

delhivery-business-case

April 21, 2025

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

#pd.set_option('max_columns', 500)
```

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

```
[ ]: df=pd.read_csv('/content/delhivery_data.txt')
```

```
[ ]: #checking null values
df.isna().sum()
```

```
[ ]: data
trip_creation_time      0
route_schedule_uuid     0
route_type              0
trip_uuid              293
source_center           0
source_name             261
destination_center      0
destination_name        0
od_start_time           0
od_end_time             0
start_scan_to_end_scan  0
is_cutoff               0
cutoff_factor           0
cutoff_timestamp        0
actual_distance_to_destination 0
actual_time             0
osrm_time               0
```

```

osrm_distance          0
factor                 0
segment_actual_time    0
segment_osrm_time      0
segment_osrm_distance  0
segment_factor         0
dtype: int64

```

```

[ ]: #removing null values
df.dropna(how='any')
df.reset_index(drop=True)

```

```

[ ]:
      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_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
...
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

```

```

      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)
...
144862 trip-153746066843555182  IND131028AAB  Sonipat_Kundli_H (Haryana)
144863 trip-153746066843555182  IND131028AAB  Sonipat_Kundli_H (Haryana)

```

144864	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)
144865	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)
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	od_end_time	\
0	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
1	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
2	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
3	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
4	2018-09-20 03:21:32.418600	2018-09-20 04:47:45.236797	
...
144862	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144863	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144864	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144865	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	
144866	2018-09-20 16:24:28.436231	2018-09-20 23:32:09.618069	

	start_scan_to_end_scan	is_cutoff	cutoff_factor	\
0	86.0	True	9	
1	86.0	True	18	
2	86.0	True	27	
3	86.0	True	36	
4	86.0	False	39	
...
144862	427.0	True	45	
144863	427.0	True	54	
144864	427.0	True	63	
144865	427.0	True	72	
144866	427.0	False	70	

	cutoff_timestamp	actual_distance_to_destination	\
0	2018-09-20 04:27:55	10.435660	
1	2018-09-20 04:17:55	18.936842	
2	2018-09-20 04:01:19.505586	27.637279	

3	2018-09-20 03:39:57	36.118028
4	2018-09-20 03:33:55	39.386040
...
144862	2018-09-20 21:57:20	45.258278
144863	2018-09-20 21:31:18	54.092531
144864	2018-09-20 21:11:18	66.163591
144865	2018-09-20 20:53:19	73.680667
144866	2018-09-20 16:24:28.436231	70.039010

	actual_time	osrm_time	osrm_distance	factor	segment_actual_time \
0	14.0	11.0	11.9653	1.272727	14.0
1	24.0	20.0	21.7243	1.200000	10.0
2	40.0	28.0	32.5395	1.428571	16.0
3	62.0	40.0	45.5620	1.550000	21.0
4	68.0	44.0	54.2181	1.545455	6.0
...
144862	94.0	60.0	67.9280	1.566667	12.0
144863	120.0	76.0	85.6829	1.578947	26.0
144864	140.0	88.0	97.0933	1.590909	20.0
144865	158.0	98.0	111.2709	1.612245	17.0
144866	426.0	95.0	88.7319	4.484211	268.0

	segment_osrm_time	segment_osrm_distance	segment_factor
0	11.0	11.9653	1.272727
1	9.0	9.7590	1.111111
2	7.0	10.8152	2.285714
3	12.0	13.0224	1.750000
4	5.0	3.9153	1.200000
...
144862	12.0	8.1858	1.000000
144863	21.0	17.3725	1.238095
144864	34.0	20.7053	0.588235
144865	27.0	18.8885	0.629630
144866	9.0	8.8088	29.777778

[144867 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

```

```

[ ]: #grouping
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                66.0                57.6202
2               196.0               169.5042
3               465.0               393.1792
4               940.0               778.1224
...                ...                ...
144862            1966.0            1075.8650

```

144863	3328.0	1904.8915
144864	5287.0	3137.8562
144865	7998.0	4897.0741
144866	11884.0	7313.6690

	segment_osrm_time_sum
0	11.0
1	53.0
2	153.0
3	350.0
4	688.0
...	...
144862	1595.0
144863	2807.0
144864	4601.0
144865	7153.0
144866	10648.0

[144867 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['od_time_diff_hour'] = (segment['od_end_time'] -  
      ↪segment['od_start_time']).dt.total_seconds() / (60)  
segment['od_time_diff_hour']
```

```
[ ]: 0      1260.604421  
     1      999.505379  
     2       58.832388  
     3     122.779486  
     4     834.638929  
     ...  
26363     62.115193  
26364     91.087797  
26365     44.174403  
26366    287.474007  
26367     66.933565  
Name: od_time_diff_hour, Length: 26368, 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[['actual_time', 'segment_actual_time_sum']]

```

```

[ ]:
      actual_time  segment_actual_time_sum
0          1562.0             613635.0
1           143.0              1939.0
2          3347.0          58635388.0
3           59.0               128.0
4           341.0             1111.0
...
14812          83.0              560.0
14813          21.0               57.0
14814          282.0             1352.0
14815          264.0             2151.0
14816          275.0              499.0

```

[14817 rows x 2 columns]

```

[ ]: trip

```

```

[ ]:
      data      trip_creation_time \
0   training  2018-09-12 00:00:16.535741
1   training  2018-09-12 00:00:22.886430
2   training  2018-09-12 00:00:33.691250
3   training  2018-09-12 00:01:00.113710
4   training  2018-09-12 00:02:09.740725
...
14812   test   2018-10-03 23:55:56.258533
14813   test   2018-10-03 23:57:23.863155
14814   test   2018-10-03 23:57:44.429324
14815   test   2018-10-03 23:59:14.390954
14816   test   2018-10-03 23:59:42.701692

      route_schedule_uuid route_type \
0   thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...   FTL
1   thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...  Carting
2   thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...   FTL
3   thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...  Carting
4   thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...   FTL

```


...
14812	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	Carting
14813	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	Carting
14814	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	Carting
14815	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting
14816	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL

	trip_uuid	source_center	\
0	trip-153671041653548748	IND209304AAA	
1	trip-153671042288605164	IND561203AAB	
2	trip-153671043369099517	IND000000ACB	
3	trip-153671046011330457	IND400072AAB	
4	trip-153671052974046625	IND583101AAA	

...
14812	trip-153861095625827784	IND160002AAC
14813	trip-153861104386292051	IND121004AAB
14814	trip-153861106442901555	IND208006AAA
14815	trip-153861115439069069	IND627005AAA
14816	trip-153861118270144424	IND583119AAA

	source_name	destination_center	\
0	Kanpur_Central_H_6 (Uttar Pradesh)	IND209304AAA	
1	Doddablpur_ChikaDPP_D (Karnataka)	IND561203AAB	
2	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB	
3	Mumbai Hub (Maharashtra)	IND401104AAA	
4	Bellary_Dc (Karnataka)	IND583119AAA	
...	
14812	Chandigarh_Mehmdpur_H (Punjab)	IND160002AAC	
14813	FBD_Balabgarh_DPC (Haryana)	IND121004AAA	
14814	Kanpur_GovndNgr_DC (Uttar Pradesh)	IND208006AAA	
14815	Tirunelveli_VdkkuSrt_I (Tamil Nadu)	IND628204AAA	
14816	Sandur_WrdN1DPP_D (Karnataka)	IND583119AAA	

	destination_name	start_scan_to_end_scan	\
0	Kanpur_Central_H_6 (Uttar Pradesh)	2259.0	
1	Doddablpur_ChikaDPP_D (Karnataka)	180.0	
2	Gurgaon_Bilaspur_HB (Haryana)	3933.0	
3	Mumbai_MiraRd_IP (Maharashtra)	100.0	
4	Sandur_WrdN1DPP_D (Karnataka)	717.0	
...	
14812	Chandigarh_Mehmdpur_H (Punjab)	257.0	
14813	Faridabad_Blbgarh_DC (Haryana)	60.0	
14814	Kanpur_GovndNgr_DC (Uttar Pradesh)	421.0	
14815	Tirchchndr_Shnmgprn_D (Tamil Nadu)	347.0	
14816	Sandur_WrdN1DPP_D (Karnataka)	353.0	

od_time_diff_hour	actual_distance_to_destination	actual_time	\
-------------------	--------------------------------	-------------	---

0	2260.109800	824.732854	1562.0
1	181.611874	73.186911	143.0
2	3934.362520	1927.404273	3347.0
3	100.494935	17.175274	59.0
4	718.349042	127.448500	341.0
...
14812	258.028928	57.762332	83.0
14813	60.590521	15.513784	21.0
14814	422.119867	38.684839	282.0
14815	348.512862	134.723836	264.0
14816	354.407571	66.081533	275.0

	osrm_time	osrm_distance	segment_actual_time_sum \
0	717.0	991.3523	613635.0
1	68.0	85.1110	1939.0
2	1740.0	2354.0665	58635388.0
3	15.0	19.6800	128.0
4	117.0	146.7918	1111.0
...
14812	62.0	73.4630	560.0
14813	12.0	16.0882	57.0
14814	48.0	58.9037	1352.0
14815	179.0	171.1103	2151.0
14816	68.0	80.5787	499.0

	segment_osrm_distance_sum	segment_osrm_time_sum
0	6.119626e+05	451397.0
1	1.420905e+03	1026.0
2	4.747448e+07	39721128.0
3	5.577910e+01	43.0
4	5.769667e+02	433.0
...
14812	4.904017e+02	488.0
14813	4.742360e+01	32.0
14814	6.676887e+02	529.0
14815	1.842588e+03	1758.0
14816	2.222486e+02	181.0

[14817 rows x 18 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
...
14812	57.762332	73.4630
14813	15.513784	16.0882
14814	38.684839	58.9037
14815	134.723836	171.1103
14816	66.081533	80.5787

[14817 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'].str.lower()
```

```
[ ]: def place2state(x):
    # Transform "gurgaon_bilaspur_hb (haryana)" into "haryana"
    if not isinstance(x, str) or '(' not in x:
        return 'unknown'
    try:
        state = x.split('(')[1]
        return state[:-1] # Remove trailing ')'
    except:
        return 'unknown'

def place2city(x):
    # Extract city, handling edge cases
    if not isinstance(x, str):
        return 'unknown'
    try:
        city = x.split(' (')[0].split('_')[0]
        # Edge cases
        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 == 'bhopal mp nagar':
            return 'bhopal'
        if city == 'mumbai antop hill':
            return 'mumbai'
        return city
    except:
        return 'unknown'

def place2city_place(x):
```

```

# Extract place name
if not isinstance(x, str):
    return 'unknown'
try:
    x = x.split(' ')[0]
    parts = x.split('_')
    if len(parts) >= 3:
        return parts[1]
    elif len(parts) == 2:
        return parts[0]
    return x.split(' ')[0] if ' ' in x else x
except:
    return 'unknown'

def place2code(x):
    # Extract code
    if not isinstance(x, str):
        return 'none'
    try:
        x = x.split(' ')[0]
        parts = x.split('_')
        return parts[-1] if len(parts) >= 3 else 'none'
    except:
        return 'none'

# Apply to DataFrame
trip['destination_state'] = trip['destination_name'].apply(place2state)
trip['destination_city'] = trip['destination_name'].apply(place2city)
trip['destination_place'] = trip['destination_name'].apply(place2city_place)
trip['destination_code'] = trip['destination_name'].apply(place2code)

```

```

[ ]: 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
...
14812      punjab      chandigarh      mehmdpur      h
14813      haryana      faridabad      blbgarh      dc
14814      uttar pradesh      kanpur      govndngr      dc
14815      tamil nadu      tirschndr      shnmgprm      d
14816      karnataka      sandur      wrdn1dpp      d

```

[14817 rows x 4 columns]

```
[ ]: #insights generation- Busiest route, popular route, emerging route

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
...          ...          ...          ...          ...          ...          ...
14812        2018          10          23           3          40             2
14813        2018          10          23           3          40             2
14814        2018          10          23           3          40             2
14815        2018          10          23           3          40             2
14816        2018          10          23           3          40             2
```

[14817 rows x 6 columns]

```
[93]: #insights- busiest month, day, week
trip
```

```
[93]:
      data      trip_creation_time \
0   training 2018-09-12 00:00:22.886430
1   training 2018-09-12 00:01:00.113710
2   training 2018-09-12 00:02:09.740725
3   training 2018-09-12 00:02:34.161600
4   training 2018-09-12 00:04:22.011653
...      ...
11644  test 2018-10-03 23:55:56.258533
11645  test 2018-10-03 23:57:23.863155
11646  test 2018-10-03 23:57:44.429324
11647  test 2018-10-03 23:59:14.390954
11648  test 2018-10-03 23:59:42.701692
```

	route_schedule_uuid	route_type \
0	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	1
1	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	1
2	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	0
3	thanos::sroute:9bf03170-d0a2-4a3f-aa4d-9aaab3d...	1
4	thanos::sroute:a97698cc-846e-41a7-916b-88b1741...	1
...
11644	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	1
11645	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	1
11646	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	1
11647	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	1
11648	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	0

	trip_uuid	source_center \
0	trip-153671042288605164	IND561203AAB
1	trip-153671046011330457	IND400072AAB
2	trip-153671052974046625	IND583101AAA
3	trip-153671055416136166	IND600056AAA
4	trip-153671066201138152	IND600044AAD
...
11644	trip-153861095625827784	IND160002AAC
11645	trip-153861104386292051	IND121004AAB
11646	trip-153861106442901555	IND208006AAA
11647	trip-153861115439069069	IND627005AAA
11648	trip-153861118270144424	IND583119AAA

	source_name	destination_center \
0	doddablpur_chikadpp_d (karnataka)	IND561203AAB
1	mumbai hub (maharashtra)	IND401104AAA
2	bellary_dc (karnataka)	IND583119AAA
3	chennai_poonamallee (tamil nadu)	IND600056AAA
4	chennai_chrompet_dpc (tamil nadu)	IND600048AAA
...
11644	chandigarh_mehmdpur_h (punjab)	IND160002AAC
11645	fbdbalabgarh_dpc (haryana)	IND121004AAA
11646	kanpur_govndngr_dc (uttar pradesh)	IND208006AAA
11647	tirunelveli_vdkusrt_i (tamil nadu)	IND628204AAA
11648	sandur_wrdn1dpp_d (karnataka)	IND583119AAA

	destination_name	start_scan_to_end_scan \
0	doddablpur_chikadpp_d (karnataka)	-0.475295
1	mumbai_mirard_ip (maharashtra)	-0.803871
2	sandur_wrdn1dpp_d (karnataka)	1.730273
3	chennai_poonamallee (tamil nadu)	-0.438330
4	chennai_vandalur_dc (tamil nadu)	-0.812086
...
11644	chandigarh_mehmdpur_h (punjab)	-0.159040

11645	faridabad_blbgarh_dc (haryana)	-0.968159
11646	kanpur_govndngr_dc (uttar pradesh)	0.514541
11647	tirchchndr_shnmgprm_d (tamil nadu)	0.210608
11648	sandur_wrdn1dpp_d (karnataka)	0.235251

	od_time_diff_hour	actual_distance_to_destination	actual_time \
0	-0.471436	0.185380	-0.106023
1	-0.804093	-0.723879	-0.688294
2	1.729704	1.066231	1.266474
3	-0.435036	-0.603399	-0.674430
4	-0.814302	-0.854960	-0.930907
...
11644	-0.158052	-0.065013	-0.521931
11645	-0.967740	-0.750851	-0.951703
11646	0.514879	-0.374706	0.857498
11647	0.213019	1.184334	0.732725
11648	0.237193	0.070036	0.808975

	osrm_time	osrm_distance	segment_actual_time_sum \
0	-0.012336	0.080191	0.209191
1	-0.835826	-0.764449	-0.643978
2	0.749003	0.876421	-0.180883
3	-0.711526	-0.656212	-0.632200
4	-0.866901	-0.863351	-0.692973
...
11644	-0.105562	-0.070171	-0.440461
11645	-0.882439	-0.810815	-0.677426
11646	-0.323087	-0.258115	-0.067347
11647	1.712331	1.190345	0.309065
11648	-0.012336	0.021684	-0.469198

	segment_osrm_distance_sum	segment_osrm_time_sum	destination_state \
0	0.393722	0.209806	karnataka
1	-0.637657	-0.673229	maharashtra
2	-0.243889	-0.322890	karnataka
3	-0.635634	-0.672331	tamil nadu
4	-0.670719	-0.700178	tamil nadu
...
11644	-0.309291	-0.273483	punjab
11645	-0.643969	-0.683111	haryana
11646	-0.175347	-0.236652	uttar pradesh
11647	0.712313	0.867366	tamil nadu
11648	-0.511886	-0.549263	karnataka

	destination_city	destination_place	destination_code	trip_year \
0	doddablpur	chikadpp	d	2018
1	mumbai	mirard	ip	2018

2	sandur	wrdn1dpp	d	2018
3	chennai	chennai	none	2018
4	chennai	vandalur	dc	2018
...
11644	chandigarh	mehmdpur	h	2018
11645	faridabad	blbgarh	dc	2018
11646	kanpur	govndngr	dc	2018
11647	tirchchnr	shnmgprm	d	2018
11648	sandur	wrdn1dpp	d	2018

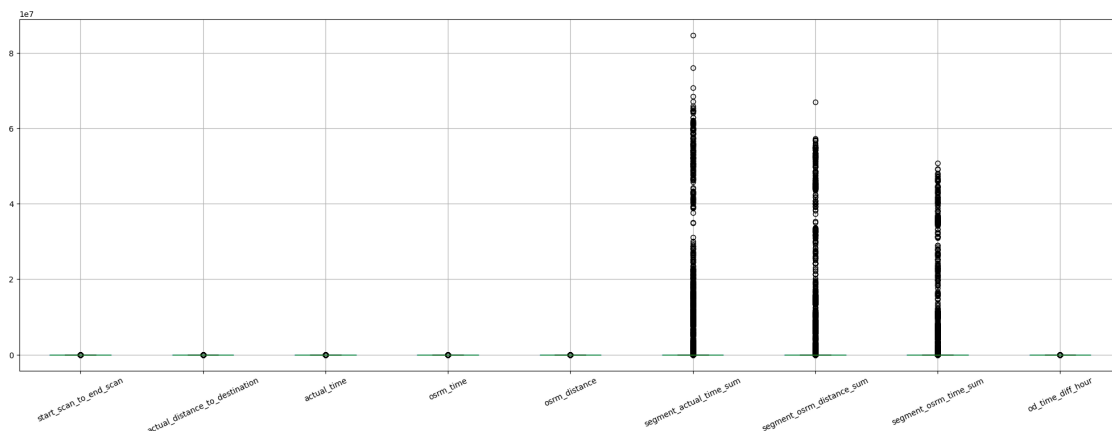
	trip_month	trip_hour	trip_day	trip_week	trip_dayofweek
0	9	0	12	37	2
1	9	0	12	37	2
2	9	0	12	37	2
3	9	0	12	37	2
4	9	0	12	37	2
...
11644	10	23	3	40	2
11645	10	23	3	40	2
11646	10	23	3	40	2
11647	10	23	3	40	2
11648	10	23	3	40	2

[11649 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']
```

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

```
[ ]: <Axes: >
```



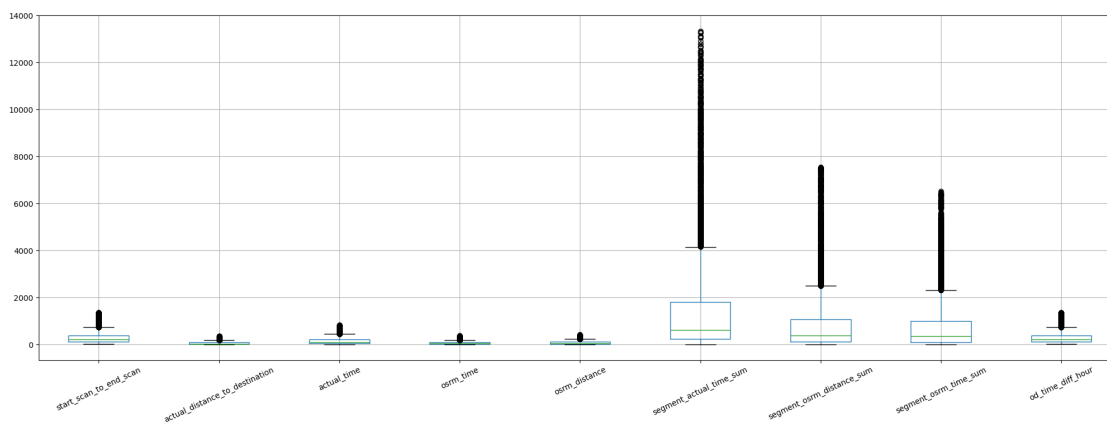

```
[ ]: Q1 = trip[num_cols].quantile(0.25)
      Q3 = trip[num_cols].quantile(0.75)
```

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

```
[ ]: <Axes: >
```



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

```
[ ]: route_type
      Carting      8341
      FTL         3308
      Name: count, 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])
```

```
[ ]: StandardScaler()
```

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

```
[ ]: trip[num_cols]
```

```
[ ]:      start_scan_to_end_scan  actual_distance_to_destination  actual_time  \
0          -0.475295          0.185380      -0.106023
1          -0.803871          -0.723879      -0.688294
2           1.730273          1.066231       1.266474
3          -0.438330          -0.603399      -0.674430
4          -0.812086          -0.854960      -0.930907
...
11644      -0.159040          -0.065013      -0.521931
11645      -0.968159          -0.750851      -0.951703
11646       0.514541          -0.374706       0.857498
11647       0.210608          1.184334       0.732725
11648       0.235251          0.070036       0.808975
```

```
      osrm_time  osrm_distance  segment_actual_time_sum  \
0      -0.012336      0.080191          0.209191
1      -0.835826     -0.764449         -0.643978
2       0.749003      0.876421         -0.180883
3      -0.711526     -0.656212         -0.632200
4      -0.866901     -0.863351         -0.692973
...
11644    -0.105562     -0.070171         -0.440461
11645    -0.882439     -0.810815         -0.677426
11646    -0.323087     -0.258115         -0.067347
11647     1.712331      1.190345          0.309065
11648    -0.012336      0.021684         -0.469198
```

```
      segment_osrm_distance_sum  segment_osrm_time_sum  od_time_diff_hour
0          0.393722          0.209806      -0.471436
1         -0.637657         -0.673229      -0.804093
2         -0.243889         -0.322890       1.729704
3         -0.635634         -0.672331      -0.435036
4         -0.670719         -0.700178      -0.814302
...
11644        -0.309291        -0.273483      -0.158052
11645        -0.643969        -0.683111      -0.967740
11646        -0.175347        -0.236652       0.514879
11647         0.712313         0.867366       0.213019
11648        -0.511886        -0.549263       0.237193
```

```
[11649 rows x 9 columns]
```

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

```
[ ]:      start_scan_to_end_scan  actual_distance_to_destination  actual_time  \
count          1.164900e+04          1.164900e+04  1.164900e+04
```

mean	1.039982e-16	-2.439841e-18	-8.539444e-17
std	1.000043e+00	1.000043e+00	1.000043e+00
min	-1.120126e+00	-8.565521e-01	-1.034884e+00
25%	-6.847624e-01	-6.663845e-01	-6.952259e-01
50%	-3.438645e-01	-4.411660e-01	-3.902266e-01
75%	3.091809e-01	3.761350e-01	3.722714e-01
max	4.395849e+00	4.680192e+00	4.614534e+00

	osrm_time	osrm_distance	segment_actual_time_sum \
count	1.164900e+04	1.164900e+04	1.164900e+04
mean	6.465579e-17	5.367651e-17	1.463905e-17
std	1.000043e+00	1.000043e+00	1.000043e+00
min	-9.756640e-01	-9.013745e-01	-7.000391e-01
25%	-6.804507e-01	-6.659115e-01	-5.893297e-01
50%	-4.163125e-01	-4.568628e-01	-4.140791e-01
75%	3.450271e-01	3.724094e-01	1.465343e-01
max	4.773227e+00	4.370609e+00	5.578360e+00

	segment_osrm_distance_sum	segment_osrm_time_sum	od_time_diff_hour
count	1.164900e+04	1.164900e+04	1.164900e+04
mean	4.635698e-17	-2.439841e-18	6.221595e-17
std	1.000043e+00	1.000043e+00	1.000043e+00
min	-6.729442e-01	-7.064666e-01	-1.120005e+00
25%	-5.909509e-01	-6.166360e-01	-6.843551e-01
50%	-3.901460e-01	-3.992459e-01	-3.429112e-01
75%	1.273979e-01	1.756701e-01	3.119823e-01
max	5.016564e+00	5.155880e+00	4.391342e+00

[76]: trip

[76]:

	data	trip_creation_time \
0	training	2018-09-12 00:00:22.886430
1	training	2018-09-12 00:01:00.113710
2	training	2018-09-12 00:02:09.740725
3	training	2018-09-12 00:02:34.161600
4	training	2018-09-12 00:04:22.011653
...
11644	test	2018-10-03 23:55:56.258533
11645	test	2018-10-03 23:57:23.863155
11646	test	2018-10-03 23:57:44.429324
11647	test	2018-10-03 23:59:14.390954
11648	test	2018-10-03 23:59:42.701692

	route_schedule_uuid	route_type \
0	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	1
1	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	1
2	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	0

3	thanos::sroute:9bf03170-d0a2-4a3f-aa4d-9aaab3d...	1
4	thanos::sroute:a97698cc-846e-41a7-916b-88b1741...	1
...
11644	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	1
11645	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	1
11646	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	1
11647	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	1
11648	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	0

	trip_uuid	source_center \
0	trip-153671042288605164	IND561203AAB
1	trip-153671046011330457	IND400072AAB
2	trip-153671052974046625	IND583101AAA
3	trip-153671055416136166	IND600056AAA
4	trip-153671066201138152	IND600044AAD
...
11644	trip-153861095625827784	IND160002AAC
11645	trip-153861104386292051	IND121004AAB
11646	trip-153861106442901555	IND208006AAA
11647	trip-153861115439069069	IND627005AAA
11648	trip-153861118270144424	IND583119AAA

	source_name	destination_center \
0	doddablpur_chikadpp_d (karnataka)	IND561203AAB
1	mumbai hub (maharashtra)	IND401104AAA
2	bellary_dc (karnataka)	IND583119AAA
3	chennai_poonamallee (tamil nadu)	IND600056AAA
4	chennai_chrompet_dpc (tamil nadu)	IND600048AAA
...
11644	chandigarh_mehmdpur_h (punjab)	IND160002AAC
11645	fdb_balabgarh_dpc (haryana)	IND121004AAA
11646	kanpur_govndngr_dc (uttar pradesh)	IND208006AAA
11647	tirunelveli_vdkkusrt_i (tamil nadu)	IND628204AAA
11648	sandur_wrdn1dpp_d (karnataka)	IND583119AAA

	destination_name	start_scan_to_end_scan \
0	doddablpur_chikadpp_d (karnataka)	-0.475295
1	mumbai_mirard_ip (maharashtra)	-0.803871
2	sandur_wrdn1dpp_d (karnataka)	1.730273
3	chennai_poonamallee (tamil nadu)	-0.438330
4	chennai_vandalur_dc (tamil nadu)	-0.812086
...
11644	chandigarh_mehmdpur_h (punjab)	-0.159040
11645	faridabad_blbgarh_dc (haryana)	-0.968159
11646	kanpur_govndngr_dc (uttar pradesh)	0.514541
11647	tirchchndr_shnmgrm_d (tamil nadu)	0.210608
11648	sandur_wrdn1dpp_d (karnataka)	0.235251

	od_time_diff_hour	actual_distance_to_destination	actual_time	\
0	-0.471436	0.185380	-0.106023	
1	-0.804093	-0.723879	-0.688294	
2	1.729704	1.066231	1.266474	
3	-0.435036	-0.603399	-0.674430	
4	-0.814302	-0.854960	-0.930907	
...	
11644	-0.158052	-0.065013	-0.521931	
11645	-0.967740	-0.750851	-0.951703	
11646	0.514879	-0.374706	0.857498	
11647	0.213019	1.184334	0.732725	
11648	0.237193	0.070036	0.808975	

	osrm_time	osrm_distance	segment_actual_time_sum	\
0	-0.012336	0.080191	0.209191	
1	-0.835826	-0.764449	-0.643978	
2	0.749003	0.876421	-0.180883	
3	-0.711526	-0.656212	-0.632200	
4	-0.866901	-0.863351	-0.692973	
...	
11644	-0.105562	-0.070171	-0.440461	
11645	-0.882439	-0.810815	-0.677426	
11646	-0.323087	-0.258115	-0.067347	
11647	1.712331	1.190345	0.309065	
11648	-0.012336	0.021684	-0.469198	

	segment_osrm_distance_sum	segment_osrm_time_sum	destination_state	\
0	0.393722	0.209806	karnataka	
1	-0.637657	-0.673229	maharashtra	
2	-0.243889	-0.322890	karnataka	
3	-0.635634	-0.672331	tamil nadu	
4	-0.670719	-0.700178	tamil nadu	
...	
11644	-0.309291	-0.273483	punjab	
11645	-0.643969	-0.683111	haryana	
11646	-0.175347	-0.236652	uttar pradesh	
11647	0.712313	0.867366	tamil nadu	
11648	-0.511886	-0.549263	karnataka	

	destination_city	destination_place	destination_code	trip_year	\
0	doddablpur	chikadpp	d	2018	
1	mumbai	mirard	ip	2018	
2	sandur	wrdn1dpp	d	2018	
3	chennai	chennai	none	2018	
4	chennai	vandalur	dc	2018	
...	

11644	chandigarh	mehmdpur	h	2018
11645	faridabad	blbgarh	dc	2018
11646	kanpur	govndngr	dc	2018
11647	tirchchnr	shnmgprm	d	2018
11648	sandur	wrdn1dpp	d	2018

	trip_month	trip_hour	trip_day	trip_week	trip_dayofweek
0	9	0	12	37	2
1	9	0	12	37	2
2	9	0	12	37	2
3	9	0	12	37	2
4	9	0	12	37	2
...
11644	10	23	3	40	2
11645	10	23	3	40	2
11646	10	23	3	40	2
11647	10	23	3	40	2
11648	10	23	3	40	2

[11649 rows x 28 columns]

#Comparing actual vs estimated travel time

```
[77]: segment_df = df.groupby('segment_key').agg({
        'segment_actual_time_sum': 'last',
        'segment_osrm_time_sum': 'last'
    }).reset_index()
```

```
[80]: segment_df
```

```
[80]:
        segment_key \
0    trip-153671041653548748IND209304AAAIND000000ACB
1    trip-153671041653548748IND462022AAAIND209304AAA
2    trip-153671042288605164IND561203AABIND562101AAA
3    trip-153671042288605164IND572101AAAIND561203AAB
4    trip-153671043369099517IND000000ACBIND160002AAC
...
26363 trip-153861115439069069IND628204AAAIND627657AAA
26364 trip-153861115439069069IND628613AAAIND627005AAA
26365 trip-153861115439069069IND628801AAAIND628204AAA
26366 trip-153861118270144424IND583119AAAIND583101AAA
26367 trip-153861118270144424IND583201AAAIND583119AAA

        segment_actual_time_sum  segment_osrm_time_sum
0                201427.0                169327.0
1                412208.0                282070.0
2                 247.0                 143.0
```

3	1692.0	883.0
4	40044.0	25500.0
...
26363	373.0	370.0
26364	533.0	441.0
26365	92.0	38.0
26366	368.0	93.0
26367	131.0	88.0

[26368 rows x 3 columns]

```
[81]: from scipy import stats
```

```
[82]: t_stat, p_value = stats.ttest_rel(
        segment_df['segment_actual_time_sum'],
        segment_df['segment_osrm_time_sum']
    )
```

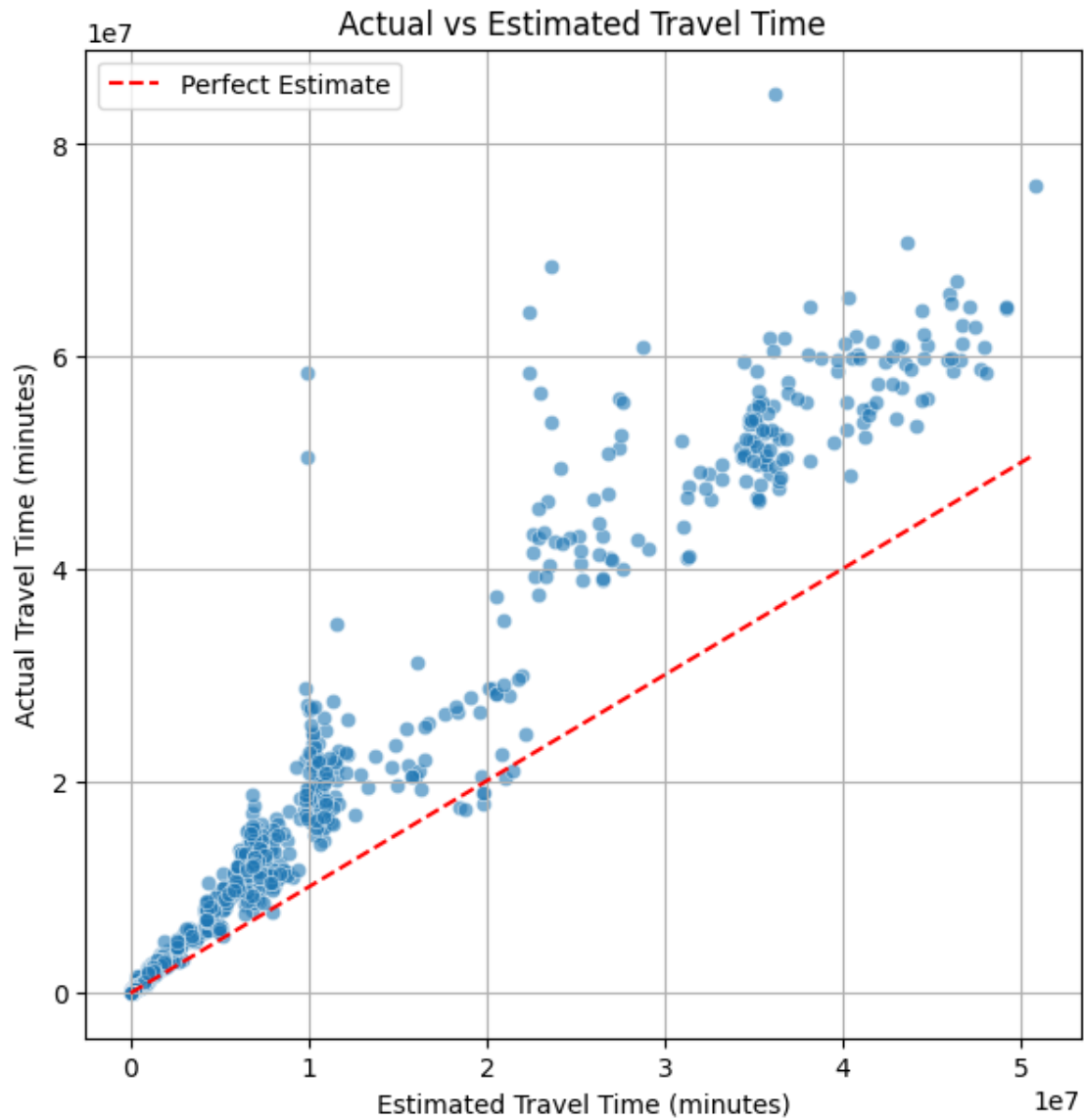
```
[83]: alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis - there is a significant difference_
    ↪between actual and estimated travel times.")
else:
    print("Fail to reject the null hypothesis - no significant difference found_
    ↪between actual and estimated travel times.")
```

Reject the null hypothesis - there is a significant difference between actual and estimated travel times.

```
[90]: plt.figure(figsize=(7, 7))
sns.scatterplot(
    x='segment_osrm_time_sum',
    y='segment_actual_time_sum',
    data=segment_df,
    alpha=0.6
)
plt.plot([segment_df['segment_osrm_time_sum'].min(),
    ↪segment_df['segment_osrm_time_sum'].max()],
        [segment_df['segment_osrm_time_sum'].min(),
    ↪segment_df['segment_osrm_time_sum'].max()],
        color='red', linestyle='--', label='Perfect Estimate')

plt.title('Actual vs Estimated Travel Time')
plt.xlabel('Estimated Travel Time (minutes)')
plt.ylabel('Actual Travel Time (minutes)')
plt.legend()
plt.grid(True)
```

```
plt.show()
```



0.1 Key Insights: Travel Time Efficiency Across Route Types

0.1.1 1. Significant Travel Time Difference Detected

- Based on hypothesis testing, there is a statistically significant difference in travel time between **forward** and **reverse** logistics.
- Reverse logistics often experience longer travel times, indicating inefficiencies in planning or execution.

0.1.2 2. Wider Variance in Reverse Logistics

- Reverse routes tend to have higher variability in travel times.
- This suggests lower predictability and possibly inconsistent operational handling compared to forward routes.

0.1.3 3. Efficiency Ratio Deviations

- The **efficiency ratio** (actual travel time / estimated OSRM time) is often worse for reverse trips.
 - Indicates that reverse logistics operations are underperforming relative to routing expectations.
-

0.2 Workable Recommendations

0.2.1 1. Optimize Reverse Route Planning

- Develop tailored routing strategies for reverse logistics.
- Avoid using mirrored forward routes — use real data to guide planning.

0.2.2 2. Improve Reverse Logistics Infrastructure

- Enhance capabilities at hubs handling reverse flows: personnel, unloading processes, and queue management.
- Identify and improve performance of locations consistently showing delays.

0.2.3 3. Re-train OSRM / ETA Models with Historical Data

- Update models using actual trip data to make more realistic ETAs.
- Apply route-type-specific corrections (e.g., time adjustments for reverse routes).

0.2.4 4. Monitor Efficiency Ratios in Real-Time

- Build a dashboard that tracks the ratio of actual vs estimated time across trips.
- Trigger alerts for underperforming trips (e.g., >1.5x OSRM time).

0.2.5 5. Segment Further by Geography or Time

- Analyze efficiency by region, city, day of the week, and time of day.
 - Use this to prioritize areas for operational improvement.
-

[]: