

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
!gdown https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/551/original/delhivery_data.csv?1642751181
```

```
Downloading...
From: https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/551/original/delhivery_data.csv?1642751181
To: /content/delhivery_data.csv?1642751181
100% 55.6M/55.6M [00:17<00:00, 3.19MB/s]
```

```
df = pd.read_csv('delhivery_data.csv?1642751181')
```

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

```
df.head()
```

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA

5 rows × 24 columns

```
df.head()
```

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AA

5 rows × 24 columns

```
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
```

```
df.describe()
```

	start_scan_to_end_scan	cutoff_factor	actual_distance_to_destination	actual_time	osrm_time	osrm_distance	factor
count	144316.000000	144316.000000	144316.000000	144316.000000	144316.000000	144316.000000	144316.000000
mean	963.697698	233.561345	234.708498	417.996237	214.437055	285.549785	2.120178
std	1038.082976	345.245823	345.480571	598.940065	308.448543	421.717826	1.717065
min	20.000000	9.000000	9.000045	9.000000	6.000000	9.008200	0.144000
25%	161.000000	22.000000	23.352027	51.000000	27.000000	29.896250	1.604545
50%	451.000000	66.000000	66.135322	132.000000	64.000000	78.624400	1.857143
75%	1645.000000	286.000000	286.919294	516.000000	259.000000	346.305400	2.212280
max	7898.000000	1927.000000	1927.447705	4532.000000	1686.000000	2326.199100	77.387097


Grouping by sub-journey in the trip



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

144316 rows × 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',

}

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'])

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	trip_creation_time	route_schedule_uuid	route_type	trip
0	0	153671041653548748IND209304AAAIND000000ACBtrip-	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	1536710416535
1	1	153671041653548748IND462022AAAIND209304AAAtrip-	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	1536710416535
2	2	153671042288605164IND561203AABIND562101AAAtrip-	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	1536710422886
3	3	153671042288605164IND572101AAAIND561203AABtrip-	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	1536710422886
4	4	153671043369099517IND000000ACBIND160002AACtrip-	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	1536710433690
...
26217	26217	153861115439069069IND628204AAAIND627657AAAtrip-	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	1538611154390
26218	26218	153861115439069069IND628613AAAIND627005AAAtrip-	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	1538611154390
26219	26219	153861115439069069IND628801AAAIND628204AAAtrip-	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	1538611154390
26220	26220	153861118270144424IND583119AAAIND583101AAAtrip-	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	1538611182701
26221	26221	153861118270144424IND583201AAAIND583119AAAtrip-	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	1538611182701

26222 rows × 21 columns

```
segment.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   segment_key                          26222 non-null  object
2   data                                26222 non-null  object
3   trip_creation_time                   26222 non-null  datetime64[ns]
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](3), float64(8), int64(1), object(9)
memory usage: 4.2+ MB
```

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

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



	od_time_diff_hour
0	1260.604421
1	999.505379
2	58.832388
3	122.779486
4	834.638929
...	...
26217	62.115193
26218	91.087797
26219	44.174403
26220	287.474007
26221	66.933565

26222 rows × 1 columns

dtype: float64

```
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
```



in_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination
8-09-12 535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	153671041653548748	trip-IND209304AAA	Kanpur_Central_H_6 (Uttar Pradesh)	IND
8-09-12 886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	153671042288605164	trip-IND561203AAB	Doddablpur_ChikaDPP_D (Karnataka)	IND
8-09-12 691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	153671043369099517	trip-IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND
8-09-12 .113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	153671046011330457	trip-IND400072AAB	Mumbai Hub (Maharashtra)	IND
8-09-12 740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	153671052974046625	trip-IND583101AAA	Bellary_Dc (Karnataka)	IND
...
8-10-03 258533	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	Carting	153861095625827784	trip-IND160002AAC	Chandigarh_Mehmdpur_H (Punjab)	IND
8-10-03 863155	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	Carting	153861104386292051	trip-IND121004AAB	FBD_Balabhgarh_DPC (Haryana)	IND
8-10-03 429324	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	Carting	153861106442901555	trip-IND208006AAA	Kanpur_GovndNgr_DC (Uttar Pradesh)	IND
8-10-03 390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	153861115439069069	trip-IND627005AAA	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	IND
8-10-03 701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	153861118270144424	trip-IND583119AAA	Sandur_WrdN1DPP_D (Karnataka)	IND

Next steps: [Generate code with trip](#) [View recommended plots](#) [New interactive sheet](#)


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





	actual_time	segment_actual_time_sum
0	1562.0	1548.0
1	143.0	141.0
2	3347.0	3308.0
3	59.0	59.0
4	341.0	340.0
...
14782	83.0	82.0
14783	21.0	21.0
14784	282.0	281.0
14785	264.0	258.0
14786	275.0	274.0

14787 rows × 2 columns

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



	actual_distance_to_destination	osrm_distance
0	824.732854	991.3523
1	73.186911	85.1110
2	1927.404273	2354.0665
3	17.175274	19.6800
4	127.448500	146.7918
...
14782	57.762332	73.4630
14783	15.513784	16.0882
14784	38.684839	58.9037
14785	134.723836	171.1103
14786	66.081533	80.5787

14787 rows × 2 columns

Hypothesis Testing

```

trip['destination_name'] = trip['destination_name'].str.lower()
trip['source_name'] = trip['source_name'].str.lower()

def place2state(x):
    state = x.split('(')[1]

    return state[:-1]

def place2city(x):
    city = x.split(' ')[0]

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

    if city == 'pnq vadgaon sheri dpc':
        return 'vadgaonsheri'

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

    if city == 'hbr layout pc' : return 'bengaluru'
    if city == 'bhupal mp nagar' : return 'bhupal'
    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', 'destination_code']]
```






	destination_state	destination_city	destination_place	destination_code	
0	uttar pradesh	kanpur	central	6	
1	karnataka	doddablpur	chikadpp	d	
2	haryana	gurgaon	bilaspur	hb	
3	maharashtra	mumbai	mirard	ip	
4	karnataka	sandur	wrdn1dpp	d	
...	
14782	punjab	chandigarh	mehmdpur	h	
14783	haryana	faridabad	blbgarh	dc	
14784	uttar pradesh	kanpur	govndngr	dc	
14785	tamil nadu	tirchchndr	shnmgprn	d	
14786	karnataka	sandur	wrdn1dpp	d	

14787 rows × 4 columns

```
trip['source_state'] = trip['source_name'].apply(lambda x: place2state(x))
trip['source_city'] = trip['source_name'].apply(lambda x: place2city(x))
trip['source_place'] = trip['source_name'].apply(lambda x: place2city_place(x))
trip['source_code'] = trip['source_name'].apply(lambda x: place2code(x))
```

```
trip[['source_state', 'source_city', 'source_place', 'source_code']]
```



	source_state	source_city	source_place	source_code	
0	uttar pradesh	kanpur	central	6	
1	karnataka	doddablpur	chikadpp	d	
2	haryana	gurgaon	bilaspur	hb	
3	maharashtra	mumbai hub	mumbai	none	
4	karnataka	bellary	bellary	none	
...	
14782	punjab	chandigarh	mehmdpur	h	
14783	haryana	fbid	balabhgarh	dpc	
14784	uttar pradesh	kanpur	govndngr	dc	
14785	tamil nadu	tirunelveli	vdkkusrt	i	
14786	karnataka	sandur	wrdn1dpp	d	

14787 rows × 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', 'trip_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	0	12	37	2	
3	2018	9	0	12	37	2	
4	2018	9	0	12	37	2	
...	
14782	2018	10	23	3	40	2	
14783	2018	10	23	3	40	2	
14784	2018	10	23	3	40	2	
14785	2018	10	23	3	40	2	
14786	2018	10	23	3	40	2	

14787 rows × 6 columns

```
trip.head()
```



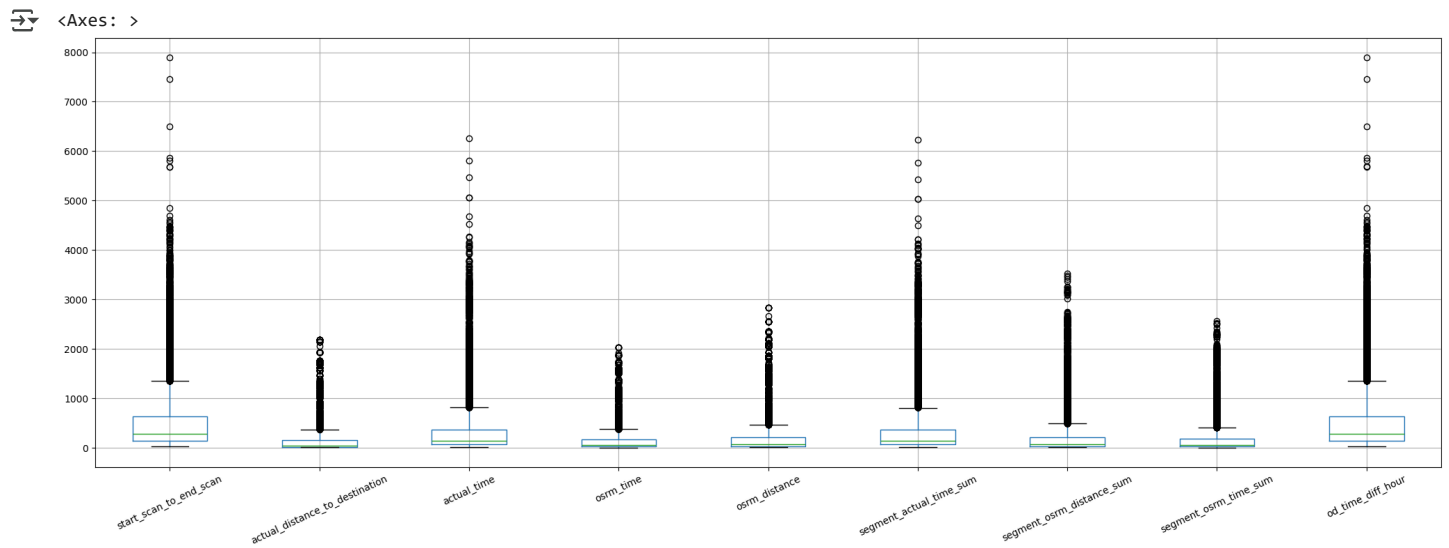
	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center
1g	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	trip-153671041653548748	IND209304AAA	kanpur_central_h_6 (uttar pradesh)	IND209304
1g	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	trip-153671042288605164	IND561203AAB	doddablpur_chikadpp_d (karnataka)	IND561203
1g	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	trip-153671043369099517	IND000000ACB	gurgaon_bilaspur_hb (haryana)	IND000000
1g	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	trip-153671046011330457	IND400072AAB	mumbai hub (maharashtra)	IND401104
1g	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	trip-153671052974046625	IND583101AAA	bellary_dc (karnataka)	IND583119

12 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 numerical 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))
```




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

```
IQR = Q3 - Q1
IQR
```

	0
start_scan_to_end_scan	483.000000
actual_distance_to_destination	140.814159
actual_time	300.000000
osrm_time	139.000000
osrm_distance	175.887300
segment_actual_time_sum	298.000000
segment_osrm_distance_sum	183.981750
segment_osrm_time_sum	154.000000
od_time_diff_hour	483.839201

dtype: float64

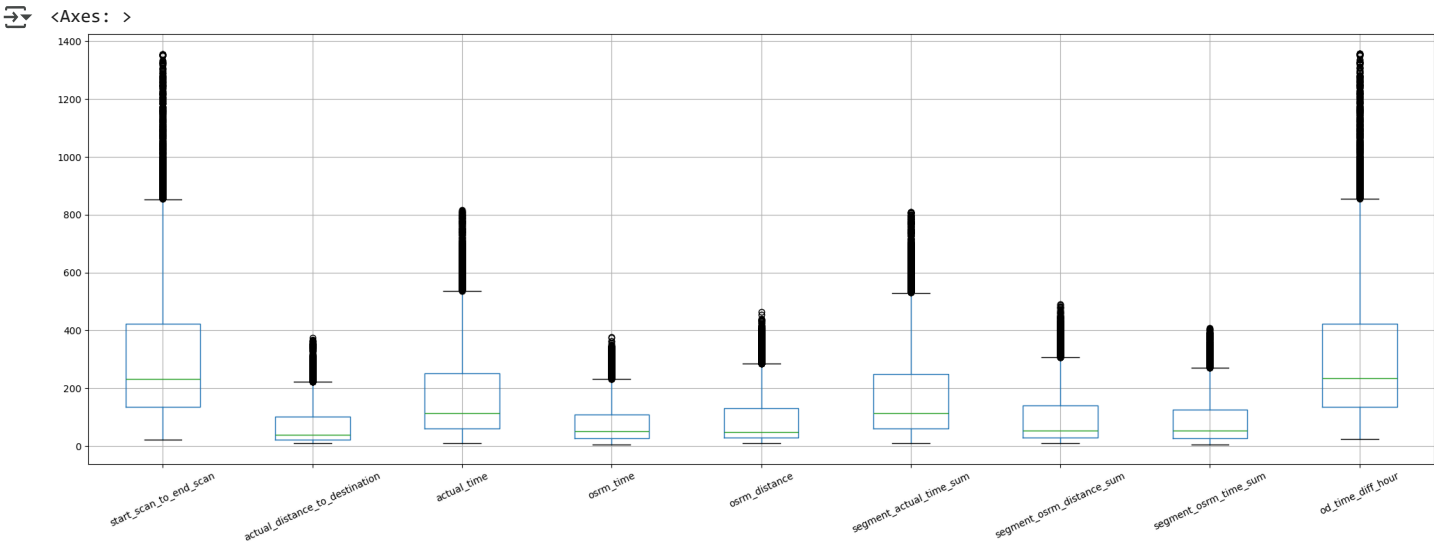
```
trip = trip[~((trip[num_cols] < (Q1 - 1.5 * IQR)) | (trip[num_cols] > (Q3 + 1.5 * IQR))).any(axis=1)]
trip = trip.reset_index(drop=True)
trip
```




	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destinat:
0	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	trip-153671042288605164	IND561203AAB	doddablpur_chikadpp_d (karnataka)	IND
1	training	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	trip-153671046011330457	IND400072AAB	mumbai hub (maharashtra)	IND
2	training	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	trip-153671052974046625	IND583101AAA	bellary_dc (karnataka)	IND
3	training	2018-09-12 00:02:34.161600	thanos::sroute:9bf03170-d0a2-4a3f-aa4d-9aab3d...	Carting	trip-153671055416136166	IND600056AAA	chennai_poonamallee (tamil nadu)	IND
4	training	2018-09-12 00:04:22.011653	thanos::sroute:a97698cc-846e-41a7-916b-88b1741...	Carting	trip-153671066201138152	IND600044AAD	chennai_chrompet_dpc (tamil nadu)	IND
...
12718	test	2018-10-03 23:55:56.258533	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	Carting	trip-153861095625827784	IND160002AAC	chandigarh_mehmdpur_h (punjab)	IND
12719	test	2018-10-03 23:57:23.863155	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	Carting	trip-153861104386292051	IND121004AAB	fbd_balabhgarh_dpc (haryana)	IND
12720	test	2018-10-03 23:57:44.429324	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	Carting	trip-153861106442901555	IND208006AAA	kanpur_govndngr_dc (uttar pradesh)	IND
12721	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	trip-153861115439069069	IND627005AAA	tirunelveli_vdkkusrt_i (tamil nadu)	IND
12722	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	trip-153861118270144424	IND583119AAA	sandur_wrdn1dpp_d (karnataka)	IND

12723 rows × 32 columns

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



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



	count
route_type	
Carting	8812
FTL	3911

dtype: int64

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

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

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