |

16633 rows × 46 columns

In [259…

```
find_outliers_IQR("segment_actual_time")
```

```
Outliers by feature name -->  segment_actual_time
```

Out[259]:

| | data | trip_creation_time | route_schedule_uuid | route_type | trip_uuid | source_center |
|---|---|---|---|---|---|---|
| **21** | training | 2018-09-13 20:44:19.424489 | thanos::sroute:76951383-1608-44e4-a284-46d92e8... | FTL | trip-153687145942424248 | IND560099AAB |
| **34** | training | 2018-09-13 20:44:19.424489 | thanos::sroute:76951383-1608-44e4-a284-46d92e8... | FTL | trip-153687145942424248 | IND560099AAB |
| **72** | test | 2018-10-01 16:00:45.719099 | thanos::sroute:182fdad4-dc74-4c7d-ada4-0a76d4c... | FTL | trip-153840964571880594 | IND131028AAB |
| **73** | test | 2018-10-01 16:00:45.719099 | thanos::sroute:182fdad4-dc74-4c7d-ada4-0a76d4c... | FTL | trip-153840964571880594 | IND247667AAB |
| **106** | training | 2018-09-25 08:53:04.377810 | thanos::sroute:4460a38d-ab9b-484e-bd4e-f4201d0... | FTL | trip-153786558437756691 | IND306401AAB |
| **...** | ... | ... | ... | ... | ... | ... |
| **144790** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB |
| **144819** | training | 2018-09-26 14:05:52.096792 | thanos::sroute:f7de4133-6bd9-4367-a7f7-ab190b6... | FTL | trip-153797075209653066 | IND413401AAA |
| **144848** | training | 2018-09-14 18:45:34.164734 | thanos::sroute:40b6dc9c-faa1-4753-8bc8-ac8c3e0... | Carting | trip-153695073416451616 | IND400102AAB |
| **144853** | training | 2018-09-22 11:30:41.399439 | thanos::sroute:d81088e2-9ccd-43e9-9260-3e85633... | FTL | trip-153761584139918815 | IND421302AAG |
| **144866** | training | 2018-09-20 16:24:28.436231 | thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5... | Carting | trip-153746066843555182 | IND131028AAB |

9298 rows × 46 columns

In [260…

```
find_outliers_IQR("actual_time")
```

```
Outliers by feature name -->  actual_time
```

Out[260]:

| | data | trip_creation_time | route_schedule_uuid | route_type | trip_uuid | source_center | |
|---|---|---|---|---|---|---|---|
| **406** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **407** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **408** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **409** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **410** | training | 2018-09-25 15:06:59.975279 | thanos::sroute:51d8aa58-9b5b-4bc2-81e9-bb284c6... | FTL | trip-153788801997503817 | IND825409AAA | Jhun |
| **...** | ... | ... | ... | ... | ... | ... | |
| **144796** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144797** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144798** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144799** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |
| **144800** | test | 2018-10-01 18:17:37.047270 | thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f... | FTL | trip-153841785704702048 | IND000000ACB | G |

16633 rows × 46 columns

- **Conclusion**
  - **Outliers impacts mean based statistical methods**
  - **Observation during outlier removal**
  - **There are outliers in almost every features**
  - **Removing** those **outliers iteratively , can result loss of significant feature**
  - **Need domain expert consultation** before removing indepndent feature based outliers
  - Hence **proceeding with Baseline analysis with outliers**
  - Moreover, certain deep learning model can work without outliers , hence skipping outlier removal

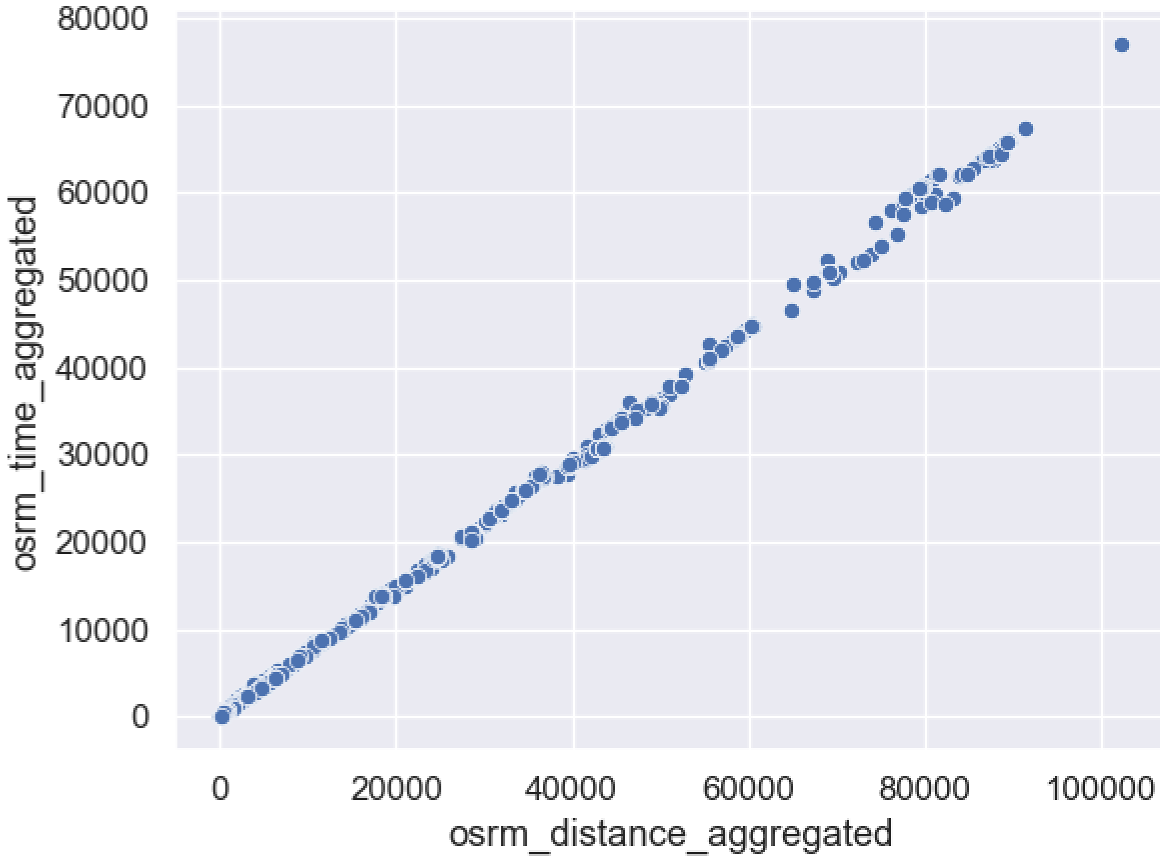
# Checking relationship between aggregated fields

In [261…

```
sns.scatterplot(x="osrm_distance_aggregated", y="osrm_time_aggregated", data=df_agg)
```

Out[261]:

```
<AxesSubplot:xlabel='osrm_distance_aggregated', ylabel='osrm_time_aggregated'>
```

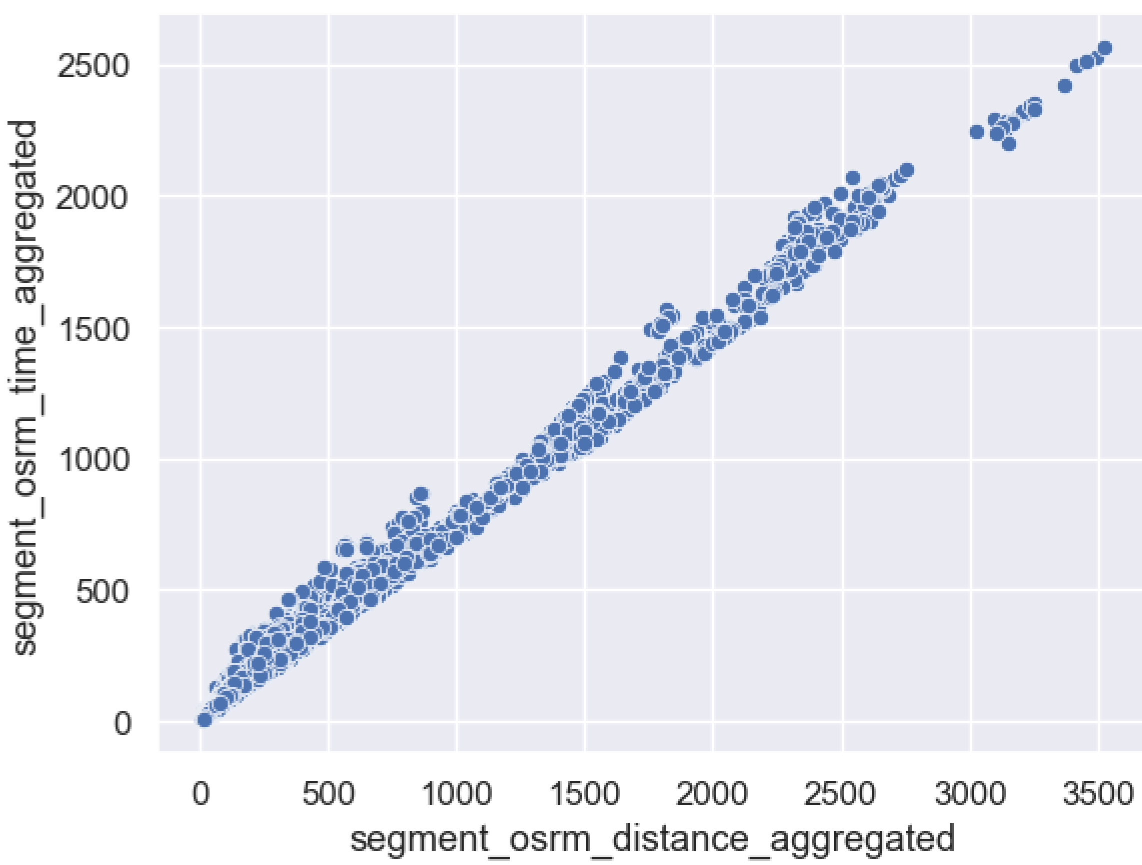



In [262…

```
sns.scatterplot(x="segment_osrm_distance_aggregated", y="segment_osrm_time_aggregated",
```

Out[262]:

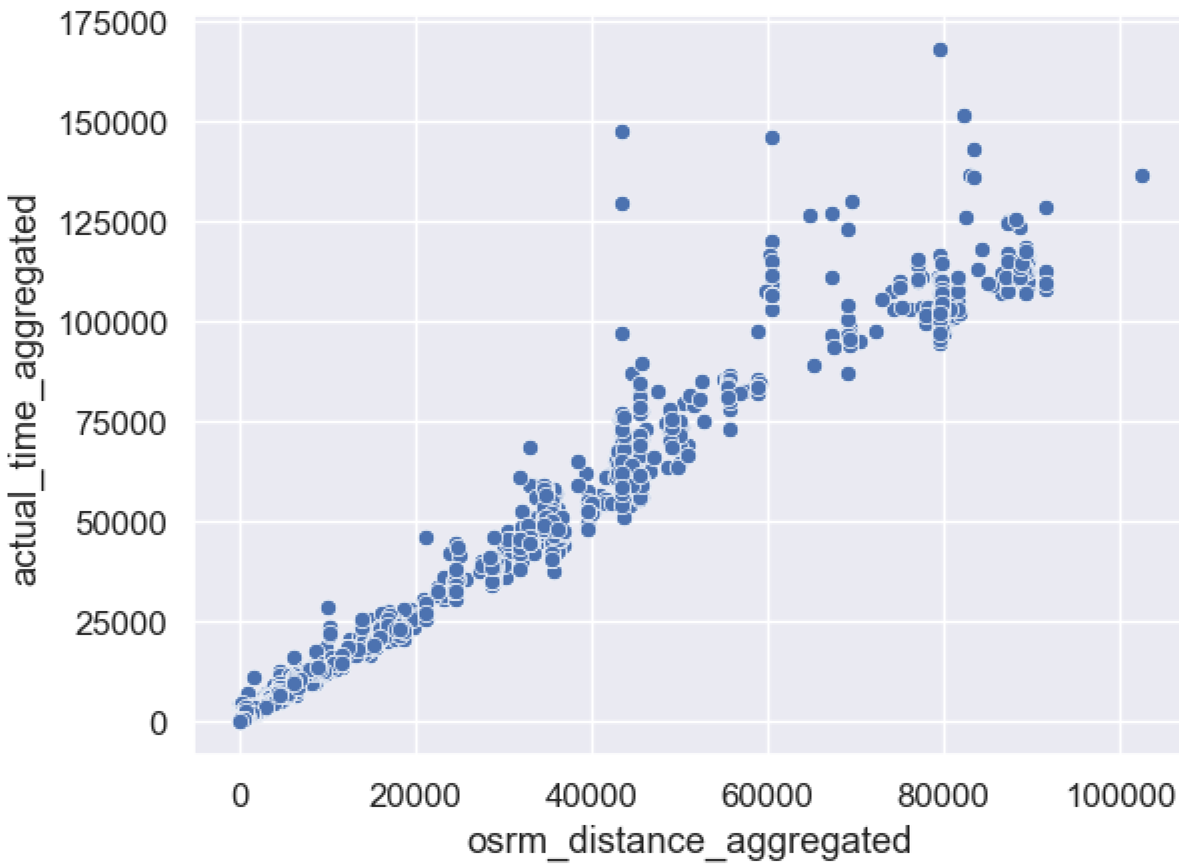
```
<AxesSubplot:xlabel='segment_osrm_distance_aggregated', ylabel='segment_osrm_time_aggregated'>
```




In [263…

```
sns.scatterplot(x="osrm_distance_aggregated", y="actual_time_aggregated", data=df_agg)
```

Out[263]: <AxesSubplot:xlabel='osrm_distance_aggregated', ylabel='actual_time_aggregated'>




- **Conclusion**
  - **Approximate Linear relationship observed** between following **aggregated features**
    - "osrm_distance_aggregated" vs "osrm_time_aggregated"
    - "segment_osrm_distance_aggregated" vs "segment_osrm_time_aggregated"
    - "osrm_distance_aggregated" vs "actual_time_aggregated"

## Handling categorical values

- Categorical to Numerical encoding

In [264…

```
df["route_type"].value_counts()
```

Out[264]:

```
FTL         99660
Carting     45207
Name: route_type, dtype: int64
```

In [265…

```
from sklearn.preprocessing import LabelEncoder
# Running level encoding for convertion from categorical to numerical data for feeding i
label_encoder = LabelEncoder()
col='route_type'
df[col] = label_encoder.fit_transform(df[col])
```

In [266…

```
df["route_type"].value_counts()
```

Out[266]:

```
1    99660
0    45207
Name: route_type, dtype: int64
```

## Column Normalization /Column Standardization

- Normalize/ Standardize the numerical
  - MinMaxScaler
  - StandardScaler.

In [267…

```
continious_features
```

Out[267]:

```
Index(['start_scan_to_end_scan', 'cutoff_factor',
       'actual_distance_to_destination', 'actual_time', 'osrm_time',
       'osrm_distance', 'factor', 'segment_actual_time', 'segment_osrm_time',
       'segment_osrm_distance', 'segment_factor', 'trip_creation_time_year',
       'trip_creation_time_weekday', 'trip_creation_time_day',
       'od_start_time_year', 'od_start_time_weekday', 'od_start_time_day',
       'od_end_time_year', 'od_end_time_weekday', 'od_end_time_day',
       'time_taken_between_od_start_and_od_end_time',
       'start_scan_to_end_scan_in_hrs'],
      dtype='object')
```

In [268…

```
from sklearn.preprocessing import StandardScaler, MinMaxScaler

scaler = StandardScaler()
std_data = scaler.fit_transform(df[continious_features])
std_data = pd.DataFrame(std_data, columns=continious_features)
std_data.head()
```

Out[268]:

| | start_scan_to_end_scan | cutoff_factor | actual_distance_to_destination | actual_time | osrm_time | osrm_distance | |
|---|---|---|---|---|---|---|---|
| **0** | -0.844026 | -0.649525 | -0.648246 | -0.673677 | -0.658642 | -0.647814 | -0.4 |
| **1** | -0.844026 | -0.623419 | -0.623604 | -0.656958 | -0.629422 | -0.624640 | -0.5 |
| **2** | -0.844026 | -0.597314 | -0.598385 | -0.630207 | -0.603449 | -0.598958 | -0.4 |
| **3** | -0.844026 | -0.571208 | -0.573802 | -0.593424 | -0.564489 | -0.568034 | -0.3 |
| **4** | -0.844026 | -0.562506 | -0.564329 | -0.583392 | -0.551502 | -0.547479 | -0.3 |

5 rows × 22 columns

In [269…

```
scaler = MinMaxScaler()
max_scaler_data = scaler.fit_transform(df[continious_features])
max_scaler_data = pd.DataFrame(max_scaler_data, columns=continious_features)
max_scaler_data.head()
```

Out[269]:

| | start_scan_to_end_scan | cutoff_factor | actual_distance_to_destination | actual_time | osrm_time | osrm_distance | f |
|---|---|---|---|---|---|---|---|
| **0** | 0.008378 | 0.000000 | 0.000748 | 0.001105 | 0.002976 | 0.001276 | 0.0 |
| **1** | 0.008378 | 0.004692 | 0.005180 | 0.003316 | 0.008333 | 0.005488 | 0.0 |
| **2** | 0.008378 | 0.009385 | 0.009715 | 0.006854 | 0.013095 | 0.010155 | 0.0 |
| **3** | 0.008378 | 0.014077 | 0.014135 | 0.011718 | 0.020238 | 0.015775 | 0.0 |
| **4** | 0.008378 | 0.015641 | 0.015839 | 0.013044 | 0.022619 | 0.019511 | 0.0 |

5 rows × 22 columns

## Business Insights - Should include patterns observed in the data along

with what you can infer from it.

- Relatively **less full truck loads** trips **starting on 5th or 6th day of the month**
- Relatively **large full truck loads** trips **ending on 5th or 6th day of the month**
- Relatively **less full truck loads** trips **ending on 7th or 8th day of the month**
- **Most Time taken route** (i.e. for route type full truck load) to deliver from source to destination are
    - **Haryana**
    - **Karnataka**
    - **Maharshtra**
    - **Delhi**
    - **Telangana**
- Actual **time taken to complete the delivery** are **high for source/destination
    - **Haryana**
    - **Karnataka**
    - **Maharshtra**
    - **Delhi**
    - **Telangana**
- Most **distanced cities** for full truck deliveries are **Gurgaon , Bengaluru , Bhiwandi**
- **Most trips** are being **booked on 12,13,15,17,18,20,21,22,25,26 days of the month**
- **osrm_time** (i.e. An open-source routing engine calculated time) is **not close to actual time** of delivery
    - **Estimation of delivery time needs more accurate prediction**
    - **Once delivery time is accurately predicted** then **more pre-booking** can be taken up
- Approximate **Linear relationship observed** between **following distance and time features**
    - "actual_distance_to_destination" vs "osrm_time"
    - "osrm_distance" vs "osrm_time"
    - "start_scan_to_end_scan" vs "actual_time"
    - "osrm_distance" vs "osrm_time"

## Check from where most orders are coming from (State, Corridor etc)

In [270…

```
df["source_name_city"].value_counts()
```

Out[270]:

```
Gurgaon          23639
Bengaluru        14206
Bhiwandi          9088
Delhi             4318
Pune              4160
                 ...
Kayamkulam           1
Chikhli              1
Kothanalloor         1
Thirthahalli         1
Soro                 1
Name: source_name_city, Length: 560, dtype: int64
```

In [271…

```
selected_features = ['route_type', 'source_center', 'source_name', 'destination_center',
df_selected = df[selected_features]
```

In [272…

```
# Show percentage values on the top of the bar
def show_values_on_bars(axs, h_v="v", space=1):
    total = float(len(df_selected))
    def _show_on_single_plot(ax):
```

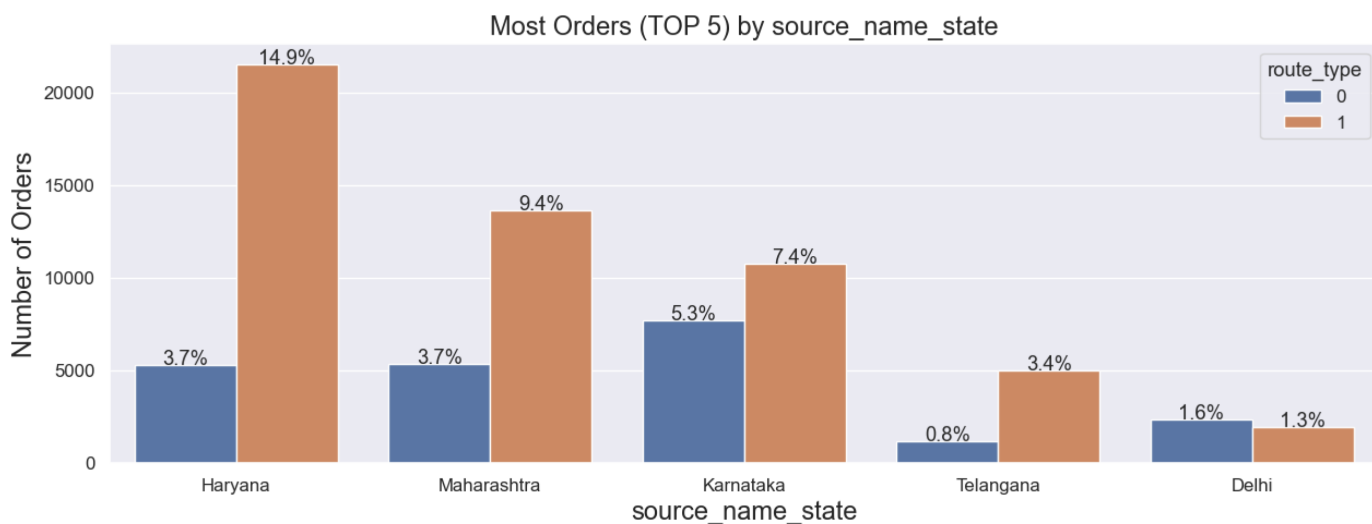
```
        if h_v == "v":
            for p in ax.patches:
                _x = p.get_x() + p.get_width() / 2
                _y = p.get_y() + p.get_height()
                value='{:.1f}%'.format(100 * p.get_height()/total)
                ax.text(_x, _y, value, ha="center")
        elif h_v == "h":
            for p in ax.patches:
                _x = p.get_x() + p.get_width() + float(space)
                _y = p.get_y() + p.get_height()
                value='{:.1f}%'.format(100 * p.get_width()/total)
                ax.text(_x, _y, value, ha="left")
    if isinstance(axs, np.ndarray):
         for idx, ax in np.ndenumerate(axs):
            _show_on_single_plot(ax)
    else:
            _show_on_single_plot(axs)
```

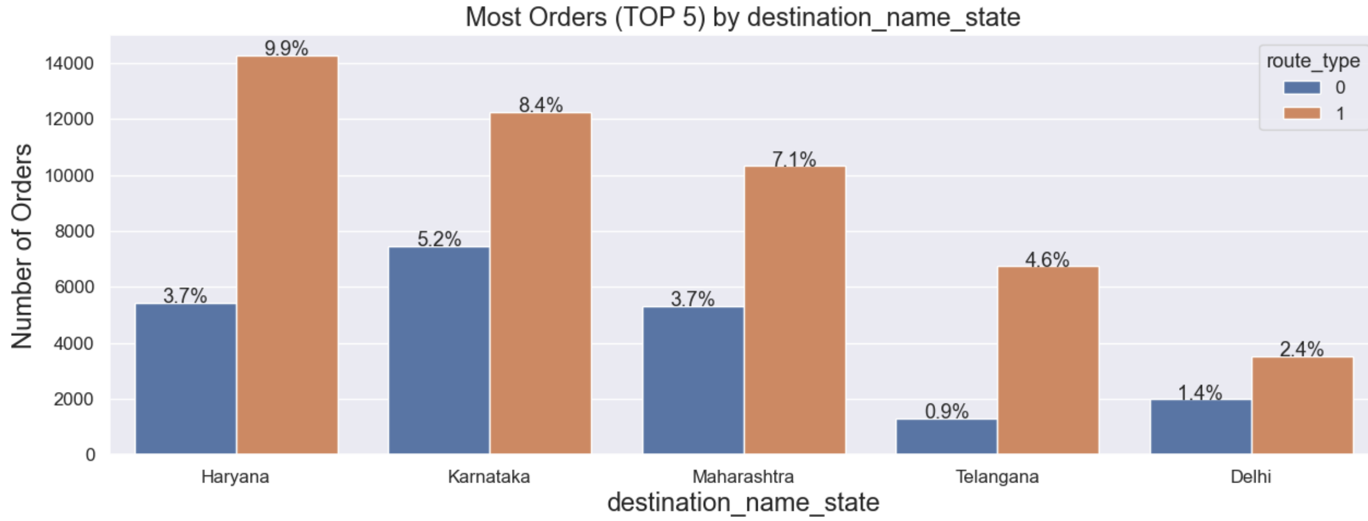
In [273…

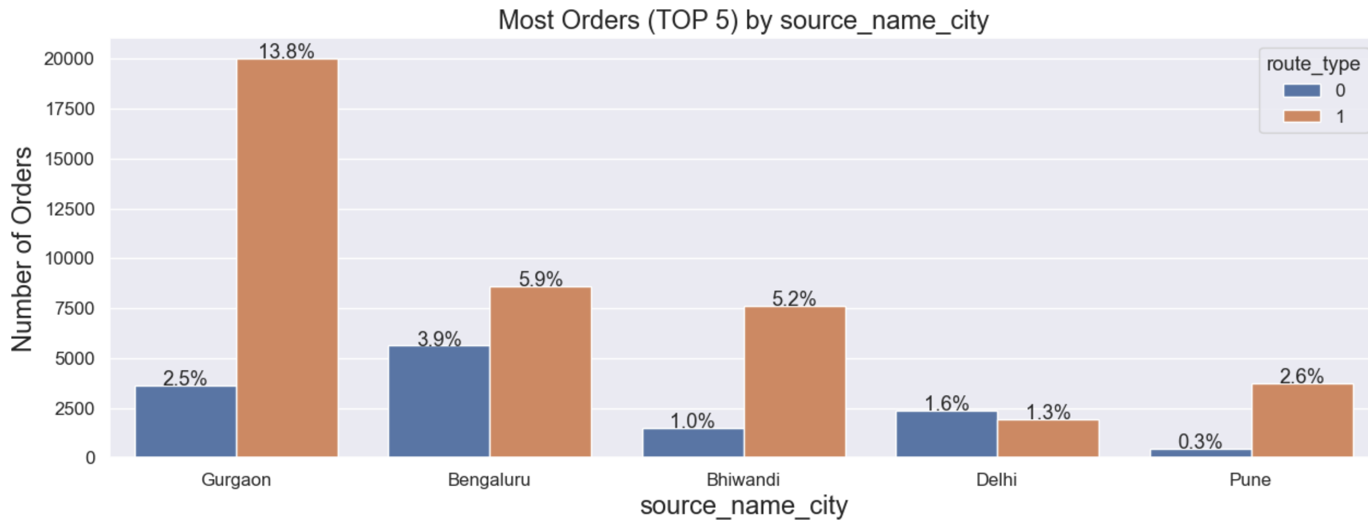
```
# plot top orders by feature name using count plot with percentage displayed on the top
# Top 5 orders are being plotted
def plot_orders_by_feature(feature_list):
    for feature_name in feature_list:
        sns.set(style="whitegrid")
        sns.set(font_scale = 1.1)
        plt.figure(figsize=(15,5))
        ax = sns.countplot(x=feature_name,data=df_selected,hue="route_type",order=df_sel
        plt.xlabel(feature_name, fontsize=17)
        plt.ylabel("Number of Orders",fontsize=17)
        plt.title("Most Orders (TOP 5) by "+feature_name,fontdict ={"fontsize": 17})
        show_values_on_bars(ax,h_v="v",space=1)
```

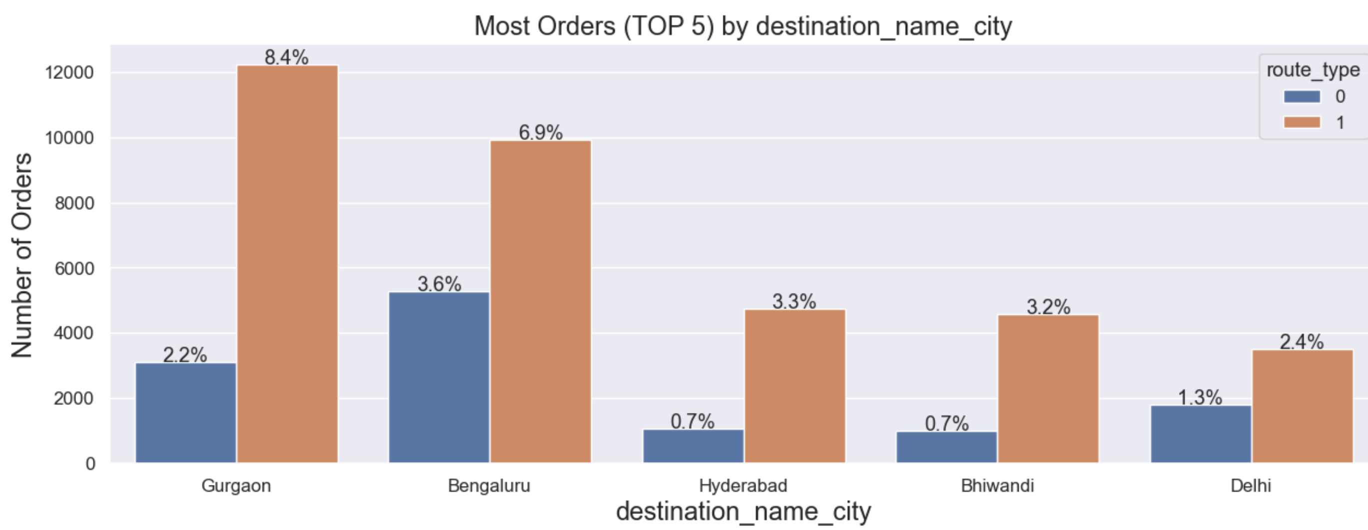
In [274…

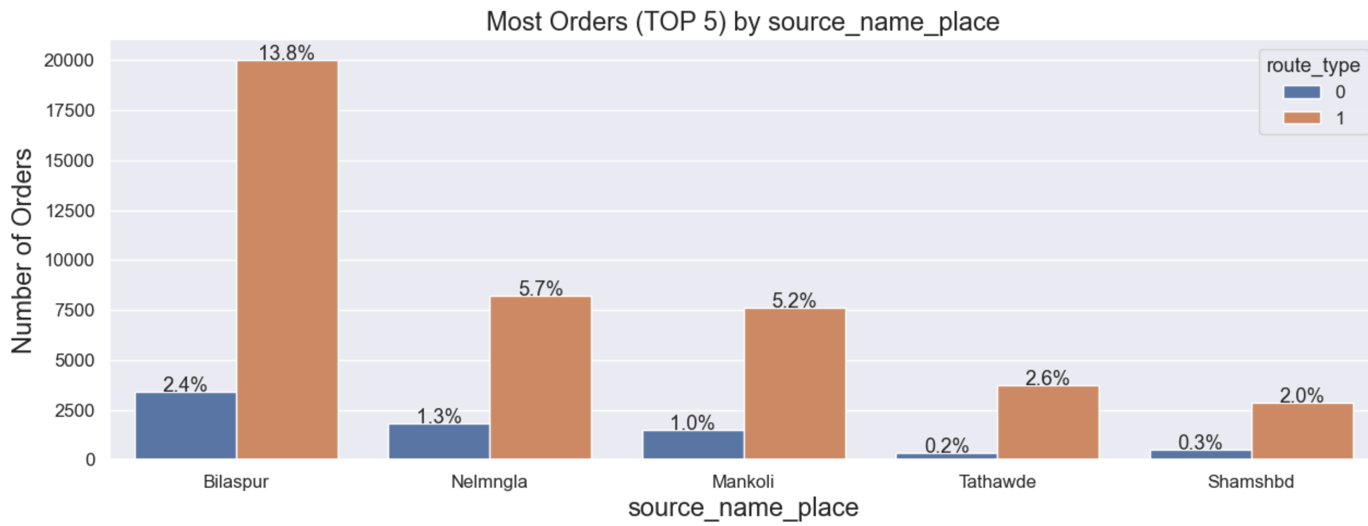
```
plot_orders_by_feature(['source_name_state','destination_name_state','source_name_city',
```

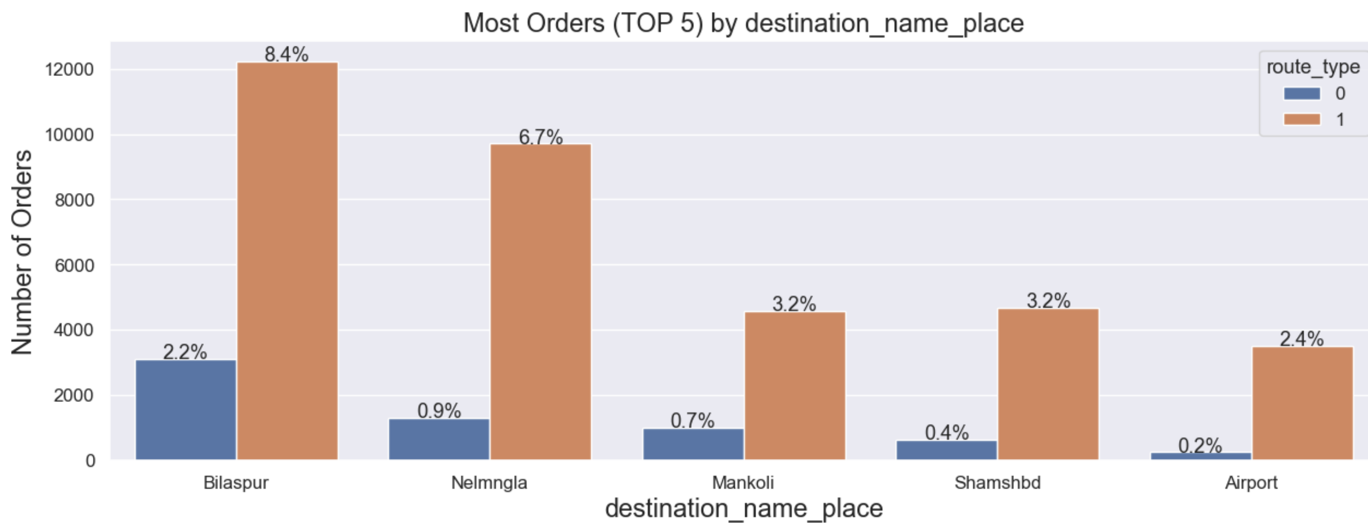



Most Orders (TOP 5) by destination_name_state
Number of Orders
destination_name_state
route_type
0
1
Haryana
Karnataka
Maharashtra
Telangana
Delhi
9.9%
3.7%
8.4%
5.2%
7.1%
3.7%
4.6%
0.9%
2.4%
1.4%
0
2000
4000
6000
8000
10000
12000
14000


Most Orders (TOP 5) by source_name_city
Number of Orders
source_name_city
route_type
0
1
Gurgaon
Bengaluru
Bhiwandi
Delhi
Pune
13.8%
2.5%
5.9%
3.9%
5.2%
1.0%
1.6%
1.3%
2.6%
0.3%
0
2500
5000
7500
10000
12500
15000
17500
20000


Most Orders (TOP 5) by destination_name_city
Number of Orders
destination_name_city
route_type
0
1
Gurgaon
Bengaluru
Hyderabad
Bhiwandi
Delhi
8.4%
2.2%
6.9%
3.6%
3.3%
0.7%
3.2%
0.7%
2.4%
1.3%
0
2000
4000
6000
8000
10000
12000




- **Insights** on Busiest State/Corridor
  - **Busiest source states** are:
    - **Haryana**(14.9% Full truck load , 3.7% Carting)
    - **Maharashtra** (9.4% Full truck load , 3.7% Carting)
    - **Karnataka** (7.4% Full truck load , 5.3% Carting)
  - **Busiest Destination states** are:
    - **Haryana**(9.9% Full truck load , 3.7% Carting)
    - **Karnataka** (8.4% Full truck load , 5.2% Carting)
    - **Maharashtra** (7.1% Full truck load , 3.7% Carting)
  - **Busiest source city** are:
    - **Gurgaon**(13.8% Full truck load , 2.5% Carting)
    - **Bengaluru** (5.9% Full truck load , 3.9% Carting)
    - **Bhiwandi** (5.2% Full truck load , 1% Carting)
  - **Busiest Destination city** are:
    - **Gurgaon**(8.4% Full truck load , 2.2% Carting)
    - **Bengaluru** (6.9% Full truck load , 3.6% Carting)
    - **Hyderabad** (3.3% Full truck load , 0.7% Carting)
  - **Busiest source Place** are:
    - **Bilaspur**(13.8% Full truck load , 2.4% Carting)
    - **Nelmngla** (5.7% Full truck load , 1.3% Carting)
    - **Mankoli** (5.2% Full truck load , 1% Carting)
  - **Busiest Destination Place** are:

- **Bilaspur**(8.4% Full truck load , 2.2% Carting)
- **Nelmngla** (6.7% Full truck load , 0.9% Carting)
- **Mankoli** (3.2% Full truck load , 0.7% Carting)

## Busiest corridor | avg distance , avg time taken

In [275…

```
stats.ttest_ind(df["start_scan_to_end_scan"],df["actual_time"])
```

Out[275]:

```
Ttest_indResult(statistic=173.06512569398697, pvalue=0.0)
```

In [276…

```
df_busiest_corridor = df[(df["destination_name_city"] == 'Gurgaon') | (df["destination_r
df_busiest_corridor = df_busiest_corridor[["actual_distance_to_destination","actual_time
```

### Calculate "avg time" of destination busiest corridor - Gurgaon , Bangaluru , Hyderabad

In [277…

```
avg_time_taken = df_busiest_corridor.groupby(['destination_name_city'])['actual_time'].a
avg_time_taken.reset_index().sort_values(['mean','destination_name_city'],ascending=Fals
```

Out[277]:

| | destination_name_city | mean |
|---|---|---|
| **2** | Gurgaon | 739.855776 |
| **0** | Bengaluru | 677.203274 |
| **1** | Bhiwandi | 667.344970 |
| **3** | Hyderabad | 512.264766 |

### Calculate "avg distance" of destination busiest corridor - Gurgaon , Bangaluru , Hyderabad

In [278…

```
avg_distance_taken = df_busiest_corridor.groupby(['destination_name_city'])['actual_dist
avg_distance_taken.reset_index().sort_values(['mean','destination_name_city'],ascending=
```

Out[278]:

| | destination_name_city | mean |
|---|---|---|
| **2** | Gurgaon | 437.276411 |
| **0** | Bengaluru | 412.491168 |
| **1** | Bhiwandi | 367.409988 |
| **3** | Hyderabad | 295.929211 |

## Recommendations - Actionable items for business. No technical jargon. No complications. Simple action items that everyone can understand.

- **Key considerations**:
  - Below recommendation will be more effective when more appropriate measures taken care wrt. outliers and more granular feature engineering are taken care as well
- **Actionable items for business**
  - **State/Place/City based marketing** would be **more effective**:
    - More **discounts on reverse routes** to **raise volume of orders**
    - **Long term Reserved booking discounts** between **top source and destinations**

- **Days based promotions when** there are **less truck loads bookings**
- **Volume based marketing** on **popular source and destinations** would be more effective
- **Delivery time estimation to be predicted accurately**
  - **More pre-booking order** can be taken up **based on estimated delivery time**
  - **More promotions on pre-booking and discounting offers**
  - **Orders on 5th, 6th and 7th day of the month** should **maximized based on delivery time estimates** as their minimal operations on those days
- **Route based promotions** can be more **effective**
  - **Full truck load promotions in states such as Haryana, Maharashtra, Karnataka, Delhi, Telangana**
  - **Carting promotions in states such as "Uttarakhand”, "Rajasthan", "Jharkhand"**

In [ ]: