

Delhivery

```
In [76]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

In [77]: df=pd.read_csv("https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/551/original/delhivery_data.csv?1642751181")

In [78]: df

Out[78]:
```

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center	destination_name	od_start_time	...	cutoff_timestamp
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	...	2018-09-20 04:27:55
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	...	2018-09-20 04:17:55
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	...	2018-09-20 04:01:19.505586
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	...	2018-09-20 03:39:57
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	...	2018-09-20 03:33:55
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	...	2018-09-20 21:57:20
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	...	2018-09-20 21:31:18
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	...	2018-09-20 21:11:18
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	...	2018-09-20 20:53:19
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	...	2018-09-20 16:24:28.436231

144867 rows × 24 columns

dropping unwanted columns

```
In [79]: df.drop( columns=['is_cutoff','cutoff_factor','cutoff_timestamp','segment_factor'],inplace=True)

In [80]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144867 entries, 0 to 144866
Data columns (total 20 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   actual_distance_to_destination      144867 non-null  float64
13   actual_time                        144867 non-null  float64
14   osrm_time                          144867 non-null  float64
15   osrm_distance                      144867 non-null  float64
16   factor                             144867 non-null  float64
17   segment_actual_time                144867 non-null  float64
18   segment_osrm_time                  144867 non-null  float64
19   segment_osrm_distance              144867 non-null  float64
dtypes: float64(9), object(11)
memory usage: 22.1+ MB

In [81]: df.shape

Out[81]: (144867, 20)

In [82]: df.describe(include="object")

Out[82]:
```

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center	destination_name	od_start_time	od_end_time
count	144867	144867	144867	144867	144867	144867	144574	144867	144606	144867	144867
unique	2	14817	1504	2	14817	1508	1498	1481	1468	26369	26369
top	training	2018-09-28 05:23:15.359220	thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...	FTL	trip-153811219535896559	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-21 18:37:09.322207	2018-09-24 09:59:15.691618
freq	104858	101	1812	99660	101	23347	23347	15192	15192	81	81

```
In [83]: df.describe()
```

Out[83]:

	start_scan_to_end_scan	actual_distance_to_destination	actual_time	osrm_time	osrm_distance	factor	segment_actual_time	segment_osrm_time	segment_osrm_distance
count	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.00000
mean	961.262986	234.073372	416.927527	213.868272	284.771297	2.120107	36.196111	18.507548	22.82902
std	1037.012769	344.990009	598.103621	308.011085	421.119294	1.715421	53.571158	14.775960	17.86066
min	20.000000	9.000045	9.000000	6.000000	9.008200	0.144000	-244.000000	0.000000	0.00000
25%	161.000000	23.355874	51.000000	27.000000	29.914700	1.604264	20.000000	11.000000	12.07010
50%	449.000000	66.126571	132.000000	64.000000	78.525800	1.857143	29.000000	17.000000	23.51300
75%	1634.000000	286.708875	513.000000	257.000000	343.193250	2.213483	40.000000	22.000000	27.81325
max	7898.000000	1927.447705	4532.000000	1686.000000	2326.199100	77.387097	3051.000000	1611.000000	2191.40370

In [84]:

```
df.describe(include="float64")
```

Out[84]:

	start_scan_to_end_scan	actual_distance_to_destination	actual_time	osrm_time	osrm_distance	factor	segment_actual_time	segment_osrm_time	segment_osrm_distance
count	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.00000
mean	961.262986	234.073372	416.927527	213.868272	284.771297	2.120107	36.196111	18.507548	22.82902
std	1037.012769	344.990009	598.103621	308.011085	421.119294	1.715421	53.571158	14.775960	17.86066
min	20.000000	9.000045	9.000000	6.000000	9.008200	0.144000	-244.000000	0.000000	0.00000
25%	161.000000	23.355874	51.000000	27.000000	29.914700	1.604264	20.000000	11.000000	12.07010
50%	449.000000	66.126571	132.000000	64.000000	78.525800	1.857143	29.000000	17.000000	23.51300
75%	1634.000000	286.708875	513.000000	257.000000	343.193250	2.213483	40.000000	22.000000	27.81325
max	7898.000000	1927.447705	4532.000000	1686.000000	2326.199100	77.387097	3051.000000	1611.000000	2191.40370

In [85]:

```
df.isnull().sum()
```

Out[85]:

	0
data	0
trip_creation_time	0
route_schedule_uuid	0
route_type	0
trip_uuid	0
source_center	0
source_name	293
destination_center	0
destination_name	261
od_start_time	0
od_end_time	0
start_scan_to_end_scan	0
actual_distance_to_destination	0
actual_time	0
osrm_time	0
osrm_distance	0
factor	0
segment_actual_time	0
segment_osrm_time	0
segment_osrm_distance	0

dtype: int64

In [86]:

```
df.fillna(value={'destination_name':'mode'},inplace=True)
df.fillna(value={'source_name':'mode'},inplace=True)
df
```

Out[86]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center	destination_name	od_start_time	od_end_time	start
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	

144867 rows × 20 columns



In [87]:

```
category_columns=df.dtypes=="object"  
numerical_columns=df.dtypes!="object"
```

In [88]:

```
for i in df.columns[category_columns]:  
    print(i,":",df[i].nunique())  
    print("-----")
```

```
data : 2  
*-----*  
trip_creation_time : 14817  
*-----*  
route_schedule_uuid : 1504  
*-----*  
route_type : 2  
*-----*  
trip_uuid : 14817  
*-----*  
source_center : 1508  
*-----*  
source_name : 1499  
*-----*  
destination_center : 1481  
*-----*  
destination_name : 1469  
*-----*  
od_start_time : 26369  
*-----*  
od_end_time : 26369  
*-----*
```

In [89]:

```
# unique count of numerical columns  
for i in df.columns[numerical_columns]:  
    print(i,":",df[i].nunique())  
    print("-----")
```

```
start_scan_to_end_scan : 1915  
*-----*  
actual_distance_to_destination : 144515  
*-----*  
actual_time : 3182  
*-----*  
osrm_time : 1531  
*-----*  
osrm_distance : 138046  
*-----*  
factor : 45641  
*-----*  
segment_actual_time : 747  
*-----*  
segment_osrm_time : 214  
*-----*  
segment_osrm_distance : 113799  
*-----*
```

In [90]:

```
df
```

Out[90]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center	destination_name	od_start_time	od_end_time	start
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	

144867 rows × 20 columns



In [91]:

category_columns[category_columns]

Out[91]:

0
dataTrue
trip_creation_timeTrue
route_schedule_uuidTrue
route_typeTrue
trip_uuidTrue
source_centerTrue
source_nameTrue
destination_centerTrue
destination_nameTrue
od_start_timeTrue
od_end_timeTrue

dtype: bool

In [92]:

numerical_columns[numerical_columns]

Out[92]:

0
start_scan_to_end_scanTrue
actual_distance_to_destinationTrue
actual_timeTrue
osrm_timeTrue
osrm_distanceTrue
factorTrue
segment_actual_timeTrue
segment_osrm_timeTrue
segment_osrm_distanceTrue

dtype: bool

Convert object type To DateTime

In [93]:

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'])

In [94]:

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144867 entries, 0 to 144866
Data columns (total 20 columns):
#   Column                                Non-Null Count  Dtype
---  -
0    data                                144867 non-null  object
1    trip_creation_time                  144867 non-null  datetime64[ns]
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                        144867 non-null  object
7    destination_center                 144867 non-null  object
8    destination_name                   144867 non-null  object
9    od_start_time                      144867 non-null  datetime64[ns]
10   od_end_time                        144867 non-null  datetime64[ns]
11   start_scan_to_end_scan              144867 non-null  float64
12   actual_distance_to_destination      144867 non-null  float64
13   actual_time                         144867 non-null  float64
14   osrm_time                          144867 non-null  float64
15   osrm_distance                      144867 non-null  float64
16   factor                             144867 non-null  float64
17   segment_actual_time                 144867 non-null  float64
18   segment_osrm_time                  144867 non-null  float64
19   segment_osrm_distance               144867 non-null  float64
dtypes: datetime64[ns](3), float64(9), object(8)
memory usage: 22.1+ MB
```

no of uniques in each column

```
In [95]: for i in df.columns:
         print(i,":",df[i].nunique())
         print("-----")
```

```
data : 2
*-----*
trip_creation_time : 14817
*-----*
route_schedule_uuid : 1504
*-----*
route_type : 2
*-----*
trip_uuid : 14817
*-----*
source_center : 1508
*-----*
source_name : 1499
*-----*
destination_center : 1481
*-----*
destination_name : 1469
*-----*
od_start_time : 26369
*-----*
od_end_time : 26369
*-----*
start_scan_to_end_scan : 1915
*-----*
actual_distance_to_destination : 144515
*-----*
actual_time : 3182
*-----*
osrm_time : 1531
*-----*
osrm_distance : 138046
*-----*
factor : 45641
*-----*
segment_actual_time : 747
*-----*
segment_osrm_time : 214
*-----*
segment_osrm_distance : 113799
*-----*
```

```
In [96]: df_merge=df.groupby(['trip_uuid','source_center','destination_center']).agg({'data':'first',
                                             'trip_creation_time':'first',
                                             'route_schedule_uuid':'first',
                                             'route_type':'first',
                                             'source_name':'first',
                                             'destination_name':'last',
                                             'od_start_time':'first',
                                             'od_end_time':'last',
                                             'start_scan_to_end_scan':'max',
                                             'actual_distance_to_destination':'max',
                                             'actual_time':'max',
                                             'osrm_time':'max',
                                             'osrm_distance':'max',
                                             'segment_actual_time':'sum',
                                             'segment_osrm_time':'sum',
                                             'segment_osrm_distance':'sum'}).reset_index()

df_merge
```


Out[96]:

	trip_uuid	source_center	destination_center	data	trip_creation_time	route_schedule_uuid	route_type	source_name	destination_name	od_start_time	od_end_time
0	trip-153671041653548748	IND209304AAA	IND000000ACB	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	Kanpur_Central_H_6 (Uttar Pradesh)	Gurgaon_Bilaspur_HB (Haryana)	2018-09-12 16:39:46.858469	2018-09-13 13:40:23.123744
1	trip-153671041653548748	IND462022AAA	IND209304AAA	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	Bhopal_Trnsport_H (Madhya Pradesh)	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12 00:00:16.535741	2018-09-12 16:39:46.858469
2	trip-153671042288605164	IND561203AAB	IND562101AAA	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	Doddablpur_ChikaDPP_D (Karnataka)	Chikblapur_ShntiSgr_D (Karnataka)	2018-09-12 02:03:09.655591	2018-09-12 03:01:59.598855
3	trip-153671042288605164	IND572101AAA	IND561203AAB	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	Tumkur_Veersagr_I (Karnataka)	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12 00:00:22.886430	2018-09-12 02:03:09.655591
4	trip-153671043369099517	IND000000ACB	IND160002AAC	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	Gurgaon_Bilaspur_HB (Haryana)	Chandigarh_Mehmdpur_H (Punjab)	2018-09-14 03:40:17.106733	2018-09-14 17:34:55.442454
...
26363	trip-153861115439069069	IND628204AAA	IND627657AAA	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	Tirchchnr_Shnmgprm_D (Tamil Nadu)	Thisayanvilai_UdnkdiRD_D (Tamil Nadu)	2018-10-04 02:29:04.272194	2018-10-04 03:31:11.183797
26364	trip-153861115439069069	IND628613AAA	IND627005AAA	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	Peikulam_SriVnktpm_D (Tamil Nadu)	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	2018-10-04 04:16:39.894872	2018-10-04 05:47:45.162682
26365	trip-153861115439069069	IND628801AAA	IND628204AAA	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	Eral_Busstand_D (Tamil Nadu)	Tirchchnr_Shnmgprm_D (Tamil Nadu)	2018-10-04 01:44:53.808000	2018-10-04 02:29:04.272194
26366	trip-153861118270144424	IND583119AAA	IND583101AAA	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	Sandur_WrdN1DPP_D (Karnataka)	Bellary_Dc (Karnataka)	2018-10-04 03:58:40.726547	2018-10-04 08:46:09.166940
26367	trip-153861118270144424	IND583201AAA	IND583119AAA	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	Hospet (Karnataka)	Sandur_WrdN1DPP_D (Karnataka)	2018-10-04 02:51:44.712656	2018-10-04 03:58:40.726547

26368 rows × 19 columns

```
In [97]: df=df_merge.groupby('trip_uuid').agg({'data':'first',
                                         'trip_creation_time':'first',
                                         'route_schedule_uuid':'first',
                                         'route_type':'first','source_center':'first',
                                         'source_name':'first',
                                         'destination_center':'last',
                                         'destination_name':'last',
                                         'od_start_time':'first',
                                         'od_end_time':'last',
                                         'start_scan_to_end_scan':'sum',
                                         'actual_distance_to_destination':'sum',
                                         'actual_time':'sum',
                                         'osrm_time':'sum',
                                         'osrm_distance':'sum',
                                         'segment_actual_time':'sum',
                                         'segment_osrm_time':'sum',
                                         'segment_osrm_distance':'sum'}).reset_index()
```

feature engineering (creating features for further exploration)

```
In [98]: od_time_diff_hour = df['od_end_time'] - df['od_start_time']
df['od_time_diff_hour'] = od_time_diff_hour.dt.components['hours']

In [99]: df[['destination','dest_state']] = df['destination_name'].str.split('(', n=1, expand=True)
df['dest_state'] = df['dest_state'].str.rstrip(')')
df[['dest_City','dest_place','dest_code']] = df['destination'].str.split('_', n=2, expand=True)
df[['source','source_state']] = df['source_name'].str.split('(', n=1, expand=True)
df['source_state'] = df['source_state'].str.rstrip(')')
df[['source_City','source_place','source_code']] = df['source'].str.split('_', n=2, expand=True)

In [100... df['total_min_diff']=(df['od_end_time']-df['od_start_time'])/pd.Timedelta(minutes=1)
df
```

Out[100...

	trip_uuid	data	trip_creation_time	route_schedule_uuid	route_type	source_center	source_name	destination_center	destination_name	od_start_time	...	dest_state
0	trip-153671041653548748	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12 16:39:46.858469	...	Uttar Pradesh
1	trip-153671042288605164	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12 02:03:09.655591	...	Karnataka
2	trip-153671043369099517	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-14 03:40:17.106733	...	Haryana
3	trip-153671046011330457	training	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	IND400072AAB	Mumbai_Hub (Maharashtra)	IND401104AAA	Mumbai_MiraRd_IP (Maharashtra)	2018-09-12 00:01:00.113710	...	Maharashtra
4	trip-153671052974046625	training	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	IND583101AAA	Bellary_Dc (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	2018-09-12 00:02:09.740725	...	Karnataka
...
14812	trip-153861095625827784	test	2018-10-03 23:55:56.258533	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	Carting	IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	2018-10-03 23:55:56.258533	...	Punjab
14813	trip-153861104386292051	test	2018-10-03 23:57:23.863155	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	Carting	IND121004AAB	FBD_Balabhgarh_DPC (Haryana)	IND121004AAA	Faridabad_Blbgarh_DC (Haryana)	2018-10-03 23:57:23.863155	...	Haryana
14814	trip-153861106442901555	test	2018-10-03 23:57:44.429324	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	Carting	IND208006AAA	Kanpur_GovndNgr_DC (Uttar Pradesh)	IND208006AAA	Kanpur_GovndNgr_DC (Uttar Pradesh)	2018-10-04 02:51:27.075797	...	Uttar Pradesh
14815	trip-153861115439069069	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	IND627005AAA	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	IND628204AAA	Tirchchndr_Shnmgprn_D (Tamil Nadu)	2018-10-03 23:59:14.390954	...	Tamil Nadu
14816	trip-153861118270144424	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	2018-10-04 03:58:40.726547	...	Karnataka

14817 rows × 31 columns



data visualisation

In [101...

df

	trip_uuid	data	trip_creation_time	route_schedule_uuid	route_type	source_center	source_name	destination_center	destination_name	od_start_time	...	dest_state
0	trip-153671041653548748	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12 16:39:46.858469	...	Uttar Pradesh
1	trip-153671042288605164	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12 02:03:09.655591	...	Karnataka
2	trip-153671043369099517	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-14 03:40:17.106733	...	Haryana
3	trip-153671046011330457	training	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	IND400072AAB	Mumbai_Hub (Maharashtra)	IND401104AAA	Mumbai_MiraRd_IP (Maharashtra)	2018-09-12 00:01:00.113710	...	Maharashtra
4	trip-153671052974046625	training	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	IND583101AAA	Bellary_Dc (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	2018-09-12 00:02:09.740725	...	Karnataka
...
14812	trip-153861095625827784	test	2018-10-03 23:55:56.258533	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	Carting	IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	2018-10-03 23:55:56.258533	...	Punjab
14813	trip-153861104386292051	test	2018-10-03 23:57:23.863155	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	Carting	IND121004AAB	FBD_Balabhgarh_DPC (Haryana)	IND121004AAA	Faridabad_Blbgarh_DC (Haryana)	2018-10-03 23:57:23.863155	...	Haryana
14814	trip-153861106442901555	test	2018-10-03 23:57:44.429324	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	Carting	IND208006AAA	Kanpur_GovndNgr_DC (Uttar Pradesh)	IND208006AAA	Kanpur_GovndNgr_DC (Uttar Pradesh)	2018-10-04 02:51:27.075797	...	Uttar Pradesh
14815	trip-153861115439069069	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	IND627005AAA	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	IND628204AAA	Tirchchndr_Shnmgprn_D (Tamil Nadu)	2018-10-03 23:59:14.390954	...	Tamil Nadu
14816	trip-153861118270144424	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	2018-10-04 03:58:40.726547	...	Karnataka

14817 rows × 31 columns

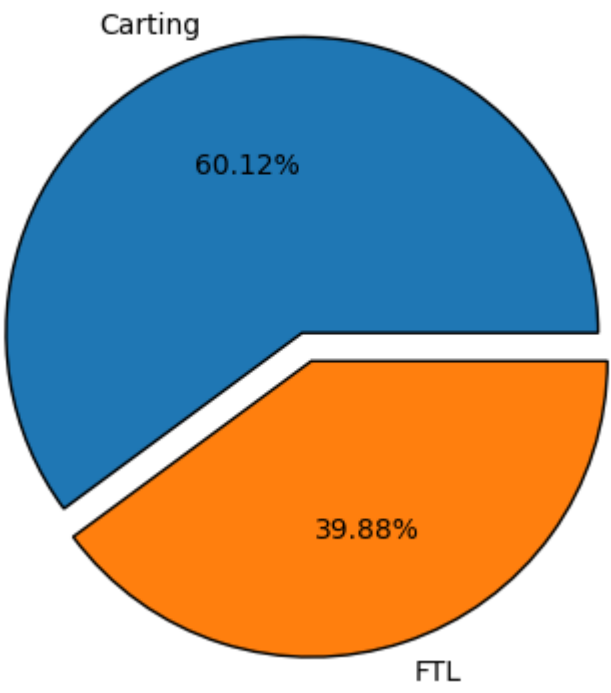


which transportation is preferred more

In [102...

```
plt.pie(df['route_type'].value_counts(),labels=df['route_type'].value_counts().index,autopct='%1.2f%',wedgeprops={'edgecolor': 'black'},explode=(0,0.1))
plt.title('Transportation is preferred more')
plt.show()
```

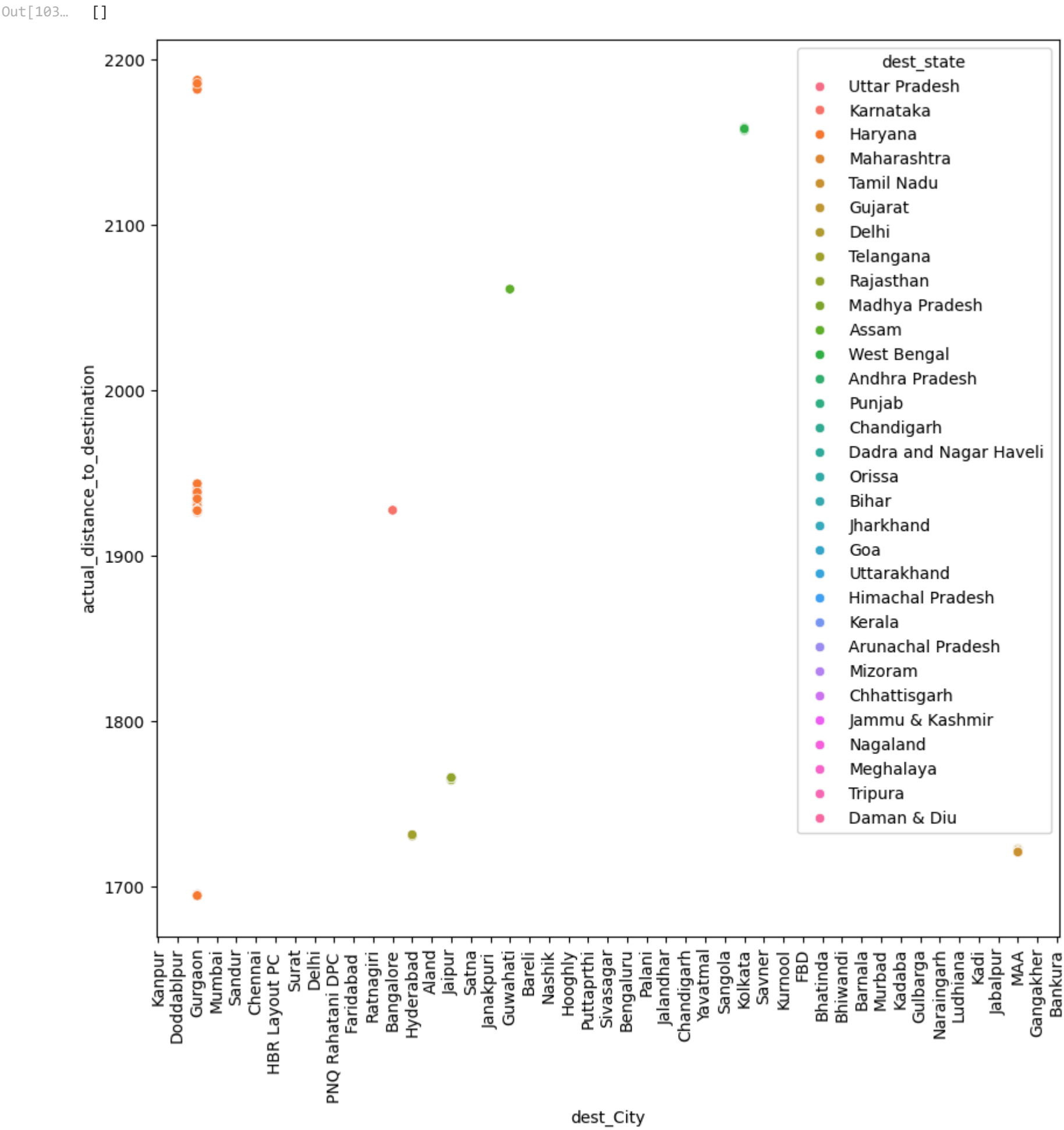
Transportation is preferred more



- Most of them using carting shipments rout type .
- FTL:39.88%
- Carting:60.12%

actual distance to destination

```
In [103... y=df['actual_distance_to_destination'].sort_values(ascending=False).head(100).to_frame()
y
plt.figure(figsize=(10,10))
sns.scatterplot(x=df['dest_City'],y=y['actual_distance_to_destination'],hue=df['dest_state'])
plt.xticks(rotation=90)
plt.plot()
```



- rom this information we can say that gurgaon in hariyana is the city which take more distance to reach destination compaired other cities in whole states.

destination state delivery count

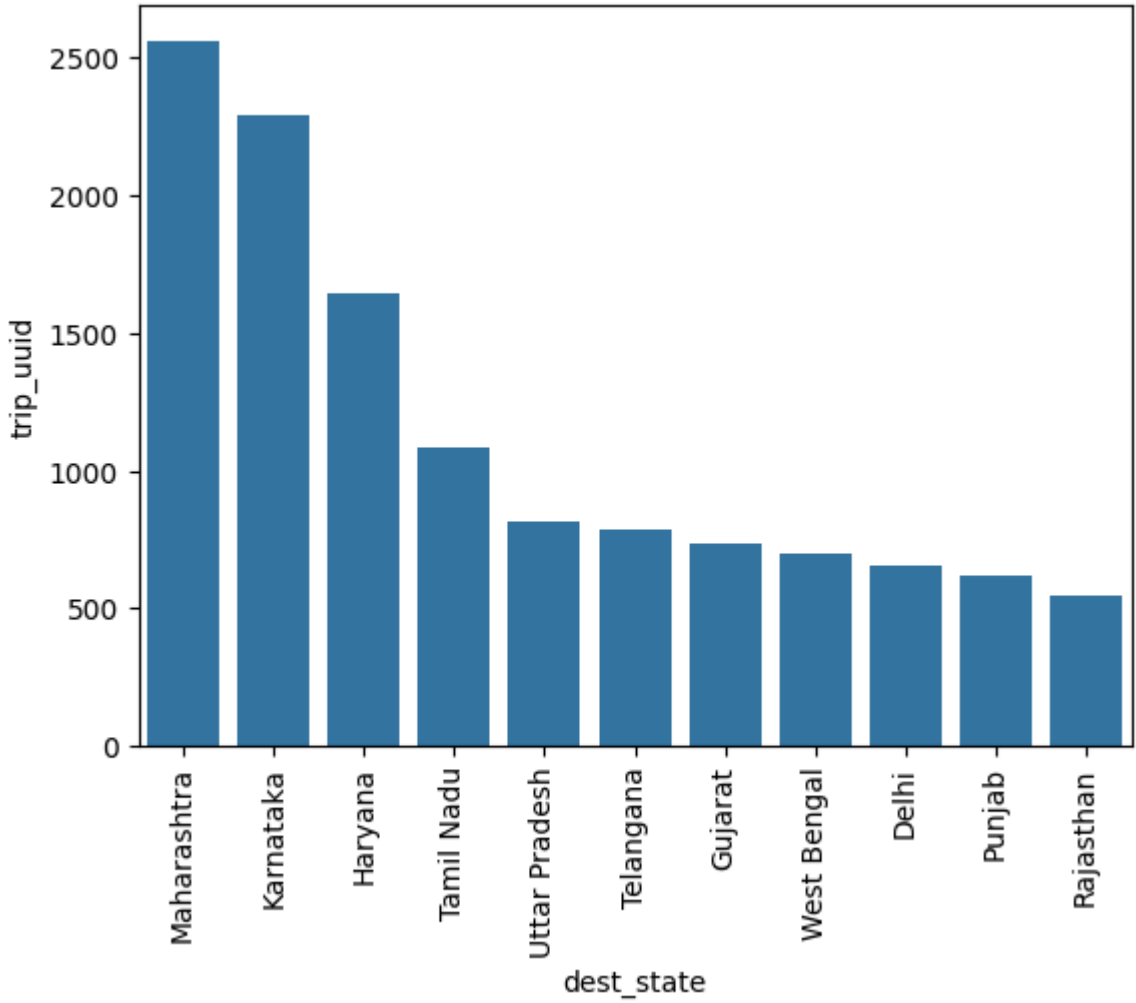
```
In [104... x=df.groupby('dest_state')['trip_uuid'].count().sort_values(ascending=False).head(11).reset_index()
x
```


Out[104...

	dest_state	trip_uuid
0	Maharashtra	2561
1	Karnataka	2294
2	Haryana	1643
3	Tamil Nadu	1084
4	Uttar Pradesh	811
5	Telangana	784
6	Gujarat	734
7	West Bengal	697
8	Delhi	652
9	Punjab	617
10	Rajasthan	543

In [105...

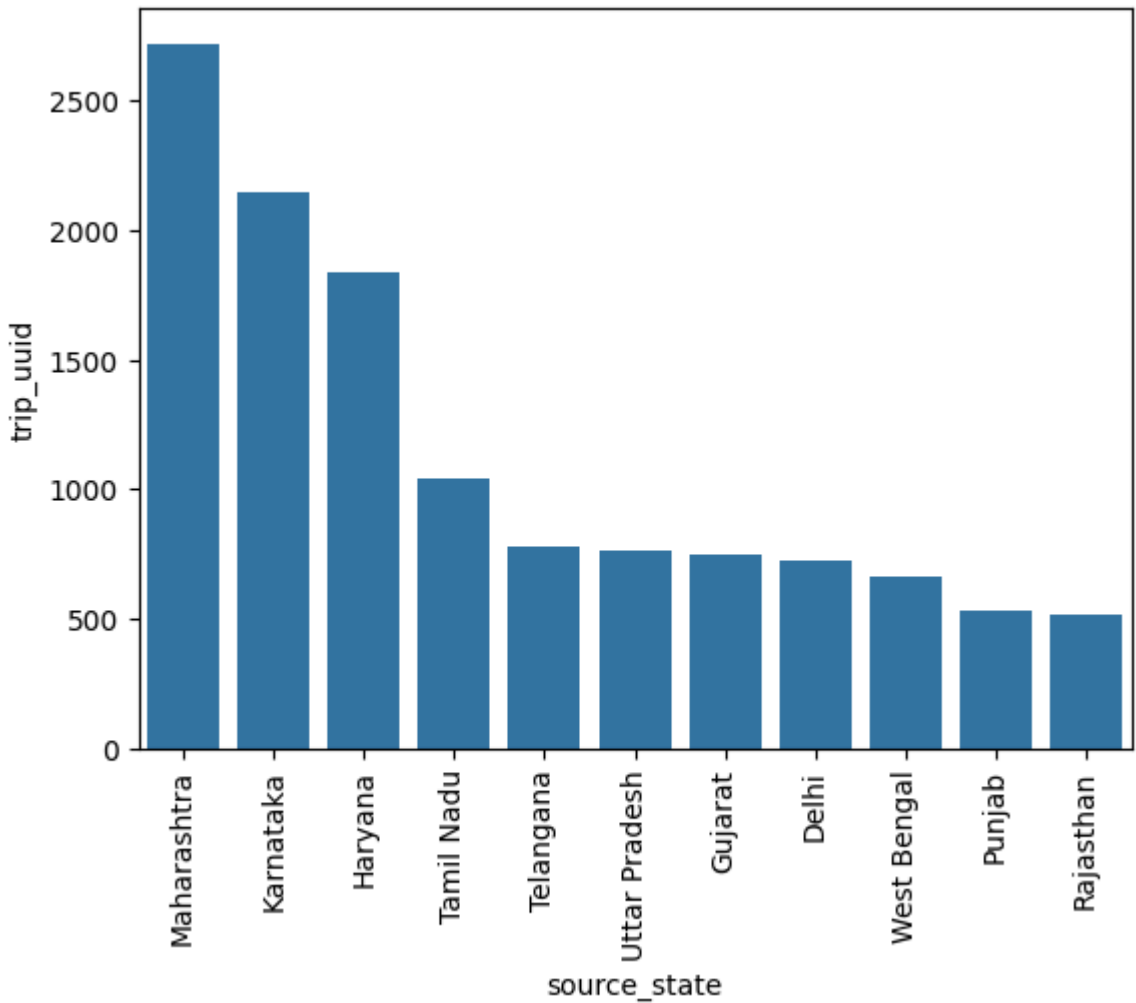
```
sns.barplot(x='dest_state',y='trip_uuid',data=x)
plt.xticks(rotation=90)
plt.show()
```



destination state by delivery count

In [106...

```
x_source = df.groupby('source_state')['trip_uuid'].count().sort_values(ascending=False).head(11).reset_index()
sns.barplot(x='source_state', y='trip_uuid', data=x_source)
plt.xticks(rotation=90)
plt.show()
```



Most deliveries done in Maharashtra state :2561

- After that karnataka : 2294
- Top 5 Indian states with most product delivery destinations are :
- 1)Maharashtra
- 2)Karnataka
- 3)Haryana
- 4)Tamil Nadu and
- 5)Utter pradesh

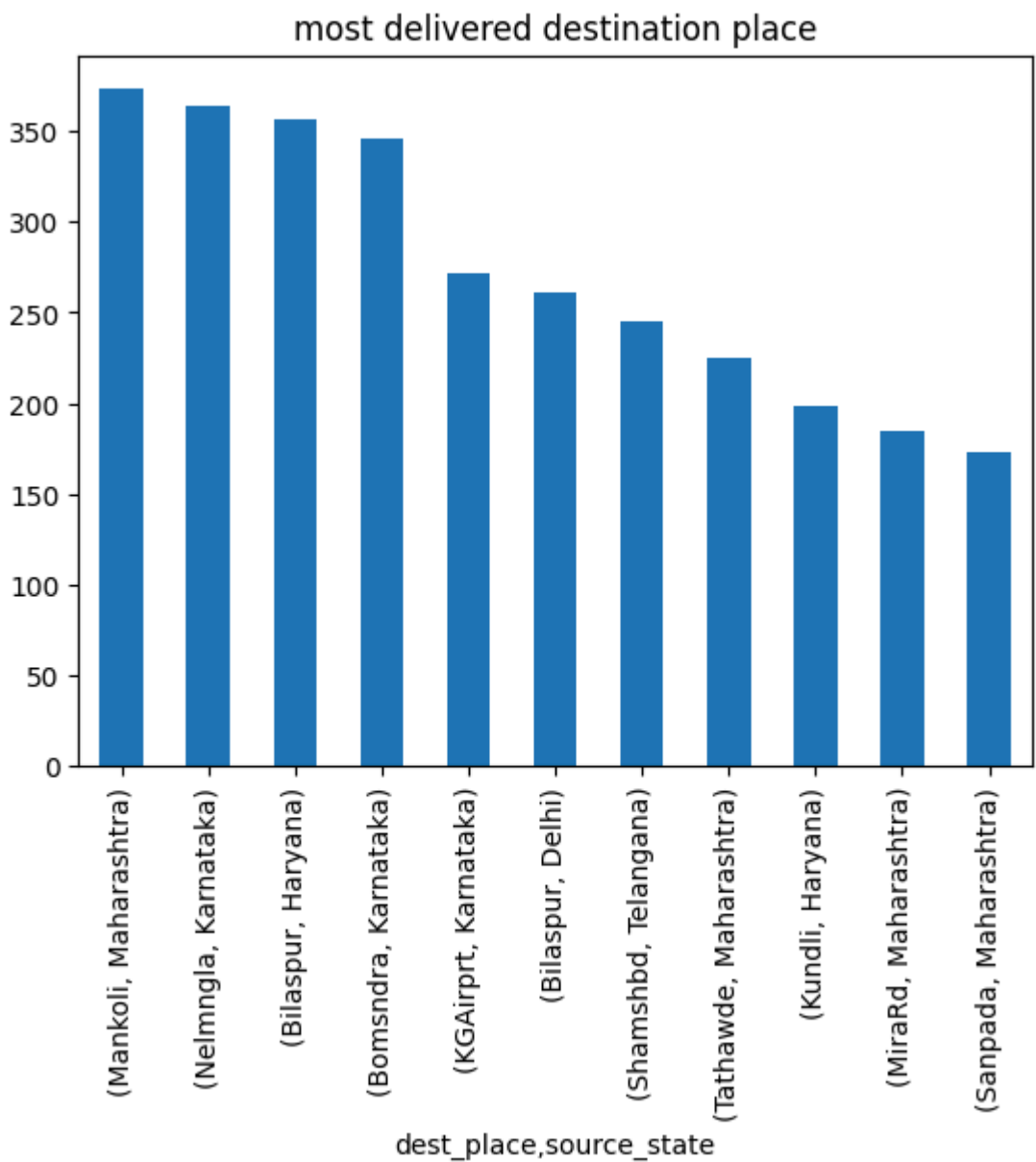
most delivered destination places

In [107...

```
df.groupby(['dest_place','source_state'])['trip_uuid'].count().sort_values(ascending=False).head(11).plot(kind='bar')
plt.title("most delivered destination place")
```

Out[107...

```
Text(0.5, 1.0, 'most delivered destination place')
```



The states with most destination places are :

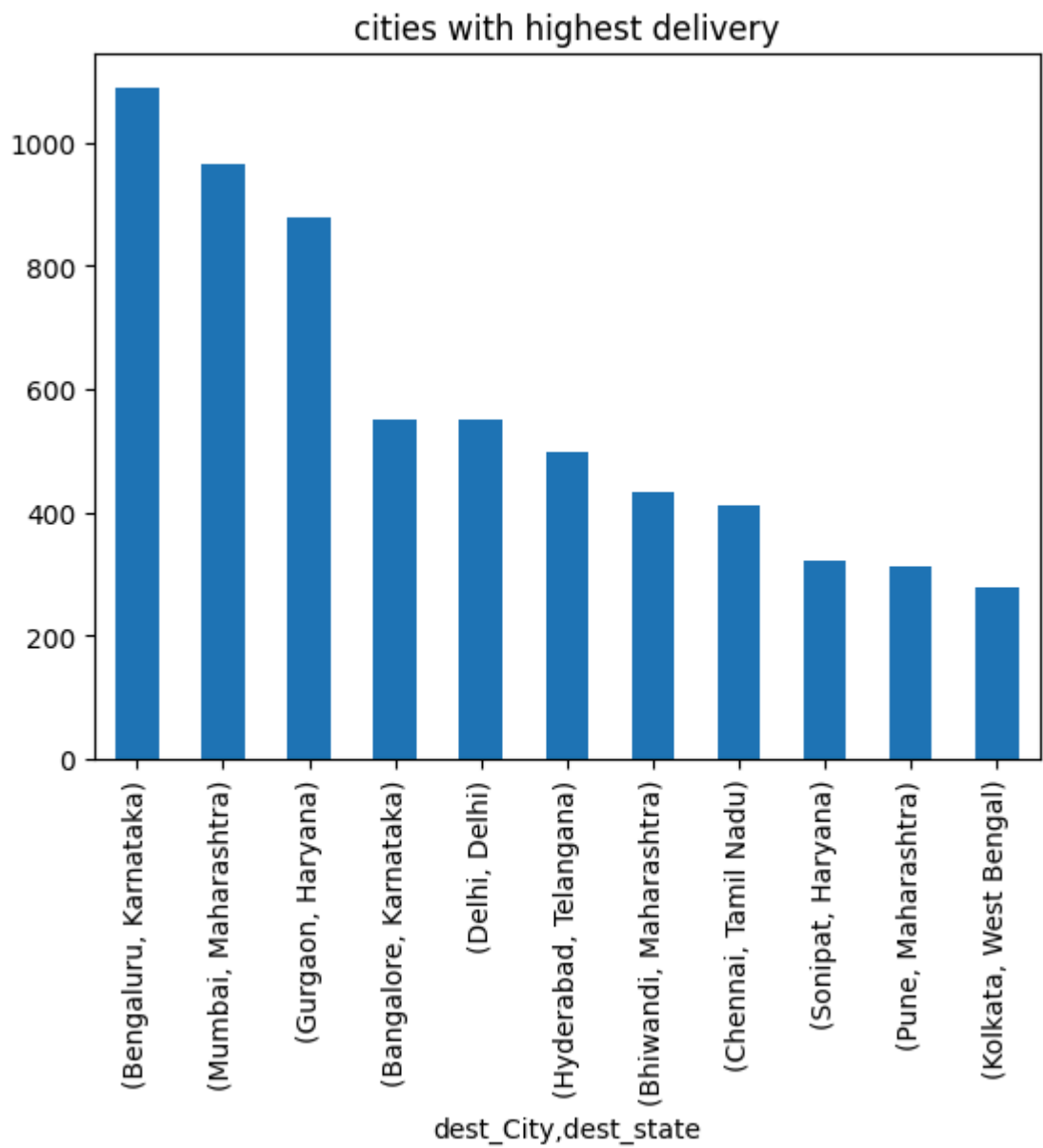
- Bilaspur :Haryana
- Mankoli :Maharashtra
- Nelmngla :karnataka
- Bomsndra : karnataka
- Tathawde :Maharashtra

cities with highest delivery

```
In [108... city=df.groupby(['dest_City','dest_state'])['trip_uuid'].count().sort_values(ascending=False).head(11).reset_index()
city
```

	dest_City	dest_state	trip_uuid
0	Bengaluru	Karnataka	1088
1	Mumbai	Maharashtra	966
2	Gurgaon	Haryana	877
3	Bangalore	Karnataka	551
4	Delhi	Delhi	549
5	Hyderabad	Telangana	499
6	Bhiwandi	Maharashtra	434
7	Chennai	Tamil Nadu	410
8	Sonipat	Haryana	322
9	Pune	Maharashtra	313
10	Kolkata	West Bengal	277

```
In [109... df.groupby(['dest_City','dest_state'])['trip_uuid'].count().sort_values(ascending=False).head(11).plot(kind='bar')
plt.title("cities with highest delivery")
plt.show()
```

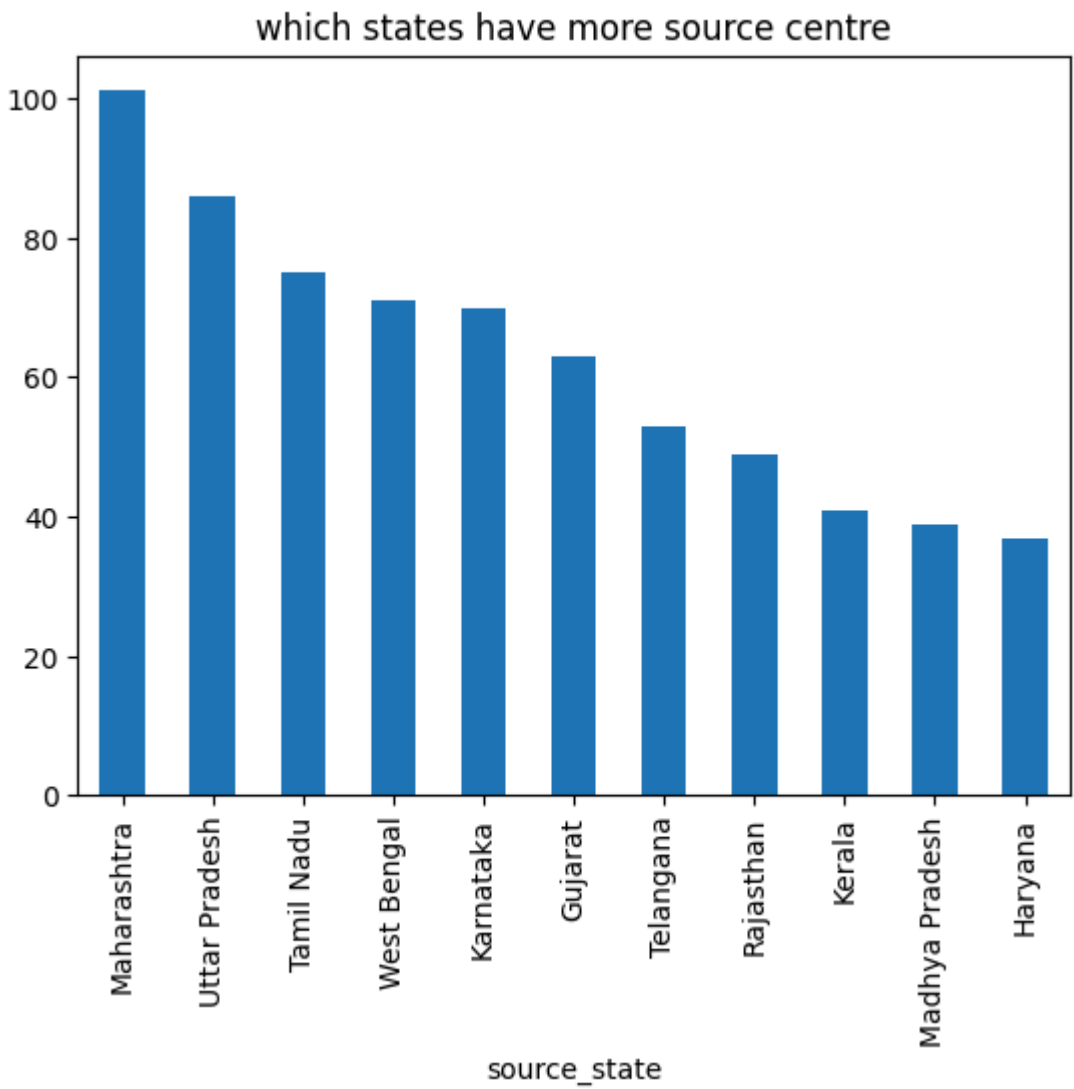


- banglore karnataka is the highest delivered city followed by mumbai and gurgaon,kolkatta being the last.

which states have more source centre

```
In [110... df.groupby('source_state')['source_center'].nunique().sort_values(ascending=False).head(11).plot(kind='bar')
plt.title("which states have more source centre")
```

```
plt.show()
```



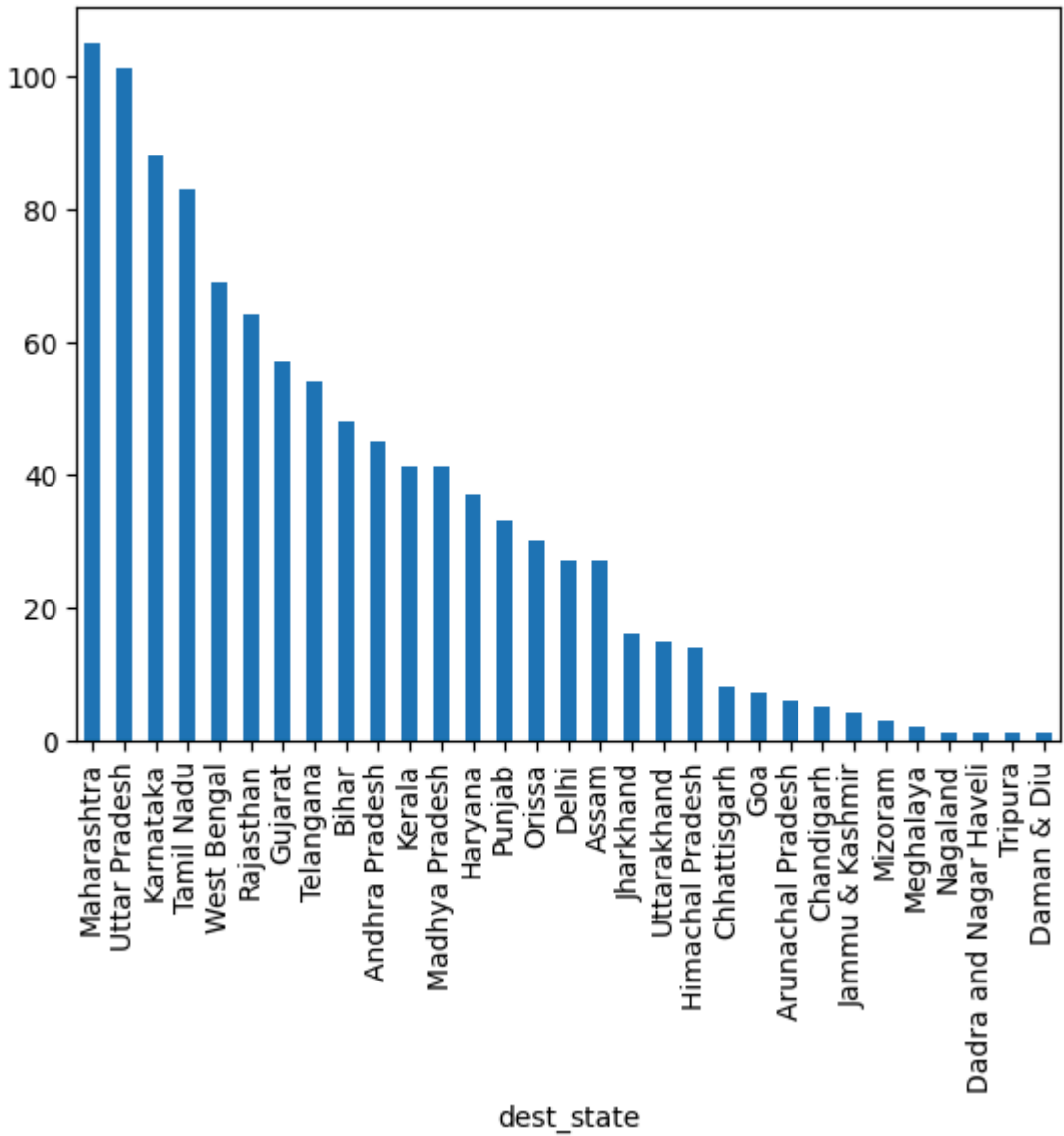
By this we can say maharashtra have more source center compaired to other states After thar uttar pradesh ,thamilnadu owned 2 nd and 3rd place of state have more centers. Top 5 Indian states with most delivery source centres are :

- Maharashtra
- Uttar Pradesh
- Tamil Nadu
- West Bengal and
- Karnataka

which state has more destination centre

```
In [111... df.groupby('dest_state')['destination_center'].nunique().sort_values(ascending=False).plot(kind='bar')
```

```
Out[111... <Axes: xlabel='dest_state'>
```

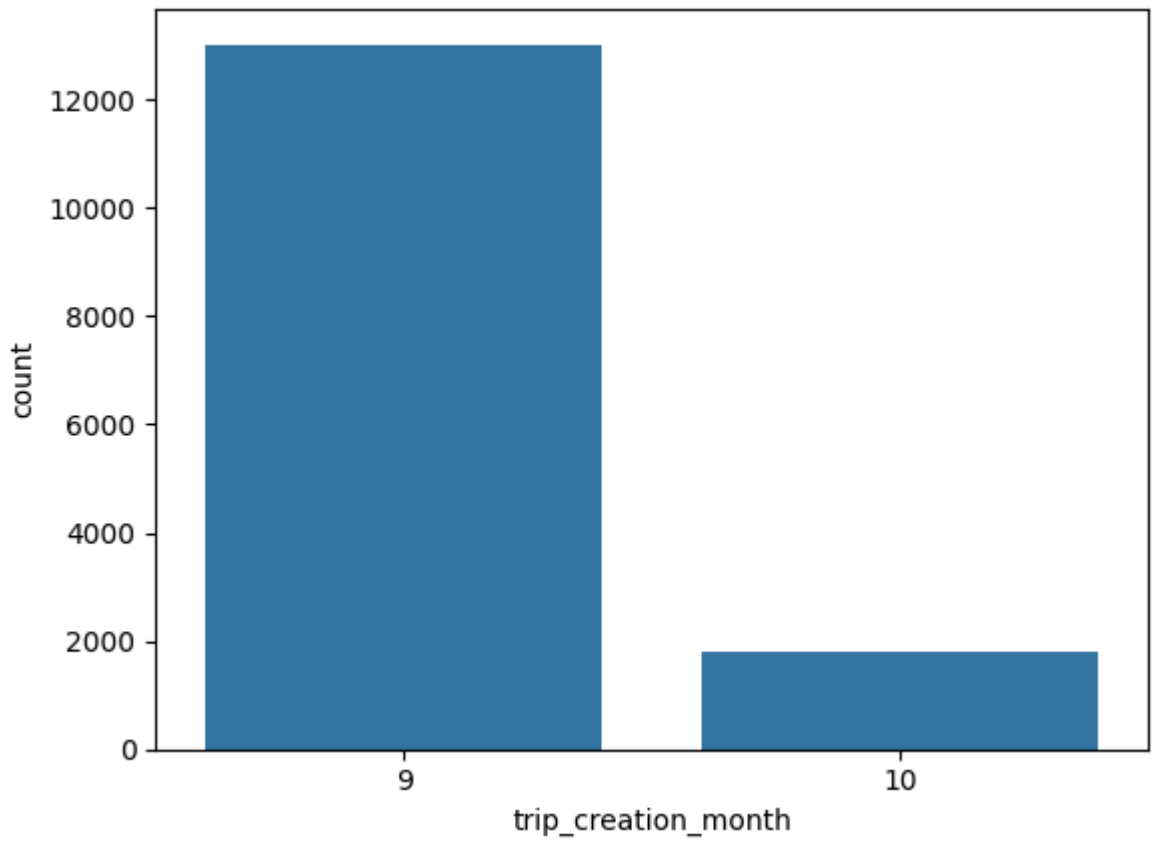


By this we can say maharashtra have more destination center compaired to other states After thar uttar pradesh ,Karnataka owned 2nd and 3rd place of state have more centers. Top 5 Indian states with most delivery destination centres are :

- Maharashtra
- Uttar Pradesh
- Haryana
- Karnataka and
- Tamil Nadu

trip according to month

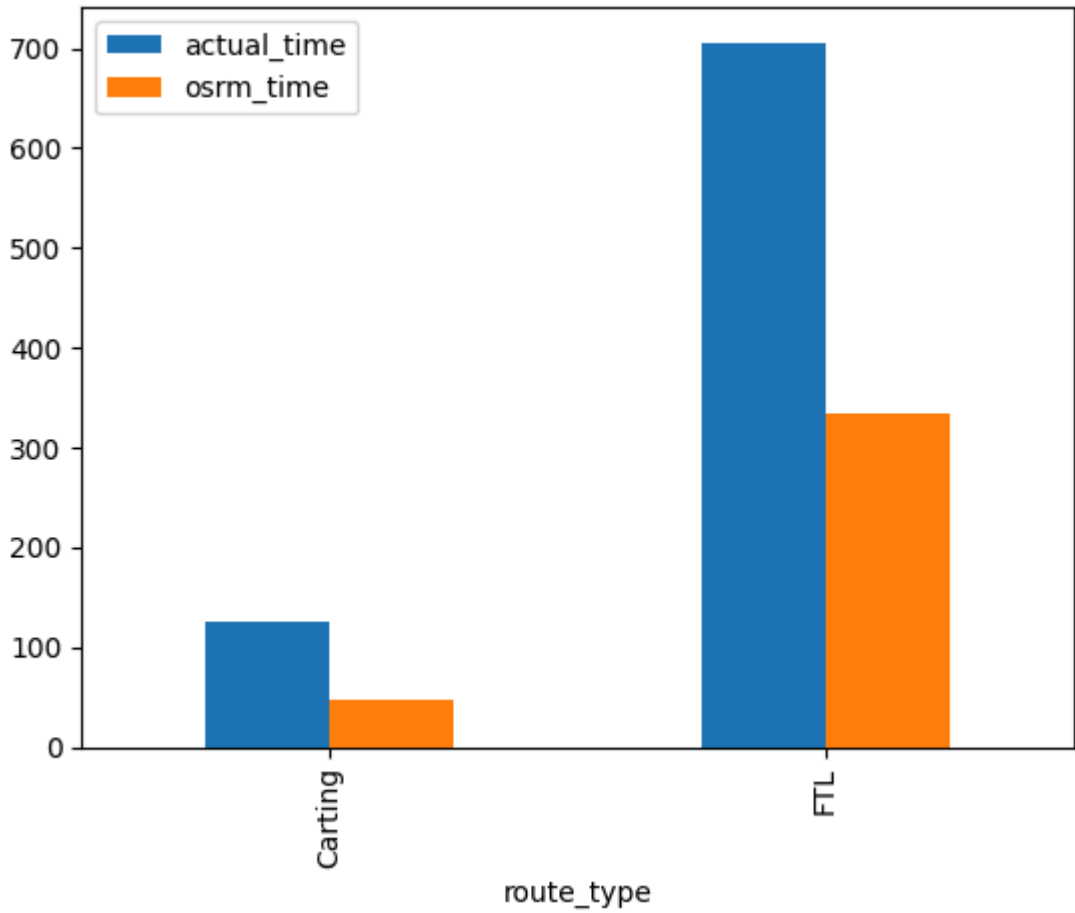
```
In [112... df['trip_creation_month'] = df['trip_creation_time'].dt.month
sns.countplot(x='trip_creation_month', data=df)
plt.show()
```



majority of trip creation occurred in 9 th month

which route type use more time to reach destination

```
In [113... df.groupby('route_type').aggregate({'actual_time': 'mean', 'osrm_time': 'mean'}).plot(kind='bar')
plt.show()
```

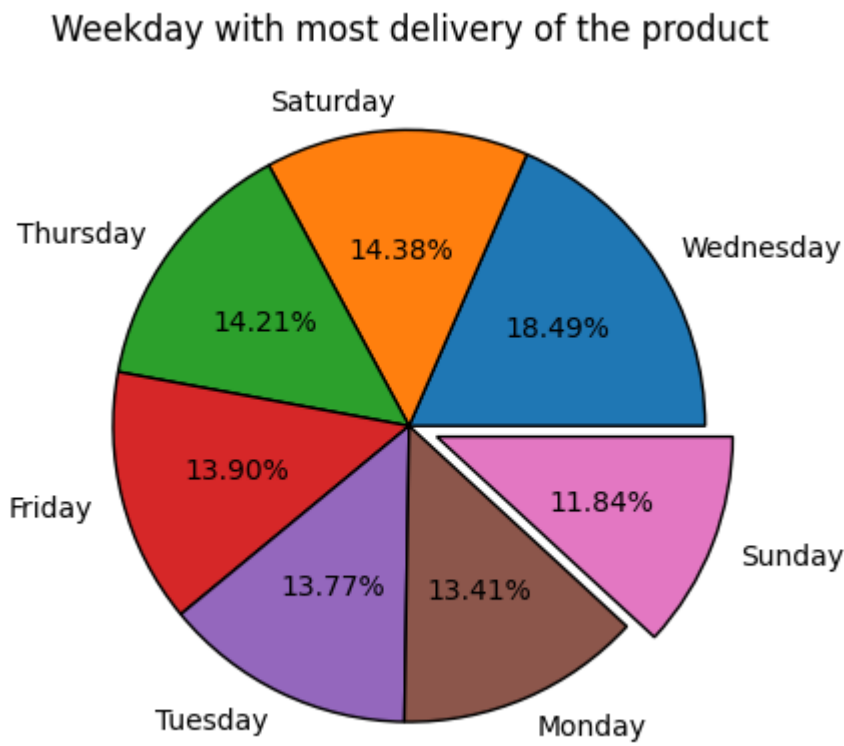


By this we can say that Full Truck Load, FTL shipments Full Truck Load takes more time to deliver product and time taken to actual and osrm higher than that of route type.

Weekday with most delivery of the product

```
In [114... df['weekday'] = df['trip_creation_time'].dt.day_name()

In [115... plt.pie(df['weekday'].value_counts(), labels=df['weekday'].value_counts().index
        , autopct='%1.2f%%', wedgeprops={'edgecolor': 'black'}, explode=(0,0,0,0,0,0,0.1))
plt.title('Weekday with most delivery of the product')
plt.show()
```



- Compaired to other days mostly deliveries done in wednesday wednesday :18.49% of trips are done
- After that thursday: 14.38%

Actual time v/s Time which computes the shortest path

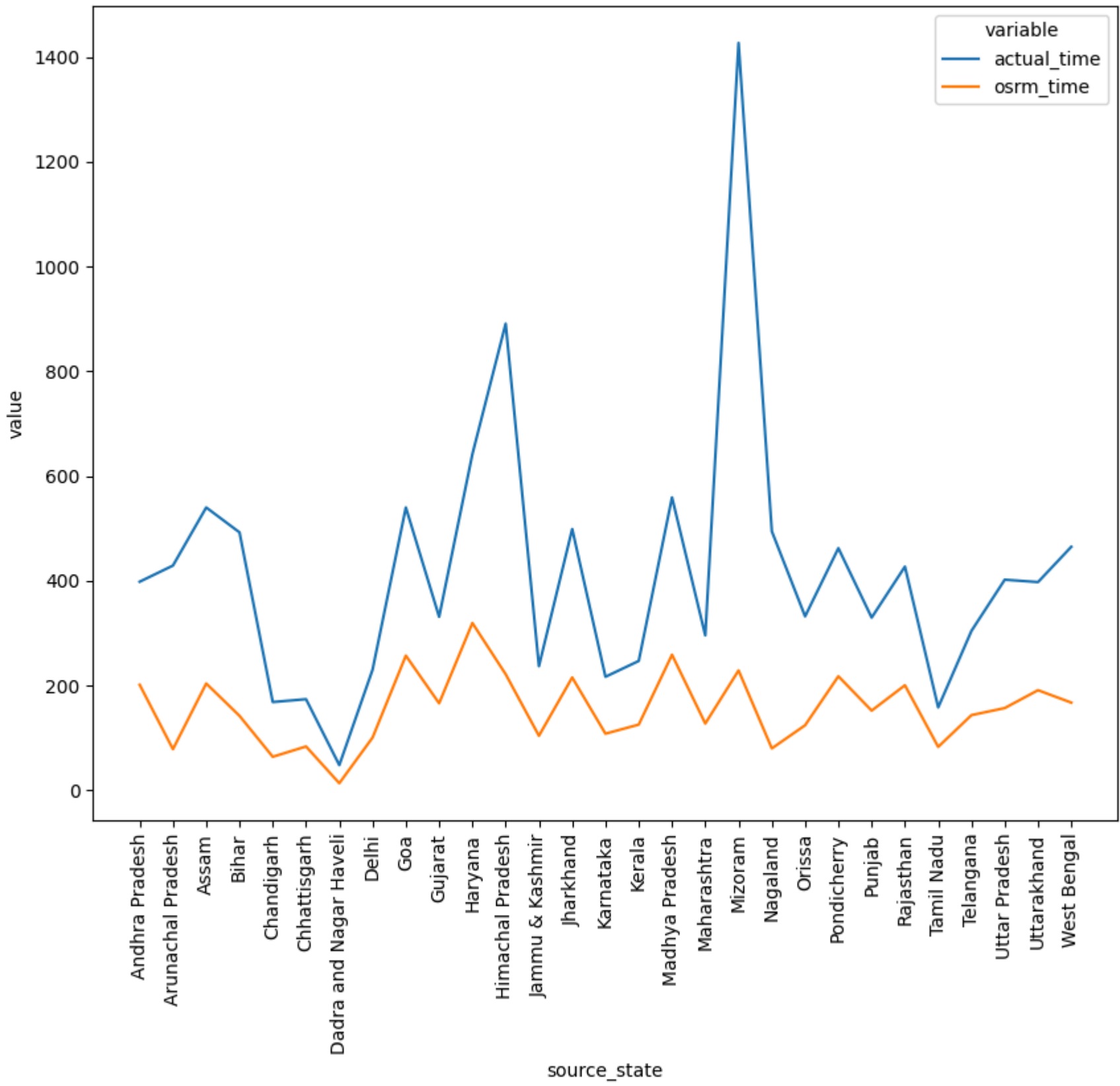
```
In [116... x1=df.groupby('source_state').agg({'actual_time': 'mean', 'osrm_time': 'mean'}).reset_index()
x1=pd.melt(x1, id_vars=['source_state'], value_vars=['actual_time', 'osrm_time'])
x1
```


Out[116...

	source_state	variable	value
0	Andhra Pradesh	actual_time	398.435484
1	Arunachal Pradesh	actual_time	429.250000
2	Assam	actual_time	540.171642
3	Bihar	actual_time	492.645714
4	Chandigarh	actual_time	168.741935
5	Chhattisgarh	actual_time	174.139535
6	Dadra and Nagar Haveli	actual_time	48.333333
7	Delhi	actual_time	230.550824
8	Goa	actual_time	540.138462
9	Gujarat	actual_time	331.470667
10	Haryana	actual_time	641.968988
11	Himachal Pradesh	actual_time	891.088235
12	Jammu & Kashmir	actual_time	237.235294
13	Jharkhand	actual_time	498.981250
14	Karnataka	actual_time	216.933738
15	Kerala	actual_time	247.173010
16	Madhya Pradesh	actual_time	559.406940
17	Maharashtra	actual_time	296.011054
18	Mizoram	actual_time	1427.000000
19	Nagaland	actual_time	494.600000
20	Orissa	actual_time	332.308411
21	Pondicherry	actual_time	462.416667
22	Punjab	actual_time	329.917910
23	Rajasthan	actual_time	427.264591
24	Tamil Nadu	actual_time	158.486044
25	Telangana	actual_time	304.624840
26	Uttar Pradesh	actual_time	402.257218
27	Uttarakhand	actual_time	397.728070
28	West Bengal	actual_time	465.124812
29	Andhra Pradesh	osrm_time	201.670507
30	Arunachal Pradesh	osrm_time	78.750000
31	Assam	osrm_time	204.085821
32	Bihar	osrm_time	142.431429
33	Chandigarh	osrm_time	64.215054
34	Chhattisgarh	osrm_time	83.906977
35	Dadra and Nagar Haveli	osrm_time	13.733333
36	Delhi	osrm_time	100.934066
37	Goa	osrm_time	257.276923
38	Gujarat	osrm_time	166.368000
39	Haryana	osrm_time	319.556039
40	Himachal Pradesh	osrm_time	221.852941
41	Jammu & Kashmir	osrm_time	104.352941
42	Jharkhand	osrm_time	215.600000
43	Karnataka	osrm_time	108.295847
44	Kerala	osrm_time	125.650519
45	Madhya Pradesh	osrm_time	258.952681
46	Maharashtra	osrm_time	127.739499
47	Mizoram	osrm_time	229.000000
48	Nagaland	osrm_time	80.200000
49	Orissa	osrm_time	124.411215
50	Pondicherry	osrm_time	217.916667
51	Punjab	osrm_time	152.354478
52	Rajasthan	osrm_time	200.708171
53	Tamil Nadu	osrm_time	83.142445
54	Telangana	osrm_time	143.732394
55	Uttar Pradesh	osrm_time	157.217848
56	Uttarakhand	osrm_time	191.324561
57	West Bengal	osrm_time	167.622556

In [117...

```
plt.figure(figsize=(10,8))
sns.lineplot(data=x1,x='source_state',y='value',hue='variable')
plt.xticks(rotation=90)
plt.show()
```



- From the above observation we can say that
- In Dadra and Nagar Haveli the actual and osrm time are almost equal
- There is not that much variation in Tamil Nadu also In Mizoram we can see there is high variation in actual and osrm time Similarly in case of Haryana and Himachalpradesh also have high variation

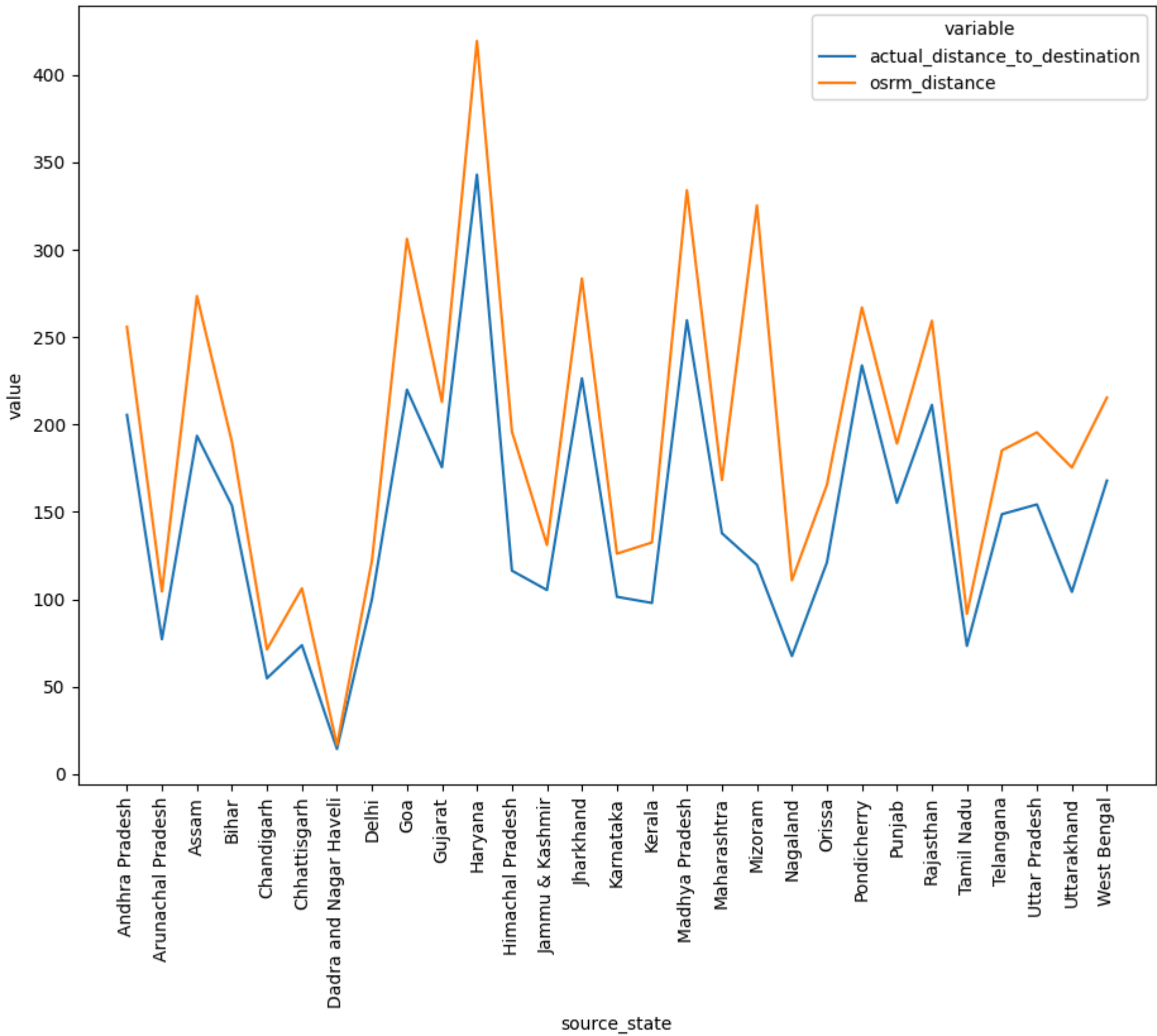
```
In [118... x1=df.groupby('source_state').agg({'actual_distance_to_destination':'mean','osrm_distance':'mean'}).reset_index()
x1=pd.melt(x1, id_vars=['source_state'], value_vars=['actual_distance_to_destination', 'osrm_distance'])
x1
```

Out[118...

	source_state	variable	value
0	Andhra Pradesh	actual_distance_to_destination	205.461342
1	Arunachal Pradesh	actual_distance_to_destination	77.150532
2	Assam	actual_distance_to_destination	193.532247
3	Bihar	actual_distance_to_destination	153.588085
4	Chandigarh	actual_distance_to_destination	54.811432
5	Chhattisgarh	actual_distance_to_destination	73.734984
6	Dadra and Nagar Haveli	actual_distance_to_destination	14.349976
7	Delhi	actual_distance_to_destination	100.054953
8	Goa	actual_distance_to_destination	219.883526
9	Gujarat	actual_distance_to_destination	175.549300
10	Haryana	actual_distance_to_destination	342.895103
11	Himachal Pradesh	actual_distance_to_destination	116.370463
12	Jammu & Kashmir	actual_distance_to_destination	105.300217
13	Jharkhand	actual_distance_to_destination	226.406040
14	Karnataka	actual_distance_to_destination	101.412585
15	Kerala	actual_distance_to_destination	97.875658
16	Madhya Pradesh	actual_distance_to_destination	259.562780
17	Maharashtra	actual_distance_to_destination	137.783489
18	Mizoram	actual_distance_to_destination	119.774782
19	Nagaland	actual_distance_to_destination	67.510835
20	Orissa	actual_distance_to_destination	121.063622
21	Pondicherry	actual_distance_to_destination	233.749084
22	Punjab	actual_distance_to_destination	155.165304
23	Rajasthan	actual_distance_to_destination	211.217983
24	Tamil Nadu	actual_distance_to_destination	73.311371
25	Telangana	actual_distance_to_destination	148.657690
26	Uttar Pradesh	actual_distance_to_destination	154.254277
27	Uttarakhand	actual_distance_to_destination	104.240000
28	West Bengal	actual_distance_to_destination	167.877726
29	Andhra Pradesh	osrm_distance	255.804854
30	Arunachal Pradesh	osrm_distance	104.452450
31	Assam	osrm_distance	273.490164
32	Bihar	osrm_distance	189.872446
33	Chandigarh	osrm_distance	71.311419
34	Chhattisgarh	osrm_distance	106.256274
35	Dadra and Nagar Haveli	osrm_distance	16.592587
36	Delhi	osrm_distance	122.053160
37	Goa	osrm_distance	306.198908
38	Gujarat	osrm_distance	212.907671
39	Haryana	osrm_distance	419.345857
40	Himachal Pradesh	osrm_distance	195.963321
41	Jammu & Kashmir	osrm_distance	131.068324
42	Jharkhand	osrm_distance	283.488592
43	Karnataka	osrm_distance	126.073458
44	Kerala	osrm_distance	132.494697
45	Madhya Pradesh	osrm_distance	333.969259
46	Maharashtra	osrm_distance	168.249902
47	Mizoram	osrm_distance	325.281675
48	Nagaland	osrm_distance	110.817940
49	Orissa	osrm_distance	165.384751
50	Pondicherry	osrm_distance	266.931242
51	Punjab	osrm_distance	189.070016
52	Rajasthan	osrm_distance	259.339060
53	Tamil Nadu	osrm_distance	91.682308
54	Telangana	osrm_distance	185.170089
55	Uttar Pradesh	osrm_distance	195.463270
56	Uttarakhand	osrm_distance	175.364052
57	West Bengal	osrm_distance	215.371641

In [119...

```
plt.figure(figsize=(11,8))
sns.lineplot(data=x1,x='source_state',y='value',hue='variable')
plt.xticks(rotation=90)
plt.show()
```



*From the above observation we can say that In Dadra and Nagar Havel,Delhi,Tamil nadu actual and osrm distance are almost equal *There is not that much variation in Bihar,Arunachal Pradesh also *In Mizoram we can see there is high variation in actual and osrm distance Similarly in case of Assam Haryana,and Goa also have high variation.

outlier detetction

```
In [120...
attrs=['start_scan_to_end_scan','actual_distance_to_destination','actual_time','osrm_time',
       'osrm_distance','segment_actual_time','segment_osrm_time','segment_osrm_distance']
fig, axs = plt.subplots(nrows=3, ncols=2, figsize=(15, 12))
fig.subplots_adjust(top=1.3)
count = 0
for row in range(3):
    for col in range(2):
        sns.boxplot(data=df, x=attrs[count], ax=axs[row, col], palette='Set3')
        axs[row,col].set_title(f" {attrs[count]}", pad=12, fontsize=13)
        count += 1
plt.show()
```

<ipython-input-120-969e0c032b94>:8: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(data=df, x=attrs[count], ax=axs[row, col], palette='Set3')
```

<ipython-input-120-969e0c032b94>:8: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(data=df, x=attrs[count], ax=axs[row, col], palette='Set3')
```

<ipython-input-120-969e0c032b94>:8: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(data=df, x=attrs[count], ax=axs[row, col], palette='Set3')
```

<ipython-input-120-969e0c032b94>:8: FutureWarning:

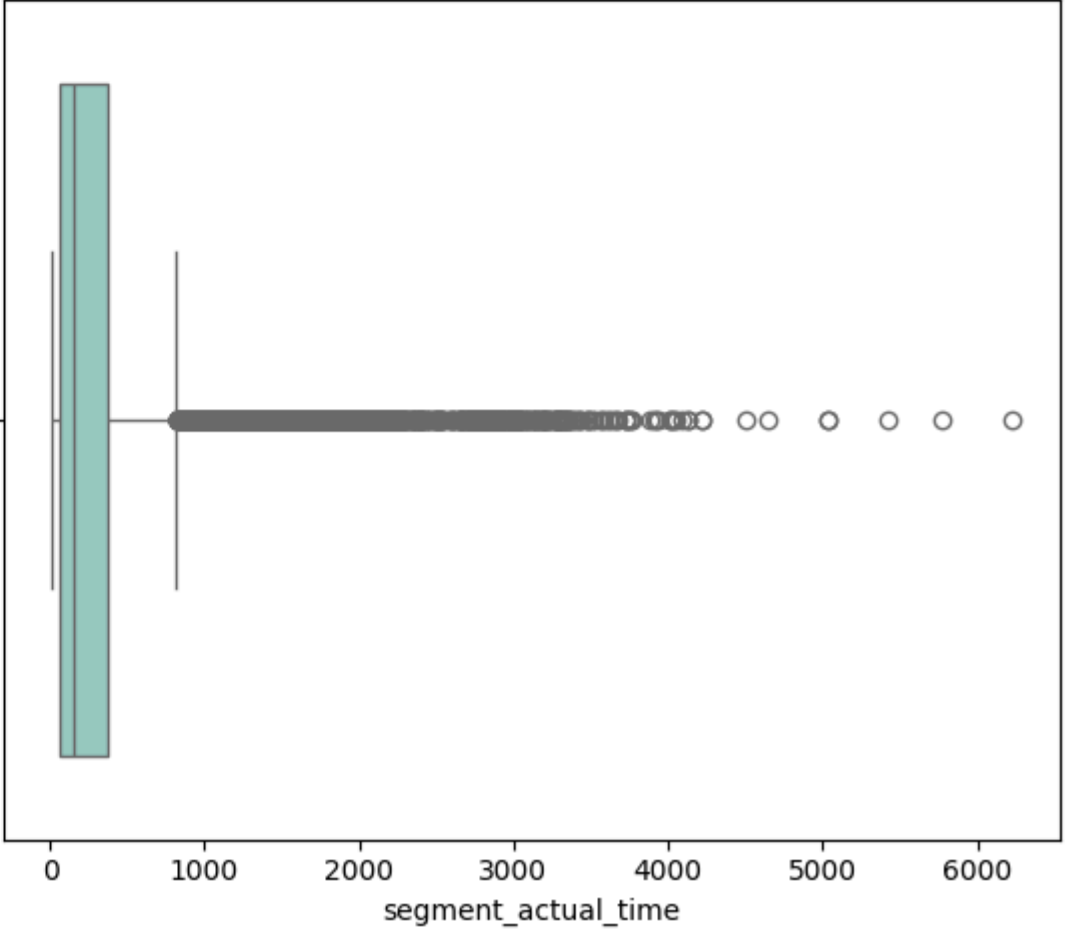
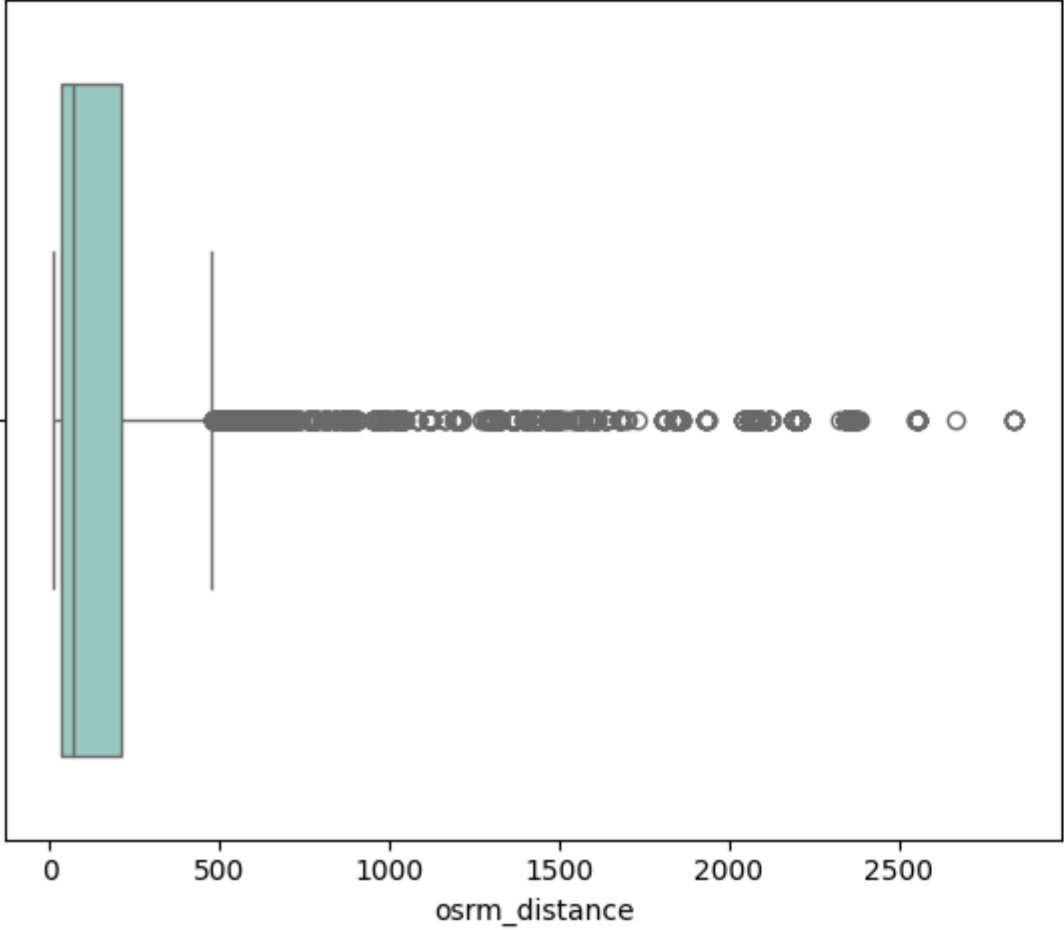
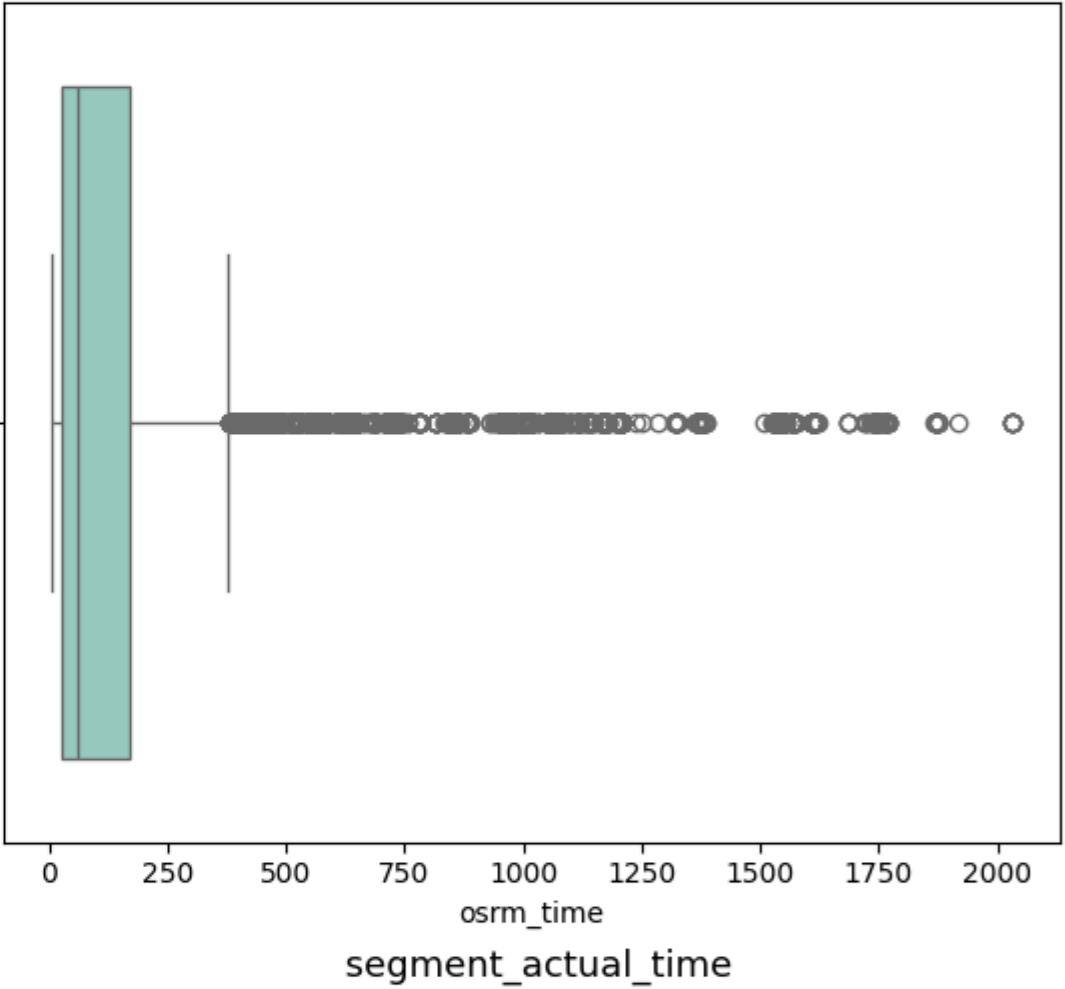
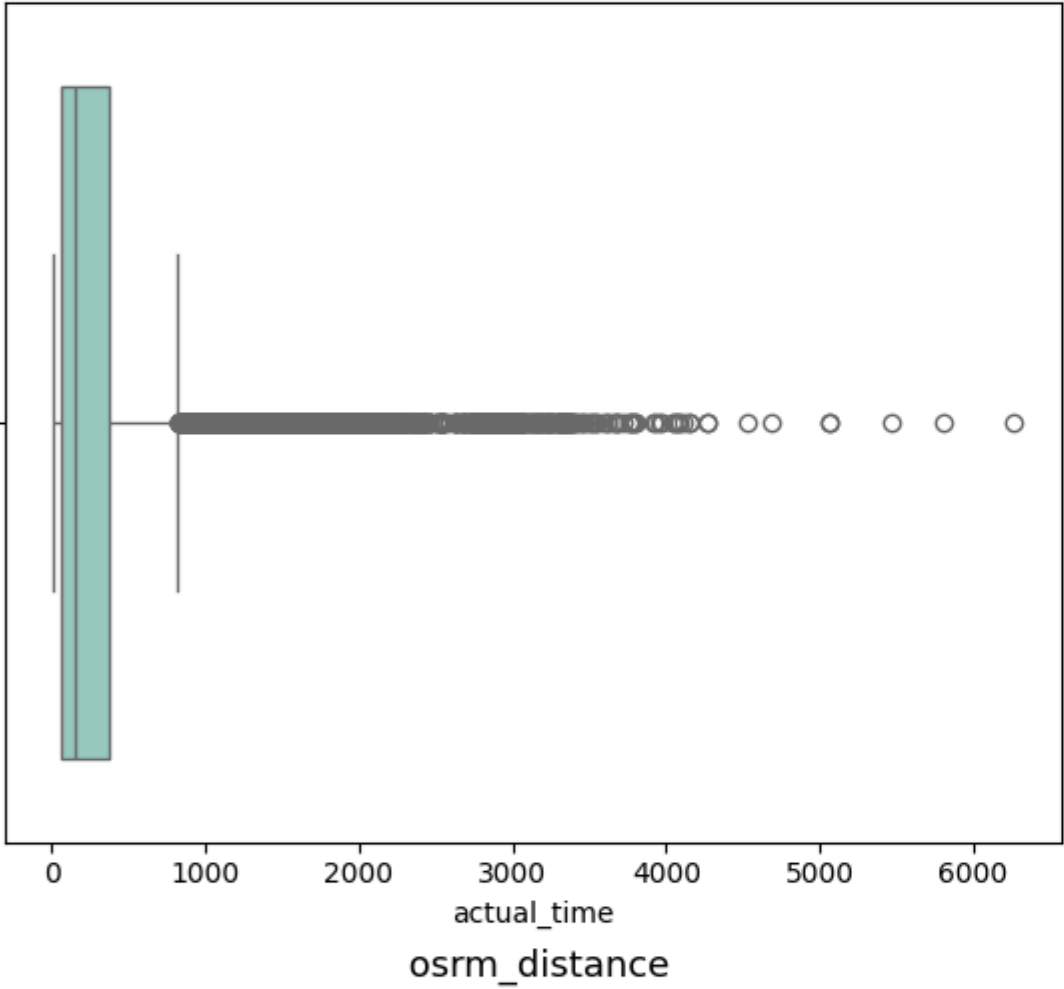
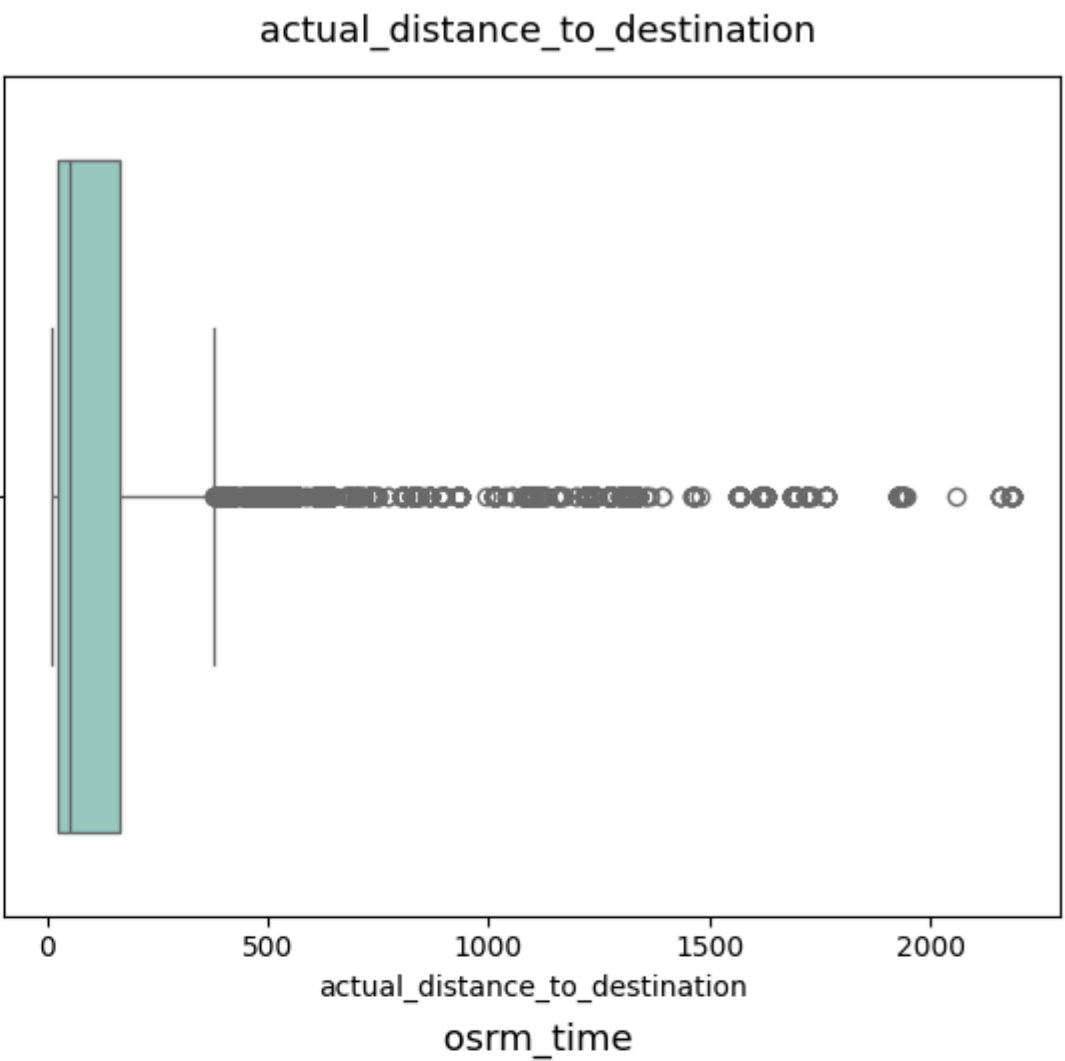
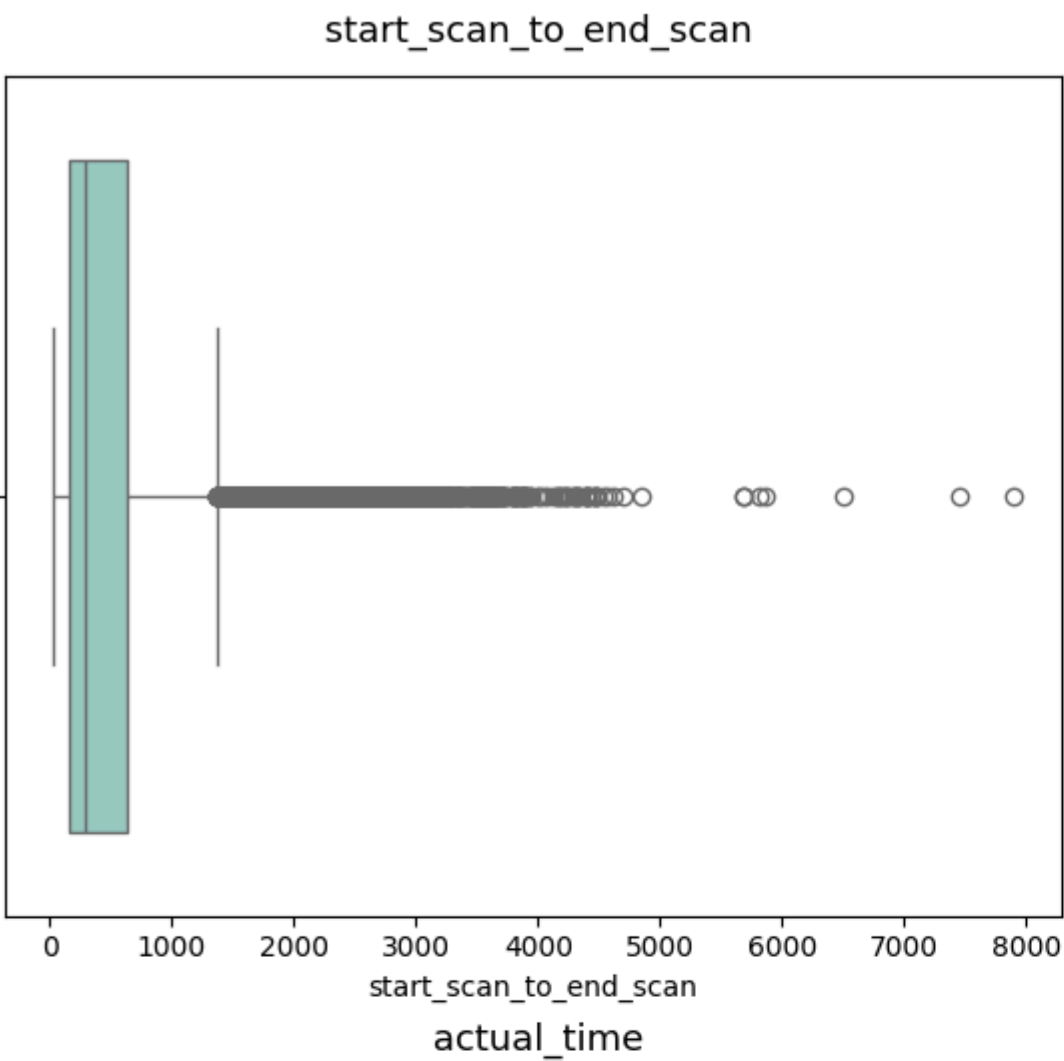
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(data=df, x=attrs[count], ax=axs[row, col], palette='Set3')
```

<ipython-input-120-969e0c032b94>:8: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(data=df, x=attrs[count], ax=axs[row, col], palette='Set3')
```

outlier treatment

```
In [121... ds=df.copy()
attrs=['start_scan_to_end_scan','actual_distance_to_destination','actual_time','osrm_time',
'osrm_distance','segment_actual_time','segment_osrm_time','segment_osrm_distance']
for i in attrs:
    q1=df[i].quantile(.25)
    q3=df[i].quantile(.75)
    iqr=q3-q1
    lower=q1-(1.5*iqr)
    upper=q3+(1.5*iqr)
    print('lower limit of',i,'=',lower)
    print('upper limit of',i,'=',upper)
    print('-----')
    ds=ds[~((ds[i]<lower)|(ds[i]>upper))]
```

```
lower limit of start_scan_to_end_scan = -583.0
upper limit of start_scan_to_end_scan = 1369.0
-----
lower limit of actual_distance_to_destination = -190.12991073629408
upper limit of actual_distance_to_destination = 377.8432652176542
-----
lower limit of actual_time = -387.5
upper limit of actual_time = 824.5
-----
lower limit of osrm_time = -181.0
upper limit of osrm_time = 379.0
-----
lower limit of osrm_distance = -236.59625
upper limit of osrm_distance = 476.83855
-----
lower limit of segment_actual_time = -385.5
upper limit of segment_actual_time = 818.5
-----
lower limit of segment_osrm_time = -200.0
upper limit of segment_osrm_time = 416.0
-----
lower limit of segment_osrm_distance = -246.56735000000003
upper limit of segment_osrm_distance = 498.02425000000005
-----
```

```
In [122... df_new = pd.get_dummies(df, columns=["route_type"])
df_new
```

Out[122...

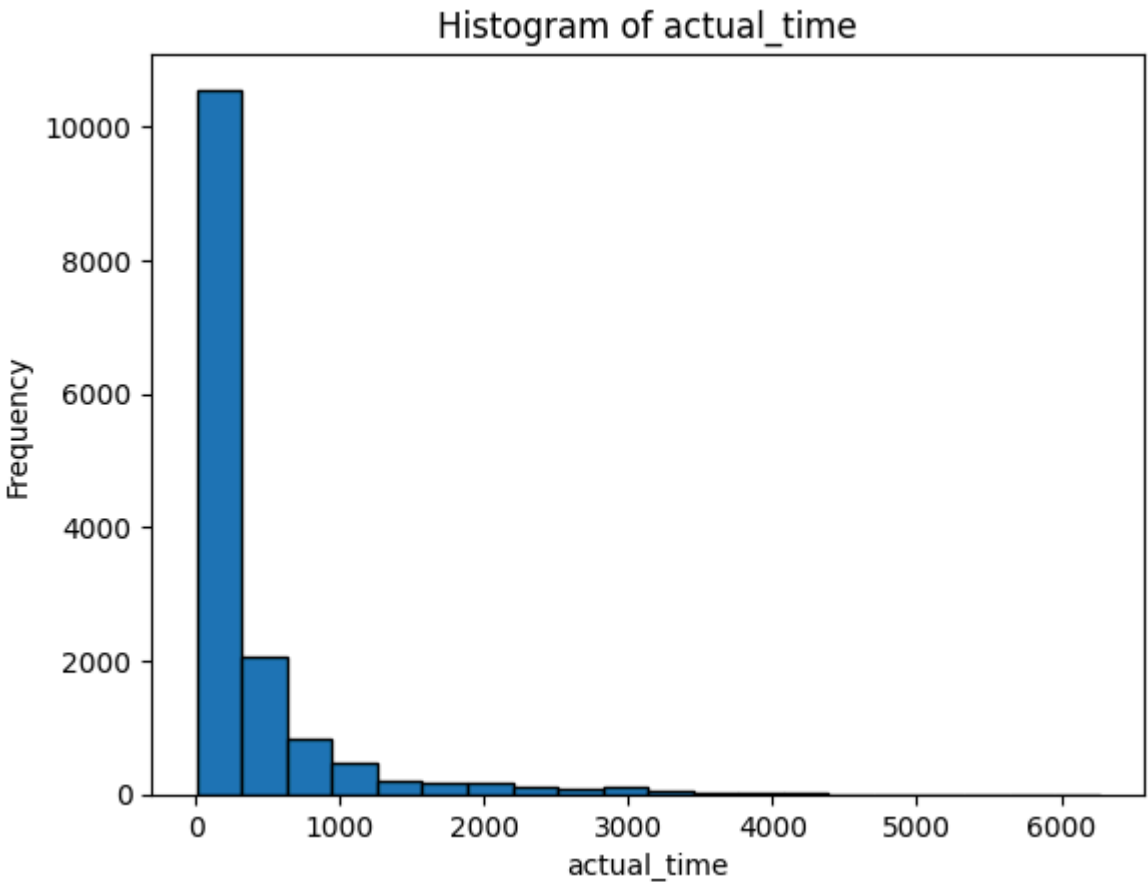
	trip_uuid	data	trip_creation_time	route_schedule_uuid	source_center	source_name	destination_center	destination_name	od_start_time	od_end_time	...	
0	trip-153671041653548748	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12 16:39:46.858469	2018-09-12 16:39:46.858469	...	Kanpur
1	trip-153671042288605164	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12 02:03:09.655591	2018-09-12 02:03:09.655591	...	Doddablpur
2	trip-153671043369099517	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-14 03:40:17.106733	2018-09-14 03:40:17.106733	...	Gurgaon
3	trip-153671046011330457	training	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	IND400072AAB	Mumbai_Hub (Maharashtra)	IND401104AAA	Mumbai_MiraRd_IP (Maharashtra)	2018-09-12 00:01:00.113710	2018-09-12 01:41:29.809822	...	Mumbai
4	trip-153671052974046625	training	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	IND583101AAA	Bellary_Dc (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	2018-09-12 00:02:09.740725	2018-09-12 03:54:43.114421	...	Sandur
...
14812	trip-153861095625827784	test	2018-10-03 23:55:56.258533	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	2018-10-03 23:55:56.258533	2018-10-04 06:41:25.409035	...	Chandigarh
14813	trip-153861104386292051	test	2018-10-03 23:57:23.863155	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	IND121004AAB	FBD_Balabhgarh_DPC (Haryana)	IND121004AAA	Faridabad_Blbgarh_DC (Haryana)	2018-10-03 23:57:23.863155	2018-10-04 00:57:59.294434	...	FBD
14814	trip-153861106442901555	test	2018-10-03 23:57:44.429324	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	IND208006AAA	Kanpur_GovndNgr_DC (Uttar Pradesh)	IND208006AAA	Kanpur_GovndNgr_DC (Uttar Pradesh)	2018-10-04 02:51:27.075797	2018-10-04 02:51:27.075797	...	Kanpur
14815	trip-153861115439069069	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	IND627005AAA	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	IND628204AAA	Tirchchndr_Shnmgprml_D (Tamil Nadu)	2018-10-03 23:59:14.390954	2018-10-04 02:29:04.272194	...	Tirunelveli
14816	trip-153861118270144424	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	2018-10-04 03:58:40.726547	2018-10-04 03:58:40.726547	...	Sandur

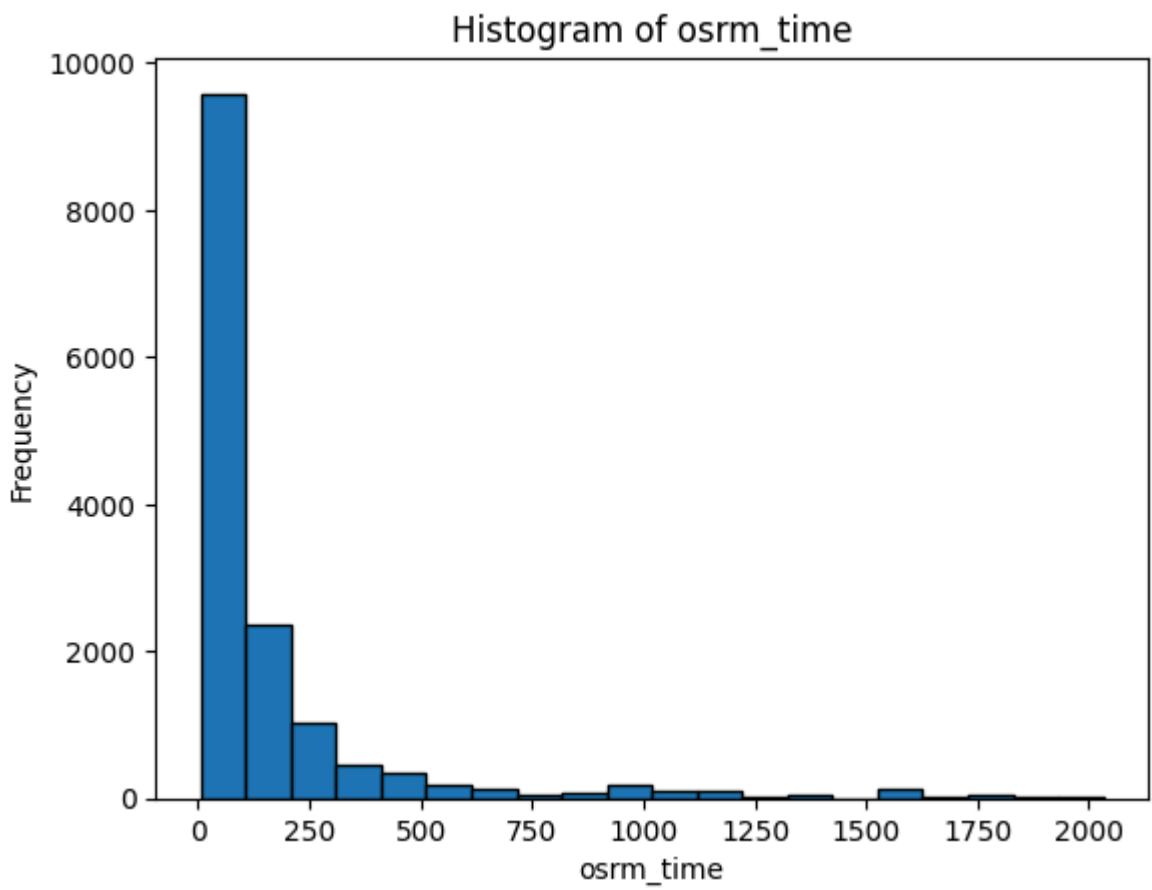
14817 rows × 34 columns

Hypothesis testing

```
In [123... from scipy.stats import ttest_ind
```

```
In [124... for column_name in ["actual_time", "osrm_time"]:
    # Create a histogram for each column using plt.histplot()
    plt.hist(df[column_name], bins=20, edgecolor='black')
    plt.xlabel(column_name) # Set the x-axis label
    plt.ylabel("Frequency") # Set the y-axis label
    plt.title(f"Histogram of {column_name}") # Set the plot title
    plt.show()
```





```
In [125... # Perform the two-sample t-test

#H0: No difference between actual_time and osrm_time
#sig Level=0.05

ttest_ind(df['actual_time'], df['osrm_time'])
```

Out[125... TtestResult(statistic=38.05620533303334, pvalue=1.934196442740037e-309, df=29632.0)

The p-value is less than the significance level of 0.05, which means we reject the null hypothesis. We can conclude that there is significant difference in the mean actual_time and osrm_time.

```
In [126... #H0: mean(actual_time) > mean(osrm_time)
ttest_ind(df['actual_time'], df['osrm_time'], alternative="greater")
```

Out[126... TtestResult(statistic=38.05620533303334, pvalue=9.6709822137002e-310, df=29632.0)

The p-value is less than the significance level of 0.05, which means we reject the null hypothesis. Mean(actual_time) < Mean(osrm_time) Estimated time between two points is greater then actual time.

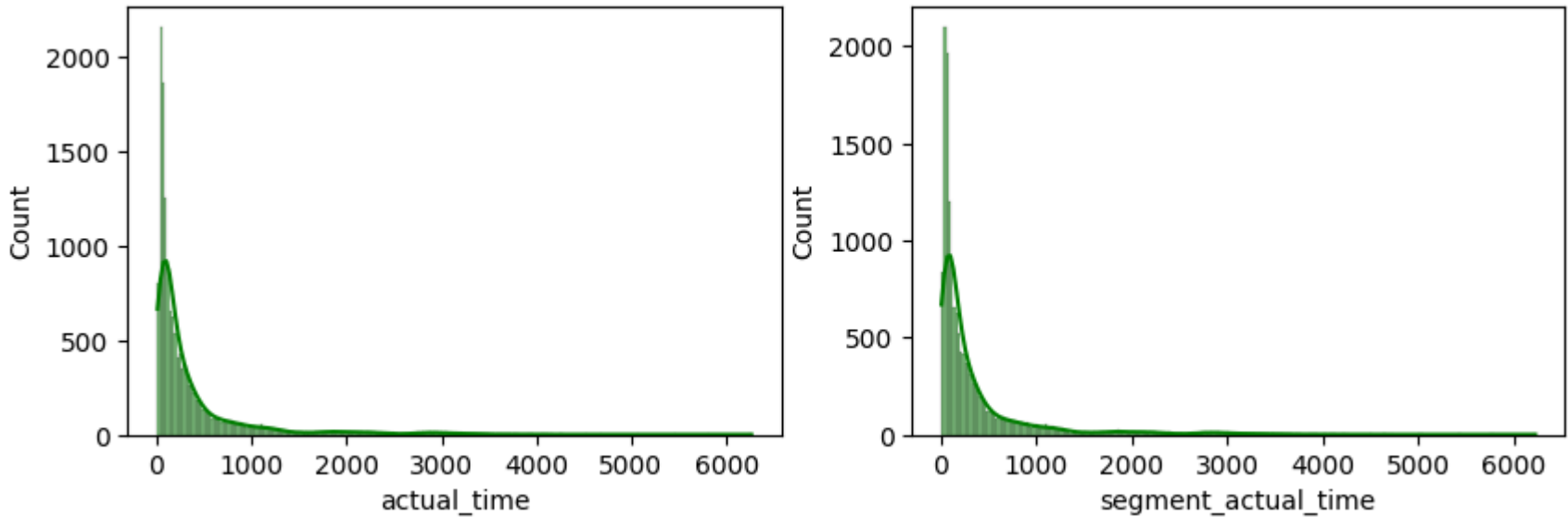
actual_time and segment_actual_time

```
In [127... plt.figure(figsize=(10,3))

plt.subplot(121)
sns.histplot(df['actual_time'], kde=True, color='green')

plt.subplot(122)
sns.histplot(df['segment_actual_time'], kde=True, color='green')

plt.show()
```



```
In [128... # Perform the two-sample t-test

#H0: No difference between actual_time and segment_actual_time

ttest_ind(df['actual_time'], df['segment_actual_time'])
```

Out[128... TtestResult(statistic=0.5008024728897531, pvalue=0.6165138648224772, df=29632.0)

The p-value is greater than the significance level of 0.05, which means we fail to reject the null hypothesis. We can conclude that there is no significant difference in the mean actual_time and cumulative segment_actual_time.

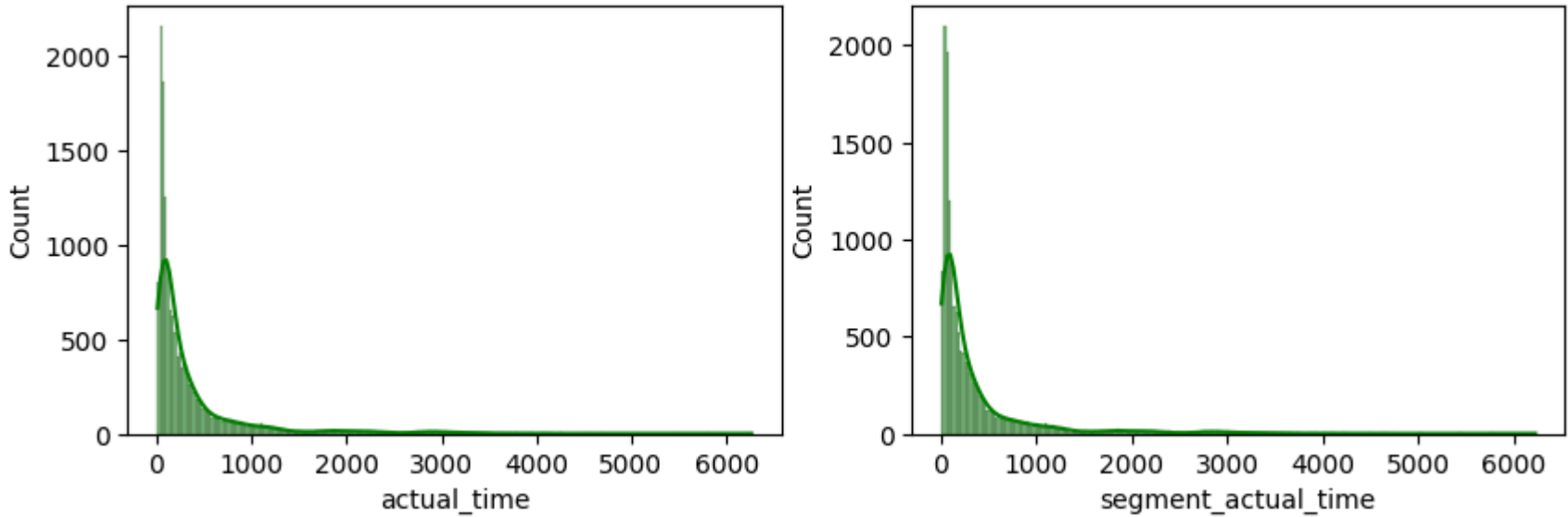
actual-time vs segment_actual_time

```
In [129... plt.figure(figsize=(10,3))

plt.subplot(121)
sns.histplot(df['actual_time'], kde=True, color='green')

plt.subplot(122)
sns.histplot(df['segment_actual_time'], kde=True, color='green')

plt.show()
```



```
In [130... # Perform the two-sample t-test

#H0: No difference between actual_time and segment_actual_time
```

```
ttest_ind(df['actual_time'], df['segment_actual_time'])
```

Out[130... TtestResult(statistic=0.5008024728897531, pvalue=0.6165138648224772, df=29632.0)

The p-value is greater than the significance level of 0.05, which means we fail to reject the null hypothesis. We can conclude that there is no significant difference in the mean actual_time and cumulative segment_actual_time

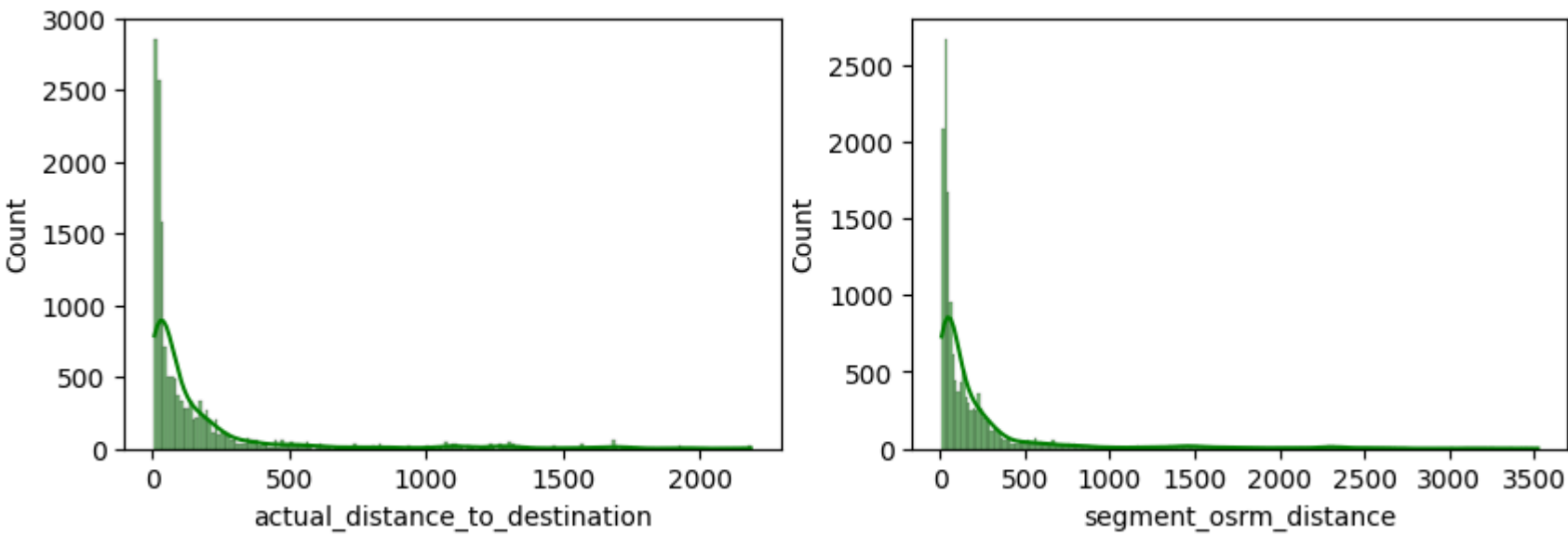
actual_distance_to_destination and segment_osrm_distance

```
In [131... plt.figure(figsize=(10,3))

plt.subplot(121)
sns.histplot(df['actual_distance_to_destination'], kde=True, color='green')

plt.subplot(122)
sns.histplot(df['segment_osrm_distance'], kde=True, color='green')

plt.show()
```



```
In [132... # Perform the two-sample t-test

#H0: No difference between cumulative actual_distance_to_destination and cumulative segment_osrm_distance

ttest_ind(df['actual_distance_to_destination'], df['segment_osrm_distance'])
```

Out[132... TtestResult(statistic=-13.786573906942445, pvalue=4.1718223728664176e-43, df=29632.0)

The p-value is less than the significance level of 0.05, which means we reject the null hypothesis. We can conclude that there is significant difference in the mean of cumulative actual_distance_to_destination and cumulative segment_osrm_distance.

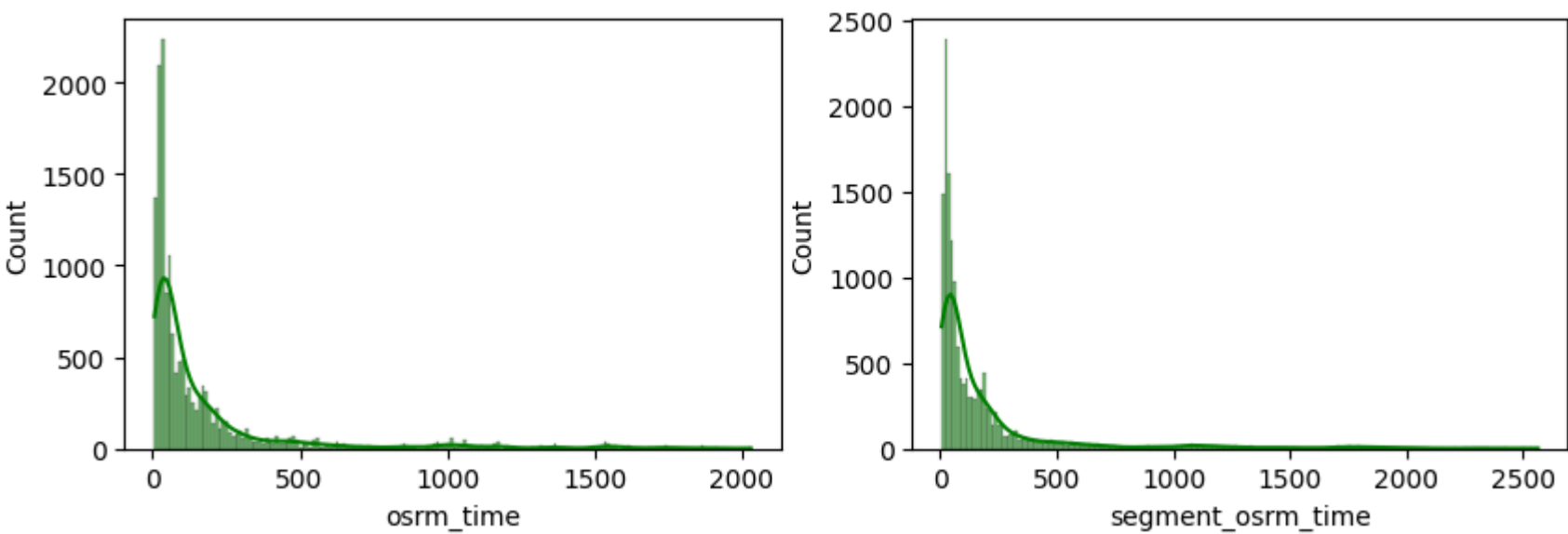
osrm_time and segment_osrm_time

```
In [133... plt.figure(figsize=(10,3))

plt.subplot(121)
sns.histplot(df['osrm_time'], kde=True, color='green')

plt.subplot(122)
sns.histplot(df['segment_osrm_time'], kde=True, color='green')

plt.show()
```



```
In [134... # Perform the two-sample t-test

#H0: No difference between osrm_distance and cumulated segment_osrm_distance

ttest_ind(df['osrm_time'], df['segment_osrm_time'])
```

Out[134... TtestResult(statistic=-5.52354601280892, pvalue=3.350128608408736e-08, df=29632.0)

The p-value is less than the significance level of 0.05, which means we reject the null hypothesis. We can conclude that there is significant difference in the mean osrm_time and cumulative segment_osrm_time.

Insights

- 1. Most Preferred Route Type: Carting shipments are used more frequently than Full Truck Load (FTL), accounting for 60.12% of deliveries. FTL shipments tend to take more time to reach destinations.
- 2)High-Delivery Locations:
 - Top 5 States for Deliveries: Maharashtra (2561) Karnataka (2294) Haryana (1643) Tamil Nadu (1084) Uttar Pradesh (811)
- Cities with Most Deliveries: Bengaluru, Karnataka (1088 deliveries) Mumbai, Maharashtra (966 deliveries) Gurgaon, Haryana (877 deliveries)
- 3. Peak Delivery Days: wednesdays see the highest number of deliveries (18.49% of total trips), followed by Thursdays (14.38%).
- 4. Time Efficiency in Different States:

Mizoram, Haryana, and Himachal Pradesh have a high variation between actual time and OSRM computed time, indicating delays. Dadra and Nagar Haveli and Tamil Nadu have almost equal actual and OSRM times, showing efficient routing.

5)Outlier Detection:

Significant outliers exist in actual time, actual distance to destination, and OSRM computed metrics, suggesting anomalies in route efficiency or data errors.

Recommendations

1)Optimize Carting vs. FTL Usage:

Since Carting shipments are used more, analyze the cost-benefit ratio of shifting certain deliveries to FTL for better efficiency.

2)Improve Route Planning for Delayed States:

Mizoram, Haryana, and Himachal Pradesh require better traffic prediction and alternative routing strategies to reduce high delivery delays. Enhance Efficiency on Peak Days (Wednesdays & Thursdays):

3)Increase fleet availability on these days to prevent bottlenecks. Implement dynamic pricing or incentives to distribute deliveries more evenly across the week. Outlier Management:

4. Further investigate anomalous trip times and distances to identify potential errors in GPS tracking or inefficient logistics. Use Machine Learning models to detect and address trip anomalies automatically. State-Specific Improvements:

5. otimize deliveries in high-density states (Maharashtra, Karnataka, Haryana) by leveraging warehouse and hub distribution strategies. Enhance infrastructure in underperforming states to match actual and computed delivery times.

In [135...

!jupyter nbconvert --to html /content/delhivery_project.ipynb

[NbConvertApp] Converting notebook /content/delhivery_project.ipynb to html

[NbConvertApp] WARNING | Alternative text is missing on 20 image(s).

[NbConvertApp] Writing 1689961 bytes to /content/delhivery_project.html