

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib as plt
from sklearn.preprocessing import LabelEncoder
from category_encoders import TargetEncoder

#!pip install category_encoders

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

```

```
df.head()
```

	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	

	source_name	trip_uuid	source_center	\
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]

df.shape

(144867, 24)

df.columns

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'],
        dtype='object')

```

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144867 entries, 0 to 144866
Data columns (total 24 columns):

```

#	Column	Non-Null Count	Dtype
0	data	144867 non-null	object
1	trip_creation_time	144867 non-null	object
2	route_schedule_uuid	144867 non-null	object
3	route_type	144867 non-null	object
4	trip_uuid	144867 non-null	object
5	source_center	144867 non-null	object
6	source_name	144574 non-null	object
7	destination_center	144867 non-null	object
8	destination_name	144606 non-null	object
9	od_start_time	144867 non-null	object
10	od_end_time	144867 non-null	object
11	start_scan_to_end_scan	144867 non-null	float64
12	is_cutoff	144867 non-null	bool
13	cutoff_factor	144867 non-null	int64
14	cutoff_timestamp	144867 non-null	object
15	actual_distance_to_destination	144867 non-null	float64
16	actual_time	144867 non-null	float64
17	osrm_time	144867 non-null	float64
18	osrm_distance	144867 non-null	float64
19	factor	144867 non-null	float64
20	segment_actual_time	144867 non-null	float64
21	segment_osrm_time	144867 non-null	float64
22	segment_osrm_distance	144867 non-null	float64
23	segment_factor	144867 non-null	float64

```
dtypes: bool(1), float64(10), int64(1), object(12)
```

```
memory usage: 25.6+ MB
```

```
df.head()
```

```

      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]

```

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

```

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144867 entries, 0 to 144866
Data columns (total 24 columns):

```

#	Column	Non-Null Count	Dtype
0	data	144867 non-null	object
1	trip_creation_time	144867 non-null	object
2	route_schedule_uuid	144867 non-null	object
3	route_type	144867 non-null	object
4	trip_uuid	144867 non-null	object
5	source_center	144867 non-null	object
6	source_name	144574 non-null	object
7	destination_center	144867 non-null	object
8	destination_name	144606 non-null	object
9	od_start_time	144867 non-null	datetime64[ns]
10	od_end_time	144867 non-null	datetime64[ns]
11	start_scan_to_end_scan	144867 non-null	float64
12	is_cutoff	144867 non-null	bool
13	cutoff_factor	144867 non-null	int64
14	cutoff_timestamp	144867 non-null	object
15	actual_distance_to_destination	144867 non-null	float64
16	actual_time	144867 non-null	float64
17	osrm_time	144867 non-null	float64
18	osrm_distance	144867 non-null	float64
19	factor	144867 non-null	float64

```

20 segment_actual_time      144867 non-null float64
21 segment_osrm_time        144867 non-null float64
22 segment_osrm_distance    144867 non-null float64
23 segment_factor            144867 non-null float64
dtypes: bool(1), datetime64[ns](2), float64(10), int64(1), object(10)
memory usage: 25.6+ MB

```

```
df.isna().sum()
```

```

data                        0
trip_creation_time          0
route_schedule_uuid         0
route_type                  0
trip_uuid                   0
source_center               0
source_name                 293
destination_center          0
destination_name            261
od_start_time               0
od_end_time                 0
start_scan_to_end_scan      0
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

```

```
df.dropna(inplace=True)
```

```
df.drop_duplicates()
```

```

      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
...
144862 training  2018-09-20 16:24:28.436231
144863 training  2018-09-20 16:24:28.436231
144864 training  2018-09-20 16:24:28.436231
144865 training  2018-09-20 16:24:28.436231
144866 training  2018-09-20 16:24:28.436231

```

	route_type \	route_schedule_uuid	
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	
...			...
144862	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	
144863	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	
144864	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	
144865	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	
144866	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	

	source_name \	trip_uuid	source_center
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
...	
...			
144862	trip-153746066843555182 (Haryana)	IND131028AAB	Sonipat_Kundli_H
144863	trip-153746066843555182 (Haryana)	IND131028AAB	Sonipat_Kundli_H
144864	trip-153746066843555182 (Haryana)	IND131028AAB	Sonipat_Kundli_H
144865	trip-153746066843555182 (Haryana)	IND131028AAB	Sonipat_Kundli_H
144866	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H

(Haryana)

	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)
...
144862	IND000000ACB	Gurgaon_Bilaspur_HB	(Haryana)
144863	IND000000ACB	Gurgaon_Bilaspur_HB	(Haryana)
144864	IND000000ACB	Gurgaon_Bilaspur_HB	(Haryana)
144865	IND000000ACB	Gurgaon_Bilaspur_HB	(Haryana)
144866	IND000000ACB	Gurgaon_Bilaspur_HB	(Haryana)

	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	
...
144862	2018-09-20 16:24:28.436231	...	2018-09-20 21:57:20	
144863	2018-09-20 16:24:28.436231	...	2018-09-20 21:31:18	
144864	2018-09-20 16:24:28.436231	...	2018-09-20 21:11:18	
144865	2018-09-20 16:24:28.436231	...	2018-09-20 20:53:19	
144866	2018-09-20 16:24:28.436231	...	2018-09-20 16:24:28.436231	

	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			
...
...			
144862	45.258278	94.0	60.0
67.9280			
144863	54.092531	120.0	76.0
85.6829			
144864	66.163591	140.0	88.0
97.0933			
144865	73.680667	158.0	98.0
111.2709			

144866	70.039010	426.0	95.0
88.7319			

	factor	segment_actual_time	segment_osrm_time	\
0	1.272727	14.0	11.0	
1	1.200000	10.0	9.0	
2	1.428571	16.0	7.0	
3	1.550000	21.0	12.0	
4	1.545455	6.0	5.0	
...	
144862	1.566667	12.0	12.0	
144863	1.578947	26.0	21.0	
144864	1.590909	20.0	34.0	
144865	1.612245	17.0	27.0	
144866	4.484211	268.0	9.0	

	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
...
144862	8.1858	1.000000
144863	17.3725	1.238095
144864	20.7053	0.588235
144865	18.8885	0.629630
144866	8.8088	29.777778

[144316 rows x 24 columns]

```
df['New_key'] = df['trip_uuid'] + df['source_center'] +
df['destination_center']
New_cols =
['segment_actual_time', 'segment_osrm_time', 'segment_osrm_distance']
for col in New_cols:
    df[col + '_sum'] = df.groupby('New_key')[col].cumsum()
df[[col + '_sum' for col in New_cols]]
```

	segment_actual_time_sum	segment_osrm_time_sum	\
0	14.0	11.0	
1	24.0	20.0	
2	40.0	27.0	
3	61.0	39.0	
4	67.0	44.0	
...	
144862	92.0	94.0	
144863	118.0	115.0	
144864	138.0	149.0	

144865	155.0	176.0
144866	423.0	185.0

	segment_osrm_distance_sum
0	11.9653
1	21.7243
2	32.5395
3	45.5619
4	49.4772
...	...
144862	65.3487
144863	82.7212
144864	103.4265
144865	122.3150
144866	131.1238

[144316 rows x 3 columns]

```
New_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',
}
```

```
new_seg = df.groupby('New_key').agg(New_dict).reset_index()
new_seg =
new_seg.sort_values(by=['New_key', 'od_end_time'], ascending=True).reset_index()
```

```
new_seg.head(2)
```

	index		New_key	data	\
0	0	trip-153671041653548748IND209304AAAIND000000ACB		training	
1	1	trip-153671041653548748IND462022AAAIND209304AAA		training	

	trip_creation_time	\
0	2018-09-12 00:00:16.535741	
1	2018-09-12 00:00:16.535741	

	route_schedule_uuid	route_type	\
0	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	
1	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	

	trip_uuid	source_center	source_name	\
0	trip-153671041653548748	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	
1	trip-153671041653548748	IND462022AAA	Bhopal_Trnsport_H (Madhya Pradesh)	

	destination_center	...	od_start_time	\
0	IND000000ACB	...	2018-09-12 16:39:46.858469	
1	IND209304AAA	...	2018-09-12 00:00:16.535741	

	od_end_time	start_scan_to_end_scan	\
0	2018-09-13 13:40:23.123744	1260.0	
1	2018-09-12 16:39:46.858469	999.0	

	actual_distance_to_destination	actual_time	osrm_time
0	446.5496	383.759164	732.0
1	544.8027	440.973689	830.0

	segment_actual_time_sum	segment_osrm_distance_sum
0	534.0	670.6205
1	474.0	649.8528

```
[2 rows x 21 columns]
```

```
new_seg.drop('index',axis=1)
```

		New_key	data	\
0	trip-153671041653548748IND209304AAAIND000000ACB		training	
1	trip-153671041653548748IND462022AAAIND209304AAA		training	

2	trip-153671042288605164IND561203AABIND562101AAA	training
3	trip-153671042288605164IND572101AAAIND561203AAB	training
4	trip-153671043369099517IND000000ACBIND160002AAC	training
...		...
26217	trip-153861115439069069IND628204AAAIND627657AAA	test
26218	trip-153861115439069069IND628613AAAIND627005AAA	test
26219	trip-153861115439069069IND628801AAAIND628204AAA	test
26220	trip-153861118270144424IND583119AAAIND583101AAA	test
26221	trip-153861118270144424IND583201AAAIND583119AAA	test

	trip_creation_time \
0	2018-09-12 00:00:16.535741
1	2018-09-12 00:00:16.535741
2	2018-09-12 00:00:22.886430
3	2018-09-12 00:00:22.886430
4	2018-09-12 00:00:33.691250
...	...
26217	2018-10-03 23:59:14.390954
26218	2018-10-03 23:59:14.390954
26219	2018-10-03 23:59:14.390954
26220	2018-10-03 23:59:42.701692
26221	2018-10-03 23:59:42.701692

	route_schedule_uuid	route_type \
0	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL
1	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL
2	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting
3	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting
4	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL
...
26217	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting
26218	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting
26219	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting
26220	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL
26221	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL

	trip_uuid	source_center \
0	trip-153671041653548748	IND209304AAA
1	trip-153671041653548748	IND462022AAA
2	trip-153671042288605164	IND561203AAB
3	trip-153671042288605164	IND572101AAA
4	trip-153671043369099517	IND000000ACB
...
26217	trip-153861115439069069	IND628204AAA
26218	trip-153861115439069069	IND628613AAA
26219	trip-153861115439069069	IND628801AAA
26220	trip-153861118270144424	IND583119AAA
26221	trip-153861118270144424	IND583201AAA

source_name	destination_center \
-------------	----------------------

0	Kanpur_Central_H_6 (Uttar Pradesh)	IND000000ACB
1	Bhopal_Trnsport_H (Madhya Pradesh)	IND209304AAA
2	Doddablpur_ChikaDPP_D (Karnataka)	IND562101AAA
3	Tumkur_Veersagr_I (Karnataka)	IND561203AAB
4	Gurgaon_Bilaspur_HB (Haryana)	IND160002AAC
...
26217	Tirchchndr_Shnmgprm_D (Tamil Nadu)	IND627657AAA
26218	Peikulam_SriVnktpm_D (Tamil Nadu)	IND627005AAA
26219	Eral_Busstand_D (Tamil Nadu)	IND628204AAA
26220	Sandur_WrdN1DPP_D (Karnataka)	IND583101AAA
26221	Hospet (Karnataka)	IND583119AAA

	destination_name	
od_start_time \		
0	Gurgaon_Bilaspur_HB (Haryana)	2018-09-12
16:39:46.858469		
1	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12
00:00:16.535741		
2	Chikblapur_ShntiSgr_D (Karnataka)	2018-09-12
02:03:09.655591		
3	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12
00:00:22.886430		
4	Chandigarh_Mehmdpur_H (Punjab)	2018-09-14
03:40:17.106733		
...
.		
26217	Thisayanvilai_UdnkdiRD_D (Tamil Nadu)	2018-10-04
02:29:04.272194		
26218	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	2018-10-04
04:16:39.894872		
26219	Tirchchndr_Shnmgprm_D (Tamil Nadu)	2018-10-04
01:44:53.808000		
26220	Bellary_Dc (Karnataka)	2018-10-04
03:58:40.726547		
26221	Sandur_WrdN1DPP_D (Karnataka)	2018-10-04
02:51:44.712656		

	od_end_time	start_scan_to_end_scan \
0	2018-09-13 13:40:23.123744	1260.0
1	2018-09-12 16:39:46.858469	999.0
2	2018-09-12 03:01:59.598855	58.0
3	2018-09-12 02:03:09.655591	122.0
4	2018-09-14 17:34:55.442454	834.0
...
26217	2018-10-04 03:31:11.183797	62.0
26218	2018-10-04 05:47:45.162682	91.0
26219	2018-10-04 02:29:04.272194	44.0
26220	2018-10-04 08:46:09.166940	287.0
26221	2018-10-04 03:58:40.726547	66.0

	actual_distance_to_destination	actual_time	osrm_time
osrm_distance \			
0	383.759164	732.0	329.0
446.5496			
1	440.973689	830.0	388.0
544.8027			
2	24.644021	47.0	26.0
28.1994			
3	48.542890	96.0	42.0
56.9116			
4	237.439610	611.0	212.0
281.2109			
...
...			
26217	33.627182	51.0	41.0
42.5213			
26218	33.673835	90.0	48.0
40.6080			
26219	12.661945	30.0	14.0
16.0185			
26220	40.546740	233.0	42.0
52.5303			
26221	25.534793	42.0	26.0
28.0484			

	segment_actual_time_sum	segment_osrm_distance_sum \
0	728.0	670.6205
1	820.0	649.8528
2	46.0	28.1995
3	95.0	55.9899
4	608.0	317.7408
...
...		
26217	49.0	42.1431
26218	89.0	78.5869
26219	29.0	16.0184
26220	233.0	52.5303
26221	41.0	28.0484

	segment_osrm_time_sum
0	534.0
1	474.0
2	26.0
3	39.0
4	231.0
...	...
...	
26217	42.0
26218	77.0
26219	14.0
26220	42.0
26221	25.0

[26222 rows x 20 columns]

```
new_seg.info()
```

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

```
RangeIndex: 26222 entries, 0 to 26221
```

```
Data columns (total 21 columns):
```

#	Column	Non-Null Count	Dtype
0	index	26222 non-null	int64
1	New_key	26222 non-null	object
2	data	26222 non-null	object
3	trip_creation_time	26222 non-null	object
4	route_schedule_uuid	26222 non-null	object
5	route_type	26222 non-null	object
6	trip_uuid	26222 non-null	object
7	source_center	26222 non-null	object
8	source_name	26222 non-null	object
9	destination_center	26222 non-null	object
10	destination_name	26222 non-null	object
11	od_start_time	26222 non-null	datetime64[ns]
12	od_end_time	26222 non-null	datetime64[ns]
13	start_scan_to_end_scan	26222 non-null	float64
14	actual_distance_to_destination	26222 non-null	float64
15	actual_time	26222 non-null	float64
16	osrm_time	26222 non-null	float64
17	osrm_distance	26222 non-null	float64
18	segment_actual_time_sum	26222 non-null	float64
19	segment_osrm_distance_sum	26222 non-null	float64
20	segment_osrm_time_sum	26222 non-null	float64

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

```
memory usage: 4.2+ MB
```

```
new_seg.head(2)
```

	index	New_key	data \
0	0	trip-153671041653548748IND209304AAAIND000000ACB	training
1	1	trip-153671041653548748IND462022AAAIND209304AAA	training

	trip_creation_time \
0	2018-09-12 00:00:16.535741
1	2018-09-12 00:00:16.535741

	route_schedule_uuid	route_type \
--	---------------------	--------------

0	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL
1	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL

	trip_uuid	source_center	source_name \
0	trip-153671041653548748	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)
1	trip-153671041653548748	IND462022AAA	Bhopal_Trnsport_H (Madhya Pradesh)

	destination_center	...	od_start_time \
0	IND000000ACB	...	2018-09-12 16:39:46.858469
1	IND209304AAA	...	2018-09-12 00:00:16.535741

	od_end_time	start_scan_to_end_scan \
0	2018-09-13 13:40:23.123744	1260.0
1	2018-09-12 16:39:46.858469	999.0

	actual_distance_to_destination	actual_time	osrm_time
osrm_distance \			
0	383.759164	732.0	329.0
446.5496			
1	440.973689	830.0	388.0
544.8027			

	segment_actual_time_sum	segment_osrm_distance_sum
segment_osrm_time_sum		
0	728.0	670.6205
534.0		
1	820.0	649.8528
474.0		

[2 rows x 21 columns]

```
new_seg.drop('data',axis=1,inplace=True)
new_seg.drop('route_schedule_uuid',axis=1,inplace=True)
```

```
new_seg.drop('index',axis=1,inplace=True)
```

new_seg

		New_key \
0	trip-153671041653548748IND209304AAAIND000000ACB	
1	trip-153671041653548748IND462022AAAIND209304AAA	
2	trip-153671042288605164IND561203AABIND562101AAA	
3	trip-153671042288605164IND572101AAAIND561203AAB	
4	trip-153671043369099517IND000000ACBIND160002AAC	
...		
26217	trip-153861115439069069IND628204AAAIND627657AAA	
26218	trip-153861115439069069IND628613AAAIND627005AAA	

26219 trip-153861115439069069IND628801AAAIND628204AAA
 26220 trip-153861118270144424IND583119AAAIND583101AAA
 26221 trip-153861118270144424IND583201AAAIND583119AAA

	trip_creation_time	route_type	trip_uuid
\			
0	2018-09-12 00:00:16.535741	FTL	trip-153671041653548748
1	2018-09-12 00:00:16.535741	FTL	trip-153671041653548748
2	2018-09-12 00:00:22.886430	Carting	trip-153671042288605164
3	2018-09-12 00:00:22.886430	Carting	trip-153671042288605164
4	2018-09-12 00:00:33.691250	FTL	trip-153671043369099517
...
26217	2018-10-03 23:59:14.390954	Carting	trip-153861115439069069
26218	2018-10-03 23:59:14.390954	Carting	trip-153861115439069069
26219	2018-10-03 23:59:14.390954	Carting	trip-153861115439069069
26220	2018-10-03 23:59:42.701692	FTL	trip-153861118270144424
26221	2018-10-03 23:59:42.701692	FTL	trip-153861118270144424

	source_center	source_name
destination_center \		
0	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)
	IND000000ACB	
1	IND462022AAA	Bhopal_Trnsport_H (Madhya Pradesh)
	IND209304AAA	
2	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)
	IND562101AAA	
3	IND572101AAA	Tumkur_Veersagr_I (Karnataka)
	IND561203AAB	
4	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)
	IND160002AAC	
...
...		
26217	IND628204AAA	Tirchchndr_Shnmgprm_D (Tamil Nadu)
	IND627657AAA	
26218	IND628613AAA	Peikulam_SriVnktm_D (Tamil Nadu)
	IND627005AAA	
26219	IND628801AAA	Eral_Busstand_D (Tamil Nadu)
	IND628204AAA	

26220 IND583119AAA Sandur_WrdN1DPP_D (Karnataka)
IND583101AAA
26221 IND583201AAA Hospet (Karnataka)
IND583119AAA

	destination_name	od_start_time \
0	Gurgaon_Bilaspur_HB (Haryana)	2018-09-12 16:39:46.858469
1	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12 00:00:16.535741
2	Chikblapur_ShntiSgr_D (Karnataka)	2018-09-12 02:03:09.655591
3	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12 00:00:22.886430
4	Chandigarh_Mehmdpur_H (Punjab)	2018-09-14 03:40:17.106733
...
26217	Thisayanvilai_UdnkdiRD_D (Tamil Nadu)	2018-10-04 02:29:04.272194
26218	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	2018-10-04 04:16:39.894872
26219	Tirchchndr_Shnmgprm_D (Tamil Nadu)	2018-10-04 01:44:53.808000
26220	Bellary_Dc (Karnataka)	2018-10-04 03:58:40.726547
26221	Sandur_WrdN1DPP_D (Karnataka)	2018-10-04 02:51:44.712656

	od_end_time	start_scan_to_end_scan \
0	2018-09-13 13:40:23.123744	1260.0
1	2018-09-12 16:39:46.858469	999.0
2	2018-09-12 03:01:59.598855	58.0
3	2018-09-12 02:03:09.655591	122.0
4	2018-09-14 17:34:55.442454	834.0
...
26217	2018-10-04 03:31:11.183797	62.0
26218	2018-10-04 05:47:45.162682	91.0
26219	2018-10-04 02:29:04.272194	44.0
26220	2018-10-04 08:46:09.166940	287.0
26221	2018-10-04 03:58:40.726547	66.0

	actual_distance_to_destination	actual_time	osrm_time
osrm_distance \			
0	383.759164	732.0	329.0
446.5496			
1	440.973689	830.0	388.0
544.8027			
2	24.644021	47.0	26.0

28.1994			
3	48.542890	96.0	42.0
56.9116			
4	237.439610	611.0	212.0
281.2109			
...
...			
26217	33.627182	51.0	41.0
42.5213			
26218	33.673835	90.0	48.0
40.6080			
26219	12.661945	30.0	14.0
16.0185			
26220	40.546740	233.0	42.0
52.5303			
26221	25.534793	42.0	26.0
28.0484			

	segment_actual_time_sum	segment_osrm_distance_sum	\
0	728.0	670.6205	
1	820.0	649.8528	
2	46.0	28.1995	
3	95.0	55.9899	
4	608.0	317.7408	
...	
26217	49.0	42.1431	
26218	89.0	78.5869	
26219	29.0	16.0184	
26220	233.0	52.5303	
26221	41.0	28.0484	

	segment_osrm_time_sum
0	534.0
1	474.0
2	26.0
3	39.0
4	231.0
...	...
26217	42.0
26218	77.0
26219	14.0
26220	42.0
26221	25.0

[26222 rows x 18 columns]

new_seg.head()

	New_key	\
0	trip-153671041653548748IND209304AAAIND000000ACB	

```

1 trip-153671041653548748IND462022AAAIND209304AAA
2 trip-153671042288605164IND561203AABIND562101AAA
3 trip-153671042288605164IND572101AAAIND561203AAB
4 trip-153671043369099517IND000000ACBIND160002AAC

```

```

      trip_creation_time route_type      trip_uuid \
0 2018-09-12 00:00:16.535741      FTL trip-153671041653548748
1 2018-09-12 00:00:16.535741      FTL trip-153671041653548748
2 2018-09-12 00:00:22.886430    Carting trip-153671042288605164
3 2018-09-12 00:00:22.886430    Carting trip-153671042288605164
4 2018-09-12 00:00:33.691250      FTL trip-153671043369099517

```

```

      source_center      source_name destination_center
\
0 IND209304AAA Kanpur_Central_H_6 (Uttar Pradesh)      IND000000ACB
1 IND462022AAA Bhopal_Trnsport_H (Madhya Pradesh)      IND209304AAA
2 IND561203AAB Doddablpur_ChikaDPP_D (Karnataka)      IND562101AAA
3 IND572101AAA Tumkur_Veersagr_I (Karnataka)      IND561203AAB
4 IND000000ACB Gurgaon_Bilaspur_HB (Haryana)      IND160002AAC

```

```

      destination_name      od_start_time \
0 Gurgaon_Bilaspur_HB (Haryana) 2018-09-12 16:39:46.858469
1 Kanpur_Central_H_6 (Uttar Pradesh) 2018-09-12 00:00:16.535741
2 Chikblapur_ShntiSgr_D (Karnataka) 2018-09-12 02:03:09.655591
3 Doddablpur_ChikaDPP_D (Karnataka) 2018-09-12 00:00:22.886430
4 Chandigarh_Mehmdpur_H (Punjab) 2018-09-14 03:40:17.106733

```

```

      od_end_time start_scan_to_end_scan \
0 2018-09-13 13:40:23.123744      1260.0
1 2018-09-12 16:39:46.858469      999.0
2 2018-09-12 03:01:59.598855      58.0
3 2018-09-12 02:03:09.655591      122.0
4 2018-09-14 17:34:55.442454      834.0

```

```

      actual_distance_to_destination actual_time osrm_time
osrm_distance \
0      383.759164      732.0      329.0
446.5496
1      440.973689      830.0      388.0
544.8027
2      24.644021      47.0      26.0
28.1994
3      48.542890      96.0      42.0
56.9116

```

```

4                237.439610          611.0          212.0
281.2109

```

```

    segment_actual_time_sum  segment_osrm_distance_sum
segment_osrm_time_sum
0                728.0                670.6205
534.0
1                820.0                649.8528
474.0
2                 46.0                 28.1995
26.0
3                 95.0                 55.9899
39.0
4                608.0                317.7408
231.0

```

```

listeddate = new_seg["source_name"].apply(lambda x:
str(x).split("(")).tolist()
listeddate_new = pd.DataFrame(listeddate,index=new_seg["trip_uuid"])
listeddate_new=listeddate_new.stack()
listeddate_new=listeddate_new.reset_index()
listeddate_new

```

```

            trip_uuid  level_1
0  trip-153671041653548748      0  Kanpur_Central_H_6
1  trip-153671041653548748      1  Uttar Pradesh)
2  trip-153671041653548748      0  Bhopal_Trnsport_H
3  trip-153671041653548748      1  Madhya Pradesh)
4  trip-153671042288605164      0  Doddablpur_ChikaDPP_D
...
52439 trip-153861115439069069      1  Tamil Nadu)
52440 trip-153861118270144424      0  Sandur_WrdN1DPP_D
52441 trip-153861118270144424      1  Karnataka)
52442 trip-153861118270144424      0  Hospet
52443 trip-153861118270144424      1  Karnataka)

```

```

[52444 rows x 3 columns]

```

```

New_df = listeddate_new.iloc[1::2]

```

```

New_df

```

```

            trip_uuid  level_1
1  trip-153671041653548748      1  Uttar Pradesh)
3  trip-153671041653548748      1  Madhya Pradesh)
5  trip-153671042288605164      1  Karnataka)
7  trip-153671042288605164      1  Karnataka)
9  trip-153671043369099517      1  Haryana)
...
52435 trip-153861115439069069      1  Tamil Nadu)
52437 trip-153861115439069069      1  Tamil Nadu)

```

52439	trip-153861115439069069	1	Tamil Nadu)
52441	trip-153861118270144424	1	Karnataka)
52443	trip-153861118270144424	1	Karnataka)

[26222 rows x 3 columns]

```
New_df.drop('level_1',inplace=True,axis=1)
```

/Users/mrunmay/opt/anaconda3/lib/python3.9/site-packages/pandas/core/frame.py:4906: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
return super().drop()

New_df

	trip_uuid	
1	trip-153671041653548748	Uttar Pradesh)
3	trip-153671041653548748	Madhya Pradesh)
5	trip-153671042288605164	Karnataka)
7	trip-153671042288605164	Karnataka)
9	trip-153671043369099517	Haryana)
...
52435	trip-153861115439069069	Tamil Nadu)
52437	trip-153861115439069069	Tamil Nadu)
52439	trip-153861115439069069	Tamil Nadu)
52441	trip-153861118270144424	Karnataka)
52443	trip-153861118270144424	Karnataka)

[26222 rows x 2 columns]

```
New_df.columns=['trip_uuid','state']
```

New_df

	trip_uuid	state
1	trip-153671041653548748	Uttar Pradesh)
3	trip-153671041653548748	Madhya Pradesh)
5	trip-153671042288605164	Karnataka)
7	trip-153671042288605164	Karnataka)
9	trip-153671043369099517	Haryana)
...
52435	trip-153861115439069069	Tamil Nadu)
52437	trip-153861115439069069	Tamil Nadu)
52439	trip-153861115439069069	Tamil Nadu)
52441	trip-153861118270144424	Karnataka)
52443	trip-153861118270144424	Karnataka)

[26222 rows x 2 columns]

```
New_df['state'].value_counts()
```

Maharashtra)	3565
Karnataka)	3445
Tamil Nadu)	2130
Haryana)	2039
Uttar Pradesh)	1818
Telangana)	1474
Gujarat)	1401
West Bengal)	1368
Andhra Pradesh)	1310
Rajasthan)	1162
Punjab)	1052
Bihar)	1048
Delhi)	818
Kerala)	762
Madhya Pradesh)	740
Assam)	484
Uttarakhand)	369
Jharkhand)	301
Orissa)	248
Himachal Pradesh)	223
Chandigarh)	160
Goa)	86
Chhattisgarh)	52
Arunachal Pradesh)	48
Jammu & Kashmir)	47
Pondicherry)	30
Dadra and Nagar Haveli)	15
Meghalaya)	13
Mizoram)	8
Nagaland)	5
Tripura)	1

Name: state, dtype: int64

```
listeddate = new_seg["destination_name"].apply(lambda x:
str(x).split("(")).tolist()
listeddate_new = pd.DataFrame(listeddate,index=new_seg["trip_uuid"])
listeddate_new=listeddate_new.stack()
listeddate_new=listeddate_new.reset_index()
listeddate_new
```

	trip_uuid	level_1		
0	trip-153671041653548748	0	Gurgaon_Bilaspur_HB	0
1	trip-153671041653548748	1	Haryana)	
2	trip-153671041653548748	0	Kanpur_Central_H_6	
3	trip-153671041653548748	1	Uttar Pradesh)	
4	trip-153671042288605164	0	Chikblapur_ShntiSgr_D	

```

...
52439 trip-153861115439069069    1    Tamil Nadu)
52440 trip-153861118270144424    0    Bellary_Dc
52441 trip-153861118270144424    1    Karnataka)
52442 trip-153861118270144424    0    Sandur_WrdN1DPP_D
52443 trip-153861118270144424    1    Karnataka)

```

[52444 rows x 3 columns]

```
New_df = listeddate_new.iloc[1::2]
```

```
New_df
```

```

          trip_uuid  level_1
1  trip-153671041653548748    1    Haryana)
3  trip-153671041653548748    1  Uttar Pradesh)
5  trip-153671042288605164    1    Karnataka)
7  trip-153671042288605164    1    Karnataka)
9  trip-153671043369099517    1    Punjab)
...
52435 trip-153861115439069069    1    Tamil Nadu)
52437 trip-153861115439069069    1    Tamil Nadu)
52439 trip-153861115439069069    1    Tamil Nadu)
52441 trip-153861118270144424    1    Karnataka)
52443 trip-153861118270144424    1    Karnataka)

```

[26222 rows x 3 columns]

```
New_df.drop(columns='level_1',inplace=True,axis=1)
```

```

/Users/mrunmay/opt/anaconda3/lib/python3.9/site-packages/pandas/core/
frame.py:4906: SettingWithCopyWarning:

```

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:

https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
    return super().drop()
```

```
New_df.columns=['trip_uuid','state']
```

```
New_df
```

```

          trip_uuid          state
1  trip-153671041653548748    Haryana)
3  trip-153671041653548748  Uttar Pradesh)
5  trip-153671042288605164    Karnataka)
7  trip-153671042288605164    Karnataka)
9  trip-153671043369099517    Punjab)
...
52435 trip-153861115439069069    Tamil Nadu)
52437 trip-153861115439069069    Tamil Nadu)

```



```

52439 trip-153861115439069069    Tamil Nadu)
52441 trip-153861118270144424    Karnataka)
52443 trip-153861118270144424    Karnataka)

```

```
[26222 rows x 2 columns]
```

```
New_df['state'].value_counts()
```

```

Karnataka)          3497
Maharashtra)        3473
Tamil Nadu)         2111
Haryana)            2014
Uttar Pradesh)      1828
Telangana)          1539
Gujarat)            1401
West Bengal)        1399
Andhra Pradesh)     1309
Rajasthan)          1201
Punjab)             1132
Bihar)              1053
Kerala)              746
Madhya Pradesh)     739
Delhi)              702
Assam)              465
Uttarakhand)        367
Jharkhand)          309
Orissa)             260
Himachal Pradesh)   241
Chandigarh)         123
Goa)                 86
Arunachal Pradesh)  55
Chhattisgarh)       52
Jammu & Kashmir)    45
Pondicherry)        31
Dadra and Nagar Haveli) 17
Meghalaya)          13
Mizoram)            10
Tripura)            2
Nagaland)           1
Daman & Diu)        1
Name: state, dtype: int64

```

```
new_seg.head()
```

```

                                New_key \
0  trip-153671041653548748IND209304AAAIND000000ACB
1  trip-153671041653548748IND462022AAAIND209304AAA
2  trip-153671042288605164IND561203AABIND562101AAA
3  trip-153671042288605164IND572101AAAIND561203AAB
4  trip-153671043369099517IND000000ACBIND160002AAC

```

	trip_creation_time	route_type	trip_uuid \
0	2018-09-12 00:00:16.535741	FTL	trip-153671041653548748
1	2018-09-12 00:00:16.535741	FTL	trip-153671041653548748
2	2018-09-12 00:00:22.886430	Carting	trip-153671042288605164
3	2018-09-12 00:00:22.886430	Carting	trip-153671042288605164
4	2018-09-12 00:00:33.691250	FTL	trip-153671043369099517

	source_center	source_name	destination_center
0	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	IND000000ACB
1	IND462022AAA	Bhopal_Trnsport_H (Madhya Pradesh)	IND209304AAA
2	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND562101AAA
3	IND572101AAA	Tumkur_Veersagr_I (Karnataka)	IND561203AAB
4	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND160002AAC

	destination_name	od_start_time \
0	Gurgaon_Bilaspur_HB (Haryana)	2018-09-12 16:39:46.858469
1	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12 00:00:16.535741
2	Chikblapur_ShntiSgr_D (Karnataka)	2018-09-12 02:03:09.655591
3	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12 00:00:22.886430
4	Chandigarh_Mehmdpur_H (Punjab)	2018-09-14 03:40:17.106733

	od_end_time	start_scan_to_end_scan \
0	2018-09-13 13:40:23.123744	1260.0
1	2018-09-12 16:39:46.858469	999.0
2	2018-09-12 03:01:59.598855	58.0
3	2018-09-12 02:03:09.655591	122.0
4	2018-09-14 17:34:55.442454	834.0

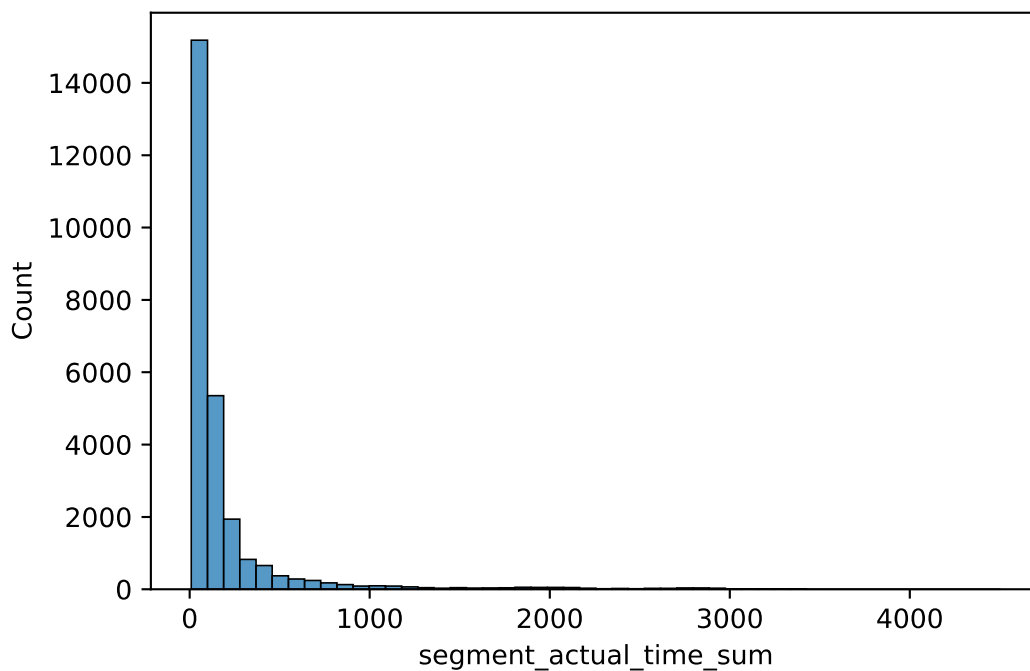
	actual_distance_to_destination	actual_time	osrm_time
0	383.759164	732.0	329.0
1	440.973689	830.0	388.0
2	24.644021	47.0	26.0
3	48.542890	96.0	42.0
4	237.439610	611.0	212.0

segment_actual_time_sum	segment_osrm_distance_sum	segment_osrm_time_sum
446.5496	544.8027	28.1994
56.9116	281.2109	

0	728.0	670.6205
534.0		
1	820.0	649.8528
474.0		
2	46.0	28.1995
26.0		
3	95.0	55.9899
39.0		
4	608.0	317.7408
231.0		

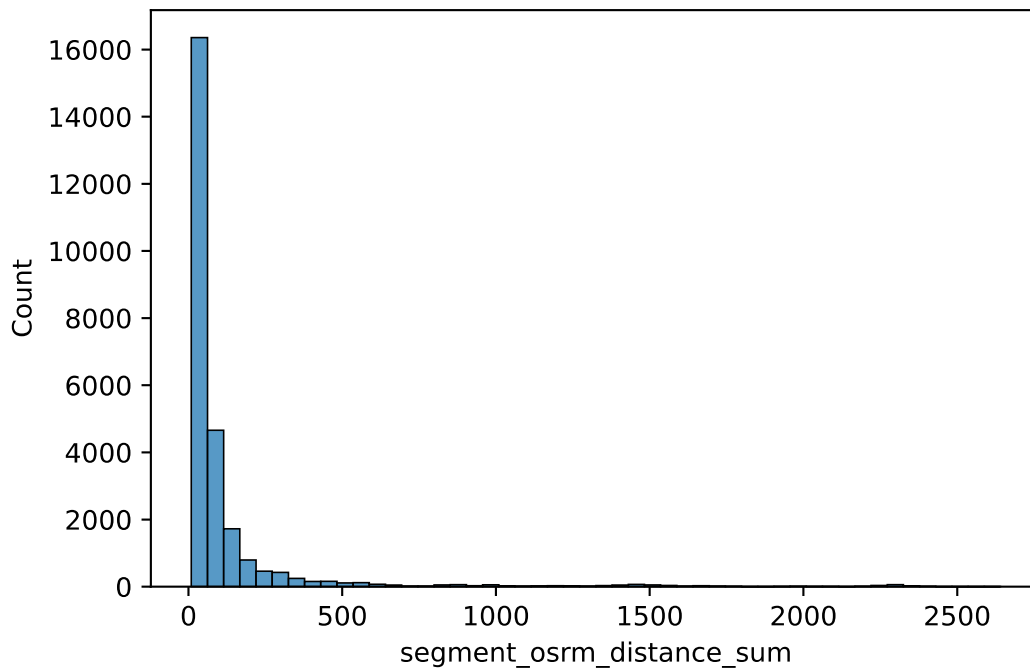
```
sns.histplot(x='segment_actual_time_sum',data=new_seg,bins=50)
```

```
<AxesSubplot:xlabel='segment_actual_time_sum', ylabel='Count'>
```



```
sns.histplot(x='segment_osrm_distance_sum',data=new_seg,bins=50)
```

```
<AxesSubplot:xlabel='segment_osrm_distance_sum', ylabel='Count'>
```



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

```
FTL      13798
Carting   12424
Name: route_type, dtype: int64
```

```
label_encoder = LabelEncoder()
col='route_type'
new_seg[col] = label_encoder.fit_transform(new_seg[col])
```

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

```
1      13798
0      12424
Name: route_type, dtype: int64
```

```
label_encoder = LabelEncoder()
col='route_type'
new_seg[col] = label_encoder.fit_transform(new_seg[col])
```

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

```
1      13798
0      12424
Name: route_type, dtype: int64
```

```
sns.heatmap(data=new_seg.corr(),cmap="Blues",annot=True)
```

```
<AxesSubplot:>
```

