

```

import numpy as np
import pandas as pd

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

df = pd.read_csv('/content/Delhivery.csv')

import matplotlib.pyplot as plt
import seaborn as sns

df = df.dropna(how='any')
df = df.reset_index(drop=True)

import numpy as np
import pandas as pd

df['od_start_time'] = pd.to_datetime(df['od_start_time'])
df['od_end_time'] = pd.to_datetime(df['od_end_time'])

df.head(20)

```

	data	trip_creation_time	\
0	training	9/20/18 2:35 AM	
1	training	9/20/18 2:35 AM	
2	training	9/20/18 2:35 AM	
3	training	9/20/18 2:35 AM	
4	training	9/20/18 2:35 AM	
5	training	9/20/18 2:35 AM	
6	training	9/20/18 2:35 AM	
7	training	9/20/18 2:35 AM	
8	training	9/20/18 2:35 AM	
9	training	9/20/18 2:35 AM	
10	training	9/23/18 6:42 AM	
11	training	9/23/18 6:42 AM	
12	training	9/23/18 6:42 AM	
13	training	9/23/18 6:42 AM	
14	training	9/23/18 6:42 AM	
15	training	9/14/18 3:42 PM	
16	training	9/14/18 3:42 PM	
17	training	9/13/18 8:44 PM	
18	training	9/13/18 8:44 PM	
19	training	9/13/18 8:44 PM	

	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 (Gujarat)	IND388121AAA	Anand_VUNagar_DC	
1 trip-153741093647649320 (Gujarat)	IND388121AAA	Anand_VUNagar_DC	
2 trip-153741093647649320 (Gujarat)	IND388121AAA	Anand_VUNagar_DC	
3 trip-153741093647649320 (Gujarat)	IND388121AAA	Anand_VUNagar_DC	
4 trip-153741093647649320 (Gujarat)	IND388121AAA	Anand_VUNagar_DC	
5 trip-153741093647649320 (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D	
6 trip-153741093647649320 (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D	
7 trip-153741093647649320 (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D	
8 trip-153741093647649320 (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D	
9 trip-153741093647649320 (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D	
10 trip-153768492602129387 (Maharashtra)	IND421302AAG	Bhiwandi_Mankoli_HB	
11 trip-153768492602129387 (Maharashtra)	IND421302AAG	Bhiwandi_Mankoli_HB	
12 trip-153768492602129387 (Maharashtra)	IND421302AAG	Bhiwandi_Mankoli_HB	
13 trip-153768492602129387 (Maharashtra)	IND421302AAG	Bhiwandi_Mankoli_HB	
14 trip-153768492602129387 (Maharashtra)	IND421302AAG	Bhiwandi_Mankoli_HB	

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	od_start_time \
0	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:00
1	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:00
2	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:00
3	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:00
4	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:00
5	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:00
6	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:00
7	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:00
8	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:00
9	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:00
10	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:00
11	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:00
12	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:00
13	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:00
14	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:00
15	IND400072AAD	Mumbai_Chndivli_PC (Maharashtra)	2018-09-14 15:42:00
16	IND400072AAD	Mumbai_Chndivli_PC (Maharashtra)	2018-09-14 15:42:00
17	IND560099AAB	Bengaluru_Bomsndra_HB (Karnataka)	2018-09-13 20:44:00

18 IND560099AAB Bengaluru_Bomsndra_HB (Karnataka) 2018-09-13
 20:44:00
 19 IND683511AAA Aluva_Peedika_H (Kerala) 2018-09-13
 23:59:00

	...	cutoff_timestamp	actual_distance_to_destination	actual_time
\				
0	...	9/20/18 4:27 AM	10.435660	14.0
1	...	9/20/18 4:17 AM	18.936842	24.0
2	...	9/20/18 4:01 AM	27.637279	40.0
3	...	9/20/18 3:39 AM	36.118028	62.0
4	...	9/20/18 3:33 AM	39.386040	68.0
5	...	9/20/18 6:15 AM	10.403038	15.0
6	...	9/20/18 5:47 AM	18.045481	44.0
7	...	9/20/18 5:25 AM	28.061896	65.0
8	...	9/20/18 5:15 AM	38.939167	76.0
9	...	9/20/18 4:49 AM	43.595802	102.0
10	...	9/23/18 11:05 AM	23.194334	38.0
11	...	9/23/18 10:27 AM	44.045659	76.0
12	...	9/23/18 9:45 AM	72.849327	117.0
13	...	9/23/18 9:21 AM	88.076599	141.0
14	...	9/23/18 8:39 AM	100.708423	183.0
15	...	9/14/18 4:29 PM	9.355852	46.0
16	...	9/14/18 4:15 PM	16.431273	60.0
17	...	9/13/18 11:25 PM	23.635811	30.0
18	...	9/13/18 10:47 PM	39.806036	67.0
19	...	9/14/18 12:45 PM	24.319864	50.0

	osrm_time	osrm_distance	factor	segment_actual_time
segment_osrm_time \				
0	11.0	11.9653	1.272727	14.0

11.0				
1	20.0	21.7243	1.200000	10.0
9.0				
2	28.0	32.5395	1.428571	16.0
7.0				
3	40.0	45.5620	1.550000	21.0
12.0				
4	44.0	54.2181	1.545455	6.0
5.0				
5	11.0	12.1171	1.363636	15.0
11.0				
6	17.0	21.2890	2.588235	28.0
6.0				
7	29.0	35.8252	2.241379	21.0
11.0				
8	39.0	47.1900	1.948718	10.0
10.0				
9	45.0	53.2334	2.266667	26.0
6.0				
10	24.0	26.8622	1.583333	38.0
24.0				
11	41.0	54.4326	1.853659	37.0
27.0				
12	68.0	89.6680	1.720588	41.0
26.0				
13	80.0	108.3939	1.762500	23.0
14.0				
14	95.0	129.3519	1.926316	41.0
15.0				
15	11.0	11.4344	4.181818	46.0
11.0				
16	16.0	18.7941	3.750000	14.0
5.0				
17	30.0	28.9765	1.000000	30.0
30.0				
18	53.0	52.1256	1.264151	37.0
26.0				
19	24.0	29.7046	2.083333	50.0
24.0				

	segment_osrm_distance	segment_factor
0	11.9653	1.272727
1	9.7590	1.111111
2	10.8152	2.285714
3	13.0224	1.750000
4	3.9153	1.200000
5	12.1171	1.363636
6	9.1719	4.666667
7	14.5362	1.909091

8	11.3648	1.000000
9	6.0434	4.333333
10	26.8622	1.583333
11	30.1058	1.370370
12	35.2353	1.576923
13	17.2476	1.642857
14	20.9580	2.733333
15	11.4344	4.181818
16	7.3597	2.800000
17	28.9765	1.000000
18	24.9545	1.423077
19	29.7046	2.083333

[20 rows x 24 columns]

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 3304 entries, 0 to 3303

Data columns (total 24 columns):

#	Column	Non-Null Count	Dtype
0	data	3304 non-null	object
1	trip_creation_time	3304 non-null	object
2	route_schedule_uuid	3304 non-null	object
3	route_type	3304 non-null	object
4	trip_uuid	3304 non-null	object
5	source_center	3304 non-null	object
6	source_name	3304 non-null	object
7	destination_center	3304 non-null	object
8	destination_name	3304 non-null	object
9	od_start_time	3304 non-null	datetime64[ns]
10	od_end_time	3304 non-null	datetime64[ns]
11	start_scan_to_end_scan	3304 non-null	float64
12	is_cutoff	3304 non-null	object
13	cutoff_factor	3304 non-null	float64
14	cutoff_timestamp	3304 non-null	object
15	actual_distance_to_destination	3304 non-null	float64
16	actual_time	3304 non-null	float64
17	osrm_time	3304 non-null	float64
18	osrm_distance	3304 non-null	float64
19	factor	3304 non-null	float64
20	segment_actual_time	3304 non-null	float64
21	segment_osrm_time	3304 non-null	float64
22	segment_osrm_distance	3304 non-null	float64
23	segment_factor	3304 non-null	float64

dtypes: datetime64[ns](2), float64(11), object(11)

memory usage: 619.6+ KB

```

df['segment_key'] = df['trip_uuid'] + df['source_center'] +
df['destination_center']

segment_cols = ['segment_actual_time', 'segment_osrm_distance',
'segment_osrm_time']

for col in segment_cols:
    df[col + '_sum'] = df.groupby('segment_key')[col].cumsum()

df[[col + '_sum' for col in segment_cols]]

```

	segment_actual_time_sum	segment_osrm_distance_sum \
0	14.0	11.9653
1	24.0	21.7243
2	40.0	32.5395
3	61.0	45.5619
4	67.0	49.4772
...
3299	54.0	32.4274
3300	71.0	44.6095
3301	86.0	51.4474
3302	104.0	61.8126
3303	122.0	73.8271

	segment_osrm_time_sum
0	11.0
1	20.0
2	27.0
3	39.0
4	44.0
...	...
3299	26.0
3300	34.0
3301	38.0
3302	45.0
3303	53.0

[3304 rows x 3 columns]

```

create_segment_dict = {
    'data' : 'first',
    'trip_creation_time' : 'first',
    'route_schedule_uuid' : 'first',
    'route_type' : 'first',
    'trip_uuid' : 'first',
    'source_center' : 'first',
    'source_name' : 'first',

    'destination_center' : 'last',

```

```

'destination_name' : 'last',
'od_start_time' : 'first',
'od_end_time' : 'first',
'start_scan_to_end_scan' : 'first',

'actual_distance_to_destination' : 'last',
'actual_time' : 'last',

'osrm_time' : 'last',
'osrm_distance' : 'last',

'segment_actual_time_sum' : 'last',
'segment_osrm_distance_sum' : 'last',
'segment_osrm_time_sum' : 'last',
}

```

```

segment =
df.groupby('segment_key').agg(create_segment_dict).reset_index()
segment = segment.sort_values(by=['segment_key', 'od_end_time'],
ascending=True).reset_index()

```

segment

	index	segment_key	data
\			
0	0	trip-153671191949943656IND487001AABIND487551AAA	training
1	1	trip-153671191949943656IND487551AAAIND464668AAA	training
2	2	trip-153671320807895983IND121004AABIND121102AAA	training
3	3	trip-153671320807895983IND121102AAAIND121004AAB	training
4	4	trip-153671602871109556IND362001AAAIND362220AAA	training
..
707	707	trip-153860909786408355IND384205AAAIND384170AAA	test
708	708	trip-153860909786408355IND384265AABIND385320AAA	test
709	709	trip-153860909786408355IND385320AAAIND385340AAB	test
710	710	trip-153860960901509071IND424006AAAIND424101AAA	test
711	711	trip-153861034802474617IND245101AAAIND201007AAA	test
trip_creation_time			
route_schedule_uuid \			
0	9/12/18 12:25 AM	thanos::sroute:0ac760f3-96cb-4046-bfd0-	


```

8bc4678...
1    9/12/18 12:25 AM  thanos::sroute:0ac760f3-96cb-4046-bfd0-
8bc4678...
2    9/12/18 12:46 AM  thanos::sroute:8c5ab716-198a-4395-b83f-
5672773...
3    9/12/18 12:46 AM  thanos::sroute:8c5ab716-198a-4395-b83f-
5672773...
4    9/12/18 1:33 AM  thanos::sroute:5f7d8d49-ae14-430e-9333-
37361e1...
..
...
707  10/3/18 11:24 PM  thanos::sroute:e7281daf-3cdf-4dc6-be95-
95266c9...
708  10/3/18 11:24 PM  thanos::sroute:e7281daf-3cdf-4dc6-be95-
95266c9...
709  10/3/18 11:24 PM  thanos::sroute:e7281daf-3cdf-4dc6-be95-
95266c9...
710  10/3/18 11:33 PM  thanos::sroute:a02b2c7d-49c4-4c7f-956e-
b4b22a1...
711  10/3/18 11:45 PM  thanos::sroute:1396edcd-faad-4029-a574-
f71a85a...

```

	route_type	trip_uuid	source_center \
0	FTL	trip-153671191949943656	IND487001AAB
1	FTL	trip-153671191949943656	IND487551AAA
2	Carting	trip-153671320807895983	IND121004AAB
3	Carting	trip-153671320807895983	IND121102AAA
4	Carting	trip-153671602871109556	IND362001AAA
..
707	FTL	trip-153860909786408355	IND384205AAA
708	FTL	trip-153860909786408355	IND384265AAB
709	FTL	trip-153860909786408355	IND385320AAA
710	Carting	trip-153860960901509071	IND424006AAA
711	Carting	trip-153861034802474617	IND245101AAA

	source_name	destination_center ... \
0	Narsinghpur_KndliDPP_D (Madhya Pradesh)	IND487551AAA ...
1	Gadarwara_MPward_D (Madhya Pradesh)	IND464668AAA ...
2	FBD_Balabgarh_DPC (Haryana)	IND121102AAA ...
3	Palwal (Haryana)	IND121004AAB ...
4	Junagadh_DPC (Gujarat)	IND362220AAA ...
..
707	Mehsana_Panchot_IP (Gujarat)	IND384170AAA ...

708	Patan_Central_D_2 (Gujarat)	IND385320AAA ...
709	Bhabhar_TirupDPP_D (Gujarat)	IND385340AAB ...
710	Dhule_MIDCAvdn_I (Maharashtra)	IND424101AAA ...
711	Hapur_Swargash_D (Uttar Pradesh)	IND201007AAA ...

	od_start_time	od_end_time	start_scan_to_end_scan \
0	2018-09-12 04:33:00	2018-09-12 05:59:00	86.0
1	2018-09-12 05:59:00	2018-09-12 09:23:00	204.0
2	2018-09-12 00:46:00	2018-09-12 01:53:00	66.0
3	2018-09-12 01:53:00	2018-09-12 05:07:00	193.0
4	2018-09-12 01:33:00	2018-09-12 02:42:00	68.0
...
707	2018-10-03 23:24:00	2018-10-04 00:21:00	57.0
708	2018-10-04 00:53:00	2018-10-04 02:53:00	119.0
709	2018-10-04 02:53:00	2018-10-04 03:30:00	36.0
710	2018-10-03 23:33:00	2018-10-04 02:01:00	148.0
711	2018-10-04 02:54:00	2018-10-04 05:23:00	148.0

	actual_distance_to_destination	actual_time	osrm_time
osrm_distance \			
0	42.630333	78.0	42.0
48.4129			
1	57.345262	175.0	63.0
76.0934			
2	37.859167	45.0	28.0
39.7245			
3	38.372339	100.0	29.0
40.2548			
4	28.340661	43.0	22.0
31.1283			
...
...			
707	22.989321	50.0	28.0
28.2656			
708	60.289747	86.0	75.0
83.5911			
709	27.199984	36.0	37.0
29.7998			
710	49.732416	128.0	41.0
57.1276			
711	44.106290	123.0	50.0
46.5093			

segment_actual_time_sum	segment_osrm_distance_sum
segment_osrm_time_sum	

0	78.0	48.4129
41.0		
1	174.0	92.1494
91.0		
2	44.0	39.7244
27.0		
3	98.0	40.4251
28.0		
4	42.0	37.8163
33.0		
..
...		
707	50.0	28.2656
28.0		
708	85.0	78.2801
82.0		
709	36.0	29.7998
36.0		
710	127.0	67.7456
47.0		
711	121.0	48.5795
50.0		

[712 rows x 21 columns]

segment[segment['trip_uuid'] == 'trip-153741093647649320']

	index	segment_key	data
\			
294	294	trip-153741093647649320IND388121AAAIND388620AAB	training
295	295	trip-153741093647649320IND388620AABIND388320AAA	training

	trip_creation_time	route_schedule_uuid	\
294	9/20/18 2:35 AM	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	
295	9/20/18 2:35 AM	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	

	route_type	trip_uuid	source_center	\
294	Carting	trip-153741093647649320	IND388121AAA	
295	Carting	trip-153741093647649320	IND388620AAB	

	source_name	destination_center	...	\
294	Anand_VUNagar_DC (Gujarat)	IND388620AAB	...	
295	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	...	

	od_start_time	od_end_time	start_scan_to_end_scan	\
--	---------------	-------------	------------------------	---

294	2018-09-20 03:21:00	2018-09-20 04:47:00	86.0
295	2018-09-20 04:47:00	2018-09-20 06:36:00	109.0

	actual_distance_to_destination	actual_time	osrm_time
osrm_distance \			
294	39.386040	68.0	44.0
54.2181			
295	43.595802	102.0	45.0
53.2334			

	segment_actual_time_sum	segment_osrm_distance_sum
segment_osrm_time_sum		
294	67.0	49.4772
44.0		
295	100.0	53.2334
44.0		

[2 rows x 21 columns]

segment.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 712 entries, 0 to 711

Data columns (total 21 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	index	712 non-null	int64
1	segment_key	712 non-null	object
2	data	712 non-null	object
3	trip_creation_time	712 non-null	object
4	route_schedule_uuid	712 non-null	object
5	route_type	712 non-null	object
6	trip_uuid	712 non-null	object
7	source_center	712 non-null	object
8	source_name	712 non-null	object
9	destination_center	712 non-null	object
10	destination_name	712 non-null	object
11	od_start_time	712 non-null	datetime64[ns]
12	od_end_time	712 non-null	datetime64[ns]
13	start_scan_to_end_scan	712 non-null	float64
14	actual_distance_to_destination	712 non-null	float64
15	actual_time	712 non-null	float64
16	osrm_time	712 non-null	float64
17	osrm_distance	712 non-null	float64
18	segment_actual_time_sum	712 non-null	float64
19	segment_osrm_distance_sum	712 non-null	float64
20	segment_osrm_time_sum	712 non-null	float64

dtypes: datetime64[ns](2), float64(8), int64(1), object(10)

memory usage: 116.9+ KB

Calculate time taken between od_start_time and od_end_time and keep it as a feature.

od_time_diff_hour is matching with start_scan_to_end_time

```
segment['od_time_diff_hour'] = (segment['od_end_time'] -
segment['od_start_time']).dt.total_seconds() /(60)
segment['od_time_diff_hour']
```

```
0      86.0
1     204.0
2      67.0
3     194.0
4      69.0
...
707    57.0
708   120.0
709    37.0
710   148.0
711   149.0
```

Name: od_time_diff_hour, Length: 712, dtype: float64

segment

	index	segment_key	data
\			
0	0	trip-153671191949943656IND487001AABIND487551AAA	training
1	1	trip-153671191949943656IND487551AAAIND464668AAA	training
2	2	trip-153671320807895983IND121004AABIND121102AAA	training
3	3	trip-153671320807895983IND121102AAAIND121004AAB	training
4	4	trip-153671602871109556IND362001AAAIND362220AAA	training
..
707	707	trip-153860909786408355IND384205AAAIND384170AAA	test
708	708	trip-153860909786408355IND384265AABIND385320AAA	test
709	709	trip-153860909786408355IND385320AAAIND385340AAB	test
710	710	trip-153860960901509071IND424006AAAIND424101AAA	test
711	711	trip-153861034802474617IND245101AAAIND201007AAA	test

```
trip_creation_time
route_schedule_uuid \
0 9/12/18 12:25 AM thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678...
```

```

1      9/12/18 12:25 AM  thanos::sroute:0ac760f3-96cb-4046-bfd0-
8bc4678...
2      9/12/18 12:46 AM  thanos::sroute:8c5ab716-198a-4395-b83f-
5672773...
3      9/12/18 12:46 AM  thanos::sroute:8c5ab716-198a-4395-b83f-
5672773...
4      9/12/18 1:33 AM  thanos::sroute:5f7d8d49-ae14-430e-9333-
37361e1...
..
...
707    10/3/18 11:24 PM  thanos::sroute:e7281daf-3cdf-4dc6-be95-
95266c9...
708    10/3/18 11:24 PM  thanos::sroute:e7281daf-3cdf-4dc6-be95-
95266c9...
709    10/3/18 11:24 PM  thanos::sroute:e7281daf-3cdf-4dc6-be95-
95266c9...
710    10/3/18 11:33 PM  thanos::sroute:a02b2c7d-49c4-4c7f-956e-
b4b22a1...
711    10/3/18 11:45 PM  thanos::sroute:1396edcd-faad-4029-a574-
f71a85a...

```

	route_type	trip_uuid	source_center \
0	FTL	trip-153671191949943656	IND487001AAB
1	FTL	trip-153671191949943656	IND487551AAA
2	Carting	trip-153671320807895983	IND121004AAB
3	Carting	trip-153671320807895983	IND121102AAA
4	Carting	trip-153671602871109556	IND362001AAA
..
707	FTL	trip-153860909786408355	IND384205AAA
708	FTL	trip-153860909786408355	IND384265AAB
709	FTL	trip-153860909786408355	IND385320AAA
710	Carting	trip-153860960901509071	IND424006AAA
711	Carting	trip-153861034802474617	IND245101AAA

	source_name	destination_center ... \
0	Narsinghpur_KndliDPP_D (Madhya Pradesh)	IND487551AAA ...
1	Gadarwara_MPward_D (Madhya Pradesh)	IND464668AAA ...
2	FBD_Balabharh_DPC (Haryana)	IND121102AAA ...
3	Palwal (Haryana)	IND121004AAB ...
4	Junagadh_DPC (Gujarat)	IND362220AAA ...
..
707	Mehsana_Panchot_IP (Gujarat)	IND384170AAA ...

708	Patan_Central_D_2 (Gujarat)	IND385320AAA ...
709	Bhabhar_TirupDPP_D (Gujarat)	IND385340AAB ...
710	Dhule_MIDCAvdn_I (Maharashtra)	IND424101AAA ...
711	Hapur_Swargash_D (Uttar Pradesh)	IND201007AAA ...

	od_end_time	start_scan_to_end_scan	actual_distance_to_destination \
0	2018-09-12 05:59:00		86.0
			42.630333
1	2018-09-12 09:23:00		204.0
			57.345262
2	2018-09-12 01:53:00		66.0
			37.859167
3	2018-09-12 05:07:00		193.0
			38.372339
4	2018-09-12 02:42:00		68.0
			28.340661
..
...			
707	2018-10-04 00:21:00		57.0
			22.989321
708	2018-10-04 02:53:00		119.0
			60.289747
709	2018-10-04 03:30:00		36.0
			27.199984
710	2018-10-04 02:01:00		148.0
			49.732416
711	2018-10-04 05:23:00		148.0
			44.106290

	actual_time	osrm_time	osrm_distance	segment_actual_time_sum \
0	78.0	42.0	48.4129	78.0
1	175.0	63.0	76.0934	174.0
2	45.0	28.0	39.7245	44.0
3	100.0	29.0	40.2548	98.0
4	43.0	22.0	31.1283	42.0
..
707	50.0	28.0	28.2656	50.0
708	86.0	75.0	83.5911	85.0
709	36.0	37.0	29.7998	36.0
710	128.0	41.0	57.1276	127.0
711	123.0	50.0	46.5093	121.0

	segment_osrm_distance_sum	segment_osrm_time_sum
od_time_diff_hour		
0	48.4129	41.0

86.0		
1	92.1494	91.0
204.0		
2	39.7244	27.0
67.0		
3	40.4251	28.0
194.0		
4	37.8163	33.0
69.0		
..
..		
707	28.2656	28.0
57.0		
708	78.2801	82.0
120.0		
709	29.7998	36.0
37.0		
710	67.7456	47.0
148.0		
711	48.5795	50.0
149.0		

[712 rows x 22 columns]

```
create_trip_dict = {
    'data' : 'first',
    'trip_creation_time' : 'first',
    'route_schedule_uuid' : 'first',
    'route_type' : 'first',
    'trip_uuid' : 'first',

    'source_center' : 'first',
    'source_name' : 'first',

    'destination_center' : 'last',
    'destination_name' : 'last',
    'start_scan_to_end_scan' : 'sum',
    'od_time_diff_hour' : 'sum',

    'actual_distance_to_destination' : 'sum',
    'actual_time' : 'sum',
    'osrm_time' : 'sum',
    'osrm_distance' : 'sum',

    'segment_actual_time_sum' : 'sum',
    'segment_osrm_distance_sum' : 'sum',
    'segment_osrm_time_sum' : 'sum',
}
```



```
trip =
segment.groupby('trip_uuid').agg(create_trip_dict).reset_index(drop =
True)
```

```
trip
```

	data	trip_creation_time	\
0	training	9/12/18 12:25 AM	
1	training	9/12/18 12:46 AM	
2	training	9/12/18 1:33 AM	
3	training	9/12/18 5:28 AM	
4	training	9/12/18 6:56 AM	
..	
386	test	10/3/18 9:44 PM	
387	test	10/3/18 9:55 PM	
388	test	10/3/18 11:24 PM	
389	test	10/3/18 11:33 PM	
390	test	10/3/18 11:45 PM	

	route_schedule_uuid	route_type	\
0	thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678...	FTL	
1	thanos::sroute:8c5ab716-198a-4395-b83f-5672773...	Carting	
2	thanos::sroute:5f7d8d49-ae14-430e-9333-37361e1...	Carting	
3	thanos::sroute:9af141fe-eb0e-4f9b-baab-4937945...	FTL	
4	thanos::sroute:8786609b-8671-4d19-ae85-0f43551...	Carting	
..	
386	thanos::sroute:c427d4a6-205a-4c62-be78-7dac143...	FTL	
387	thanos::sroute:c03e37c2-6c4b-4eee-9704-15fd85f...	Carting	
388	thanos::sroute:e7281daf-3cdf-4dc6-be95-95266c9...	FTL	
389	thanos::sroute:a02b2c7d-49c4-4c7f-956e-b4b22a1...	Carting	
390	thanos::sroute:1396edcd-faad-4029-a574-f71a85a...	Carting	

	trip_uuid	source_center	\
0	trip-153671191949943656	IND487001AAB	
1	trip-153671320807895983	IND121004AAB	
2	trip-153671602871109556	IND362001AAA	
3	trip-153673012557297456	IND637301AAA	
4	trip-153673541039568251	IND421302AAN	
..	
386	trip-153860305272724787	IND842001AAA	
387	trip-153860373904534393	IND380006AAA	
388	trip-153860909786408355	IND384170AAA	
389	trip-153860960901509071	IND424006AAA	
390	trip-153861034802474617	IND245101AAA	

	source_name	destination_center	\
0	Narsinghpur_KndliDPP_D (Madhya Pradesh)	IND464668AAA	
1	FBD_Balabgarh_DPC (Haryana)	IND121004AAB	
2	Junagadh_DPC (Gujarat)	IND362560AAA	
3	Sankari_Trchngrd_D (Tamil Nadu)	IND638001AAA	

4	GreaterThane_Anjur_C (Maharashtra)	IND400072AAB
...
386	Muzaffrpur_Bbganj_I (Bihar)	IND842001AAA
387	Ahmedabad_Paldi_D (Gujarat)	IND382430AAB
388	Unjha_DC (Gujarat)	IND385340AAB
389	Dhule_MIDCAvdn_I (Maharashtra)	IND424101AAA
390	Hapur_Swargash_D (Uttar Pradesh)	IND201007AAA
	destination_name	start_scan_to_end_scan \
0	Bareli_SourvDPP_D (Madhya Pradesh)	290.0
1	FBD_Balabhgarh_DPC (Haryana)	259.0
2	Una_Mamlatdr_DC (Gujarat)	473.0
3	Erode (Tamil Nadu)	36.0
4	Mumbai Hub (Maharashtra)	122.0
...
386	Muzaffrpur_Bbganj_I (Bihar)	930.0
387	Ahmedabad_East_H_1 (Gujarat)	46.0
388	Radhanpur_Santalpr_D (Gujarat)	243.0
389	Chalisgaon_BhadgDPP_D (Maharashtra)	148.0
390	GZB_Mohan_Nagar_DPC (Uttar Pradesh)	148.0
	od_time_diff_hour	actual_distance_to_destination actual_time \
0	290.0	99.975595 253.0
1	261.0	76.231506 145.0
2	476.0	208.394342 398.0
3	36.0	20.086304 36.0
4	123.0	20.410300 52.0
...
386	931.0	305.777339 584.0
387	46.0	11.211915 16.0
388	246.0	135.029643 203.0
389	148.0	49.732416 128.0
390	149.0	44.106290 123.0
	osrm_time	osrm_distance segment_actual_time_sum \
0	105.0	124.5063 252.0
1	57.0	79.9793 142.0
2	209.0	255.2054 391.0
3	17.0	21.0327 36.0
4	22.0	28.1662 51.0
...
386	293.0	380.0932 581.0
387	11.0	14.7688 16.0
388	164.0	168.7758 202.0
389	41.0	57.1276 127.0
390	50.0	46.5093 121.0
	segment_osrm_distance_sum	segment_osrm_time_sum
0	140.5623	132.0
1	80.1495	55.0

2	257.6815	215.0
3	21.0327	17.0
4	28.1662	21.0
..
386	428.2587	322.0
387	15.1622	13.0
388	163.4647	170.0
389	67.7456	47.0
390	48.5795	50.0

[391 rows x 18 columns]

trip[['actual_time', 'segment_actual_time_sum']]

	actual_time	segment_actual_time_sum
0	253.0	252.0
1	145.0	142.0
2	398.0	391.0
3	36.0	36.0
4	52.0	51.0
..
386	584.0	581.0
387	16.0	16.0
388	203.0	202.0
389	128.0	127.0
390	123.0	121.0

[391 rows x 2 columns]

trip

	data	trip_creation_time	\
0	training	9/12/18 12:25 AM	
1	training	9/12/18 12:46 AM	
2	training	9/12/18 1:33 AM	
3	training	9/12/18 5:28 AM	
4	training	9/12/18 6:56 AM	
..	
386	test	10/3/18 9:44 PM	
387	test	10/3/18 9:55 PM	
388	test	10/3/18 11:24 PM	
389	test	10/3/18 11:33 PM	
390	test	10/3/18 11:45 PM	

	route_schedule_uuid	route_type	\
0	thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678...	FTL	
1	thanos::sroute:8c5ab716-198a-4395-b83f-5672773...	Carting	
2	thanos::sroute:5f7d8d49-ae14-430e-9333-37361e1...	Carting	
3	thanos::sroute:9af141fe-eb0e-4f9b-baab-4937945...	FTL	
4	thanos::sroute:8786609b-8671-4d19-ae85-0f43551...	Carting	

```

..
386 thanos::sroute:c427d4a6-205a-4c62-be78-7dac143... FTL
387 thanos::sroute:c03e37c2-6c4b-4eee-9704-15fd85f... Carting
388 thanos::sroute:e7281daf-3cdf-4dc6-be95-95266c9... FTL
389 thanos::sroute:a02b2c7d-49c4-4c7f-956e-b4b22a1... Carting
390 thanos::sroute:1396edcd-faad-4029-a574-f71a85a... Carting

```

```

trip_uuid source_center \
0 trip-153671191949943656 IND487001AAB
1 trip-153671320807895983 IND121004AAB
2 trip-153671602871109556 IND362001AAA
3 trip-153673012557297456 IND637301AAA
4 trip-153673541039568251 IND421302AAN
..
386 trip-153860305272724787 IND842001AAA
387 trip-153860373904534393 IND380006AAA
388 trip-153860909786408355 IND384170AAA
389 trip-153860960901509071 IND424006AAA
390 trip-153861034802474617 IND245101AAA

```

```

source_name destination_center \
0 Narsinghpur_KndliDPP_D (Madhya Pradesh) IND464668AAA
1 FBD_Balabgarh_DPC (Haryana) IND121004AAB
2 Junagadh_DPC (Gujarat) IND362560AAA
3 Sankari_Trchngrd_D (Tamil Nadu) IND638001AAA
4 GreaterThane_Anjur_C (Maharashtra) IND400072AAB
..
386 Muzaffrpur_Bbganj_I (Bihar) IND842001AAA
387 Ahmedabad_Paldi_D (Gujarat) IND382430AAB
388 Unjha_DC (Gujarat) IND385340AAB
389 Dhule_MIDCAvdn_I (Maharashtra) IND424101AAA
390 Hapur_Swargash_D (Uttar Pradesh) IND201007AAA

```

```

destination_name start_scan_to_end_scan \
0 Bareli_SourvDPP_D (Madhya Pradesh) 290.0
1 FBD_Balabgarh_DPC (Haryana) 259.0
2 Una_Mamlatdr_DC (Gujarat) 473.0
3 Erode (Tamil Nadu) 36.0
4 Mumbai Hub (Maharashtra) 122.0
..
386 Muzaffrpur_Bbganj_I (Bihar) 930.0
387 Ahmedabad_East_H_1 (Gujarat) 46.0
388 Radhanpur_Santalpr_D (Gujarat) 243.0
389 Chalisgaon_BhadgDPP_D (Maharashtra) 148.0
390 GZB_Mohan_Nagar_DPC (Uttar Pradesh) 148.0

```

```

od_time_diff_hour actual_distance_to_destination actual_time \
0 290.0 99.975595 253.0
1 261.0 76.231506 145.0
2 476.0 208.394342 398.0

```

3	36.0	20.086304	36.0
4	123.0	20.410300	52.0
..
386	931.0	305.777339	584.0
387	46.0	11.211915	16.0
388	246.0	135.029643	203.0
389	148.0	49.732416	128.0
390	149.0	44.106290	123.0

	osrm_time	osrm_distance	segment_actual_time_sum \
0	105.0	124.5063	252.0
1	57.0	79.9793	142.0
2	209.0	255.2054	391.0
3	17.0	21.0327	36.0
4	22.0	28.1662	51.0
..
386	293.0	380.0932	581.0
387	11.0	14.7688	16.0
388	164.0	168.7758	202.0
389	41.0	57.1276	127.0
390	50.0	46.5093	121.0

	segment_osrm_distance_sum	segment_osrm_time_sum
0	140.5623	132.0
1	80.1495	55.0
2	257.6815	215.0
3	21.0327	17.0
4	28.1662	21.0
..
386	428.2587	322.0
387	15.1622	13.0
388	163.4647	170.0
389	67.7456	47.0
390	48.5795	50.0

[391 rows x 18 columns]

```
trip[trip['trip_uuid'] == 'trip-153741093647649320' ]
```

	data	trip_creation_time \
162	training	9/20/18 2:35 AM

	route_schedule_uuid	route_type \
162	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting

	trip_uuid	source_center	source_name
\			
162	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)

	destination_center	destination_name
start_scan_to_end_scan \		
162	IND388320AAA	Anand_Vaghasi_IP (Gujarat)
195.0		

	od_time_diff_hour	actual_distance_to_destination	actual_time \
162	195.0	82.981842	170.0

	osrm_time	osrm_distance	segment_actual_time_sum \
162	89.0	107.4515	167.0

	segment_osrm_distance_sum	segment_osrm_time_sum
162	102.7106	88.0


```
trip[['actual_distance_to_destination', 'osrm_distance']]
```

	actual_distance_to_destination	osrm_distance
0	99.975595	124.5063
1	76.231506	79.9793
2	208.394342	255.2054
3	20.086304	21.0327
4	20.410300	28.1662
..
386	305.777339	380.0932
387	11.211915	14.7688
388	135.029643	168.7758
389	49.732416	57.1276
390	44.106290	46.5093


```
[391 rows x 2 columns]
```

Hypothesis Testing --> Recommendation to the business¶

```
trip['destination_name'] = trip['destination_name'].str.lower() #
lowering all columns
trip['source_name'] = trip['source_name']

def place2state(x):
    # transform "gurgaon_bilaspur_hb (haryana)" into "haryana"
    state = x.split('(')[1]

    return state[:-1] #removing ')' from ending

def place2city(x):
    #we will remove state
    city = x.split(' ')[0]

    city = city.split('_')[0]

    # Now daling with edge cases
```

```

if city == 'pnq vadgaon sheri dpc': return 'vadgaonsheri'
# ['PNQ Pashan DPC', 'Bhopal MP Nagar', 'HBR Layout PC',
#  'PNQ Rahatani DPC', 'Pune Balaji Nagar', 'Mumbai Antop Hill']

if city in ['pnq pashan dpc', 'pnq rahatani dpc', 'pune balaji
nagar']:
    return 'pune'

if city == 'hbr layout pc' :
    return 'bengaluru'
if city == 'bhopal mp nagar':
    return 'bhopal'
if city == 'mumbai antop hill':
    return 'mumbai'

return city
def place2city_place(x):

    # we will remove state
    x = x.split('(')[0]

    len_ = len(x.split('_'))

    if len_ >= 3:
        return x.split('_')[1]

    # small cities have same city and place name
    if len_ == 2:
        return x.split('_')[0]

    # now we need to deal with edge cases or improper name convention
    # if len(x.split('_')) == 2:

    return x.split(' ')[0]

def place2code(x):
    # we will remove state
    x = x.split('(')[0]
    if len(x.split('_')) >= 3:
        return x.split('_')[-1]

    return 'none'

trip['destination_state'] = trip['destination_name'].apply(lambda x:
place2state(x))
trip['destination_city'] = trip['destination_name'].apply(lambda x:
place2city(x))
trip['destination_place'] = trip['destination_name'].apply(lambda x:
place2city_place(x))

```

```
trip['destination_code'] = trip['destination_name'].apply(lambda x:
place2code(x))
```

```
trip[['destination_state','destination_city','destination_place','dest
ination_code']]
```

	destination_state	destination_city	destination_place
destination_code			
0	madhya pradesh	bareli	sourvdpp
d			
1	haryana	fbd	balabhgarh
dpc			
2	gujarat	una	mamlatdr
dc			
3	tamil nadu	erode	erode
none			
4	maharashtra	mumbai hub	mumbai
none			
..
...			
386	bihar	muzaffrpur	bbganj
i			
387	gujarat	ahmedabad	east
1			
388	gujarat	radhanpur	santalpr
d			
389	maharashtra	chalisgaon	bhadgdpp
d			
390	uttar pradesh	gzb	mohan
dpc			

```
[391 rows x 4 columns]
```

```
trip['trip_creation_time'] =
pd.to_datetime(trip['trip_creation_time'])
```

```
trip['trip_year'] = trip['trip_creation_time'].dt.year
trip['trip_month'] = trip['trip_creation_time'].dt.month
trip['trip_hour'] = trip['trip_creation_time'].dt.hour
trip['trip_day'] = trip['trip_creation_time'].dt.day
trip['trip_week'] = trip['trip_creation_time'].dt.isocalendar().week
trip['trip_dayofweek'] = trip['trip_creation_time'].dt.dayofweek
```

```
trip[['trip_year','trip_month','trip_hour','trip_day','trip_week','tri
p_dayofweek']]
```

	trip_year	trip_month	trip_hour	trip_day	trip_week
trip_dayofweek					
0	2018	9	0	12	37
2					
1	2018	9	0	12	37


```

2
2      2018      9      1      12      37
2
3      2018      9      5      12      37
2
4      2018      9      6      12      37
2
..      ...      ...      ...      ...      ...
...
386     2018      10     21      3      40
2
387     2018      10     21      3      40
2
388     2018      10     23      3      40
2
389     2018      10     23      3      40
2
390     2018      10     23      3      40
2

```

[391 rows x 6 columns]

trip.head()

```

      data  trip_creation_time  \
0  training 2018-09-12 00:25:00
1  training 2018-09-12 00:46:00
2  training 2018-09-12 01:33:00
3  training 2018-09-12 05:28:00
4  training 2018-09-12 06:56:00

```

```

                                route_schedule_uuid route_type  \
0  thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678...      FTL
1  thanos::sroute:8c5ab716-198a-4395-b83f-5672773...    Carting
2  thanos::sroute:5f7d8d49-ae14-430e-9333-37361e1...    Carting
3  thanos::sroute:9af141fe-eb0e-4f9b-baab-4937945...      FTL
4  thanos::sroute:8786609b-8671-4d19-ae85-0f43551...    Carting

```

```

      trip_uuid source_center  \
0  trip-153671191949943656  IND487001AAB
1  trip-153671320807895983  IND121004AAB
2  trip-153671602871109556  IND362001AAA
3  trip-153673012557297456  IND637301AAA
4  trip-153673541039568251  IND421302AAN

```

```

                                source_name destination_center  \
0  Narsinghpur_KndliDPP_D (Madhya Pradesh)  IND464668AAA
1  FBD_Balabharhgarh_DPC (Haryana)        IND121004AAB
2  Junagadh_DPC (Gujarat)                 IND362560AAA
3  Sankari_Trchngrd_D (Tamil Nadu)        IND638001AAA

```

```

4      GreaterThane_Anjur_C (Maharashtra)      IND400072AAB

      destination_name  start_scan_to_end_scan  ...  \
0  bareli_sourvdpp_d (madhya pradesh)          290.0  ...
1    fbd_balabhgarh_dpc (haryana)              259.0  ...
2    una_mamlatdr_dc (gujarat)                 473.0  ...
3      erode (tamil nadu)                      36.0  ...
4    mumbai hub (maharashtra)                 122.0  ...

      destination_state  destination_city  destination_place
destination_code  \
0      madhya pradesh      bareli      sourvdpp
d
1      haryana      fbd      balabhgarh
dpc
2      gujarat      una      mamlatdr
dc
3      tamil nadu      erode      erode
none
4      maharashtra      mumbai hub      mumbai
none

      trip_year  trip_month  trip_hour  trip_day  trip_week  trip_dayofweek
0      2018      9      0      12      37      2
1      2018      9      0      12      37      2
2      2018      9      1      12      37      2
3      2018      9      5      12      37      2
4      2018      9      6      12      37      2

[5 rows x 28 columns]

num_cols =
['start_scan_to_end_scan', 'actual_distance_to_destination', 'actual_time',
'osrm_time',

'osrm_distance', 'segment_actual_time_sum', 'segment_osrm_distance_sum',
'segment_osrm_time_sum', 'od_time_diff_hour']

```

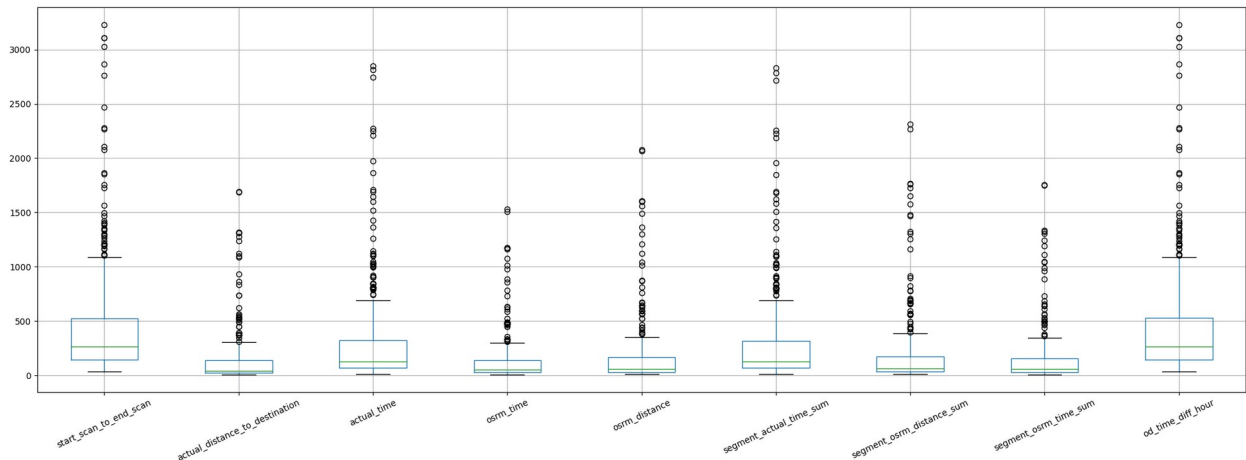
Find outliers in numericals variable (you might find outliers in almost all the variables), and visualize it using visual analysis

```

trip[num_cols].boxplot(rot=25, figsize=(25,8))

<Axes: >

```



Handle the outliers using IQR method

```
Q1 = trip[num_cols].quantile(0.25)
Q3 = trip[num_cols].quantile(0.75)

IQR = Q3 - Q1

trip = trip[-((trip[num_cols] < (Q1 - 1.5 * IQR)) | (trip[num_cols] >
(Q3 + 1.5 * IQR))).any(axis=1)]
trip = trip.reset_index(drop=True)
```

trip

	data	trip_creation_time	\
0	training	2018-09-12 00:25:00	
1	training	2018-09-12 00:46:00	
2	training	2018-09-12 01:33:00	
3	training	2018-09-12 05:28:00	
4	training	2018-09-12 06:56:00	
...	
328	test	2018-10-03 21:13:00	
329	test	2018-10-03 21:55:00	
330	test	2018-10-03 23:24:00	
331	test	2018-10-03 23:33:00	
332	test	2018-10-03 23:45:00	

	route_schedule_uuid	route_type	\
0	thanos::sroute:0ac760f3-96cb-4046-bfd0-8bc4678...	FTL	
1	thanos::sroute:8c5ab716-198a-4395-b83f-5672773...	Carting	
2	thanos::sroute:5f7d8d49-ae14-430e-9333-37361e1...	Carting	
3	thanos::sroute:9af141fe-eb0e-4f9b-baab-4937945...	FTL	
4	thanos::sroute:8786609b-8671-4d19-ae85-0f43551...	Carting	
...	
328	thanos::sroute:88356412-d28b-47fe-94ce-847266b...	Carting	
329	thanos::sroute:c03e37c2-6c4b-4eee-9704-15fd85f...	Carting	
330	thanos::sroute:e7281daf-3cdf-4dc6-be95-95266c9...	FTL	

331	thanos::sroute:a02b2c7d-49c4-4c7f-956e-b4b22a1...	Carting
332	thanos::sroute:1396edcd-faad-4029-a574-f71a85a...	Carting

	trip_uuid	source_center	\
0	trip-153671191949943656	IND487001AAB	
1	trip-153671320807895983	IND121004AAB	
2	trip-153671602871109556	IND362001AAA	
3	trip-153673012557297456	IND637301AAA	
4	trip-153673541039568251	IND421302AAN	
..	
328	trip-153860120404625095	IND000000AFJ	
329	trip-153860373904534393	IND380006AAA	
330	trip-153860909786408355	IND384170AAA	
331	trip-153860960901509071	IND424006AAA	
332	trip-153861034802474617	IND245101AAA	

	source_name	destination_center	\
0	Narsinghpur_KndliDPP_D (Madhya Pradesh)	IND464668AAA	
1	FBD_Balabharh_DPC (Haryana)	IND121004AAB	
2	Junagadh_DPC (Gujarat)	IND362560AAA	
3	Sankari_Trchngrd_D (Tamil Nadu)	IND638001AAA	
4	GreaterThane_Anjur_C (Maharashtra)	IND400072AAB	
..	
328	Trivandrum_Pettah_D (Kerala)	IND695541AAA	
329	Ahmedabad_Paldi_D (Gujarat)	IND382430AAB	
330	Unjha_DC (Gujarat)	IND385340AAB	
331	Dhule_MIDCAvdn_I (Maharashtra)	IND424101AAA	
332	Hapur_Swargash_D (Uttar Pradesh)	IND201007AAA	

	destination_name	start_scan_to_end_scan	...
\			
0	bareli_sourvdpp_d (madhya pradesh)	290.0	...
1	fbd_balabharh_dpc (haryana)	259.0	...
2	una_mamlatdr_dc (gujarat)	473.0	...
3	erode (tamil nadu)	36.0	...
4	mumbai hub (maharashtra)	122.0	...
..
328	nedumangad_arsprmbu_d (kerala)	181.0	...
329	ahmedabad_east_h_1 (gujarat)	46.0	...
330	radhanpur_santalpr_d (gujarat)	243.0	...
331	chalisgaon_bhadgdpp_d (maharashtra)	148.0	...

332 gzb_mohan_nagar_dpc (uttar pradesh) 148.0 ...

	destination_state	destination_city	destination_place
destination_code \			
0	madhya pradesh	bareli	sourvdpp
d			
1	haryana	fbd	balabhgarh
dpc			
2	gujarat	una	mamlatdr
dc			
3	tamil nadu	erode	erode
none			
4	maharashtra	mumbai hub	mumbai
none			
..
...			
328	kerala	nedumangad	arsprmbu
d			
329	gujarat	ahmedabad	east
1			
330	gujarat	radhanpur	santalpr
d			
331	maharashtra	chalisgaon	bhadgdpp
d			
332	uttar pradesh	gzb	mohan
dpc			

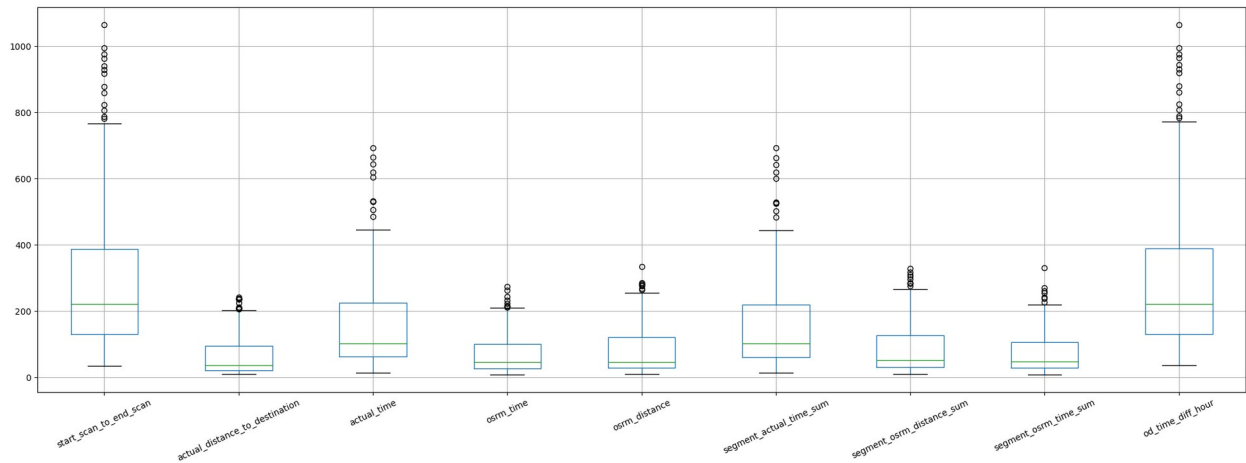
	trip_year	trip_month	trip_hour	trip_day	trip_week
trip_dayofweek					
0	2018	9	0	12	37
2					
1	2018	9	0	12	37
2					
2	2018	9	1	12	37
2					
3	2018	9	5	12	37
2					
4	2018	9	6	12	37
2					
..
..					
328	2018	10	21	3	40
2					
329	2018	10	21	3	40
2					
330	2018	10	23	3	40
2					
331	2018	10	23	3	40

```
2
332      2018      10      23      3      40
2
```

```
[333 rows x 28 columns]
```

```
trip[num_cols].boxplot(rot=25, figsize=(25,8))
```

```
<Axes: >
```



Handling Categorical Variables¶

Only two route_type-Do one hot encoding

```
trip['route_type'].value_counts()
```

```
Carting      245
```

```
FTL          88
```

```
Name: route_type, dtype: int64
```

```
trip['route_type'] = trip['route_type'].map({'FTL':0, 'Carting':1})
```

Normalize/Standarize the numerical features using MinMaxScaler or StandardScaler

```
from sklearn.preprocessing import StandardScaler
```

```
scaler = StandardScaler()
```

```
scaler.fit(trip[num_cols])
```

```
StandardScaler()
```

```
StandardScaler()
```

```
StandardScaler()
```

```
trip[num_cols] = scaler.transform(trip[num_cols])
```

trip[num_cols]

	start_scan_to_end_scan	actual_distance_to_destination	
actual_time \			
0	0.033685	0.650845	
0.737959			
1	-0.117304	0.238822	-
0.083432			
2	0.925004	2.532195	
1.840751			
3	-1.203448	-0.735444	-
0.912428			
4	-0.784576	-0.729822	-
0.790740			
..
..			
328	-0.497211	-0.390416	-
0.554971			
329	-1.154742	-0.889438	-
1.064537			
330	-0.195234	1.259124	
0.357685			
331	-0.657940	-0.221006	-
0.212725			
332	-0.657940	-0.318635	-
0.250752			

	osrm_time	osrm_distance	segment_actual_time_sum \
0	0.671847	0.666418	0.745997
1	-0.182133	0.028259	-0.095550
2	2.522136	2.539592	1.809406
3	-0.893783	-0.816561	-0.906495
4	-0.804827	-0.714324	-0.791738
..
328	-0.146551	-0.417360	-0.554575
329	-1.000530	-0.906335	-1.059503
330	1.721530	1.300887	0.363476
331	-0.466793	-0.299250	-0.210306
332	-0.306672	-0.451431	-0.256209

	segment_osrm_distance_sum	segment_osrm_time_sum	
od_time_diff_hour			
0	0.804396	0.972963	
0.029510			
1	-0.026977	-0.286897	-
0.111500			
2	2.416136	2.330995	
0.933924			
3	-0.840514	-0.908646	-
1.205549			

4	-0.742347	-0.843199	-
0.782517			
..
..			
328	-0.474470	-0.139641	-
0.490771			
329	-0.921302	-0.974093	-
1.156925			
330	1.119568	1.594713	-
0.184437			
331	-0.197673	-0.417792	-
0.660956			
332	-0.461428	-0.368706	-
0.656093			

[333 rows x 9 columns]

trip[num_cols].describe()

	start_scan_to_end_scan	actual_distance_to_destination
actual_time \		
count	3.330000e+02	3.330000e+02
3.330000e+02		
mean	-1.386945e-16	-5.467765e-17
6.401286e-17		
std	1.001505e+00	1.001505e+00
1.001505e+00		
min	-1.213189e+00	-9.248867e-01
1.087353e+00		
25%	-7.456113e-01	-7.191429e-01
7.146855e-01		
50%	-3.023867e-01	-4.608571e-01
4.104668e-01		
75%	5.061328e-01	5.623651e-01
5.174002e-01		
max	3.803529e+00	3.097330e+00
4.076759e+00		

	osrm_time	osrm_distance	segment_actual_time_sum \
count	3.330000e+02	3.330000e+02	3.330000e+02
mean	-1.733682e-17	1.013537e-16	-8.535048e-17
std	1.001505e+00	1.001505e+00	1.001505e+00
min	-1.053904e+00	-9.764595e-01	-1.082455e+00
25%	-7.158706e-01	-7.143237e-01	-7.228846e-01
50%	-3.778369e-01	-4.658019e-01	-4.015668e-01
75%	5.828904e-01	6.091763e-01	4.858823e-01
max	3.678567e+00	3.672716e+00	4.112183e+00

	segment_osrm_distance_sum	segment_osrm_time_sum
od_time_diff_hour		

count	3.330000e+02	3.330000e+02	
3.330000e+02			
mean	6.934726e-17	6.934726e-17	
8.001607e-17			
std	1.001505e+00	1.001505e+00	
1.001505e+00			
min	-9.940492e-01	-1.055903e+00	-
1.210412e+00			
25%	-7.223648e-01	-7.286662e-01	-
7.484798e-01			
50%	-4.324231e-01	-4.014297e-01	-
3.059978e-01			
75%	5.996850e-01	5.475560e-01	
5.060295e-01			
max	3.390920e+00	4.212604e+00	
3.797901e+00			

Recommendation Examples:

There is a significant difference between OSRM and actual parameters.

There is a need to: Revisit information fed to routing engine for trip planning. Check for discrepancies with transporters, if the routing engine is configured for optimum results.

North, South and West Zones corridors have significant traffic of orders. But, we have a smaller presence in Central, Eastern and North-Eastern zone. However it would be difficult to conclude this, by looking at just 2 months data. It is worth investigating and increasing our presence in these regions.

From state point of view, we have heavy traffic in Maharashtra followed by Karnataka. This is a good indicator that we need to plan for resources on ground in these 2 states on priority. Especially, during festive seasons.