In [258…

```python
import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
from sklearn import preprocessing


import warnings
warnings.filterwarnings('ignore')
```

In [259…

```python
df = pd.read_csv('delhivery_data.csv')
```

In [260…

```python
df.shape
```

Out[260…

```
(144867, 24)
```

In [261…

```python
pd.set_option("display.max_columns", None)
df.head(10)
```

Out[261…

	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_d
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	9	2018-09-20 04:27:55	
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	18	2018-09-20 04:17:55	
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	27	2018-09-20 04:01:19.505586	
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	36	2018-09-20 03:39:57	
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	False	39	2018-09-20 03:33:55	
5	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	9	2018-09-20 06:15:58	
6	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	18	2018-09-20 05:47:29	
7	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	27	2018-09-20 05:25:58	
8	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	36	2018-09-20 05:15:56	
9	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	False	43	2018-09-20 04:49:20	

In [262…

```python
# Missing value detection
df.columns[df.isna().any()].tolist()
```

Out[262…

```
['source_name', 'destination_name']
```

In [263…

```python
#There would be a mapping with Delhivery for source center to source name and destination center to destination
#name. That can be used to fill in the missing values. We do not have that mapping

# IND577116AAA appears as both source center and destination center and has no source or destination name
# corresponding to it. Using most frequent values also will lead to false data. Hence it is best to either delete such
# rows or map them to Other (India)

#To conclude given just Source Center WE CANNOT DEDUCE correct and accurate Source Name from it
```

In [264…

```python
df = df.dropna(axis=0)

#144316 rows × 24 new shape of data
```

In [265…

```python
df = df.reset_index(drop = True)
```

In [266…

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144316 entries, 0 to 144315
Data columns (total 24 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   data                            144316 non-null  object
 1   trip_creation_time              144316 non-null  object
 2   route_schedule_uuid             144316 non-null  object
 3   route_type                      144316 non-null  object
 4   trip_uuid                       144316 non-null  object
 5   source_center                   144316 non-null  object
 6   source_name                     144316 non-null  object
 7   destination_center              144316 non-null  object
 8   destination_name                144316 non-null  object
 9   od_start_time                   144316 non-null  object
 10  od_end_time                     144316 non-null  object
 11  start_scan_to_end_scan          144316 non-null  float64
 12  is_cutoff                       144316 non-null  bool
 13  cutoff_factor                   144316 non-null  int64
 14  cutoff_timestamp                144316 non-null  object
 15  actual_distance_to_destination  144316 non-null  float64
 16  actual_time                     144316 non-null  float64
 17  osrm_time                       144316 non-null  float64
 18  osrm_distance                   144316 non-null  float64
 19  factor                          144316 non-null  float64
 20  segment_actual_time             144316 non-null  float64
 21  segment_osrm_time               144316 non-null  float64
 22  segment_osrm_distance           144316 non-null  float64
 23  segment_factor                  144316 non-null  float64
dtypes: bool(1), float64(10), int64(1), object(12)
memory usage: 25.5+ MB
```

In [267…

```python
#object to datetime conversion

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

In [268…

```python
#Categorical variables are data, route_type

df['data'] = df['data'].map({'training':1, 'test':0})
df['route_type'] = df['route_type'].map({'FTL':1, 'Carting':0})
```

In [269…

```python
df.head(20)
```

Out[269…

	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_
0	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	9	2018-09-20 04:27:55	
1	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	18	2018-09-20 04:17:55	
2	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	27	2018-09-20 04:01:19.505586	
3	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	True	36	2018-09-20 03:39:57	
4	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	False	39	2018-09-20 03:33:55	
5	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	9	2018-09-20 06:15:58	
6	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	18	2018-09-20 05:47:29	
7	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	27	2018-09-20 05:25:58	
8	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	True	36	2018-09-20 05:15:56	
9	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	0	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	False	43	2018-09-20 04:49:20	
10	1	2018-09-23 06:42:06.021680	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172...	1	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:06.021680	2018-09-23 11:44:28.365845	302.0	True	22	2018-09-23 11:05:19	
11	1	2018-09-23 06:42:06.021680	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172...	1	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:06.021680	2018-09-23 11:44:28.365845	302.0	True	44	2018-09-23 10:27:22	
12	1	2018-09-23 06:42:06.021680	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172...	1	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:06.021680	2018-09-23 11:44:28.365845	302.0	True	66	2018-09-23 09:45:25	
13	1	2018-09-23 06:42:06.021680	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172...	1	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:06.021680	2018-09-23 11:44:28.365845	302.0	True	88	2018-09-23 09:21:27	
14	1	2018-09-23 06:42:06.021680	thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc21172...	1	trip-153768492602129387	IND421302AAG	Bhiwandi_Mankoli_HB (Maharashtra)	IND411033AAA	Pune_Tathawde_H (Maharashtra)	2018-09-23 06:42:06.021680	2018-09-23 11:44:28.365845	302.0	False	100	2018-09-23 08:39:31	
15	1	2018-09-14 15:42:46.437249	thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d...	0	trip-153693976643699843	IND400011AAA	LowerParel_CP (Maharashtra)	IND400072AAD	Mumbai_Chndivli_PC (Maharashtra)	2018-09-14 15:42:46.437249	2018-09-14 17:31:45.368791	108.0	True	9	2018-09-14 16:29:54	
16	1	2018-09-14 15:42:46.437249	thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d...	0	trip-153693976643699843	IND400011AAA	LowerParel_CP (Maharashtra)	IND400072AAD	Mumbai_Chndivli_PC (Maharashtra)	2018-09-14 15:42:46.437249	2018-09-14 17:31:45.368791	108.0	False	16	2018-09-14 16:15:53	
17	1	2018-09-13 20:44:19.424489	thanos::sroute:76951383-1608-44e4-a284-46d92e8...	1	trip-153687145942424248	IND562132AAA	Bangalore_Nelmngla_H (Karnataka)	IND560099AAB	Bengaluru_Bomsndra_HB (Karnataka)	2018-09-13 20:44:19.424489	2018-09-13 23:59:56.061158	195.0	True	22	2018-09-13 23:25:20	
18	1	2018-09-13 20:44:19.424489	thanos::sroute:76951383-1608-44e4-a284-46d92e8...	1	trip-153687145942424248	IND562132AAA	Bangalore_Nelmngla_H (Karnataka)	IND560099AAB	Bengaluru_Bomsndra_HB (Karnataka)	2018-09-13 20:44:19.424489	2018-09-13 23:59:56.061158	195.0	False	39	2018-09-13 22:47:26	
19	1	2018-09-13 20:44:19.424489	thanos::sroute:76951383-1608-44e4-a284-46d92e8...	1	trip-153687145942424248	IND560099AAB	Bengaluru_Bomsndra_HB (Karnataka)	IND683511AAA	Aluva_Peedika_H (Kerala)	2018-09-13 23:59:56.061158	2018-09-14 13:55:58.765334	836.0	True	22	2018-09-14 12:45:25	

In [270…

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

In [271…

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

In [272…

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

In [273…

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

Out[273…

	segment_actual_time_sum	segment_osrm_time_sum	segment_osrm_distance_sum
0	14.0	11.0	11.9653
1	24.0	20.0	21.7243
2	40.0	27.0	32.5395
3	61.0	39.0	45.5619
4	67.0	44.0	49.4772
...	...	...	...
144311	92.0	94.0	65.3487
144312	118.0	115.0	82.7212
144313	138.0	149.0	103.4265
144314	155.0	176.0	122.3150
144315	423.0	185.0	131.1238

144316 rows × 3 columns

In [274…

```python
create_segment_dict = {

    'data' : 'first',
    'trip_creation_time' : 'first',
    'route_schedule_uuid' : '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_time_sum' : 'last',
    'segment_osrm_distance_sum' : 'last',

}
```

In [275…

```python
seg = df.groupby('segment_key').agg(create_segment_dict).reset_index()
```

In [276…

```python
seg =  seg.sort_values(by = ['segment_key', 'od_end_time'], ascending=True)
```

In [277…

```python
seg
```

Out[277…

	segment_key	data	trip_creation_time	route_schedule_uuid	trip_uuid	source_center	source_name	destination_center	destination_name	od_start_time	od_end_time	start_scan_to_end_scan	actual_distance_to_d
0	trip-153671041653548748IND209304AAAIND000000ACB	1	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	trip-153671041653548748	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2018-09-12 16:39:46.858469	2018-09-13 13:40:23.123744	1260.0	3
1	trip-153671041653548748IND462022AAAIND209304AAA	1	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	trip-153671041653548748	IND462022AAA	Bhopal_Trnsport_H (Madhya Pradesh)	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	2018-09-12 00:00:16.535741	2018-09-12 16:39:46.858469	999.0	4
2	trip-153671042288605164IND561203AABIND562101AAA	1	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	trip-153671042288605164	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND562101AAA	Chikblapur_ShntiSgr_D (Karnataka)	2018-09-12 02:03:09.655591	2018-09-12 03:01:59.598855	58.0	
3	trip-153671042288605164IND572101AAAIND561203AAB	1	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	trip-153671042288605164	IND572101AAA	Tumkur_Veersagr_I (Karnataka)	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	2018-09-12 00:00:22.886430	2018-09-12 02:03:09.655591	122.0	
4	trip-153671043369099517IND000000ACBIND160002AAC	1	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	trip-153671043369099517	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	2018-09-14 03:40:17.106733	2018-09-14 17:34:55.442454	834.0	2
...	...	...	...	...	...	...	...	...	...	...	...	...	
26217	trip-153861115439069069IND628204AAAIND627657AAA	0	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	trip-153861115439069069	IND628204AAA	Tirchchndr_Shnmgprm_D (Tamil Nadu)	IND627657AAA	Thisayanvilai_UdnkdiRD_D (Tamil Nadu)	2018-10-04 02:29:04.272194	2018-10-04 03:31:11.183797	62.0	
26218	trip-153861115439069069IND628613AAAIND627005AAA	0	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	trip-153861115439069069	IND628613AAA	Peikulam_SriVnktpm_D (Tamil Nadu)	IND627005AAA	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	2018-10-04 04:16:39.894872	2018-10-04 05:47:45.162682	91.0	
26219	trip-153861115439069069IND628801AAAIND628204AAA	0	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	trip-153861115439069069	IND628801AAA	Eral_Busstand_D (Tamil Nadu)	IND628204AAA	Tirchchndr_Shnmgprm_D (Tamil Nadu)	2018-10-04 01:44:53.808000	2018-10-04 02:29:04.272194	44.0	
26220	trip-153861118270144424IND583119AAAIND583101AAA	0	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	trip-153861118270144424	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	IND583101AAA	Bellary_Dc (Karnataka)	2018-10-04 03:58:40.726547	2018-10-04 08:46:09.166940	287.0	
26221	trip-153861118270144424IND583201AAAIND583119AAA	0	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	trip-153861118270144424	IND583201AAA	Hospet (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	2018-10-04 02:51:44.712656	2018-10-04 03:58:40.726547	66.0	

26222 rows × 19 columns

In [278…

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

Out[278…

	segment_key	data	trip_creation_time	route_schedule_uuid	trip_uuid	source_center	source_name	destination_center	destination_name	od_start_time	od_end_time	start_scan_to_end_scan	actual_distance_to_desti
10370	trip-153741093647649320IND388121AAAIND388620AAB	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	86.0	39.3
10371	trip-153741093647649320IND388620AABIND388320AAA	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	2018-09-20 04:47:45.236797	2018-09-20 06:36:55.627764	109.0	43.5

In [279…

```python
seg['od_time_diff_hour'] =  seg['od_end_time'] - seg['od_start_time']
```

In [280…

```python
seg['od_time_diff_hour']
```

Out[280…

```
0       0 days 21:00:36.265275
1       0 days 16:39:30.322728
2       0 days 00:58:49.943264
3       0 days 02:02:46.769161
4       0 days 13:54:38.335721
                 ...
26217   0 days 01:02:06.911603
26218   0 days 01:31:05.267810
26219   0 days 00:44:10.464194
26220   0 days 04:47:28.440393
26221   0 days 01:06:56.013891
Name: od_time_diff_hour, Length: 26222, dtype: timedelta64[ns]
```

In [281…

```python
create_trip_dict = {

    'data' : 'first',
    'trip_creation_time' : 'first',
    'route_schedule_uuid' : 'first',
    'trip_uuid' : 'first',
    'source_center' : 'first',
    'source_name' : 'first',
    'destination_center' : 'last',
    'destination_name' : 'last',
    'od_time_diff_hour' : 'sum',
    'start_scan_to_end_scan' : 'sum',
    'actual_distance_to_destination' : 'sum',
    'actual_time' : 'sum',
    'osrm_time' : 'sum',
    'osrm_distance' : 'sum',
    'segment_actual_time_sum' : 'sum',
    'segment_osrm_time_sum' : 'sum',
    'segment_osrm_distance_sum' : 'sum',
}
```

In [282…

```python
trip = seg.groupby('trip_uuid').agg(create_trip_dict)
```

In [283…

```python
trip.head(20)
```

Out[283…

trip_uuid	data	trip_creation_time	route_schedule_uuid	trip_uuid	source_center	source_name	destination_center	destination_name	od_time_diff_hour	start_scan_to_end_scan	actual_distance_to_destination	actual_time	osrm_time	osrm_distance
trip-153671041653548748	1	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	trip-153671041653548748	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	1 days 13:40:06.588003	2259.0	824.732854	1562.0	717.0	991.3523
trip-153671042288605164	1	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	trip-153671042288605164	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	0 days 03:01:36.712425	180.0	73.186911	143.0	68.0	85.1110
trip-153671043369099517	1	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	trip-153671043369099517	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	2 days 17:34:21.751204	3933.0	1927.404273	3347.0	1740.0	2354.0665
trip-153671046011330457	1	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	trip-153671046011330457	IND400072AAB	Mumbai Hub (Maharashtra)	IND401104AAA	Mumbai_MiraRd_IP (Maharashtra)	0 days 01:40:29.696112	100.0	17.175274	59.0	15.0	19.6800
trip-153671052974046625	1	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	trip-153671052974046625	IND583101AAA	Bellary_Dc (Karnataka)	IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	0 days 11:58:20.942506	717.0	127.448500	341.0	117.0	146.7918
trip-153671055416136166	1	2018-09-12 00:02:34.161600	thanos::sroute:9bf03170-d0a2-4a3f-aa4d-9aaab3d...	trip-153671055416136166	IND600056AAA	Chennai_Poonamallee (Tamil Nadu)	IND600056AAA	Chennai_Poonamallee (Tamil Nadu)	0 days 03:10:29.270932	189.0	24.597048	61.0	23.0	28.0647
trip-153671066201138152	1	2018-09-12 00:04:22.011653	thanos::sroute:a97698cc-846e-41a7-916b-88b1741...	trip-153671066201138152	IND600044AAD	Chennai_Chrompet_DPC (Tamil Nadu)	IND600048AAA	Chennai_Vandalur_Dc (Tamil Nadu)	0 days 01:38:00.338041	98.0	9.100510	24.0	13.0	12.0184
trip-153671066826362165	1	2018-09-12 00:04:28.263977	thanos::sroute:d5b71ae9-a11a-4f52-bcb7-274b65e...	trip-153671066826362165	IND560043AAC	HBR Layout PC (Karnataka)	IND560043AAC	HBR Layout PC (Karnataka)	0 days 02:26:50.303804	146.0	22.424210	64.0	34.0	28.9203
trip-153671074033284934	1	2018-09-12 00:05:40.333071	thanos::sroute:a0e60427-16ad-4b17-b3b0-6a06643...	trip-153671074033284934	IND395009AAA	Surat_Central_D_12 (Gujarat)	IND395004AAB	Surat_Central_D_3 (Gujarat)	0 days 04:40:50.639847	280.0	25.454848	161.0	29.0	30.9359
trip-153671079956500691	1	2018-09-12 00:06:39.565253	thanos::sroute:a10888ff-f794-41e1-9b7a-7f62ef6...	trip-153671079956500691	IND110024AAA	Delhi_Lajpat_IP (Delhi)	IND110014AAA	Delhi_Bhogal (Delhi)	0 days 00:49:20.003392	49.0	9.872146	23.0	8.0	9.9566
trip-153671090980523004	1	2018-09-12 00:08:29.805514	thanos::sroute:580c788b-ff17-4c1b-9bbd-c59e7b0...	trip-153671090980523004	IND412105AAC	Pune_North_D_3 (Maharashtra)	IND411017AAA	PNQ Rahatani DPC (Maharashtra)	0 days 01:12:57.326210	72.0	13.583706	60.0	17.0	18.5886
trip-153671110078355292	1	2018-09-12 00:11:40.783923	thanos::sroute:c2ee580f-f4b2-4fa5-98ab-0c5b327...	trip-153671110078355292	IND121004AAB	FBD_Balabhgarh_DPC (Haryana)	IND121001AAA	Faridabad (Haryana)	0 days 00:38:30.030476	38.0	9.396526	17.0	9.0	10.8159
trip-153671117375890592	1	2018-09-12 00:12:53.759171	thanos::sroute:f7deffcd-f92a-409b-b86e-c477997...	trip-153671117375890592	IND415408AAC	Shirala_Central_DPP_3 (Maharashtra)	IND415612AAA	Ratnagiri_MjgaonRd_D (Maharashtra)	0 days 17:19:26.867662	1038.0	219.424984	762.0	227.0	297.1036
trip-153671121411074590	1	2018-09-12 00:13:34.111131	thanos::sroute:fd23d4f1-012f-4247-80e9-506d32e...	trip-153671121411074590	IND501359AAE	Hyderabad_Shamshbd_H (Telangana)	IND562132AAA	Bangalore_Nelmngla_H (Karnataka)	0 days 16:36:13.021832	995.0	474.446597	809.0	428.0	542.0812
trip-153671122262300031	1	2018-09-12 00:13:42.623227	thanos::sroute:1d848e84-b73b-47b3-90ec-4a12509...	trip-153671122262300031	IND508223AAA	Thirumalagiri_Xroad_D (Telangana)	IND500039AAC	Hyderabad_Uppal_I (Telangana)	0 days 04:18:48.024440	258.0	94.035128	243.0	88.0	122.8876
trip-153671127071609610	1	2018-09-12 00:14:30.716368	thanos::sroute:13c4e60a-d850-4737-9bac-0cc4a40...	trip-153671127071609610	IND585104AAA	Gulbarga_Nehrugnj_I (Karnataka)	IND585314AAA	Aland_RazaviRd_D (Karnataka)	0 days 15:30:40.474480	928.0	230.413870	561.0	253.0	284.4582
trip-153671128962918389	1	2018-09-12 00:14:49.629525	thanos::sroute:b62ab3ed-c60b-47d2-8c91-fe62135...	trip-153671128962918389	IND302014AAB	Jaipur_Central_I_7 (Rajasthan)	IND302014AAB	Jaipur_Central_I_7 (Rajasthan)	0 days 02:31:10.298476	150.0	24.417683	109.0	27.0	30.4880
trip-153671143043841452	1	2018-09-12 00:17:10.438683	thanos::sroute:93b4aac4-c6e3-4b26-9db0-246d5e6...	trip-153671143043841452	IND211002AAB	Allahabad_Central_H_1 (Uttar Pradesh)	IND485001AAC	Satna_Central_I_2 (Madhya Pradesh)	0 days 08:34:36.988628	514.0	129.749996	289.0	132.0	184.6407
trip-153671173668736946	1	2018-09-12 00:22:16.687619	thanos::sroute:7734a02c-9376-4767-a1b5-15897a7...	trip-153671173668736946	IND110043AAA	Delhi_Nangli_IP (Delhi)	IND110078AAA	Janakpuri (Delhi)	0 days 01:07:02.589793	67.0	12.756768	44.0	21.0	18.7668
trip-153671174968648046	1	2018-09-12 00:22:29.686740	thanos::sroute:83cc6bc3-6aea-456b-879b-e99a903...	trip-153671174968648046	IND781005AAA	Guwahati_North (Assam)	IND781018AAB	Guwahati_Hub (Assam)	0 days 02:18:44.031516	138.0	9.941525	110.0	18.0	19.4925

In [284…

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

Out[284…

trip_uuid	data	trip_creation_time	route_schedule_uuid	trip_uuid	source_center	source_name	destination_center	destination_name	od_time_diff_hour	start_scan_to_end_scan	actual_distance_to_destination	actual_time	osrm_time	osrm_distance	segment_a
trip-153741093647649320	1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388320AAA	Anand_Vaghasi_IP (Gujarat)	0 days 03:15:23.209164	195.0	82.981842	170.0	89.0	107.4515	

In [285…

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

Out[285…

trip_uuid	actual_distance_to_destination	osrm_distance
trip-153671041653548748	824.732854	991.3523
trip-153671042288605164	73.186911	85.1110
trip-153671043369099517	1927.404273	2354.0665
trip-153671046011330457	17.175274	19.6800
trip-153671052974046625	127.448500	146.7918
...	...	...
trip-153861095625827784	57.762332	73.4630
trip-153861104386292051	15.513784	16.0882
trip-153861106442901555	38.684839	58.9037
trip-153861115439069069	134.723836	171.1103
trip-153861118270144424	66.081533	80.5787

14787 rows × 2 columns

In [286…

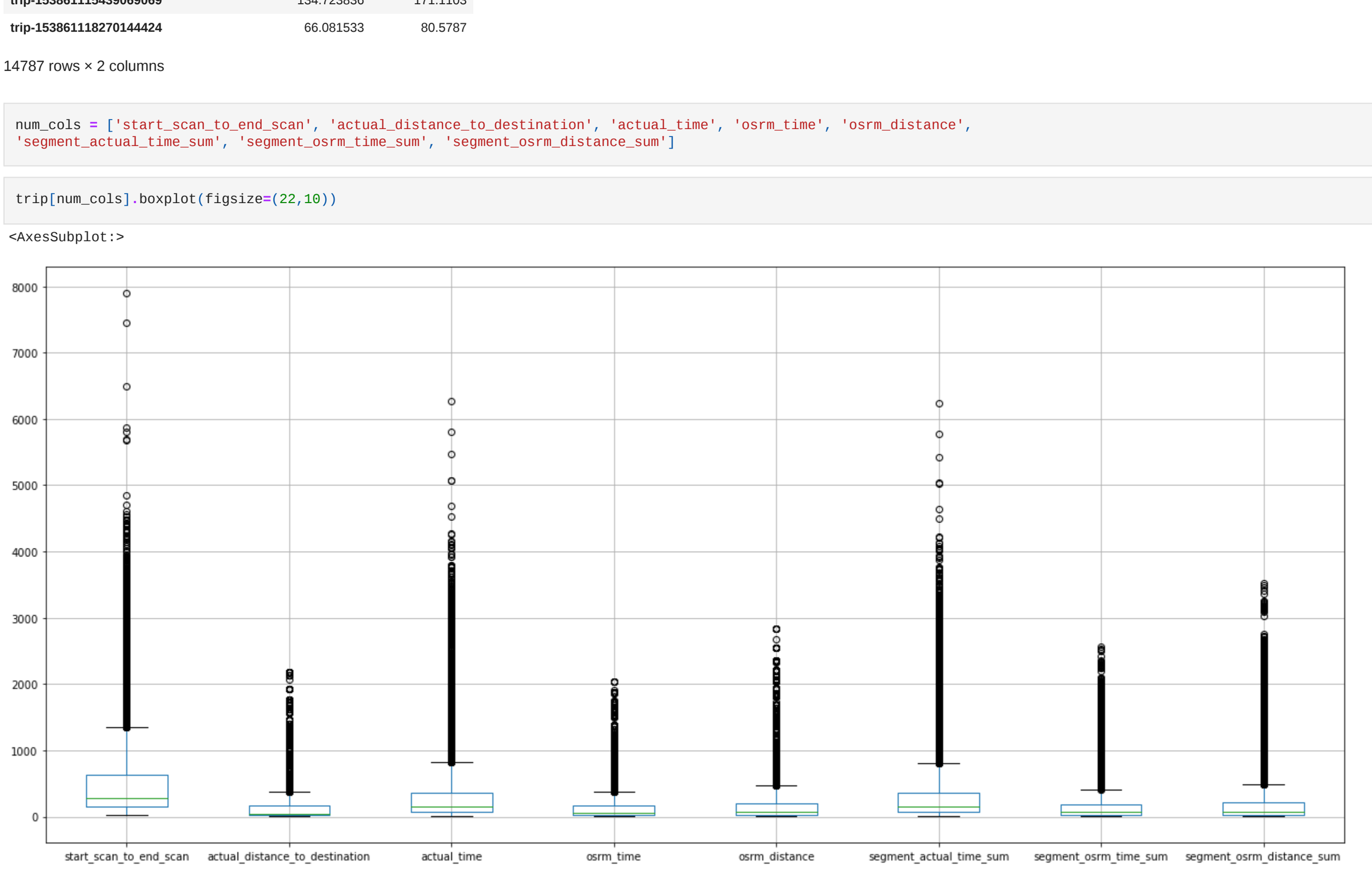
```python
num_cols = ['start_scan_to_end_scan', 'actual_distance_to_destination', 'actual_time', 'osrm_time', 'osrm_distance',
'segment_actual_time_sum', 'segment_osrm_time_sum', 'segment_osrm_distance_sum']
```

In [287…

```python
trip[num_cols].boxplot(figsize=(22,10))
```

Out[287…

```
<AxesSubplot:>
```



In [288…

```python
trip[num_cols].describe()
```

Out[288…

	start_scan_to_end_scan	actual_distance_to_destination	actual_time	osrm_time	osrm_distance	segment_actual_time_sum	segment_osrm_time_sum	segment_osrm_distance_sum
count	14787.000000	14787.000000	14787.000000	14787.000000	14787.000000	14787.000000	14787.000000	14787.000000
mean	529.429025	164.090196	356.306012	160.990938	203.887411	353.059174	180.511598	222.705466
std	658.254936	305.502982	561.517936	271.459495	370.565564	556.365911	314.679279	416.846279
min	23.000000	9.002461	9.000000	6.000000	9.072900	9.000000	6.000000	9.072900
25%	149.000000	22.777099	67.000000	29.000000	30.756900	66.000000	30.000000	32.578850
50%	279.000000	48.287894	148.000000	60.000000	65.302800	147.000000	65.000000	69.784200
75%	632.000000	163.591258	367.000000	168.000000	206.644200	364.000000	184.000000	216.560600
max	7898.000000	2186.531787	6265.000000	2032.000000	2840.081000	6230.000000	2564.000000	3523.632400

In [297…

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

IQR = Q3 - Q1

for i in num_cols:
    Q1ss = trip[i].quantile(0.25)
    Q3ss = trip[i].quantile(0.75)
    IQRss = Q3ss - Q1ss
    trip = trip[ ~ ((trip[i] < (Q1ss -  1.5 * IQRss)) | (trip[i] > (Q3ss + 1.5 * IQRss))) ]



#Q1ss = trip['start_scan_to_end_scan'].quantile(0.25)
#Q3ss = trip['start_scan_to_end_scan'].quantile(0.75)
#IQRss = Q3ss - Q1ss
```

In [298…

```python
#trip = trip[~ ((trip[num_cols] < (Q1 -  1.5 * IQR)) | (trip[num_cols] > (Q3 + 1.5 * IQR)))]
#trip = trip.reset_index(drop=True)

#Somehow giving an error
```

In [299…

```python
trip.head()
```

Out[299…

trip_uuid	data	trip_creation_time	route_schedule_uuid	trip_uuid	source_center	source_name	destination_center	destination_name	od_time_diff_hour	start_scan_to_end_scan	actual_distance_to_destination	actual_time	osrm_time	osrm_distance
trip-153671042288605164	1	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	trip-153671042288605164	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	0 days 03:01:36.712425	180.0	73.186911	143.0	68.0	85.1110
trip-153671046011330457	1	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	trip-153671046011330457	IND400072AAB	Mumbai Hub (Maharashtra)	IND401104AAA	Mumbai_MiraRd_IP (Maharashtra)	0 days 01:40:29.696112	100.0	17.175274	59.0	15.0	19.6800
trip-153671055416136166	1	2018-09-12 00:02:34.161600	thanos::sroute:9bf03170-d0a2-4a3f-aa4d-9aaab3d...	trip-153671055416136166	IND600056AAA	Chennai_Poonamallee (Tamil Nadu)	IND600056AAA	Chennai_Poonamallee (Tamil Nadu)	0 days 03:10:29.270932	189.0	24.597048	61.0	23.0	28.0647
trip-153671066201138152	1	2018-09-12 00:04:22.011653	thanos::sroute:a97698cc-846e-41a7-916b-88b1741...	trip-153671066201138152	IND600044AAD	Chennai_Chrompet_DPC (Tamil Nadu)	IND600048AAA	Chennai_Vandalur_Dc (Tamil Nadu)	0 days 01:38:00.338041	98.0	9.100510	24.0	13.0	12.0184
trip-153671066826362165	1	2018-09-12 00:04:28.263977	thanos::sroute:d5b71ae9-a11a-4f52-bcb7-274b65e...	trip-153671066826362165	IND560043AAC	HBR Layout PC (Karnataka)	IND560043AAC	HBR Layout PC (Karnataka)	0 days 02:26:50.303804	146.0	22.424210	64.0	34.0	28.9203

In [ ]:

In [ ]: