Delhivery_Business_CASE_

September 24, 2024

[6]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:100% !important; }</style>"))
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:100% !important; }</style>"))
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

[7]:
```
df=pd.read_csv('/content/sample_data/delhivery_data.csv')
df.head(20)
```

[7]:

	data	trip_creation_time	\
0	training	2018-09-20 02:35:36.476840	
1	training	2018-09-20 02:35:36.476840	
2	training	2018-09-20 02:35:36.476840	
3	training	2018-09-20 02:35:36.476840	
4	training	2018-09-20 02:35:36.476840	
5	training	2018-09-20 02:35:36.476840	
6	training	2018-09-20 02:35:36.476840	
7	training	2018-09-20 02:35:36.476840	
8	training	2018-09-20 02:35:36.476840	
9	training	2018-09-20 02:35:36.476840	
10	training	2018-09-23 06:42:06.021680	
11	training	2018-09-23 06:42:06.021680	
12	training	2018-09-23 06:42:06.021680	
13	training	2018-09-23 06:42:06.021680	
14	training	2018-09-23 06:42:06.021680	
15	training	2018-09-14 15:42:46.437249	
16	training	2018-09-14 15:42:46.437249	
17	training	2018-09-13 20:44:19.424489	
18	training	2018-09-13 20:44:19.424489	
19	training	2018-09-13 20:44:19.424489	

	route_schedule_uuid	route_type
0	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
1	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
2	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
3	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
4	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
5	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
6	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
7	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
8	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
9	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
10	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
11	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
12	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
13	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
14	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
15	thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d…	Carting
16	thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d…	Carting
17	thanos::sroute:76951383-1608-44e4-a284-46d92e8…	FTL
18	thanos::sroute:76951383-1608-44e4-a284-46d92e8…	FTL
19	thanos::sroute:76951383-1608-44e4-a284-46d92e8…	FTL

	trip_uuid	source_center	source_name
0	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)
1	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)
2	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)
3	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)
4	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)
5	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)
6	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)
7	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)
8	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)
9	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)
10	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)
11	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)
12	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)
13	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)
14	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)
15	trip-153693976643699843	IND400011AAA	LowerParel_CP (Maharashtra)
16	trip-153693976643699843	IND400011AAA	LowerParel_CP (Maharashtra)
17	trip-153687145942424248	IND562132AAA	Bangalore_Nelmngla_H (Karnataka)
18	trip-153687145942424248	IND562132AAA	Bangalore_Nelmngla_H (Karnataka)
19	trip-153687145942424248	IND560099AAB	Bengaluru_Bomsndra_HB (Karnataka)

	destination_center	destination_name
0	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)
1	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)

```
2        IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
3        IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
4        IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
5        IND388320AAA         Anand_Vaghasi_IP (Gujarat)
6        IND388320AAA         Anand_Vaghasi_IP (Gujarat)
7        IND388320AAA         Anand_Vaghasi_IP (Gujarat)
8        IND388320AAA         Anand_Vaghasi_IP (Gujarat)
9        IND388320AAA         Anand_Vaghasi_IP (Gujarat)
10       IND411033AAA      Pune_Tathawde_H (Maharashtra)
11       IND411033AAA      Pune_Tathawde_H (Maharashtra)
12       IND411033AAA      Pune_Tathawde_H (Maharashtra)
13       IND411033AAA      Pune_Tathawde_H (Maharashtra)
14       IND411033AAA      Pune_Tathawde_H (Maharashtra)
15       IND400072AAD   Mumbai_Chndivli_PC (Maharashtra)
16       IND400072AAD   Mumbai_Chndivli_PC (Maharashtra)
17       IND560099AAB  Bengaluru_Bomsndra_HB (Karnataka)
18       IND560099AAB  Bengaluru_Bomsndra_HB (Karnataka)
19       IND683511AAA           Aluva_Peedika_H (Kerala)

                 od_start_time  …            cutoff_timestamp  \
0   2018-09-20 03:21:32.418600  …         2018-09-20 04:27:55
1   2018-09-20 03:21:32.418600  …         2018-09-20 04:17:55
2   2018-09-20 03:21:32.418600  …  2018-09-20 04:01:19.505586
3   2018-09-20 03:21:32.418600  …         2018-09-20 03:39:57
4   2018-09-20 03:21:32.418600  …         2018-09-20 03:33:55
5   2018-09-20 04:47:45.236797  …         2018-09-20 06:15:58
6   2018-09-20 04:47:45.236797  …         2018-09-20 05:47:29
7   2018-09-20 04:47:45.236797  …         2018-09-20 05:25:58
8   2018-09-20 04:47:45.236797  …         2018-09-20 05:15:56
9   2018-09-20 04:47:45.236797  …         2018-09-20 04:49:20
10  2018-09-23 06:42:06.021680  …         2018-09-23 11:05:19
11  2018-09-23 06:42:06.021680  …         2018-09-23 10:27:22
12  2018-09-23 06:42:06.021680  …         2018-09-23 09:45:25
13  2018-09-23 06:42:06.021680  …         2018-09-23 09:21:27
14  2018-09-23 06:42:06.021680  …         2018-09-23 08:39:31
15  2018-09-14 15:42:46.437249  …         2018-09-14 16:29:54
16  2018-09-14 15:42:46.437249  …         2018-09-14 16:15:53
17  2018-09-13 20:44:19.424489  …         2018-09-13 23:25:20
18  2018-09-13 20:44:19.424489  …         2018-09-13 22:47:26
19  2018-09-13 23:59:56.061158  …         2018-09-14 12:45:25

    actual_distance_to_destination actual_time  osrm_time osrm_distance  \
0                        10.435660        14.0       11.0       11.9653
1                        18.936842        24.0       20.0       21.7243
2                        27.637279        40.0       28.0       32.5395
3                        36.118028        62.0       40.0       45.5620
4                        39.386040        68.0       44.0       54.2181
```

5	10.403038	15.0	11.0	12.1171
6	18.045481	44.0	17.0	21.2890
7	28.061896	65.0	29.0	35.8252
8	38.939167	76.0	39.0	47.1900
9	43.595802	102.0	45.0	53.2334
10	23.194334	38.0	24.0	26.8622
11	44.045659	76.0	41.0	54.4326
12	72.849327	117.0	68.0	89.6680
13	88.076599	141.0	80.0	108.3939
14	100.708423	183.0	95.0	129.3519
15	9.355852	46.0	11.0	11.4344
16	16.431273	60.0	16.0	18.7941
17	23.635811	30.0	30.0	28.9765
18	39.806036	67.0	53.0	52.1256
19	24.319864	50.0	24.0	29.7046

	factor	segment_actual_time	segment_osrm_time	segment_osrm_distance	\
0	1.272727	14.0	11.0	11.9653	
1	1.200000	10.0	9.0	9.7590	
2	1.428571	16.0	7.0	10.8152	
3	1.550000	21.0	12.0	13.0224	
4	1.545455	6.0	5.0	3.9153	
5	1.363636	15.0	11.0	12.1171	
6	2.588235	28.0	6.0	9.1719	
7	2.241379	21.0	11.0	14.5362	
8	1.948718	10.0	10.0	11.3648	
9	2.266667	26.0	6.0	6.0434	
10	1.583333	38.0	24.0	26.8622	
11	1.853659	37.0	27.0	30.1058	
12	1.720588	41.0	26.0	35.2353	
13	1.762500	23.0	14.0	17.2476	
14	1.926316	41.0	15.0	20.9580	
15	4.181818	46.0	11.0	11.4344	
16	3.750000	14.0	5.0	7.3597	
17	1.000000	30.0	30.0	28.9765	
18	1.264151	37.0	26.0	24.9545	
19	2.083333	50.0	24.0	29.7046	

	segment_factor
0	1.272727
1	1.111111
2	2.285714
3	1.750000
4	1.200000
5	1.363636
6	4.666667
7	1.909091

```
8         1.000000
9         4.333333
10        1.583333
11        1.370370
12        1.576923
13        1.642857
14        2.733333
15        4.181818
16        2.800000
17        1.000000
18        1.423077
19        2.083333

[20 rows x 24 columns]
```

[8]:
```
df[['route_schedule_uuid','route_type','trip_uuid','source_center','source_name',
   'destination_center','destination_name','od_start_time','od_end_time',
 ␣
↪'start_scan_to_end_scan','is_cutoff','actual_distance_to_destination','actual_time','segmen
↪head(20)
```

[8]:

	route_schedule_uuid	route_type
0	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
1	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
2	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
3	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
4	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
5	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
6	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
7	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
8	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
9	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…	Carting
10	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
11	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
12	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
13	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
14	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172…	FTL
15	thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d…	Carting
16	thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d…	Carting
17	thanos::sroute:76951383-1608-44e4-a284-46d92e8…	FTL
18	thanos::sroute:76951383-1608-44e4-a284-46d92e8…	FTL
19	thanos::sroute:76951383-1608-44e4-a284-46d92e8…	FTL

	trip_uuid	source_center	source_name
0	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)
1	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)
2	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)

```
3   trip-153741093647649320  IND388121AAA         Anand_VUNagar_DC (Gujarat)
4   trip-153741093647649320  IND388121AAA         Anand_VUNagar_DC (Gujarat)
5   trip-153741093647649320  IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
6   trip-153741093647649320  IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
7   trip-153741093647649320  IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
8   trip-153741093647649320  IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
9   trip-153741093647649320  IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
10  trip-153768492602129387  IND421302AAG  Bhiwandi_Mankoli_HB (Maharashtra)
11  trip-153768492602129387  IND421302AAG  Bhiwandi_Mankoli_HB (Maharashtra)
12  trip-153768492602129387  IND421302AAG  Bhiwandi_Mankoli_HB (Maharashtra)
13  trip-153768492602129387  IND421302AAG  Bhiwandi_Mankoli_HB (Maharashtra)
14  trip-153768492602129387  IND421302AAG  Bhiwandi_Mankoli_HB (Maharashtra)
15  trip-153693976643699843  IND400011AAA        LowerParel_CP (Maharashtra)
16  trip-153693976643699843  IND400011AAA        LowerParel_CP (Maharashtra)
17  trip-153687145942424248  IND562132AAA   Bangalore_Nelmngla_H (Karnataka)
18  trip-153687145942424248  IND562132AAA   Bangalore_Nelmngla_H (Karnataka)
19  trip-153687145942424248  IND560099AAB  Bengaluru_Bomsndra_HB (Karnataka)

   destination_center                  destination_name  \
0         IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
1         IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
2         IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
3         IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
4         IND388620AAB      Khambhat_MotvdDPP_D (Gujarat)
5         IND388320AAA         Anand_Vaghasi_IP (Gujarat)
6         IND388320AAA         Anand_Vaghasi_IP (Gujarat)
7         IND388320AAA         Anand_Vaghasi_IP (Gujarat)
8         IND388320AAA         Anand_Vaghasi_IP (Gujarat)
9         IND388320AAA         Anand_Vaghasi_IP (Gujarat)
10        IND411033AAA      Pune_Tathawde_H (Maharashtra)
11        IND411033AAA      Pune_Tathawde_H (Maharashtra)
12        IND411033AAA      Pune_Tathawde_H (Maharashtra)
13        IND411033AAA      Pune_Tathawde_H (Maharashtra)
14        IND411033AAA      Pune_Tathawde_H (Maharashtra)
15        IND400072AAD   Mumbai_Chndivli_PC (Maharashtra)
16        IND400072AAD   Mumbai_Chndivli_PC (Maharashtra)
17        IND560099AAB  Bengaluru_Bomsndra_HB (Karnataka)
18        IND560099AAB  Bengaluru_Bomsndra_HB (Karnataka)
19        IND683511AAA           Aluva_Peedika_H (Kerala)

                 od_start_time                 od_end_time  \
0   2018-09-20 03:21:32.418600  2018-09-20 04:47:45.236797
1   2018-09-20 03:21:32.418600  2018-09-20 04:47:45.236797
2   2018-09-20 03:21:32.418600  2018-09-20 04:47:45.236797
3   2018-09-20 03:21:32.418600  2018-09-20 04:47:45.236797
4   2018-09-20 03:21:32.418600  2018-09-20 04:47:45.236797
5   2018-09-20 04:47:45.236797  2018-09-20 06:36:55.627764
```

```
6   2018-09-20 04:47:45.236797  2018-09-20 06:36:55.627764
7   2018-09-20 04:47:45.236797  2018-09-20 06:36:55.627764
8   2018-09-20 04:47:45.236797  2018-09-20 06:36:55.627764
9   2018-09-20 04:47:45.236797  2018-09-20 06:36:55.627764
10  2018-09-23 06:42:06.021680  2018-09-23 11:44:28.365845
11  2018-09-23 06:42:06.021680  2018-09-23 11:44:28.365845
12  2018-09-23 06:42:06.021680  2018-09-23 11:44:28.365845
13  2018-09-23 06:42:06.021680  2018-09-23 11:44:28.365845
14  2018-09-23 06:42:06.021680  2018-09-23 11:44:28.365845
15  2018-09-14 15:42:46.437249  2018-09-14 17:31:45.368791
16  2018-09-14 15:42:46.437249  2018-09-14 17:31:45.368791
17  2018-09-13 20:44:19.424489  2018-09-13 23:59:56.061158
18  2018-09-13 20:44:19.424489  2018-09-13 23:59:56.061158
19  2018-09-13 23:59:56.061158  2018-09-14 13:55:58.765334

    start_scan_to_end_scan is_cutoff  actual_distance_to_destination  \
0                     86.0      True                       10.435660
1                     86.0      True                       18.936842
2                     86.0      True                       27.637279
3                     86.0      True                       36.118028
4                     86.0     False                       39.386040
5                    109.0      True                       10.403038
6                    109.0      True                       18.045481
7                    109.0      True                       28.061896
8                    109.0      True                       38.939167
9                    109.0     False                       43.595802
10                   302.0      True                       23.194334
11                   302.0      True                       44.045659
12                   302.0      True                       72.849327
13                   302.0      True                       88.076599
14                   302.0     False                      100.708423
15                   108.0      True                        9.355852
16                   108.0     False                       16.431273
17                   195.0      True                       23.635811
18                   195.0     False                       39.806036
19                   836.0      True                       24.319864

    actual_time  segment_osrm_time  segment_osrm_distance
0          14.0               11.0                11.9653
1          24.0                9.0                 9.7590
2          40.0                7.0                10.8152
3          62.0               12.0                13.0224
4          68.0                5.0                 3.9153
5          15.0               11.0                12.1171
6          44.0                6.0                 9.1719
7          65.0               11.0                14.5362
8          76.0               10.0                11.3648
```

```
9         102.0                6.0                 6.0434
10         38.0               24.0                26.8622
11         76.0               27.0                30.1058
12        117.0               26.0                35.2353
13        141.0               14.0                17.2476
14        183.0               15.0                20.9580
15         46.0               11.0                11.4344
16         60.0                5.0                 7.3597
17         30.0               30.0                28.9765
18         67.0               26.0                24.9545
19         50.0               24.0                29.7046
```

[9]:
```
sample=df[(df['route_schedule_uuid']=='thanos::sroute:
 ↪eb7bfc78-b351-4c0e-a951-fa3d5c3297ef') &␣
 ↪(df['trip_uuid']=='trip-153741093647649320')]
sample[['source_name','destination_name','od_start_time','od_start_time','od_end_time','start_
```

[9]:
```
                    source_name               destination_name  \
0     Anand_VUNagar_DC (Gujarat)  Khambhat_MotvdDPP_D (Gujarat)
1     Anand_VUNagar_DC (Gujarat)  Khambhat_MotvdDPP_D (Gujarat)
2     Anand_VUNagar_DC (Gujarat)  Khambhat_MotvdDPP_D (Gujarat)
3     Anand_VUNagar_DC (Gujarat)  Khambhat_MotvdDPP_D (Gujarat)
4     Anand_VUNagar_DC (Gujarat)  Khambhat_MotvdDPP_D (Gujarat)
5  Khambhat_MotvdDPP_D (Gujarat)     Anand_Vaghasi_IP (Gujarat)
6  Khambhat_MotvdDPP_D (Gujarat)     Anand_Vaghasi_IP (Gujarat)
7  Khambhat_MotvdDPP_D (Gujarat)     Anand_Vaghasi_IP (Gujarat)
8  Khambhat_MotvdDPP_D (Gujarat)     Anand_Vaghasi_IP (Gujarat)
9  Khambhat_MotvdDPP_D (Gujarat)     Anand_Vaghasi_IP (Gujarat)

                od_start_time               od_start_time  \
0  2018-09-20 03:21:32.418600  2018-09-20 03:21:32.418600
1  2018-09-20 03:21:32.418600  2018-09-20 03:21:32.418600
2  2018-09-20 03:21:32.418600  2018-09-20 03:21:32.418600
3  2018-09-20 03:21:32.418600  2018-09-20 03:21:32.418600
4  2018-09-20 03:21:32.418600  2018-09-20 03:21:32.418600
5  2018-09-20 04:47:45.236797  2018-09-20 04:47:45.236797
6  2018-09-20 04:47:45.236797  2018-09-20 04:47:45.236797
7  2018-09-20 04:47:45.236797  2018-09-20 04:47:45.236797
8  2018-09-20 04:47:45.236797  2018-09-20 04:47:45.236797
9  2018-09-20 04:47:45.236797  2018-09-20 04:47:45.236797

                 od_end_time  start_scan_to_end_scan is_cutoff  \
0  2018-09-20 04:47:45.236797                    86.0      True
1  2018-09-20 04:47:45.236797                    86.0      True
2  2018-09-20 04:47:45.236797                    86.0      True
3  2018-09-20 04:47:45.236797                    86.0      True
4  2018-09-20 04:47:45.236797                    86.0     False
```

```
5  2018-09-20 06:36:55.627764                  109.0      True
6  2018-09-20 06:36:55.627764                  109.0      True
7  2018-09-20 06:36:55.627764                  109.0      True
8  2018-09-20 06:36:55.627764                  109.0      True
9  2018-09-20 06:36:55.627764                  109.0     False

   actual_distance_to_destination  actual_time  segment_osrm_time  \
0                       10.435660         14.0               11.0
1                       18.936842         24.0                9.0
2                       27.637279         40.0                7.0
3                       36.118028         62.0               12.0
4                       39.386040         68.0                5.0
5                       10.403038         15.0               11.0
6                       18.045481         44.0                6.0
7                       28.061896         65.0               11.0
8                       38.939167         76.0               10.0
9                       43.595802        102.0                6.0

   segment_osrm_distance
0                11.9653
1                 9.7590
2                10.8152
3                13.0224
4                 3.9153
5                12.1171
6                 9.1719
7                14.5362
8                11.3648
9                 6.0434
```

[10]:
```
df.shape
```

[10]:
```
(27323, 24)
```

[11]:
```
df.dtypes
```

[11]:
```
data                            object
trip_creation_time              object
route_schedule_uuid             object
route_type                      object
trip_uuid                       object
source_center                   object
source_name                     object
destination_center              object
destination_name                object
od_start_time                   object
od_end_time                     object
```

```
start_scan_to_end_scan            float64
is_cutoff                          object
cutoff_factor                     float64
cutoff_timestamp                   object
actual_distance_to_destination    float64
actual_time                       float64
osrm_time                         float64
osrm_distance                     float64
factor                            float64
segment_actual_time               float64
segment_osrm_time                 float64
segment_osrm_distance             float64
segment_factor                    float64
dtype: object
```

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27323 entries, 0 to 27322
Data columns (total 24 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   data                            27323 non-null  object
 1   trip_creation_time              27323 non-null  object
 2   route_schedule_uuid             27323 non-null  object
 3   route_type                      27323 non-null  object
 4   trip_uuid                       27323 non-null  object
 5   source_center                   27323 non-null  object
 6   source_name                     27244 non-null  object
 7   destination_center              27322 non-null  object
 8   destination_name                27267 non-null  object
 9   od_start_time                   27322 non-null  object
 10  od_end_time                     27322 non-null  object
 11  start_scan_to_end_scan          27322 non-null  float64
 12  is_cutoff                       27322 non-null  object
 13  cutoff_factor                   27322 non-null  float64
 14  cutoff_timestamp                27322 non-null  object
 15  actual_distance_to_destination  27322 non-null  float64
 16  actual_time                     27322 non-null  float64
 17  osrm_time                       27322 non-null  float64
 18  osrm_distance                   27322 non-null  float64
 19  factor                          27322 non-null  float64
 20  segment_actual_time             27322 non-null  float64
 21  segment_osrm_time               27322 non-null  float64
 22  segment_osrm_distance           27322 non-null  float64
 23  segment_factor                  27322 non-null  float64
dtypes: float64(11), object(13)
```

```
memory usage: 5.0+ MB
```

[13]:
```
df.describe(include=object)
```

[13]:
```
            data          trip_creation_time  \
count      27323                       27323
unique         2                        2852
top     training  2018-09-28 05:23:15.359220
freq       19801                         101

                                   route_schedule_uuid route_type  \
count                                            27323      27323
unique                                            1102          2
top     thanos::sroute:0456b740-1dad-4929-bbe0-87d8843…        FTL
freq                                               462      18559

                     trip_uuid source_center                    source_name  \
count                    27323         27323                          27244
unique                    2852          1169                           1164
top     trip-153811219535896559  IND000000ACB  Gurgaon_Bilaspur_HB (Haryana)
freq                       101          4699                           4699

       destination_center              destination_name  \
count               27322                         27267
unique               1155                          1148
top          IND000000ACB  Gurgaon_Bilaspur_HB (Haryana)
freq                 2231                          2231

                     od_start_time                 od_end_time is_cutoff  \
count                        27322                       27322     27322
unique                        5188                        5188         2
top     2018-10-03 04:55:30.039225  2018-10-05 11:15:01.115906      True
freq                            79                          79     22183

           cutoff_timestamp
count                 27322
unique                23824
top     2018-09-24 05:19:20
freq                      8
```

[14]:
```
df.head()
```

[14]:
```
       data          trip_creation_time  \
0  training  2018-09-20 02:35:36.476840
1  training  2018-09-20 02:35:36.476840
2  training  2018-09-20 02:35:36.476840
3  training  2018-09-20 02:35:36.476840
```

```
4  training  2018-09-20 02:35:36.476840

                                route_schedule_uuid route_type  \
0  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
1  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
2  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
3  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
4  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting

                trip_uuid source_center                source_name  \
0  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
1  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
2  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
3  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
4  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)

  destination_center              destination_name  \
0       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
1       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
2       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
3       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
4       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)

               od_start_time  …            cutoff_timestamp  \
0  2018-09-20 03:21:32.418600  …         2018-09-20 04:27:55
1  2018-09-20 03:21:32.418600  …         2018-09-20 04:17:55
2  2018-09-20 03:21:32.418600  …  2018-09-20 04:01:19.505586
3  2018-09-20 03:21:32.418600  …         2018-09-20 03:39:57
4  2018-09-20 03:21:32.418600  …         2018-09-20 03:33:55

   actual_distance_to_destination actual_time  osrm_time osrm_distance  \
0                       10.435660        14.0       11.0       11.9653
1                       18.936842        24.0       20.0       21.7243
2                       27.637279        40.0       28.0       32.5395
3                       36.118028        62.0       40.0       45.5620
4                       39.386040        68.0       44.0       54.2181

     factor  segment_actual_time  segment_osrm_time  segment_osrm_distance  \
0  1.272727                 14.0               11.0                11.9653
1  1.200000                 10.0                9.0                 9.7590
2  1.428571                 16.0                7.0                10.8152
3  1.550000                 21.0               12.0                13.0224
4  1.545455                  6.0                5.0                 3.9153

   segment_factor
0        1.272727
1        1.111111
```

```
2        2.285714
3        1.750000
4        1.200000

[5 rows x 24 columns]
```

Datetime Formatting for datetime columns

[15]:
```
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'], format='mixed')
```

[16]:
```
df.head()
```

[16]:
```
       data         trip_creation_time  \
0  training 2018-09-20 02:35:36.476840
1  training 2018-09-20 02:35:36.476840
2  training 2018-09-20 02:35:36.476840
3  training 2018-09-20 02:35:36.476840
4  training 2018-09-20 02:35:36.476840

                               route_schedule_uuid route_type  \
0  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
1  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
2  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
3  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting
4  thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3…    Carting

                 trip_uuid source_center                source_name  \
0  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
1  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
2  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
3  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)
4  trip-153741093647649320  IND388121AAA  Anand_VUNagar_DC (Gujarat)

  destination_center             destination_name  \
0       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
1       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
2       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
3       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)
4       IND388620AAB  Khambhat_MotvdDPP_D (Gujarat)

              od_start_time  …           cutoff_timestamp  \
0 2018-09-20 03:21:32.418600  … 2018-09-20 04:27:55.000000
1 2018-09-20 03:21:32.418600  … 2018-09-20 04:17:55.000000
2 2018-09-20 03:21:32.418600  … 2018-09-20 04:01:19.505586
```

```
3 2018-09-20 03:21:32.418600  … 2018-09-20 03:39:57.000000
4 2018-09-20 03:21:32.418600  … 2018-09-20 03:33:55.000000

   actual_distance_to_destination actual_time  osrm_time osrm_distance  \
0                       10.435660        14.0       11.0       11.9653
1                       18.936842        24.0       20.0       21.7243
2                       27.637279        40.0       28.0       32.5395
3                       36.118028        62.0       40.0       45.5620
4                       39.386040        68.0       44.0       54.2181

     factor  segment_actual_time  segment_osrm_time  segment_osrm_distance  \
0  1.272727                 14.0               11.0                11.9653
1  1.200000                 10.0                9.0                 9.7590
2  1.428571                 16.0                7.0                10.8152
3  1.550000                 21.0               12.0                13.0224
4  1.545455                  6.0                5.0                 3.9153

   segment_factor
0        1.272727
1        1.111111
2        2.285714
3        1.750000
4        1.200000

[5 rows x 24 columns]
```

Missing Value Analysis

[17]:
```
df.isna().sum()
```

[17]:
```
data                            0
trip_creation_time              0
route_schedule_uuid             0
route_type                      0
trip_uuid                       0
source_center                   0
source_name                    79
destination_center              1
destination_name               56
od_start_time                   1
od_end_time                     1
start_scan_to_end_scan          1
is_cutoff                       1
cutoff_factor                   1
cutoff_timestamp                1
actual_distance_to_destination  1
actual_time                     1
```

```
osrm_time                       1
osrm_distance                   1
factor                          1
segment_actual_time             1
segment_osrm_time               1
segment_osrm_distance           1
segment_factor                  1
dtype: int64
```

Observations¶ Source name and Destination name have missing values as we have very records having missing value, we can drop those rows from analysis

[18]:
```
# drop missing values
df.drop(df[df['source_name'].isna() | df['destination_name'].isna()].index ,␣
 ↪inplace=True)
```

[19]:
```
df.isna().sum()
```

[19]:
```
data                            0
trip_creation_time              0
route_schedule_uuid             0
route_type                      0
trip_uuid                       0
source_center                   0
source_name                     0
destination_center              0
destination_name                0
od_start_time                   0
od_end_time                     0
start_scan_to_end_scan          0
is_cutoff                       0
cutoff_factor                   0
cutoff_timestamp                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
segment_factor                  0
dtype: int64
```

2. Trip Segment Analysis delivery details of one package is divided into several rows (think of it as connecting flights to reach a particular destination) we should find way to identify unique trip details for source_destination to understand delivery data. we can create unique identifier for different segment of a trip based on combination of trip_uuid, source_center

and destination_center, we will call it as segment key

Based on segment_key we will create new aggregated columns segment_actual_time,segment_osrm_distance,segment_osrm_time

[20]:
```
df['segment_key']=df.apply(lambda x:"#".
 ↪join([x['trip_uuid'],x['source_center'],x['destination_center']]), axis=1)
```

[21]:
```
df['segment_actual_time_cumsum']=df.
 ↪groupby('segment_key')['segment_actual_time'].transform(lambda x:x.cumsum())

df['segment_osrm_time_cumsum']=df.groupby('segment_key')['segment_osrm_time'].
 ↪transform(lambda x:x.cumsum())

df['segment_osrm_distance_cumsum']=df.
 ↪groupby('segment_key')['segment_osrm_distance'].transform(lambda x:x.
 ↪cumsum())
```

[22]:
```
# df.groupby('segment_key')['segment_actual_time_sum'].nth(-1)
```

[23]:
```
df['segment_actual_time_sum']=df.
 ↪groupby('segment_key')['segment_actual_time_cumsum'].transform(lambda x:x.
 ↪iloc[-1])
df['segment_osrm_time_sum']=df.
 ↪groupby('segment_key')['segment_osrm_time_cumsum'].transform(lambda x:x.
 ↪iloc[-1])
df['segment_osrm_distance_sum']=df.
 ↪groupby('segment_key')['segment_osrm_distance_cumsum'].transform(lambda x:x.
 ↪iloc[-1])
```

[24]:
```
df=df.sort_values(by=['segment_key','od_end_time'])
df
```

[24]:

```
           data         trip_creation_time  \
21691  training 2018-09-12 00:06:39.565253
21692  training 2018-09-12 00:06:39.565253
20850  training 2018-09-12 00:11:40.783923
1546   training 2018-09-12 00:25:19.499696
1547   training 2018-09-12 00:25:19.499696
…           …                         …
18729      test 2018-10-03 23:57:44.429324
11572      test 2018-10-03 23:59:42.701692
11573      test 2018-10-03 23:59:42.701692
11570      test 2018-10-03 23:59:42.701692
11571      test 2018-10-03 23:59:42.701692

                                      route_schedule_uuid route_type  \
21691  thanos::sroute:a10888ff-f794-41e1-9b7a-7f62ef6…    Carting
```

```
21692  thanos::sroute:a10888ff-f794-41e1-9b7a-7f62ef6…    Carting
20850  thanos::sroute:c2ee580f-f4b2-4fa5-98ab-0c5b327…    Carting
1546   thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678…        FTL
1547   thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678…        FTL
…                                                    …       …
18729  thanos::sroute:5609c268-e436-4e0a-8180-3db4a74…    Carting
11572  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
11573  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
11570  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
11571  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL

                    trip_uuid source_center  \
21691  trip-153671079956500691  IND110024AAA
21692  trip-153671079956500691  IND110024AAA
20850  trip-153671110078355292  IND121004AAB
1546   trip-153671191949943656  IND487001AAB
1547   trip-153671191949943656  IND487001AAB
…                          …           …
18729  trip-153861106442901555  IND209304AAA
11572  trip-153861118270144424  IND583119AAA
11573  trip-153861118270144424  IND583119AAA
11570  trip-153861118270144424  IND583201AAA
11571  trip-153861118270144424  IND583201AAA

                                source_name destination_center  \
21691                 Delhi_Lajpat_IP (Delhi)       IND110014AAA
21692                 Delhi_Lajpat_IP (Delhi)       IND110014AAA
20850            FBD_Balabhgarh_DPC (Haryana)       IND121001AAA
1546   Narsinghpur_KndliDPP_D (Madhya Pradesh)       IND487551AAA
1547   Narsinghpur_KndliDPP_D (Madhya Pradesh)       IND487551AAA
…                                          …              …
18729      Kanpur_Central_H_6 (Uttar Pradesh)       IND208006AAA
11572           Sandur_WrdN1DPP_D (Karnataka)       IND583101AAA
11573           Sandur_WrdN1DPP_D (Karnataka)       IND583101AAA
11570                      Hospet (Karnataka)       IND583119AAA
11571                      Hospet (Karnataka)       IND583119AAA

                        destination_name             od_start_time  …  \
21691                Delhi_Bhogal (Delhi) 2018-09-12 00:06:39.565253  …
21692                Delhi_Bhogal (Delhi) 2018-09-12 00:06:39.565253  …
20850                 Faridabad (Haryana) 2018-09-12 00:11:40.783923  …
1546   Gadarwara_MPward_D (Madhya Pradesh) 2018-09-12 04:33:09.726898  …
1547   Gadarwara_MPward_D (Madhya Pradesh) 2018-09-12 04:33:09.726898  …
…                                     …                        …  …
18729  Kanpur_GovndNgr_DC (Uttar Pradesh) 2018-10-03 23:57:44.429324  …
11572              Bellary_Dc (Karnataka) 2018-10-04 03:58:40.726547  …
11573              Bellary_Dc (Karnataka) 2018-10-04 03:58:40.726547  …
```

```
11570        Sandur_WrdN1DPP_D (Karnataka) 2018-10-04 02:51:44.712656  …
11571        Sandur_WrdN1DPP_D (Karnataka) 2018-10-04 02:51:44.712656  …

      segment_osrm_time  segment_osrm_distance segment_factor  \
21691               8.0                10.0040       1.500000
21692               6.0                 6.0820       1.833333
20850               9.0                10.8159       1.888889
1546               25.0                27.9776       1.920000
1547               16.0                20.4353       1.875000
…                  …                   …             …
18729               0.0                 0.1847      -1.000000
11572              17.0                23.9866       2.647059
11573              25.0                28.5437       7.520000
11570              21.0                23.2367       1.428571
11571               4.0                 4.8117       2.750000

                                                segment_key  \
21691  trip-15367107995650069 1#IND110024AAA#IND110014AAA
21692  trip-15367107995650069 1#IND110024AAA#IND110014AAA
20850  trip-15367111007835529 2#IND121004AAB#IND121001AAA
1546   trip-15367119194994365 6#IND487001AAB#IND487551AAA
1547   trip-15367119194994365 6#IND487001AAB#IND487551AAA
…                                                      …
18729  trip-15386110644290155 5#IND209304AAA#IND208006AAA
11572  trip-15386111827014442 4#IND583119AAA#IND583101AAA
11573  trip-15386111827014442 4#IND583119AAA#IND583101AAA
11570  trip-15386111827014442 4#IND583201AAA#IND583119AAA
11571  trip-15386111827014442 4#IND583201AAA#IND583119AAA

      segment_actual_time_cumsum  segment_osrm_time_cumsum  \
21691                       12.0                       8.0
21692                       23.0                      14.0
20850                       17.0                       9.0
1546                        48.0                      25.0
1547                        78.0                      41.0
…                           …                         …
18729                       91.0                      52.0
11572                       45.0                      17.0
11573                      233.0                      42.0
11570                       30.0                      21.0
11571                       41.0                      25.0

       segment_osrm_distance_cumsum  segment_actual_time_sum  \
21691                       10.0040                     23.0
21692                       16.0860                     23.0
20850                       10.8159                     17.0
1546                        27.9776                     78.0
```

```
1547                          48.4129                      78.0
…                               …                          …
18729                         69.4497                      91.0
11572                         23.9866                     233.0
11573                         52.5303                     233.0
11570                         23.2367                      41.0
11571                         28.0484                      41.0

       segment_osrm_time_sum  segment_osrm_distance_sum
21691                   14.0                    16.0860
21692                   14.0                    16.0860
20850                    9.0                    10.8159
1546                    41.0                    48.4129
1547                    41.0                    48.4129
…                          …                          …
18729                   52.0                    69.4497
11572                   42.0                    52.5303
11573                   42.0                    52.5303
11570                   25.0                    28.0484
11571                   25.0                    28.0484

[27188 rows x 31 columns]
```

3. Feature Engineering

[25]:
```
df['od_time_diff_hour']=(df['od_end_time']-df['od_start_time'])/pd.
↪Timedelta(hours=1)
```

Extract City, Place, Code and State Information¶

[26]:
```
df.columns
```

[26]:
```
Index(['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_scan_to_end_scan', 'is_cutoff', 'cutoff_factor',
       'cutoff_timestamp', 'actual_distance_to_destination', 'actual_time',
       'osrm_time', 'osrm_distance', 'factor', 'segment_actual_time',
       'segment_osrm_time', 'segment_osrm_distance', 'segment_factor',
       'segment_key', 'segment_actual_time_cumsum', 'segment_osrm_time_cumsum',
       'segment_osrm_distance_cumsum', 'segment_actual_time_sum',
       'segment_osrm_time_sum', 'segment_osrm_distance_sum',
       'od_time_diff_hour'],
      dtype='object')
```

[27]:
```
df['trip_creation_year']=df['trip_creation_time'].dt.year
df['trip_creation_month']=df['trip_creation_time'].dt.month
df['trip_creation_day']=df['trip_creation_time'].dt.day
```

[28]:
```
def get_state(name):
    pattern="\([A-Za-z]+\s?\w+\)"
    pattern="\([A-Za-z &]+\s?\w+\)"
    state=re.findall(pattern, name)[0]
    state=state.replace("(","")
    state=state.replace(")","")
    return state
```

[29]:
```
def get_city(name):
    pattern="\([A-Za-z &]+\s?\w+\)"
    state=re.findall(pattern, name)[0]
    city_place_code=name.replace(state,'')
    city_place_code_parts=city_place_code.split("_")
    if len(city_place_code_parts)==1 :
        city=city_place_code_parts[0].strip()
    elif len(city_place_code_parts)==2:
        city=city_place_code.strip()
    elif len(city_place_code_parts)==3 or len(city_place_code_parts)==4:
        city=city_place_code_parts[0].strip()
    else:
        city=city_place_code
    return city
```

[34]:
```
def get_place(name):
    pattern="\([A-Za-z &]+\s?\w+\)"
    try:
        state=re.findall(pattern, name)[0]
        city_place_code=name.replace(state,'')
        city_place_code_parts=city_place_code.split("_")
        if len(city_place_code_parts)==3 or len(city_place_code_parts)==4:
            place=city_place_code_parts[1].strip()
        else:
            place=None
        return place
    except Exception as exp:
        return None
```

[30]:
```
def get_code(name):
    pattern="\([A-Za-z &]+\s?\w+\)"
    try:
        state=re.findall(pattern, name)[0]
        city_place_code=name.replace(state,'')
        city_place_code_parts=city_place_code.split("_")
        if len(city_place_code_parts)==3 :
            code=city_place_code_parts[2].strip()
        elif  len(city_place_code_parts)==4:
            code="_".join(city_place_code_parts[2:]).strip()
```

```
        else:
            code=None
        return code
    except Exception as exp:
        return None
```

[35]:
```
from time import time
import re
```

[36]:
```
df['destination_state']=df['destination_name'].map(get_state)
df['source_state']=df['source_name'].map(get_state)
df['destination_city']=df['destination_name'].map(get_city)
df['source_city']=df['source_name'].map(get_state)
df['destination_place']=df['destination_name'].map(get_place)
df['source_place']=df['source_name'].map(get_place)
df['destination_code']=df['destination_name'].map(get_code)
df['source_code']=df['source_name'].map(get_code)
```

4. In-depth analysis: Grouping and Aggregating at Trip-level¶

a. Groups the segment data by the trip_uuid column to focus on aggregating data at the trip level.

b. Apply suitable aggregation functions like first, last, and sum specified in the create_trip_dict dictionary to calculate summary statistics for each trip.

[37]:
```
trip_actual_time_sum_df=df.groupby(['trip_uuid','segment_key']).nth(-1).
↪groupby(['trip_uuid']).agg(trip_actual_time_sum=('actual_time','sum')).
↪reset_index()
trip_actual_time_sum_df
```

[37]:

	trip_uuid	trip_actual_time_sum
0	trip-153671079956500691	23.0
1	trip-153671110078355292	17.0
2	trip-153671191949943656	253.0
3	trip-153671237597058150	241.0
4	trip-153671262893947351	195.0
…	…	…
2840	trip-153861034802474617	123.0
2841	trip-153861052318770017	47.0
2842	trip-153861089872028474	62.0
2843	trip-153861106442901555	282.0
2844	trip-153861118270144424	275.0

[2845 rows x 2 columns]

[38]:
```

```

```
trip_segment_actual_time_sum_df=df.groupby(['trip_uuid','segment_key']).nth(-1).
 ↪groupby(['trip_uuid']).
 ↪agg(trip_segment_actual_time_sum=('segment_actual_time_sum','sum')).
 ↪reset_index()
trip_segment_actual_time_sum_df
```

[38]:

	trip_uuid	trip_segment_actual_time_sum
0	trip-153671079956500691	23.0
1	trip-153671110078355292	17.0
2	trip-153671191949943656	252.0
3	trip-153671237597058150	240.0
4	trip-153671262893947351	195.0
…	…	…
2840	trip-153861034802474617	121.0
2841	trip-153861052318770017	47.0
2842	trip-153861089872028474	61.0
2843	trip-153861106442901555	281.0
2844	trip-153861118270144424	274.0

[2845 rows x 2 columns]

[39]:
```
trip_segment_osrm_time_sum_df=df.groupby(['trip_uuid','segment_key']).nth(-1).
 ↪groupby(['trip_uuid']).
 ↪agg(trip_segment_osrm_time_sum=('segment_osrm_time_sum','sum')).reset_index()
trip_segment_osrm_time_sum_df
```

[39]:

	trip_uuid	trip_segment_osrm_time_sum
0	trip-153671079956500691	14.0
1	trip-153671110078355292	9.0
2	trip-153671191949943656	132.0
3	trip-153671237597058150	49.0
4	trip-153671262893947351	26.0
…	…	…
2840	trip-153861034802474617	50.0
2841	trip-153861052318770017	13.0
2842	trip-153861089872028474	33.0
2843	trip-153861106442901555	88.0
2844	trip-153861118270144424	67.0

[2845 rows x 2 columns]

[40]:
```
trip_osrm_time_sum_df=df.groupby(['trip_uuid','segment_key']).nth(-1).
 ↪groupby(['trip_uuid']).agg(trip_osrm_time_sum=('osrm_time','sum')).
 ↪reset_index()
trip_osrm_time_sum_df
```

[40]:
```
                    trip_uuid  trip_osrm_time_sum
0     trip-153671079956500691                 8.0
1     trip-153671110078355292                 9.0
2     trip-153671191949943656               105.0
3     trip-153671237597058150                33.0
4     trip-153671262893947351                26.0
…                          …                   …
2840  trip-153861034802474617                50.0
2841  trip-153861052318770017                13.0
2842  trip-153861089872028474                28.0
2843  trip-153861106442901555                48.0
2844  trip-153861118270144424                68.0

[2845 rows x 2 columns]
```

[41]:
```
trip_osrm_distance_sum_df=df.groupby(['trip_uuid','segment_key']).nth(-1).
 ↪groupby(['trip_uuid']).agg(trip_osrm_distance_sum=('osrm_distance','sum')).
 ↪reset_index()
trip_osrm_distance_sum_df
```

[41]:
```
                    trip_uuid  trip_osrm_distance_sum
0     trip-153671079956500691                  9.9566
1     trip-153671110078355292                 10.8159
2     trip-153671191949943656                124.5063
3     trip-153671237597058150                 46.9087
4     trip-153671262893947351                 30.4646
…                          …                       …
2840  trip-153861034802474617                 46.5093
2841  trip-153861052318770017                 15.7803
2842  trip-153861089872028474                 38.2867
2843  trip-153861106442901555                 58.9037
2844  trip-153861118270144424                 80.5787

[2845 rows x 2 columns]
```

[42]:
```
trip_segment_osrm_distance_sum_df=df.groupby(['trip_uuid','segment_key']).
 ↪nth(-1).groupby(['trip_uuid']).
 ↪agg(trip_segment_osrm_distance_sum=('segment_osrm_distance_sum','sum')).
 ↪reset_index()
trip_segment_osrm_distance_sum_df
```

[42]:
```
                    trip_uuid  trip_segment_osrm_distance_sum
0     trip-153671079956500691                         16.0860
1     trip-153671110078355292                         10.8159
2     trip-153671191949943656                        140.5623
3     trip-153671237597058150                         56.7577
4     trip-153671262893947351                         30.4646
```

```
…                          …                                  …
2840  trip-153861034802474617                          48.5795
2841  trip-153861052318770017                          15.7803
2842  trip-153861089872028474                          33.6400
2843  trip-153861106442901555                         104.8866
2844  trip-153861118270144424                          80.5787

[2845 rows x 2 columns]
```

[43]:
```
trip_total_time_in_hrs_df=df.groupby(['trip_uuid','segment_key']).nth(-1).
 ↪groupby(['trip_uuid']).
 ↪agg(trip_total_time_in_hrs=('od_time_diff_hour','sum')).reset_index()
trip_total_time_in_hrs_df
```

[43]:
```
                    trip_uuid  trip_total_time_in_hrs
0     trip-153671079956500691                0.822223
1     trip-153671110078355292                0.641675
2     trip-153671191949943656                4.844371
3     trip-153671237597058150                4.201283
4     trip-153671262893947351                4.500402
…                          …                       …
2840  trip-153861034802474617                2.479912
2841  trip-153861052318770017                1.395993
2842  trip-153861089872028474                3.040278
2843  trip-153861106442901555                7.035331
2844  trip-153861118270144424                5.906793

[2845 rows x 2 columns]
```

Combine all trip aggregated dataframes to single trip aggregations dataframe

[44]:
```
trip_aggregated_df=pd.concat(
    objs=(
            iDF.set_index('trip_uuid') for iDF in␣
 ↪(trip_actual_time_sum_df,trip_osrm_time_sum_df, trip_osrm_distance_sum_df,
                                                  ␣
 ↪trip_segment_actual_time_sum_df,trip_segment_osrm_time_sum_df,trip_segment_osrm_distance_su
                                                    trip_total_time_in_hrs_df)
        ),
    axis=1,
    join='inner'
).reset_index()
trip_aggregated_df
```

[44]:
```
                    trip_uuid  trip_actual_time_sum  trip_osrm_time_sum  \
0     trip-153671079956500691                  23.0                 8.0
1     trip-153671110078355292                  17.0                 9.0
```

```
2     trip-153671191949943656                253.0               105.0
3     trip-153671237597058150                241.0                33.0
4     trip-153671262893947351                195.0                26.0
…                        …                    …                   …
2840  trip-153861034802474617                123.0                50.0
2841  trip-153861052318770017                 47.0                13.0
2842  trip-153861089872028474                 62.0                28.0
2843  trip-153861106442901555                282.0                48.0
2844  trip-153861118270144424                275.0                68.0

      trip_osrm_distance_sum  trip_segment_actual_time_sum  \
0                     9.9566                          23.0
1                    10.8159                          17.0
2                   124.5063                         252.0
3                    46.9087                         240.0
4                    30.4646                         195.0
…                        …                           …
2840                 46.5093                         121.0
2841                 15.7803                          47.0
2842                 38.2867                          61.0
2843                 58.9037                         281.0
2844                 80.5787                         274.0

      trip_segment_osrm_time_sum  trip_segment_osrm_distance_sum  \
0                           14.0                         16.0860
1                            9.0                         10.8159
2                          132.0                        140.5623
3                           49.0                         56.7577
4                           26.0                         30.4646
…                              …                            …
2840                        50.0                         48.5795
2841                        13.0                         15.7803
2842                        33.0                         33.6400
2843                        88.0                        104.8866
2844                        67.0                         80.5787

      trip_total_time_in_hrs
0                   0.822223
1                   0.641675
2                   4.844371
3                   4.201283
4                   4.500402
…                      …
2840                2.479912
2841                1.395993
2842                3.040278
2843                7.035331
```

```
2844                5.906793

[2845 rows x 8 columns]
```

[45]:
```
df=pd.merge(df,trip_aggregated_df, on='trip_uuid' )
df
```

[45]:
```
           data         trip_creation_time  \
0      training 2018-09-12 00:06:39.565253
1      training 2018-09-12 00:06:39.565253
2      training 2018-09-12 00:11:40.783923
3      training 2018-09-12 00:25:19.499696
4      training 2018-09-12 00:25:19.499696
…         …                          …
27183      test 2018-10-03 23:57:44.429324
27184      test 2018-10-03 23:59:42.701692
27185      test 2018-10-03 23:59:42.701692
27186      test 2018-10-03 23:59:42.701692
27187      test 2018-10-03 23:59:42.701692

                                    route_schedule_uuid route_type  \
0      thanos::sroute:a10888ff-f794-41e1-9b7a-7f62ef6…    Carting
1      thanos::sroute:a10888ff-f794-41e1-9b7a-7f62ef6…    Carting
2      thanos::sroute:c2ee580f-f4b2-4fa5-98ab-0c5b327…    Carting
3      thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678…        FTL
4      thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678…        FTL
…                                                  …          …
27183  thanos::sroute:5609c268-e436-4e0a-8180-3db4a74…    Carting
27184  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
27185  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
27186  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
27187  thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL

                     trip_uuid source_center  \
0      trip-153671079956500691  IND110024AAA
1      trip-153671079956500691  IND110024AAA
2      trip-153671110078355292  IND121004AAB
3      trip-153671191949943656  IND487001AAB
4      trip-153671191949943656  IND487001AAB
…                          …           …
27183  trip-153861106442901555  IND209304AAA
27184  trip-153861118270144424  IND583119AAA
27185  trip-153861118270144424  IND583119AAA
27186  trip-153861118270144424  IND583201AAA
27187  trip-153861118270144424  IND583201AAA

                                   source_name destination_center  \
```

```
0                     Delhi_Lajpat_IP (Delhi)       IND110014AAA
1                     Delhi_Lajpat_IP (Delhi)       IND110014AAA
2                FBD_Balabhgarh_DPC (Haryana)       IND121001AAA
3      Narsinghpur_KndliDPP_D (Madhya Pradesh)      IND487551AAA
4      Narsinghpur_KndliDPP_D (Madhya Pradesh)      IND487551AAA
…                                        …              …
27183       Kanpur_Central_H_6 (Uttar Pradesh)      IND208006AAA
27184            Sandur_WrdN1DPP_D (Karnataka)      IND583101AAA
27185            Sandur_WrdN1DPP_D (Karnataka)      IND583101AAA
27186                       Hospet (Karnataka)      IND583119AAA
27187                       Hospet (Karnataka)      IND583119AAA

                         destination_name              od_start_time  …  \
0                     Delhi_Bhogal (Delhi) 2018-09-12 00:06:39.565253  …
1                     Delhi_Bhogal (Delhi) 2018-09-12 00:06:39.565253  …
2                      Faridabad (Haryana) 2018-09-12 00:11:40.783923  …
3      Gadarwara_MPward_D (Madhya Pradesh) 2018-09-12 04:33:09.726898  …
4      Gadarwara_MPward_D (Madhya Pradesh) 2018-09-12 04:33:09.726898  …
…                                     …                        …  …
27183   Kanpur_GovndNgr_DC (Uttar Pradesh) 2018-10-03 23:57:44.429324  …
27184               Bellary_Dc (Karnataka) 2018-10-04 03:58:40.726547  …
27185               Bellary_Dc (Karnataka) 2018-10-04 03:58:40.726547  …
27186        Sandur_WrdN1DPP_D (Karnataka) 2018-10-04 02:51:44.712656  …
27187        Sandur_WrdN1DPP_D (Karnataka) 2018-10-04 02:51:44.712656  …

      source_place  destination_code source_code  trip_actual_time_sum  \
0           Lajpat              None          IP                  23.0
1           Lajpat              None          IP                  23.0
2        Balabhgarh             None         DPC                  17.0
3          KndliDPP                D           D                 253.0
4          KndliDPP                D           D                 253.0
…             …                 …         …                    …
27183       Central               DC         H_6                 282.0
27184     WrdN1DPP              None           D                 275.0
27185     WrdN1DPP              None           D                 275.0
27186         None                 D        None                 275.0
27187         None                 D        None                 275.0

      trip_osrm_time_sum  trip_osrm_distance_sum  \
0                    8.0                  9.9566
1                    8.0                  9.9566
2                    9.0                 10.8159
3                  105.0                124.5063
4                  105.0                124.5063
…                    …                    …
27183               48.0                 58.9037
27184               68.0                 80.5787
```

```
27185                68.0                  80.5787
27186                68.0                  80.5787
27187                68.0                  80.5787

       trip_segment_actual_time_sum  trip_segment_osrm_time_sum  \
0                              23.0                        14.0
1                              23.0                        14.0
2                              17.0                         9.0
3                             252.0                       132.0
4                             252.0                       132.0
…                               …                           …
27183                         281.0                        88.0
27184                         274.0                        67.0
27185                         274.0                        67.0
27186                         274.0                        67.0
27187                         274.0                        67.0

       trip_segment_osrm_distance_sum  trip_total_time_in_hrs
0                             16.0860                0.822223
1                             16.0860                0.822223
2                             10.8159                0.641675
3                            140.5623                4.844371
4                            140.5623                4.844371
…                               …                       …
27183                        104.8866                7.035331
27184                         80.5787                5.906793
27185                         80.5787                5.906793
27186                         80.5787                5.906793
27187                         80.5787                5.906793

[27188 rows x 50 columns]
```

Outlier Detection & Treatment¶ a. Find any existing outliers in numerical features.

b. Visualize the outlier values using Boxplot.

c. Handle the outliers using the IQR method.

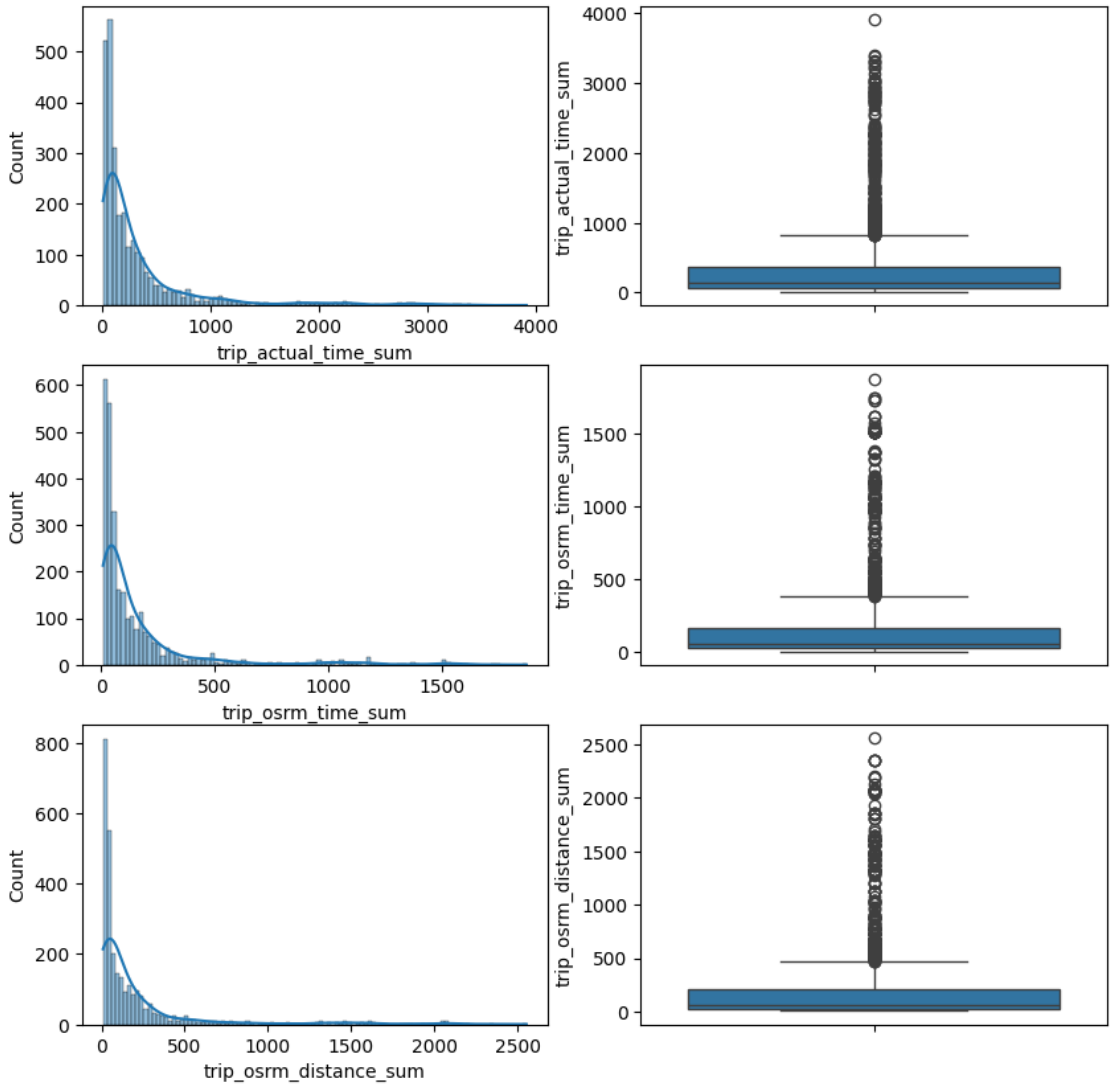
Perform one-hot encoding on categorical features. Normalize/ Standardize the numerical features using MinMaxScaler or StandardScaler.

[46]:
```
trip_aggregated_df
```

[46]:
```
                    trip_uuid  trip_actual_time_sum  trip_osrm_time_sum  \
0     trip-153671079956500691                  23.0                 8.0
1     trip-153671110078355292                  17.0                 9.0
2     trip-153671191949943656                 253.0               105.0
3     trip-153671237597058150                 241.0                33.0
4     trip-153671262893947351                 195.0                26.0
```

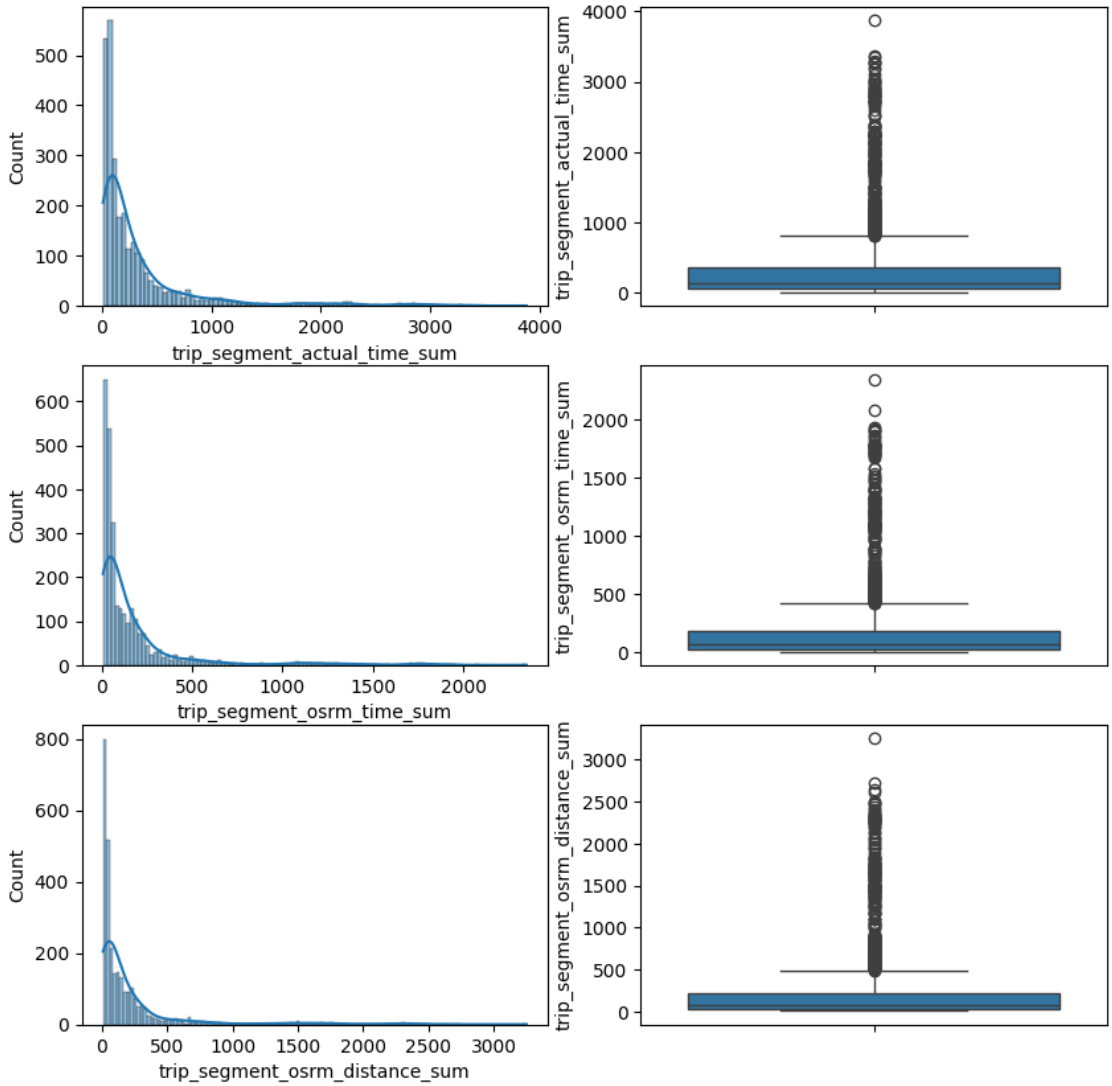
```
…                          …                    …                  …
2840  trip-153861034802474617                123.0                50.0
2841  trip-153861052318770017                 47.0                13.0
2842  trip-153861089872028474                 62.0                28.0
2843  trip-153861106442901555                282.0                48.0
2844  trip-153861118270144424                275.0                68.0

      trip_osrm_distance_sum  trip_segment_actual_time_sum  \
0                     9.9566                          23.0
1                    10.8159                          17.0
2                   124.5063                         252.0
3                    46.9087                         240.0
4                    30.4646                         195.0
…                        …                             …
2840                 46.5093                         121.0
2841                 15.7803                          47.0
2842                 38.2867                          61.0
2843                 58.9037                         281.0
2844                 80.5787                         274.0

      trip_segment_osrm_time_sum  trip_segment_osrm_distance_sum  \
0                           14.0                         16.0860
1                            9.0                         10.8159
2                          132.0                        140.5623
3                           49.0                         56.7577
4                           26.0                         30.4646
…                              …                             …
2840                        50.0                         48.5795
2841                        13.0                         15.7803
2842                        33.0                         33.6400
2843                        88.0                        104.8866
2844                        67.0                         80.5787

      trip_total_time_in_hrs
0                   0.822223
1                   0.641675
2                   4.844371
3                   4.201283
4                   4.500402
…                      …
2840                2.479912
2841                1.395993
2842                3.040278
2843                7.035331
2844                5.906793

[2845 rows x 8 columns]
```

[47]:
```
import warnings
warnings.filterwarnings("ignore")
fig, axs = plt.subplots(3, 2,  figsize=(10,10))
sns.histplot(ax=axs[0,0],data=␣
 ↪trip_aggregated_df['trip_actual_time_sum'],kde=True)
sns.boxplot(ax=axs[0,1],data=trip_aggregated_df['trip_actual_time_sum'])

sns.histplot(ax=axs[1,0],data=␣
 ↪trip_aggregated_df['trip_osrm_time_sum'],kde=True)
sns.boxplot(ax=axs[1,1],data= trip_aggregated_df['trip_osrm_time_sum'])

sns.histplot(ax=axs[2,0],data=␣
 ↪trip_aggregated_df['trip_osrm_distance_sum'],kde=True)
sns.boxplot(ax=axs[2,1],data= trip_aggregated_df['trip_osrm_distance_sum'])


plt.show()
```



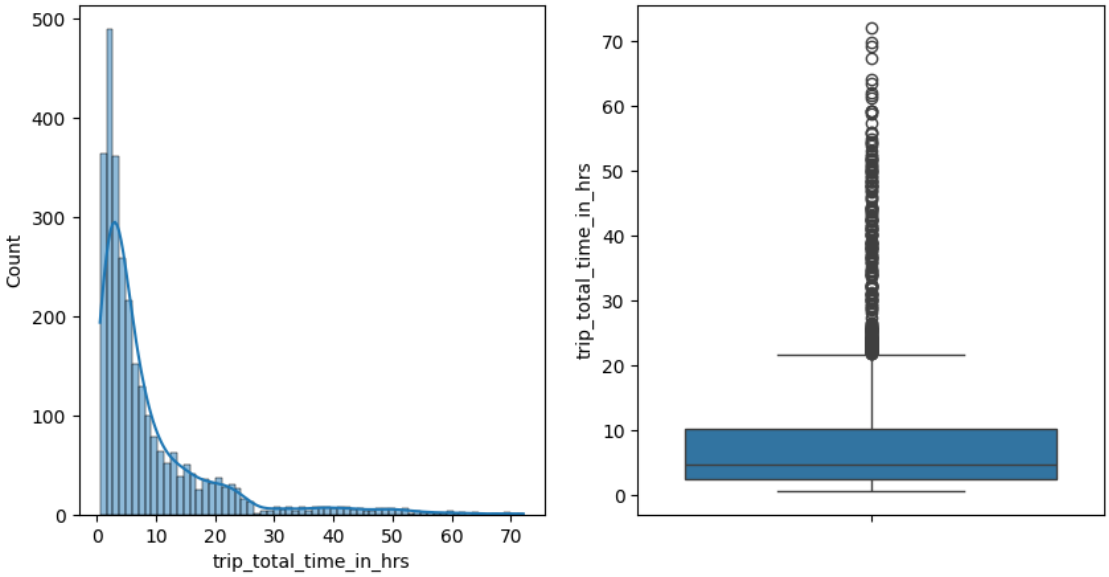
[48]:
```
import warnings
warnings.filterwarnings("ignore")
fig, axs = plt.subplots(3, 2,  figsize=(10,10))

sns.histplot(ax=axs[0,0],data=␣
 ↪trip_aggregated_df['trip_segment_actual_time_sum'],kde=True)
sns.boxplot(ax=axs[0,1],data=␣
 ↪trip_aggregated_df['trip_segment_actual_time_sum'])


sns.histplot(ax=axs[1,0],data=␣
 ↪trip_aggregated_df['trip_segment_osrm_time_sum'],kde=True)
sns.boxplot(ax=axs[1,1],data= trip_aggregated_df['trip_segment_osrm_time_sum'])
```

```
sns.histplot(ax=axs[2,0],data=␣
  ↪trip_aggregated_df['trip_segment_osrm_distance_sum'],kde=True)
sns.boxplot(ax=axs[2,1],data=␣
  ↪trip_aggregated_df['trip_segment_osrm_distance_sum'])


plt.show()
```



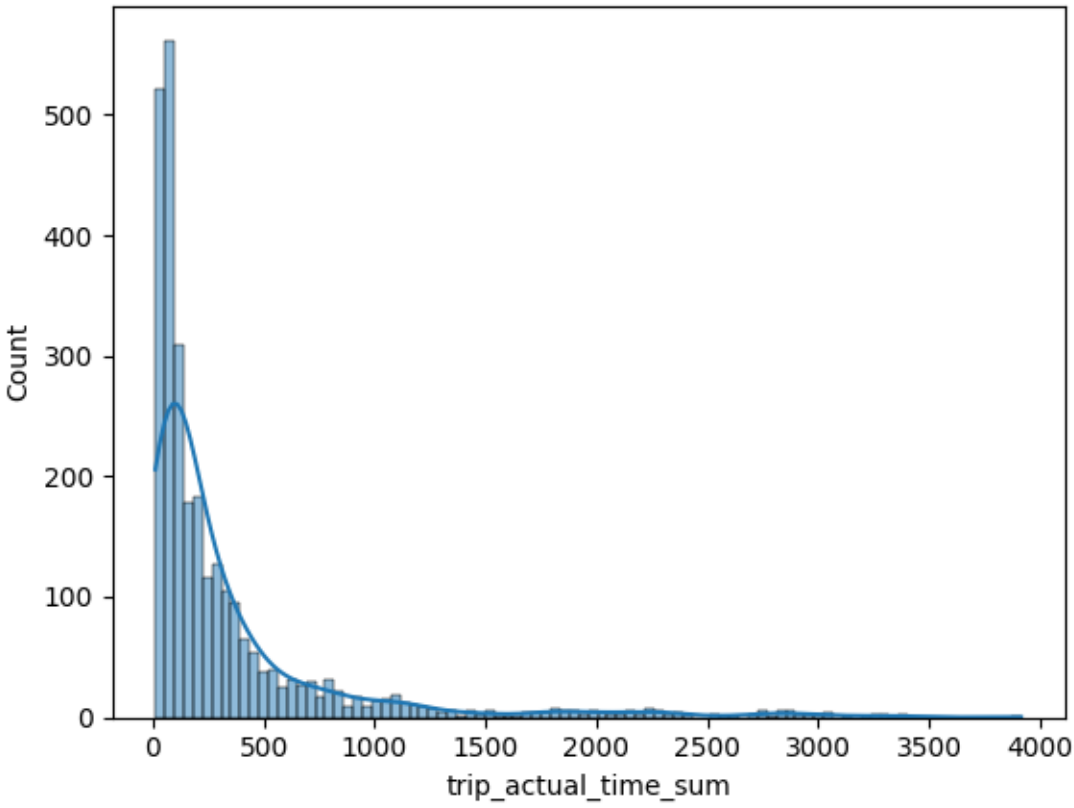
[49]:
```
import warnings
warnings.filterwarnings("ignore")
fig, axs = plt.subplots(ncols= 2,  figsize=(10,5))
```

```
sns.histplot(ax=axs[0],data=␣
 ↪trip_aggregated_df['trip_total_time_in_hrs'],kde=True)
sns.boxplot(ax=axs[1],data= trip_aggregated_df['trip_total_time_in_hrs'])
plt.show()
```



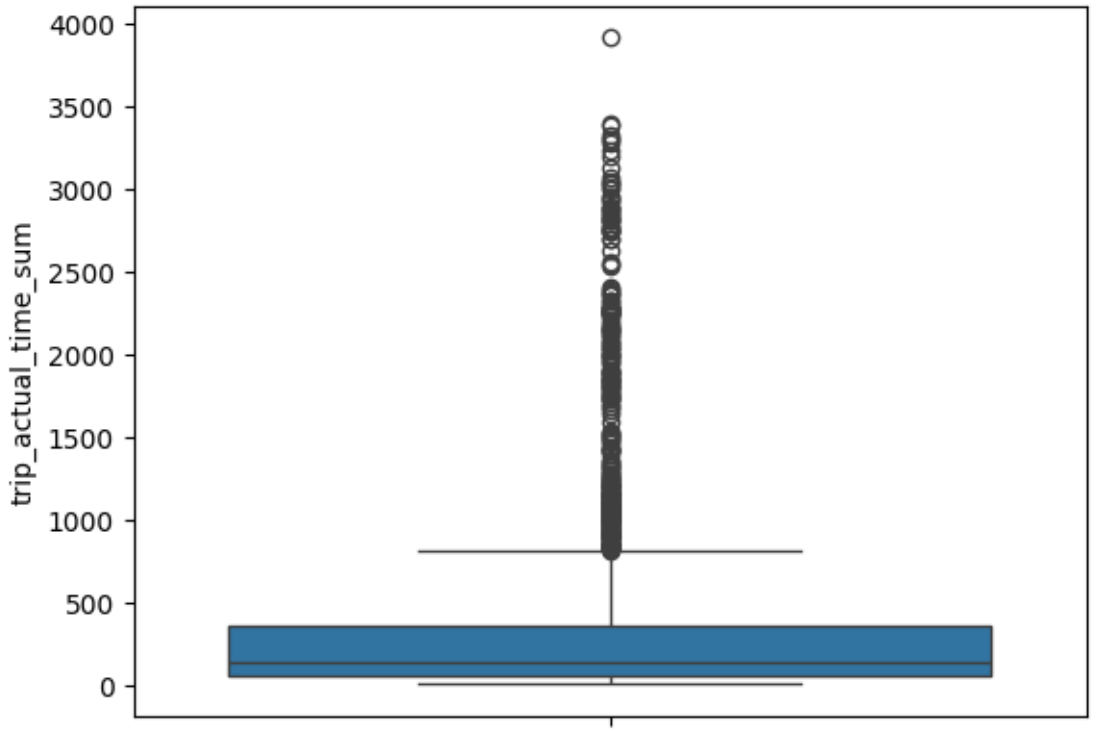
[50]:
```
sns.histplot(data=trip_aggregated_df['trip_actual_time_sum'],kde=True)
```

[50]: <Axes: xlabel='trip_actual_time_sum', ylabel='Count'>



[51]:
```
sns.boxplot(data=trip_aggregated_df['trip_actual_time_sum'])
```

[51]:
```
<Axes: ylabel='trip_actual_time_sum'>
```



Observations we have lots of outliers let's remove those outliers Based on IQR range, we will maxout values based on IQR range

[52]:
```
def clip_value_helper(row,cl, Q1,Q3, minval, maxval):
    #     Q1=row[cl].quantile(0.25)
    #     Q3=row[cl].quantile(0.75)
    #     minval=min(row[cl])
    #     maxval=max(row[cl])
    IQR=Q3-Q1
    if row[cl]<Q1-1.5*IQR:
        return min(minval, Q1-1.5*IQR)
    elif row[cl] > Q3+1.5*IQR:
        return min(maxval, Q3+1.5*IQR)
    else:
        return row[cl]
```

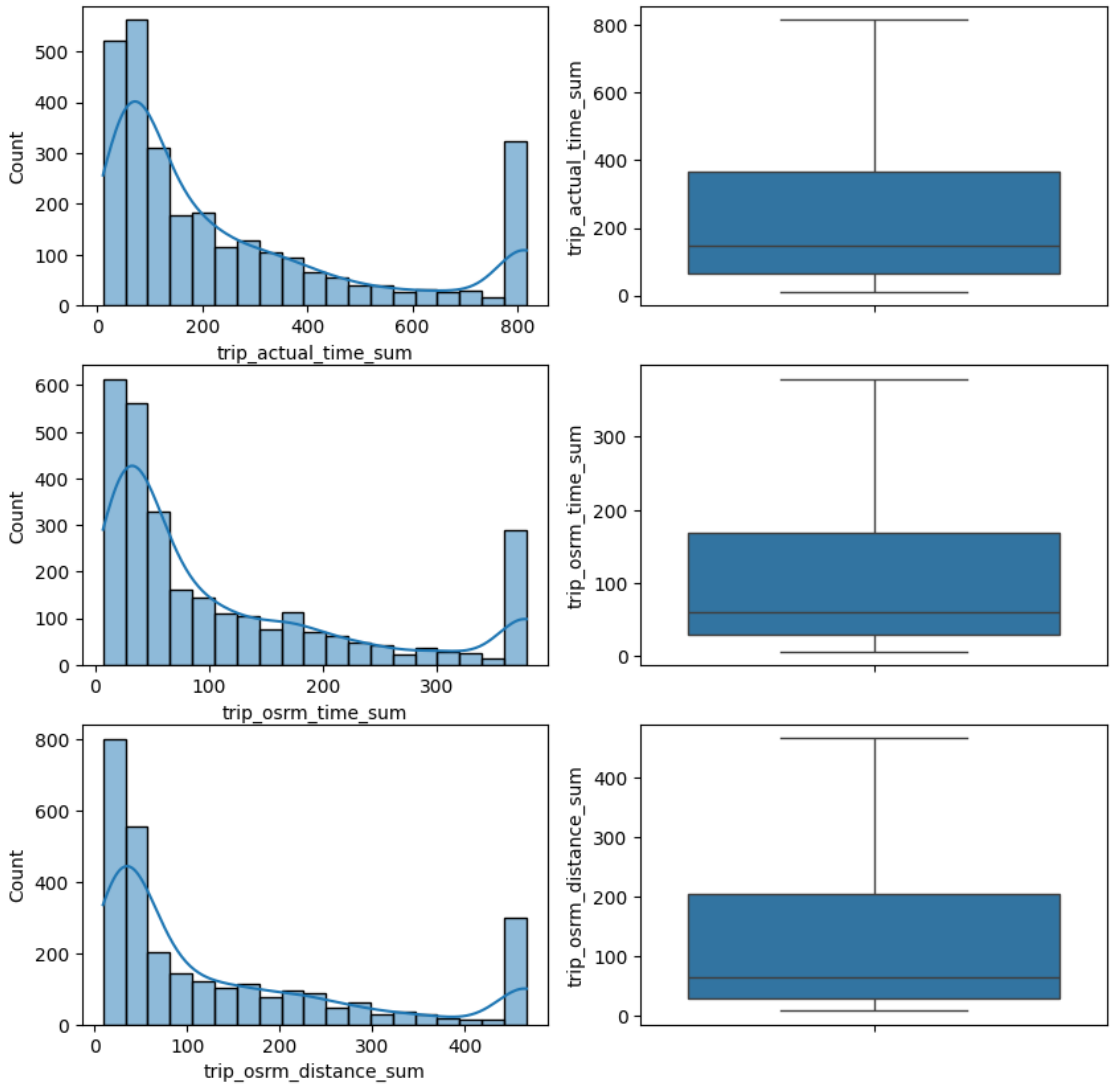
[53]:
```
trip_aggregated_df.columns
```

[53]:
```
Index(['trip_uuid', 'trip_actual_time_sum', 'trip_osrm_time_sum',
       'trip_osrm_distance_sum', 'trip_segment_actual_time_sum',
       'trip_segment_osrm_time_sum', 'trip_segment_osrm_distance_sum',
       'trip_total_time_in_hrs'],
      dtype='object')
```

[54]:
```
for cl in ['trip_actual_time_sum', 'trip_osrm_time_sum',␣
 ↪'trip_osrm_distance_sum',␣
 ↪'trip_segment_actual_time_sum','trip_segment_osrm_time_sum',␣
 ↪'trip_segment_osrm_distance_sum','trip_total_time_in_hrs']:
    Q1=trip_aggregated_df[cl].quantile(0.25)
    Q3=trip_aggregated_df[cl].quantile(0.75)
    minval=min(trip_aggregated_df[cl])
    maxval=max(trip_aggregated_df[cl])
    trip_aggregated_df[cl]=trip_aggregated_df.apply(lambda row:
 ↪clip_value_helper(row,cl,
                                                                  ␣
 ↪Q1,Q3,minval, maxval) , axis=1)
    trip_aggregated_df
```

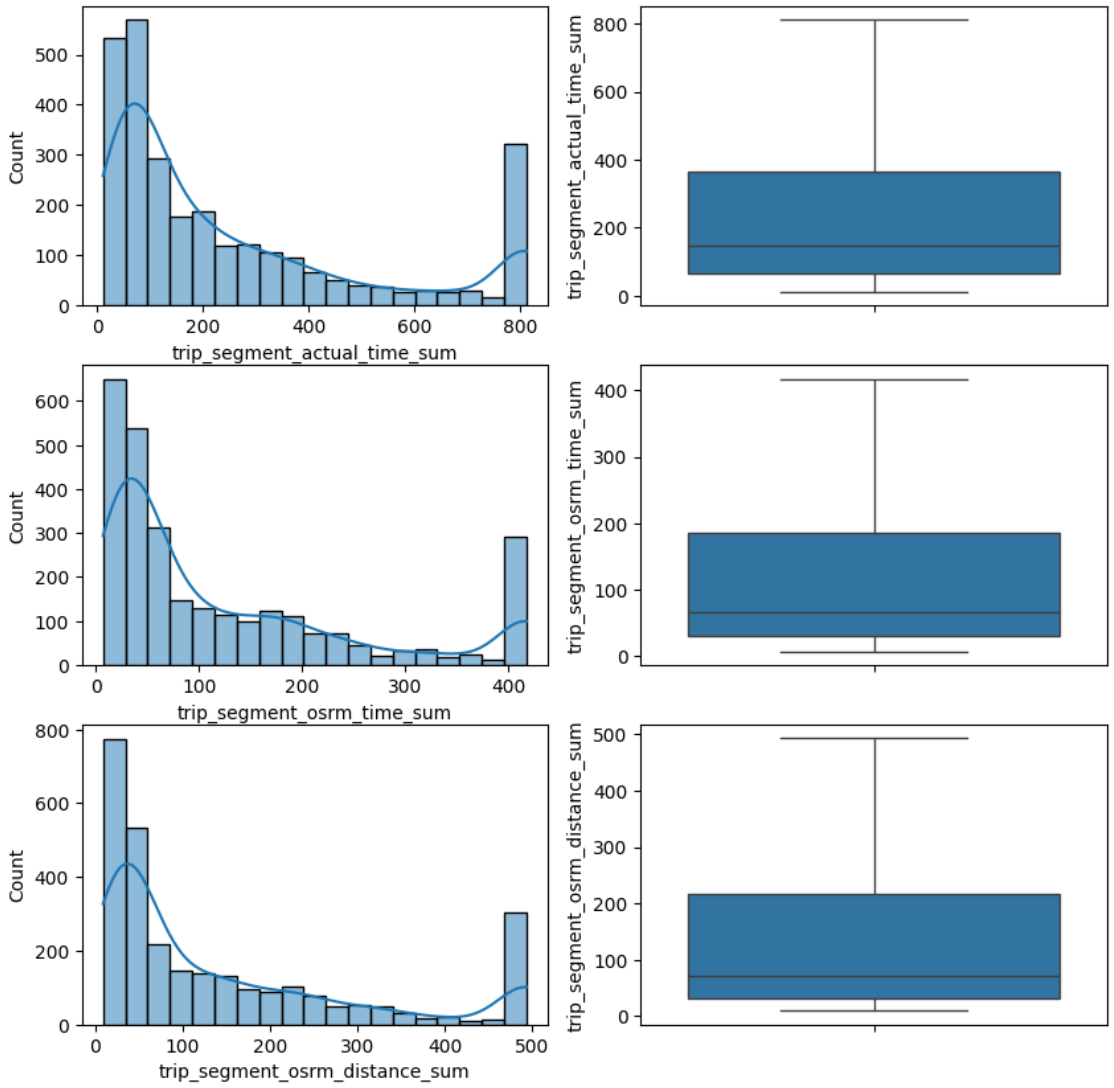
After Clipping Outliers Based On IQR Distribution¶

[55]:
```
import warnings
warnings.filterwarnings("ignore")
fig, axs = plt.subplots(3, 2,  figsize=(10,10))
sns.histplot(ax=axs[0,0],data=␣
 ↪trip_aggregated_df['trip_actual_time_sum'],kde=True)
sns.boxplot(ax=axs[0,1],data=trip_aggregated_df['trip_actual_time_sum'])

sns.histplot(ax=axs[1,0],data=␣
 ↪trip_aggregated_df['trip_osrm_time_sum'],kde=True)
sns.boxplot(ax=axs[1,1],data= trip_aggregated_df['trip_osrm_time_sum'])

sns.histplot(ax=axs[2,0],data=␣
 ↪trip_aggregated_df['trip_osrm_distance_sum'],kde=True)
sns.boxplot(ax=axs[2,1],data= trip_aggregated_df['trip_osrm_distance_sum'])


plt.show()
```



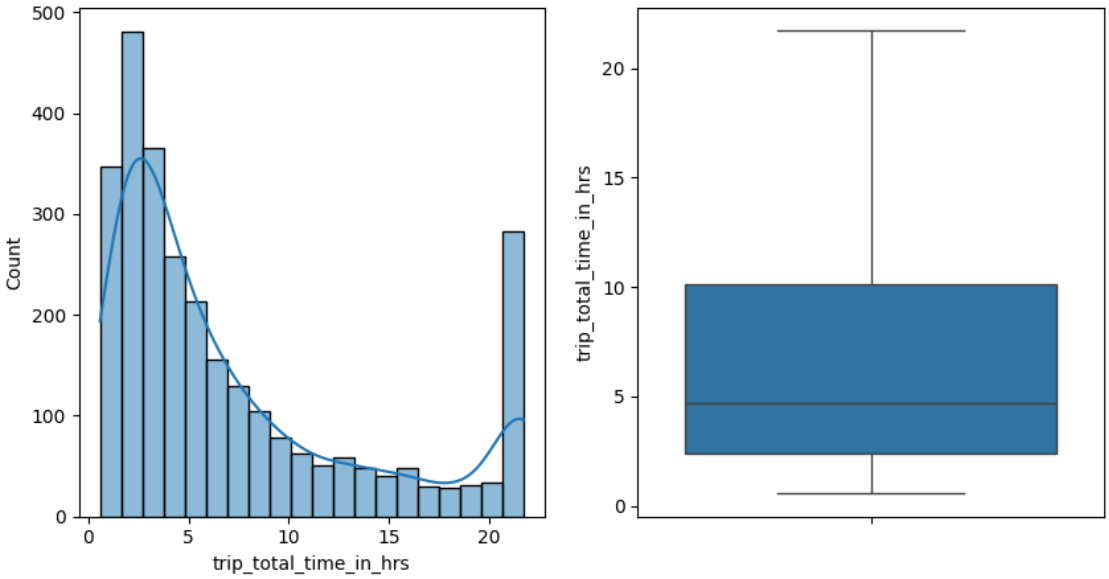
[56]:
```
import warnings
warnings.filterwarnings("ignore")
fig, axs = plt.subplots(3, 2,  figsize=(10,10))

sns.histplot(ax=axs[0,0],data=␣
 ↪trip_aggregated_df['trip_segment_actual_time_sum'],kde=True)
sns.boxplot(ax=axs[0,1],data=␣
 ↪trip_aggregated_df['trip_segment_actual_time_sum'])


sns.histplot(ax=axs[1,0],data=␣
 ↪trip_aggregated_df['trip_segment_osrm_time_sum'],kde=True)
sns.boxplot(ax=axs[1,1],data= trip_aggregated_df['trip_segment_osrm_time_sum'])
```

```
sns.histplot(ax=axs[2,0],data=␣
 ↪trip_aggregated_df['trip_segment_osrm_distance_sum'],kde=True)
sns.boxplot(ax=axs[2,1],data=␣
 ↪trip_aggregated_df['trip_segment_osrm_distance_sum'])


plt.show()
```



[57]:
```
import warnings
warnings.filterwarnings("ignore")
fig, axs = plt.subplots(ncols= 2,  figsize=(10,5))
```

```
sns.histplot(ax=axs[0],data=␣
 ↪trip_aggregated_df['trip_total_time_in_hrs'],kde=True)
sns.boxplot(ax=axs[1],data= trip_aggregated_df['trip_total_time_in_hrs'])
plt.show()
```



5. Hypothesis Testing

. actual_time aggregated value and OSRM time aggregated value.

we will use ttest paired sample test to know if there is significant difference in actual trip aggregated time and OSRM trip aggregated time for each trip

HO : mean Actual time to deliver package from source to destination is lesser than OSRM time for entire trip

HA: mean Actual time to deliver package from source to destination is greater than OSRM time

[58]:
```
from scipy.stats import ttest_rel, ttest_ind
stat, pval=ttest_rel(trip_aggregated_df['trip_actual_time_sum'],
                     trip_aggregated_df['trip_osrm_time_sum'],
                    alternative='greater')
print(f"stat {stat} pval {pval}")
```

```
stat 49.54905609405925 pval 0.0
```

[59]:
```
if pval <0.05:
    print("We will reject HO :")
    print("mean Actual time to deliver package from source to destination is␣
 ↪greater than OSRM time")
else:
```

```
    print("We fail to  reject H0 :")
    print("mean Actual time to deliver package from source to destination is␣
 ↪lesser or equal  than OSRM time")
```

```
We will reject H0 :
mean Actual time to deliver package from source to destination is greater than
OSRM time
```

[60]:
```
np.mean(trip_aggregated_df['trip_actual_time_sum']), np.
 ↪mean(trip_aggregated_df['trip_osrm_time_sum'])
```

[60]:
```
(258.7082601054482, 115.17223198594024)
```

b. actual_time aggregated value and segment actual time aggregated value. we will use ttest paired sample test to know if there is significant difference in actual trip aggregated time and segment actual time aggregated value for each trip

HO : mean Actual aggregated trip time to deliver package from source to destination is lesser than segment actual time aggregated value for entire trip

HA: mean Actual aggregated trip time to deliver package from source to destination is greater than segment actual time aggregated value for entire trip

[61]:
```
stat, pval=ttest_rel(trip_aggregated_df['trip_actual_time_sum'],
                     trip_aggregated_df['trip_segment_actual_time_sum'],
                    alternative='greater')
print(f"stat {stat} pval {pval}")
```

```
stat 55.445752235969664 pval 0.0
```

[62]:
```
if pval <0.05:
    print("We will reject H0 :")
    print("mean Actual trip aggregated time   to deliver package from source to␣
 ↪destination is greater than segment actual aggregated time")
else:
    print("We fail to  reject H0 :")
    print("mean Actual trip aggregated time to deliver package from source to␣
 ↪destination is lesser or equal  than segment actual aggregated time")
```

```
We will reject H0 :
mean Actual trip aggregated time   to deliver package from source to destination
is greater than segment actual aggregated time
```

c. OSRM distance aggregated value and segment OSRM distance aggregated value. ¶ we will use ttest paired sample test to know if there is significant difference in OSRM distance aggregated value and segment actual aggregated distance for each trip

HO : mean Actual aggregated OSRM distance for trip to deliver package from source to destination is lesser than segment actual OSRM distance aggregated value for entire trip

HA: mean Actual aggregated OSRM distance for trip to deliver package from source to destination is greater than segment actual OSRM distance aggregated value for entire trip

[63]:
```
stat, pval=ttest_rel(trip_aggregated_df['trip_osrm_distance_sum'],
                     trip_aggregated_df['trip_segment_osrm_distance_sum'],
                    alternative='greater')
print(f"stat {stat} pval {pval}")
```

```
stat -21.773765031064443 pval 1.0
```

[64]:
```
if pval <0.05:
    print("We will reject H0 :")
    print("mean Actual trip OSRM distance for trip  to deliver package from␣
 ↪source to destination is greater than segment actual aggregated OSRM␣
 ↪distance for trip")
else:
    print("We fail to  reject H0 :")
    print("mean Actual trip OSRM distance for trip to deliver package from␣
 ↪source to destination is lesser or equal  than segment actual aggregated␣
 ↪OSRM distance for trip")
```

```
We fail to  reject H0 :
mean Actual trip OSRM distance for trip to deliver package from source to
destination is lesser or equal  than segment actual aggregated OSRM distance for
trip
```

[65]:
```
trip_aggregated_df
```

[65]:

	trip_uuid	trip_actual_time_sum	trip_osrm_time_sum	\
0	trip-153671079956500691	23.0	8.0	
1	trip-153671110078355292	17.0	9.0	
2	trip-153671191949943656	253.0	105.0	
3	trip-153671237597058150	241.0	33.0	
4	trip-153671262893947351	195.0	26.0	
…	…	…	…	
2840	trip-153861034802474617	123.0	50.0	
2841	trip-153861052318770017	47.0	13.0	
2842	trip-153861089872028474	62.0	28.0	
2843	trip-153861106442901555	282.0	48.0	
2844	trip-153861118270144424	275.0	68.0	

	trip_osrm_distance_sum	trip_segment_actual_time_sum	\
0	9.9566	23.0	
1	10.8159	17.0	
2	124.5063	252.0	
3	46.9087	240.0	
4	30.4646	195.0	
…	…	…	

```
2840                46.5093                          121.0
2841                15.7803                           47.0
2842                38.2867                           61.0
2843                58.9037                          281.0
2844                80.5787                          274.0

      trip_segment_osrm_time_sum  trip_segment_osrm_distance_sum  \
0                           14.0                         16.0860
1                            9.0                         10.8159
2                          132.0                        140.5623
3                           49.0                         56.7577
4                           26.0                         30.4646
…                              …                               …
2840                        50.0                         48.5795
2841                        13.0                         15.7803
2842                        33.0                         33.6400
2843                        88.0                        104.8866
2844                        67.0                         80.5787

      trip_total_time_in_hrs
0                   0.822223
1                   0.641675
2                   4.844371
3                   4.201283
4                   4.500402
…                          …
2840                2.479912
2841                1.395993
2842                3.040278
2843                7.035331
2844                5.906793

[2845 rows x 8 columns]
```

[66]:
```
np.max(trip_aggregated_df['trip_actual_time_sum']), np.
↪max(trip_aggregated_df['trip_segment_actual_time_sum'])
```

[66]:
```
(817.5, 813.0)
```

OSRM time aggregated value and segment OSRM time aggregated value. ¶ we will use ttest paired sample test to know if there is significant difference in OSRM time aggregated value and segment OSRM aggregated time for each trip

HO : mean Actual aggregated OSRM time aggreated for trip to deliver package from source to destination is lesser than segment OSRM aggregated time value for entire trip

HA: mean Actual aggregated OSRM time aggregated for trip to deliver package from source to destination is greater than segment OSRM aggregated time value for entire trip

[67]:
```
stat, pval=ttest_rel(trip_aggregated_df['trip_osrm_time_sum'],
                     trip_aggregated_df['trip_segment_osrm_time_sum'],
                    alternative='greater')
print(f"stat {stat} pval {pval}")
```

```
stat -27.296301433442657 pval 1.0
```

[68]:
```
if pval <0.05:
    print("We will reject H0 :")
    print("mean Actual trip aggregated  OSRM time  for trip  to deliver package␣
 ↪from source to destination is greater than segment actual aggregated OSRM␣
 ↪time for trip")
else:
    print("We fail to  reject H0 :")
    print("mean Actual trip aggregated OSRM time  for trip to deliver package␣
 ↪from source to destination is lesser or equal  than segment actual␣
 ↪aggregated OSRM time for trip")
```

```
We fail to  reject H0 :
mean Actual trip aggregated OSRM time  for trip to deliver package from source
to destination is lesser or equal  than segment actual aggregated OSRM time for
trip
```

Business Insights & Recommendations

From Where the Most Orders are coming from

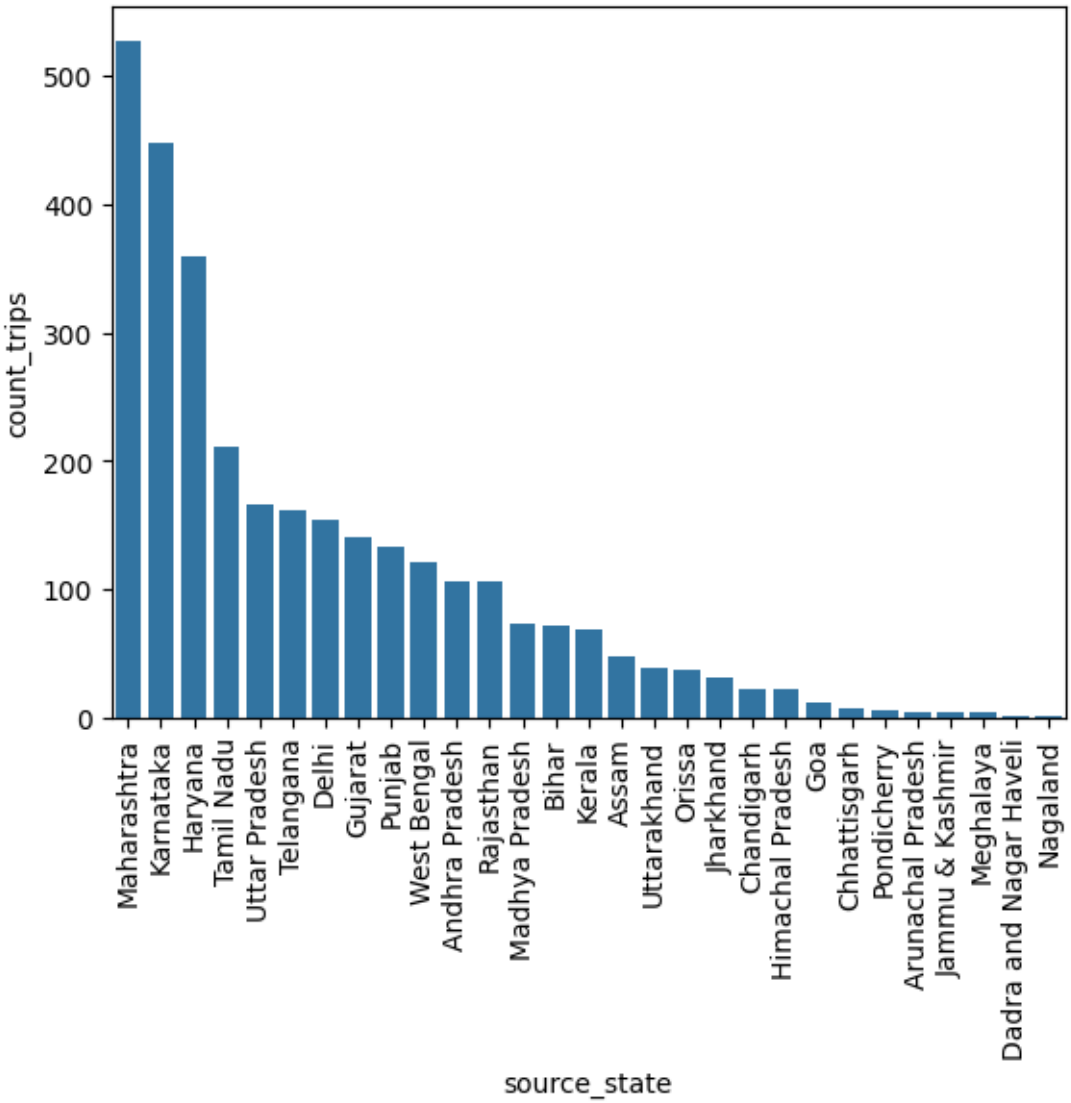
[69]:
```
# Check from where most orders are coming from (State, Corridor, etc.)


orders_from_df=df.groupby('source_state').
 ↪agg(count_trips=('trip_uuid','nunique')).reset_index()
orders_from_df=orders_from_df.sort_values(by='count_trips', ascending=False)
orders_from_df
```

[69]:

	source_state	count_trips
17	Maharashtra	527
14	Karnataka	448
10	Haryana	359
24	Tamil Nadu	211
26	Uttar Pradesh	166
25	Telangana	161
7	Delhi	154
9	Gujarat	141
22	Punjab	133
28	West Bengal	121
0	Andhra Pradesh	107
23	Rajasthan	107

```
16          Madhya Pradesh           73
3                    Bihar           72
15                  Kerala           69
2                    Assam           48
27             Uttarakhand           39
20                  Orissa           38
13               Jharkhand           31
4                Chandigarh           23
11         Himachal Pradesh           23
8                       Goa           12
5              Chhattisgarh            7
21             Pondicherry             6
1         Arunachal Pradesh            5
12          Jammu & Kashmir            5
18                Meghalaya            4
6   Dadra and Nagar Haveli             2
19                 Nagaland            2
```

[70]:
```
sns.barplot(data=orders_from_df,x='source_state',y='count_trips')
plt.xticks(rotation=90)
plt.show()
```



To Which State Most orders are going

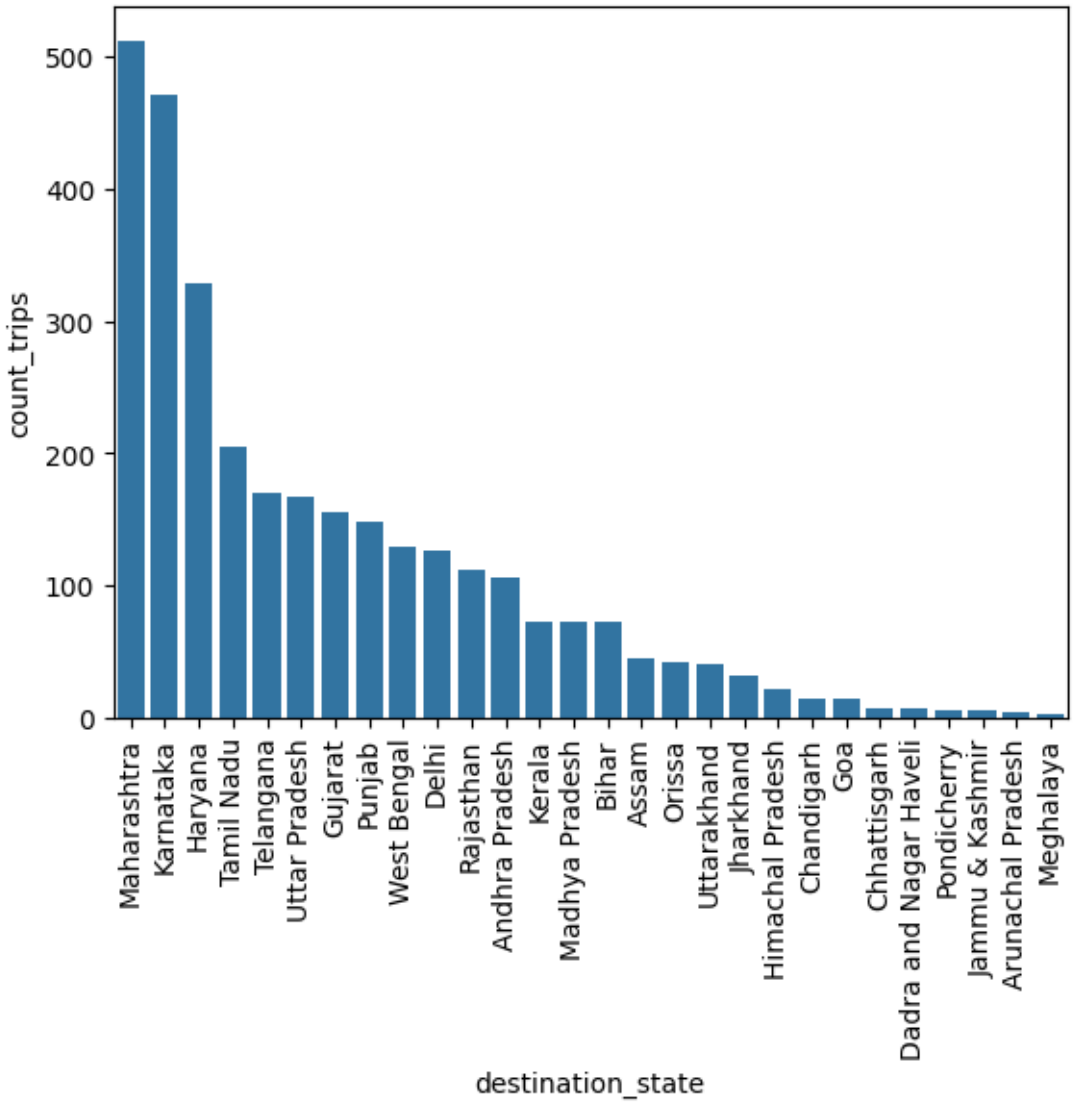
[71]:
```
orders_to_df=df.groupby('destination_state').
  ↪agg(count_trips=('trip_uuid','nunique')).reset_index()
orders_to_df=orders_to_df.sort_values(by='count_trips', ascending=False)
orders_to_df
```

[71]:

	destination_state	count_trips
17	Maharashtra	512
14	Karnataka	471
10	Haryana	329
23	Tamil Nadu	205
24	Telangana	170

25	Uttar Pradesh	167
9	Gujarat	155
21	Punjab	148
27	West Bengal	129
7	Delhi	126
22	Rajasthan	112
0	Andhra Pradesh	106
15	Kerala	73
16	Madhya Pradesh	73
3	Bihar	72
2	Assam	45
19	Orissa	42
26	Uttarakhand	40
13	Jharkhand	32
11	Himachal Pradesh	22
4	Chandigarh	15
8	Goa	14
5	Chhattisgarh	7
6	Dadra and Nagar Haveli	7
20	Pondicherry	6
12	Jammu & Kashmir	5
1	Arunachal Pradesh	4
18	Meghalaya	3

[72]:
```
sns.barplot(data=orders_to_df,x='destination_state',y='count_trips')
plt.xticks(rotation=90)
plt.show()
```



Most Busiest corridor

[73]:
```
#␣
 ↪corriodor_df=df[['source_center','destination_center','actual_time','osrm_time',␣
 ↪'osrm_distance']]

df['corridor']=df.apply(lambda x:"#".
 ↪join([x['source_center'],x['destination_center']]), axis=1)
df
```

[73]:
```
            data         trip_creation_time  \
0       training 2018-09-12 00:06:39.565253
1       training 2018-09-12 00:06:39.565253
```

```
2       training 2018-09-12 00:11:40.783923
3       training 2018-09-12 00:25:19.499696
4       training 2018-09-12 00:25:19.499696
…          …                          …
27183     test 2018-10-03 23:57:44.429324
27184     test 2018-10-03 23:59:42.701692
27185     test 2018-10-03 23:59:42.701692
27186     test 2018-10-03 23:59:42.701692
27187     test 2018-10-03 23:59:42.701692

                                    route_schedule_uuid route_type  \
0       thanos::sroute:a10888ff-f794-41e1-9b7a-7f62ef6…    Carting
1       thanos::sroute:a10888ff-f794-41e1-9b7a-7f62ef6…    Carting
2       thanos::sroute:c2ee580f-f4b2-4fa5-98ab-0c5b327…    Carting
3       thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678…        FTL
4       thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678…        FTL
…                                                    …        …
27183   thanos::sroute:5609c268-e436-4e0a-8180-3db4a74…    Carting
27184   thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
27185   thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
27186   thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL
27187   thanos::sroute:412fea14-6d1f-4222-8a5f-a517042…        FTL

                     trip_uuid source_center  \
0       trip-153671079956500691  IND110024AAA
1       trip-153671079956500691  IND110024AAA
2       trip-153671110078355292  IND121004AAB
3       trip-153671191949943656  IND487001AAB
4       trip-153671191949943656  IND487001AAB
…                           …           …
27183   trip-153861106442901555  IND209304AAA
27184   trip-153861118270144424  IND583119AAA
27185   trip-153861118270144424  IND583119AAA
27186   trip-153861118270144424  IND583201AAA
27187   trip-153861118270144424  IND583201AAA

                                    source_name destination_center  \
0                     Delhi_Lajpat_IP (Delhi)       IND110014AAA
1                     Delhi_Lajpat_IP (Delhi)       IND110014AAA
2                 FBD_Balabhgarh_DPC (Haryana)       IND121001AAA
3       Narsinghpur_KndliDPP_D (Madhya Pradesh)       IND487551AAA
4       Narsinghpur_KndliDPP_D (Madhya Pradesh)       IND487551AAA
…                                            …              …
27183       Kanpur_Central_H_6 (Uttar Pradesh)       IND208006AAA
27184            Sandur_WrdN1DPP_D (Karnataka)       IND583101AAA
27185            Sandur_WrdN1DPP_D (Karnataka)       IND583101AAA
27186                       Hospet (Karnataka)       IND583119AAA
```

```
27187                        Hospet (Karnataka)       IND583119AAA

                          destination_name              od_start_time  …  \
0                      Delhi_Bhogal (Delhi) 2018-09-12 00:06:39.565253  …
1                      Delhi_Bhogal (Delhi) 2018-09-12 00:06:39.565253  …
2                       Faridabad (Haryana) 2018-09-12 00:11:40.783923  …
3       Gadarwara_MPward_D (Madhya Pradesh) 2018-09-12 04:33:09.726898  …
4       Gadarwara_MPward_D (Madhya Pradesh) 2018-09-12 04:33:09.726898  …
…                                       …                         …  …
27183    Kanpur_GovndNgr_DC (Uttar Pradesh) 2018-10-03 23:57:44.429324  …
27184                Bellary_Dc (Karnataka) 2018-10-04 03:58:40.726547  …
27185                Bellary_Dc (Karnataka) 2018-10-04 03:58:40.726547  …
27186         Sandur_WrdN1DPP_D (Karnataka) 2018-10-04 02:51:44.712656  …
27187         Sandur_WrdN1DPP_D (Karnataka) 2018-10-04 02:51:44.712656  …

      destination_code  source_code trip_actual_time_sum  trip_osrm_time_sum  \
0                 None           IP                 23.0                 8.0
1                 None           IP                 23.0                 8.0
2                 None          DPC                 17.0                 9.0
3                    D            D                253.0               105.0
4                    D            D                253.0               105.0
…                  …            …                    …                   …
27183               DC          H_6                282.0                48.0
27184             None            D                275.0                68.0
27185             None            D                275.0                68.0
27186                D         None                275.0                68.0
27187                D         None                275.0                68.0

      trip_osrm_distance_sum  trip_segment_actual_time_sum  \
0                     9.9566                          23.0
1                     9.9566                          23.0
2                    10.8159                          17.0
3                   124.5063                         252.0
4                   124.5063                         252.0
…                        …                             …
27183               58.9037                          281.0
27184               80.5787                          274.0
27185               80.5787                          274.0
27186               80.5787                          274.0
27187               80.5787                          274.0

       trip_segment_osrm_time_sum  trip_segment_osrm_distance_sum  \
0                            14.0                         16.0860
1                            14.0                         16.0860
2                             9.0                         10.8159
3                           132.0                        140.5623
4                           132.0                        140.5623
```

```
…                          …                          …
27183                       88.0                   104.8866
27184                       67.0                    80.5787
27185                       67.0                    80.5787
27186                       67.0                    80.5787
27187                       67.0                    80.5787

       trip_total_time_in_hrs                   corridor
0                    0.822223  IND110024AAA#IND110014AAA
1                    0.822223  IND110024AAA#IND110014AAA
2                    0.641675  IND121004AAB#IND121001AAA
3                    4.844371  IND487001AAB#IND487551AAA
4                    4.844371  IND487001AAB#IND487551AAA
…                         …                          …
27183                7.035331  IND209304AAA#IND208006AAA
27184                5.906793  IND583119AAA#IND583101AAA
27185                5.906793  IND583119AAA#IND583101AAA
27186                5.906793  IND583201AAA#IND583119AAA
27187                5.906793  IND583201AAA#IND583119AAA

[27188 rows x 51 columns]
```

[74]:
```
corridor_total_trips=df.groupby(['trip_uuid','corridor']).nth(-1).
 ↪groupby(['corridor']).agg(total_trips=('trip_uuid','nunique')).reset_index()
corridor_total_trips
```

[74]:
```
                      corridor  total_trips
0     IND000000AAL#IND411033AAA            4
1     IND000000AAS#IND783370AAC            2
2     IND000000ABA#IND683565AAA            2
3     IND000000ABD#IND562132AAA            2
4     IND000000ABG#IND501359AAF            3
…                           …          …
1791  IND854326AAB#IND854105AAB            3
1792  IND854326AAB#IND854311AAA            5
1793  IND854334AAA#IND852118AAA            2
1794  IND854335AAA#IND852111AAA            3
1795  IND854335AAA#IND854326AAB            1

[1796 rows x 2 columns]
```

[75]:
```
corridor_actual_time_mean_df=df.groupby(['trip_uuid','corridor']).nth(-1).
 ↪groupby(['corridor']).
 ↪agg(corridor_actual_time_mean=('segment_actual_time_cumsum','mean')).
 ↪reset_index()
corridor_actual_time_mean_df
```

[75]:
```
                       corridor  corridor_actual_time_mean
0     IND000000AAL#IND411033AAA                  81.750000
1     IND000000AAS#IND783370AAC                  63.000000
2     IND000000ABA#IND683565AAA                  48.000000
3     IND000000ABD#IND562132AAA                 620.500000
4     IND000000ABG#IND501359AAF                 208.666667
…                           …                         …
1791  IND854326AAB#IND854105AAB                  56.333333
1792  IND854326AAB#IND854311AAA                  68.400000
1793  IND854334AAA#IND852118AAA                  26.000000
1794  IND854335AAA#IND852111AAA                  37.666667
1795  IND854335AAA#IND854326AAB                 197.000000

[1796 rows x 2 columns]
```

[76]:
```
corridor_osrm_time_mean_df=df.groupby(['trip_uuid','corridor']).nth(-1).
 ↪groupby(['corridor']).
 ↪agg(corridor_dsrm_time_mean=('segment_osrm_time_cumsum','mean')).
 ↪reset_index()
corridor_osrm_time_mean_df
```

[76]:
```
                       corridor  corridor_dsrm_time_mean
0     IND000000AAL#IND411033AAA                28.500000
1     IND000000AAS#IND783370AAC                30.000000
2     IND000000ABA#IND683565AAA                20.000000
3     IND000000ABD#IND562132AAA               402.500000
4     IND000000ABG#IND501359AAF                33.666667
…                           …                       …
1791  IND854326AAB#IND854105AAB                19.666667
1792  IND854326AAB#IND854311AAA                29.400000
1793  IND854334AAA#IND852118AAA                21.000000
1794  IND854335AAA#IND852111AAA                19.000000
1795  IND854335AAA#IND854326AAB                82.000000

[1796 rows x 2 columns]
```

[77]:
```
corridor_osrm_distance_mean_df=df.groupby(['trip_uuid','corridor']).nth(-1).
 ↪groupby(['corridor']).
 ↪agg(corridor_osrm_distance_mean=('segment_osrm_distance_cumsum','mean')).
 ↪reset_index()
corridor_osrm_distance_mean_df
```

[77]:
```
                       corridor  corridor_osrm_distance_mean
0     IND000000AAL#IND411033AAA                    28.299475
1     IND000000AAS#IND783370AAC                    42.577350
2     IND000000ABA#IND683565AAA                    16.638750
3     IND000000ABD#IND562132AAA                   443.786100
```

```
4     IND000000ABG#IND501359AAF                    35.057133
…                           …                          …
1791  IND854326AAB#IND854105AAB                    29.209900
1792  IND854326AAB#IND854311AAA                    42.575020
1793  IND854334AAA#IND852118AAA                    23.600650
1794  IND854335AAA#IND852111AAA                    27.882833
1795  IND854335AAA#IND854326AAB                   109.264800

[1796 rows x 2 columns]
```

[78]:
```
corridor_aggregated_df=pd.concat(
    objs=(
            iDF.set_index('corridor') for iDF in (corridor_total_trips,
                                                  corridor_actual_time_mean_df,
                                                  ␣
 ↪corridor_osrm_distance_mean_df,
                                                    corridor_osrm_time_mean_df
                                                  )
        ),
    axis=1,
    join='inner'
).reset_index()
corridor_aggregated_df
```

[78]:
```
                      corridor  total_trips  corridor_actual_time_mean  \
0     IND000000AAL#IND411033AAA            4                  81.750000
1     IND000000AAS#IND783370AAC            2                  63.000000
2     IND000000ABA#IND683565AAA            2                  48.000000
3     IND000000ABD#IND562132AAA            2                 620.500000
4     IND000000ABG#IND501359AAF            3                 208.666667
…                           …          …                          …
1791  IND854326AAB#IND854105AAB            3                  56.333333
1792  IND854326AAB#IND854311AAA            5                  68.400000
1793  IND854334AAA#IND852118AAA            2                  26.000000
1794  IND854335AAA#IND852111AAA            3                  37.666667
1795  IND854335AAA#IND854326AAB            1                 197.000000

      corridor_osrm_distance_mean  corridor_dsrm_time_mean
0                       28.299475                28.500000
1                       42.577350                30.000000
2                       16.638750                20.000000
3                      443.786100               402.500000
4                       35.057133                33.666667
…                             …                        …
1791                    29.209900                19.666667
1792                    42.575020                29.400000
1793                    23.600650                21.000000
```

```
1794                    27.882833                19.000000
1795                   109.264800                82.000000

[1796 rows x 5 columns]
```

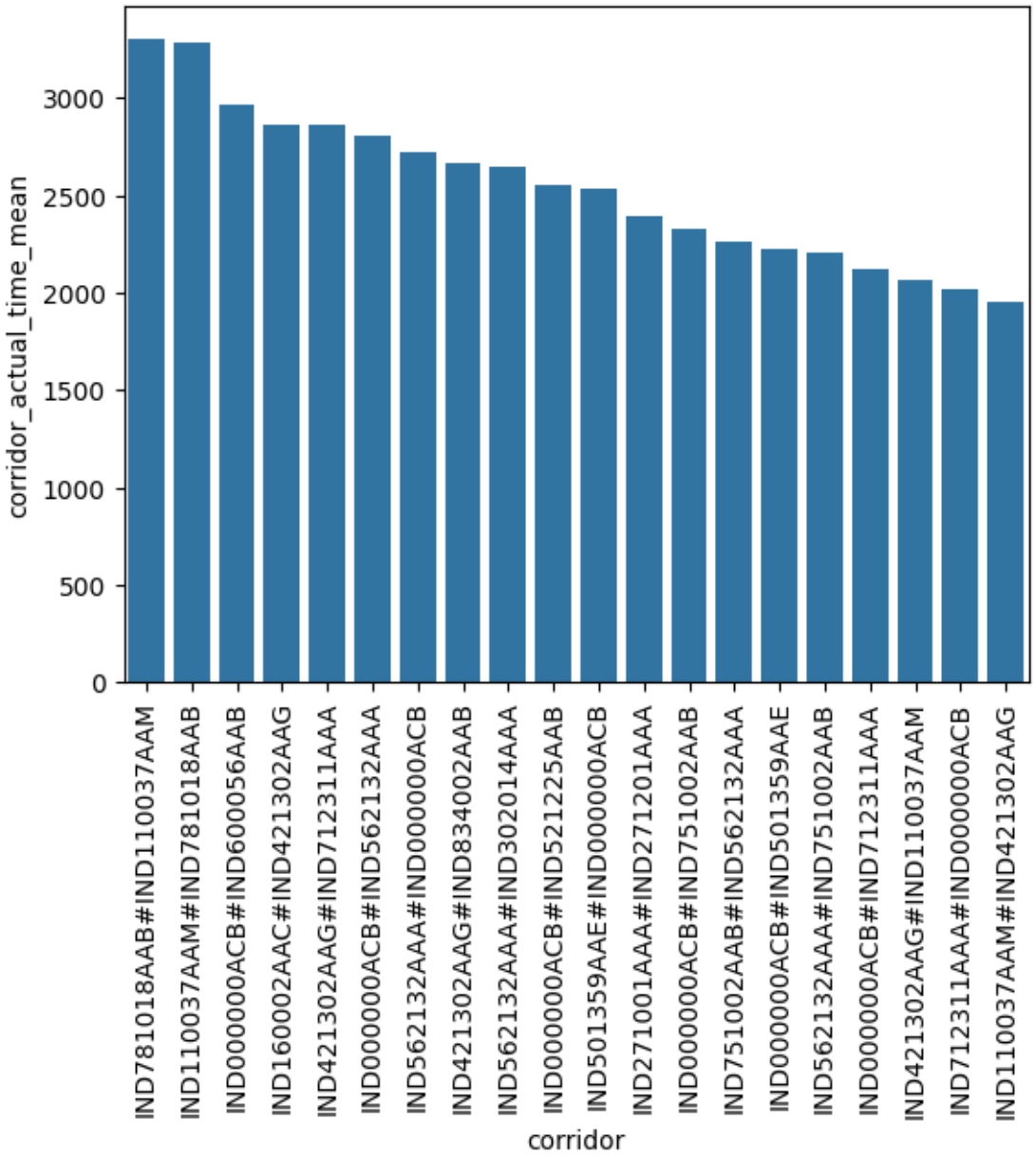
Busiest corridor Actual time¶

[79]:
```
top_20trip_time_corridor_df=corridor_aggregated_df[['corridor','corridor_actual_time_mean']].
 ↪sort_values(by='corridor_actual_time_mean', ascending=False)[:20]
top_20trip_time_corridor_df
```

[79]:

	corridor	corridor_actual_time_mean
1638	IND781018AAB#IND110037AAM	3302.500000
86	IND110037AAM#IND781018AAB	3281.000000
44	IND000000ACB#IND600056AAB	2963.000000
222	IND160002AAC#IND421302AAG	2867.000000
765	IND421302AAG#IND712311AAA	2867.000000
43	IND000000ACB#IND562132AAA	2808.642857
1130	IND562132AAA#IND000000ACB	2721.333333
766	IND421302AAG#IND834002AAB	2667.000000
1132	IND562132AAA#IND302014AAA	2651.000000
42	IND000000ACB#IND521225AAB	2550.333333
899	IND501359AAE#IND000000ACB	2536.000000
402	IND271001AAA#IND271201AAA	2398.000000
46	IND000000ACB#IND751002AAB	2323.750000
1601	IND751002AAB#IND562132AAA	2264.000000
41	IND000000ACB#IND501359AAE	2222.500000
1158	IND562132AAA#IND751002AAB	2208.000000
45	IND000000ACB#IND712311AAA	2121.600000
742	IND421302AAG#IND110037AAM	2064.500000
1515	IND712311AAA#IND000000ACB	2021.000000
85	IND110037AAM#IND421302AAG	1952.666667

[80]:
```
sns.
 ↪barplot(data=top_20trip_time_corridor_df,x='corridor',y='corridor_actual_time_mean')
plt.xticks(rotation=90)
plt.show()
```



[81]:
```
corridor_aggregated_df[['corridor_actual_time_mean','corridor_osrm_distance_mean','corridor_ds
↪corr()
```

[81]:

	corridor_actual_time_mean
corridor_actual_time_mean	1.000000
corridor_osrm_distance_mean	0.937734
corridor_dsrm_time_mean	0.934818
total_trips	-0.011093

	corridor_osrm_distance_mean

```
corridor_actual_time_mean                    0.937734
corridor_osrm_distance_mean                  1.000000
corridor_dsrm_time_mean                      0.995915
total_trips                                  0.007006
```

	corridor_dsrm_time_mean	total_trips
corridor_actual_time_mean	0.934818	-0.011093
corridor_osrm_distance_mean	0.995915	0.007006
corridor_dsrm_time_mean	1.000000	0.008240
total_trips	0.008240	1.000000

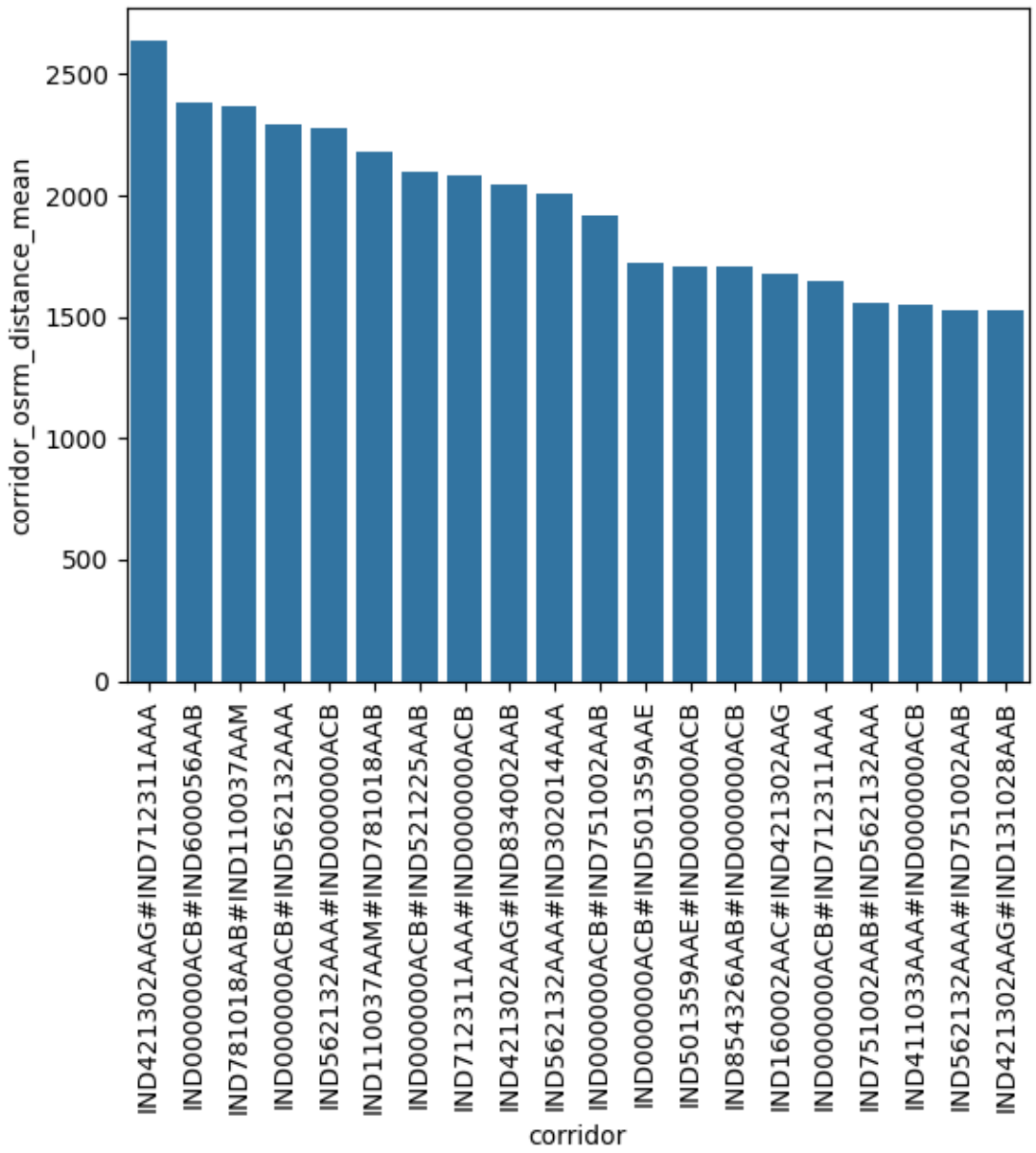
Busiest Corridor by Distance

[82]:
```
top_20trip_dist_corridor_df=corridor_aggregated_df[['corridor','corridor_osrm_distance_mean']]
  ↪sort_values(by='corridor_osrm_distance_mean', ascending=False)[:20]
top_20trip_dist_corridor_df
```

[82]:

	corridor	corridor_osrm_distance_mean
765	IND421302AAG#IND712311AAA	2640.924700
44	IND000000ACB#IND600056AAB	2384.514700
1638	IND781018AAB#IND110037AAM	2367.129350
43	IND000000ACB#IND562132AAA	2289.301729
1130	IND562132AAA#IND000000ACB	2276.539317
86	IND110037AAM#IND781018AAB	2181.460700
42	IND000000ACB#IND521225AAB	2099.734933
1515	IND712311AAA#IND000000ACB	2079.612400
766	IND421302AAG#IND834002AAB	2046.679900
1132	IND562132AAA#IND302014AAA	2010.760400
46	IND000000ACB#IND751002AAB	1916.699800
41	IND000000ACB#IND501359AAE	1721.693467
899	IND501359AAE#IND000000ACB	1708.819400
1786	IND854326AAB#IND000000ACB	1707.401300
222	IND160002AAC#IND421302AAG	1679.658000
45	IND000000ACB#IND712311AAA	1645.840390
1601	IND751002AAB#IND562132AAA	1557.065950
690	IND411033AAA#IND000000ACB	1553.552800
1158	IND562132AAA#IND751002AAB	1527.861300
743	IND421302AAG#IND131028AAB	1524.396400

[83]:
```
sns.
  ↪barplot(data=top_20trip_dist_corridor_df,x='corridor',y='corridor_osrm_distance_mean')
plt.xticks(rotation=90)
plt.show()
```



Business Recommendations

Maharashtra, Karnataka, Tamil Nadu, Utter Pradesh, Telangana and Gujarat States are states where most delivery trips are done. Most Bussiest corridor are in these states.

Business should focus on identifying best corridors to move packages very quickly, they should focus on potential reasons for difference in actual delivery time and osrm delivery time value.

If Actual delivery time is higher than osrm time then should focus on hops which are causing delays, if delays are related to processing or logistic that should be quickly fixed.

If Issue is not related to delivery and logistic process then should focus on identifying best route to

move packages quickly.