

```
In [50]: !pip install numpy
```

Requirement already satisfied: numpy in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (1.26.3)

[notice] A new release of pip is available: 23.2.1 -> 24.1.1

[notice] To update, run: python.exe -m pip install --upgrade pip

```
In [51]: !pip install pandas
```

Requirement already satisfied: pandas in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (2.2.0)

Requirement already satisfied: numpy<2,>=1.26.0 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from pandas) (1.26.3)

Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2.8.2)

Requirement already satisfied: pytz>=2020.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2023.3.post1)

Requirement already satisfied: tzdata>=2022.7 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from pandas) (2023.4)

Requirement already satisfied: six>=1.5 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

[notice] A new release of pip is available: 23.2.1 -> 24.1.1

[notice] To update, run: python.exe -m pip install --upgrade pip

```
In [52]: !pip install seaborn
```

Requirement already satisfied: seaborn in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (1.26.3)
Requirement already satisfied: pandas>=1.2 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (2.2.0)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from seaborn) (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.47.2)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=8 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from pandas>=1.2->seaborn) (2023.4)
Requirement already satisfied: six>=1.5 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
[notice] A new release of pip is available: 23.2.1 -> 24.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip

In [53]: !pip install matplotlib

Requirement already satisfied: matplotlib in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (3.8.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (4.47.2)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (1.4.5)
Requirement already satisfied: numpy<2,>=1.21 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (1.26.3)
Requirement already satisfied: packaging>=20.0 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (23.2)
Requirement already satisfied: pillow>=8 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)

[notice] A new release of pip is available: 23.2.1 -> 24.1.1

[notice] To update, run: python.exe -m pip install --upgrade pip

```
In [54]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

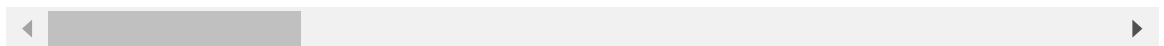
```
In [55]: df=pd.read_csv("delhivery_data.csv")
```

```
In [56]: df
```

Out[56]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355

144867 rows × 24 columns



1. Basic data cleaning and exploration:

In [57]: `df.isna().sum()`

```
Out[57]: data
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
```

```
In [58]: df['source_name'].fillna('no source name', inplace=True)
```

C:\Users\rahul\AppData\Local\Temp\ipykernel_3576\1465308693.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method. The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df['source_name'].fillna('no source name', inplace=True)
```

```
In [59]: df['destination_name'].fillna('no destination name', inplace=True)
```

C:\Users\rahul\AppData\Local\Temp\ipykernel_3576\626190445.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method. The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df['destination_name'].fillna('no destination name', inplace=True)
```

```
In [60]: df.isna().sum()
```

```

Out[60]: data
trip_creation_time      0
route_schedule_uuid     0
route_type              0
trip_uuid              0
source_center           0
source_name             0
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

```

2.merging the rows

```

In [130... df = pd.DataFrame(df)

df['segment_key'] = df['trip_uuid'] + '_' + df['source_center'] + '_' + df['dest

aggregated_df = df.groupby('segment_key').agg(
    segment_actual_time_sum=('segment_actual_time', 'sum'),
    segment_osrm_distance_sum=('segment_osrm_distance', 'sum'),
    segment_osrm_time_sum=('segment_osrm_time', 'sum')
).reset_index()

df_merged = df.merge(aggregated_df, on='segment_key')

print(df_merged)

```

	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)	

	start_scan_to_end_scan	...	hour	minute	second	weekday	state	\
0	86.0	...	2	35	36	3	Gujarat	
1	86.0	...	2	35	36	3	Gujarat	
2	86.0	...	2	35	36	3	Gujarat	
3	86.0	...	2	35	36	3	Gujarat	
4	86.0	...	2	35	36	3	Gujarat	
...	
144862	427.0	...	16	24	28	3	Haryana	

144863	427.0	...	16	24	28	3	Haryana
144864	427.0	...	16	24	28	3	Haryana
144865	427.0	...	16	24	28	3	Haryana
144866	427.0	...	16	24	28	3	Haryana

	city	corridor	segment_actual_time_sum \
0	Anand	IND388121AAA to IND388620AAB	67.0
1	Anand	IND388121AAA to IND388620AAB	67.0
2	Anand	IND388121AAA to IND388620AAB	67.0
3	Anand	IND388121AAA to IND388620AAB	67.0
4	Anand	IND388121AAA to IND388620AAB	67.0
...
144862	Sonipat	IND131028AAB to IND000000ACB	423.0
144863	Sonipat	IND131028AAB to IND000000ACB	423.0
144864	Sonipat	IND131028AAB to IND000000ACB	423.0
144865	Sonipat	IND131028AAB to IND000000ACB	423.0
144866	Sonipat	IND131028AAB to IND000000ACB	423.0

	segment_osrm_distance_sum	segment_osrm_time_sum
0	49.4772	44.0
1	49.4772	44.0
2	49.4772	44.0
3	49.4772	44.0
4	49.4772	44.0
...
144862	131.1238	185.0
144863	131.1238	185.0
144864	131.1238	185.0
144865	131.1238	185.0
144866	131.1238	185.0

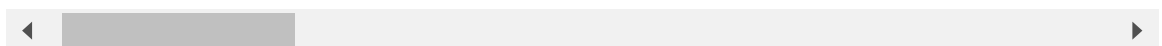
[144867 rows x 43 columns]

In [62]: df

Out[62]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320
...
62	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182
63	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182
64	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182
65	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182
66	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182

67 rows × 25 columns



In [63]:

```
df = pd.DataFrame(df)

df['segment_key'] = df['trip_uuid'] + '_' + df['source_center'] + '_' + df['dest

create_segment_dict = {
    'segment_actual_time': 'sum',
    'segment_osrm_distance': 'sum',
    'segment_osrm_time': 'sum',
    'od_end_time': 'last',
}
```

```

aggregated_df = df.groupby('segment_key').agg(create_segment_dict).reset_index()

sorted_df = aggregated_df.sort_values(by=['segment_key', 'od_end_time']).reset_i

print(sorted_df)

```

	segment_key	segment_actual_time \
0	trip-153671041653548748_IND209304AAA_IND000000ACB	728.0
1	trip-153671041653548748_IND462022AAA_IND209304AAA	820.0
2	trip-153671042288605164_IND561203AAB_IND562101AAA	46.0
3	trip-153671042288605164_IND572101AAA_IND561203AAB	95.0
4	trip-153671043369099517_IND000000ACB_IND160002AAC	608.0
...
26363	trip-153861115439069069_IND628204AAA_IND627657AAA	49.0
26364	trip-153861115439069069_IND628613AAA_IND627005AAA	89.0
26365	trip-153861115439069069_IND628801AAA_IND628204AAA	29.0
26366	trip-153861118270144424_IND583119AAA_IND583101AAA	233.0
26367	trip-153861118270144424_IND583201AAA_IND583119AAA	41.0

	segment_osrm_distance	segment_osrm_time	od_end_time
0	670.6205	534.0	2018-09-13 13:40:23.123744
1	649.8528	474.0	2018-09-12 16:39:46.858469
2	28.1995	26.0	2018-09-12 03:01:59.598855
3	55.9899	39.0	2018-09-12 02:03:09.655591
4	317.7408	231.0	2018-09-14 17:34:55.442454
...
26363	42.1431	42.0	2018-10-04 03:31:11.183797
26364	78.5869	77.0	2018-10-04 05:47:45.162682
26365	16.0184	14.0	2018-10-04 02:29:04.272194
26366	52.5303	42.0	2018-10-04 08:46:09.166940
26367	28.0484	25.0	2018-10-04 03:58:40.726547

[26368 rows x 5 columns]

3.Feature Engineering

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

.

```

In [64]: df = pd.DataFrame(df)

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

df['od_time_diff_hour'] = (df['od_end_time'] - df['od_start_time']).dt.total_sec

df.drop(columns=['od_start_time', 'od_end_time'], inplace=True)

```

```
print(df)
```

	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)	

	start_scan_to_end_scan	...	actual_time	osrm_time	osrm_distance	\
0	86.0	...	14.0	11.0	11.9653	
1	86.0	...	24.0	20.0	21.7243	
2	86.0	...	40.0	28.0	32.5395	
3	86.0	...	62.0	40.0	45.5620	
4	86.0	...	68.0	44.0	54.2181	
...	
144862	427.0	...	94.0	60.0	67.9280	

144863	427.0	...	120.0	76.0	85.6829
144864	427.0	...	140.0	88.0	97.0933
144865	427.0	...	158.0	98.0	111.2709
144866	427.0	...	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	

	segment_key	od_time_diff_hour
0	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
1	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
2	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
3	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
4	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
...
144862	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144863	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144864	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144865	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144866	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106

[144867 rows x 24 columns]

In [65]: df

Out[65]:

source_name	destination_center	destination_name	start_scan_to_end_scan	...	ac
-------------	--------------------	------------------	------------------------	-----	----

nd_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86.0	...	
----------------------------	--------------	----------------------------------	------	-----	--

nd_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86.0	...	
----------------------------	--------------	----------------------------------	------	-----	--

nd_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86.0	...	
----------------------------	--------------	----------------------------------	------	-----	--

nd_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86.0	...	
----------------------------	--------------	----------------------------------	------	-----	--

nd_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86.0	...	
----------------------------	--------------	----------------------------------	------	-----	--

...
-----	-----	-----	-----	-----	-----

ionipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427.0	...	
-------------------------------	--------------	----------------------------------	-------	-----	--

ionipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427.0	...	
-------------------------------	--------------	----------------------------------	-------	-----	--

ionipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427.0	...	
-------------------------------	--------------	----------------------------------	-------	-----	--

ionipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427.0	...	
-------------------------------	--------------	----------------------------------	-------	-----	--

ionipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427.0	...	
-------------------------------	--------------	----------------------------------	-------	-----	--



2. Destination Name: Split and extract features out of destination. City-place-code(state)

```
In [66]: df = pd.DataFrame(df)

print("Original DataFrame:")
print(df)

def extract_features(destination_name):
```

```
import re
match = re.match(r'^([_]+)_([_]+)(.+?)\s\((\w+)\)', destination_name)
if match:
    city = match.group(1)
    place_code = match.group(2) + '_' + match.group(3)
    state = match.group(4)
    return city, place_code, state
else:
    return None, None, None

df[['destination_city', 'destination_place_code', 'destination_state']] = df['de

print("\nDataFrame with extracted destination features:")
print(df)
```

Original DataFrame:

	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)	

	start_scan_to_end_scan	...	actual_time	osrm_time	osrm_distance	\
0	86.0	...	14.0	11.0	11.9653	
1	86.0	...	24.0	20.0	21.7243	
2	86.0	...	40.0	28.0	32.5395	
3	86.0	...	62.0	40.0	45.5620	
4	86.0	...	68.0	44.0	54.2181	
...	

144862	427.0	...	94.0	60.0	67.9280
144863	427.0	...	120.0	76.0	85.6829
144864	427.0	...	140.0	88.0	97.0933
144865	427.0	...	158.0	98.0	111.2709
144866	427.0	...	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	

	segment_key	od_time_diff_hour
0	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
1	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
2	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
3	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
4	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
...
144862	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144863	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144864	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144865	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144866	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106

[144867 rows x 24 columns]

DataFrame with extracted destination features:

	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)

	start_scan_to_end_scan	...	factor	segment_actual_time	\
0	86.0	...	1.272727	14.0	
1	86.0	...	1.200000	10.0	
2	86.0	...	1.428571	16.0	
3	86.0	...	1.550000	21.0	
4	86.0	...	1.545455	6.0	
...	
144862	427.0	...	1.566667	12.0	
144863	427.0	...	1.578947	26.0	
144864	427.0	...	1.590909	20.0	
144865	427.0	...	1.612245	17.0	
144866	427.0	...	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

	segment_key	od_time_diff_hour	\
0	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
1	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
2	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
3	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
4	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
...	
144862	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106	
144863	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106	
144864	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106	
144865	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106	
144866	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106	

	destination_city	destination_place_code	destination_state
0	Khambhat	MotvdDPP_D	Gujarat
1	Khambhat	MotvdDPP_D	Gujarat
2	Khambhat	MotvdDPP_D	Gujarat
3	Khambhat	MotvdDPP_D	Gujarat
4	Khambhat	MotvdDPP_D	Gujarat
...
144862	Gurgaon	Bilaspur_HB	Haryana
144863	Gurgaon	Bilaspur_HB	Haryana
144864	Gurgaon	Bilaspur_HB	Haryana
144865	Gurgaon	Bilaspur_HB	Haryana
144866	Gurgaon	Bilaspur_HB	Haryana

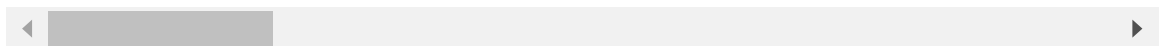
[144867 rows x 27 columns]

In [67]: df

Out[67]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355

144867 rows × 27 columns



3. Source Name: Split and extract features out of destination. City-place-code

(State)

```
In [68]: print("Original DataFrame:")
print(df)

def extract_features(source_name):
    import re
    match = re.match(r'([^_]+)_([^_]+)(.+(?))s\(((w+)\))', source_name)
```

```
if match:
    city = match.group(1)
    place_code = match.group(2) + '_' + match.group(3)
    state = match.group(4) # Enclosed in parentheses
    return city, place_code, state
else:
    return None, None, None

df[['source_city', 'source_place_code', 'source_state']] = df['source_name'].app

print("\nDataFrame with extracted source features:")
print(df)
```

Original DataFrame:

	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)	

	start_scan_to_end_scan	...	factor	segment_actual_time	\
0	86.0	...	1.272727	14.0	
1	86.0	...	1.200000	10.0	
2	86.0	...	1.428571	16.0	
3	86.0	...	1.550000	21.0	
4	86.0	...	1.545455	6.0	
...	

144862	427.0	...	1.566667	12.0
144863	427.0	...	1.578947	26.0
144864	427.0	...	1.590909	20.0
144865	427.0	...	1.612245	17.0
144866	427.0	...	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	

	segment_key	od_time_diff_hour	\
0	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
1	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
2	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
3	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
4	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894	
...	
144862	trip-153746066843555182_IND131028AAB_IND0000000ACB	7.128106	
144863	trip-153746066843555182_IND131028AAB_IND0000000ACB	7.128106	
144864	trip-153746066843555182_IND131028AAB_IND0000000ACB	7.128106	
144865	trip-153746066843555182_IND131028AAB_IND0000000ACB	7.128106	
144866	trip-153746066843555182_IND131028AAB_IND0000000ACB	7.128106	

	destination_city	destination_place_code	destination_state
0	Khambhat	MotvdDPP_D	Gujarat
1	Khambhat	MotvdDPP_D	Gujarat
2	Khambhat	MotvdDPP_D	Gujarat
3	Khambhat	MotvdDPP_D	Gujarat
4	Khambhat	MotvdDPP_D	Gujarat
...
144862	Gurgaon	Bilaspur_HB	Haryana
144863	Gurgaon	Bilaspur_HB	Haryana
144864	Gurgaon	Bilaspur_HB	Haryana
144865	Gurgaon	Bilaspur_HB	Haryana
144866	Gurgaon	Bilaspur_HB	Haryana

[144867 rows x 27 columns]

DataFrame with extracted source features:

	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)

	start_scan_to_end_scan	...	segment_osrm_distance	segment_factor \
0	86.0	...	11.9653	1.272727
1	86.0	...	9.7590	1.111111
2	86.0	...	10.8152	2.285714
3	86.0	...	13.0224	1.750000
4	86.0	...	3.9153	1.200000
...
144862	427.0	...	8.1858	1.000000
144863	427.0	...	17.3725	1.238095
144864	427.0	...	20.7053	0.588235
144865	427.0	...	18.8885	0.629630
144866	427.0	...	8.8088	29.777778

	segment_key	od_time_diff_hour \
0	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
1	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
2	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
3	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
4	trip-153741093647649320_IND388121AAA_IND388620AAB	1.436894
...

144862	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144863	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144864	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144865	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106
144866	trip-153746066843555182_IND131028AAB_IND000000ACB	7.128106

	destination_city	destination_place_code	destination_state	\
0	Khambhat	MotvdDPP_D	Gujarat	
1	Khambhat	MotvdDPP_D	Gujarat	
2	Khambhat	MotvdDPP_D	Gujarat	
3	Khambhat	MotvdDPP_D	Gujarat	
4	Khambhat	MotvdDPP_D	Gujarat	
...	
144862	Gurgaon	Bilaspur_HB	Haryana	
144863	Gurgaon	Bilaspur_HB	Haryana	
144864	Gurgaon	Bilaspur_HB	Haryana	
144865	Gurgaon	Bilaspur_HB	Haryana	
144866	Gurgaon	Bilaspur_HB	Haryana	

	source_city	source_place_code	source_state
0	Anand	VUNagar_DC	Gujarat
1	Anand	VUNagar_DC	Gujarat
2	Anand	VUNagar_DC	Gujarat
3	Anand	VUNagar_DC	Gujarat
4	Anand	VUNagar_DC	Gujarat
...
144862	Sonipat	Kundli_H	Haryana
144863	Sonipat	Kundli_H	Haryana
144864	Sonipat	Kundli_H	Haryana
144865	Sonipat	Kundli_H	Haryana
144866	Sonipat	Kundli_H	Haryana

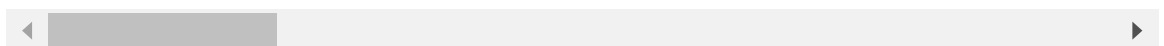
[144867 rows x 30 columns]

In [69]: df

Out[69]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355

144867 rows × 30 columns



4. Trip_creation_time: Extract features like month, year, day, etc.

In [131...]

```
df = pd.DataFrame(df)

df['trip_creation_time'] = pd.to_datetime(df['trip_creation_time'])

print("Original DataFrame:")
print(df)
```

```
df['year'] = df['trip_creation_time'].dt.year
df['month'] = df['trip_creation_time'].dt.month
df['day'] = df['trip_creation_time'].dt.day
df['hour'] = df['trip_creation_time'].dt.hour
df['minute'] = df['trip_creation_time'].dt.minute
df['second'] = df['trip_creation_time'].dt.second

df['weekday'] = df['trip_creation_time'].dt.weekday # Monday=0, Sunday=6

print("\nDataFrame with extracted time features:")
print(df)
```

Original DataFrame:

	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)	

	start_scan_to_end_scan	...	year	month	day	hour	minute	second	\
0	86.0	...	2018	9	20	2	35	36	
1	86.0	...	2018	9	20	2	35	36	
2	86.0	...	2018	9	20	2	35	36	
3	86.0	...	2018	9	20	2	35	36	
4	86.0	...	2018	9	20	2	35	36	
...	

144862	427.0	...	2018	9	20	16	24	28
144863	427.0	...	2018	9	20	16	24	28
144864	427.0	...	2018	9	20	16	24	28
144865	427.0	...	2018	9	20	16	24	28
144866	427.0	...	2018	9	20	16	24	28

	weekday	state	city	corridor
0	3	Gujarat	Anand	IND388121AAA to IND388620AAB
1	3	Gujarat	Anand	IND388121AAA to IND388620AAB
2	3	Gujarat	Anand	IND388121AAA to IND388620AAB
3	3	Gujarat	Anand	IND388121AAA to IND388620AAB
4	3	Gujarat	Anand	IND388121AAA to IND388620AAB
...
144862	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144863	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144864	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144865	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144866	3	Haryana	Sonipat	IND131028AAB to IND000000ACB

[144867 rows x 40 columns]

DataFrame with extracted time features:

	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)

	start_scan_to_end_scan	...	year	month	day	hour	minute	second	\
0	86.0	...	2018	9	20	2	35	36	
1	86.0	...	2018	9	20	2	35	36	
2	86.0	...	2018	9	20	2	35	36	
3	86.0	...	2018	9	20	2	35	36	
4	86.0	...	2018	9	20	2	35	36	
...	
144862	427.0	...	2018	9	20	16	24	28	
144863	427.0	...	2018	9	20	16	24	28	
144864	427.0	...	2018	9	20	16	24	28	
144865	427.0	...	2018	9	20	16	24	28	
144866	427.0	...	2018	9	20	16	24	28	

	weekday	state	city	corridor
0	3	Gujarat	Anand	IND388121AAA to IND388620AAB
1	3	Gujarat	Anand	IND388121AAA to IND388620AAB
2	3	Gujarat	Anand	IND388121AAA to IND388620AAB
3	3	Gujarat	Anand	IND388121AAA to IND388620AAB
4	3	Gujarat	Anand	IND388121AAA to IND388620AAB
...
144862	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144863	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144864	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144865	3	Haryana	Sonipat	IND131028AAB to IND000000ACB
144866	3	Haryana	Sonipat	IND131028AAB to IND000000ACB

[144867 rows x 40 columns]

In [71]: df

Out[71]:

inter	source_name	destination_center	destination_name	start_scan_to_end_scan
-------	-------------	--------------------	------------------	------------------------

AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86
AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86
AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86
AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86
AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	86
...
IAAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427
IAAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427
IAAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427
IAAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427
IAAB	Sonipat_Kundli_H (Haryana)	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	427



4. In-depth analysis:

1. Grouping and Aggregating at Trip-level

a. Groups the segment data by the trip_uuid column to focus on aggregating data at the trip level.

```
In [72]: df = pd.DataFrame(df)

print("Original DataFrame:")
```

```
print(df)

trip_agg = df.groupby('trip_uuid').agg({
    'segment_actual_time': 'sum',
    'segment_osrm_distance': 'sum',
    'segment_osrm_time': 'sum'
}).reset_index()

print("\nAggregated DataFrame at Trip-level:")
print(trip_agg)
```


Original DataFrame:

	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)	

	start_scan_to_end_scan	...	source_city	source_place_code	\
0	86.0	...	Anand	VUNagar_DC	
1	86.0	...	Anand	VUNagar_DC	
2	86.0	...	Anand	VUNagar_DC	
3	86.0	...	Anand	VUNagar_DC	
4	86.0	...	Anand	VUNagar_DC	
...	

144862	427.0	...	Sonipat	Kundli_H
144863	427.0	...	Sonipat	Kundli_H
144864	427.0	...	Sonipat	Kundli_H
144865	427.0	...	Sonipat	Kundli_H
144866	427.0	...	Sonipat	Kundli_H

	source_state	year	month	day	hour	minute	second	weekday
0	Gujarat	2018	9	20	2	35	36	3
1	Gujarat	2018	9	20	2	35	36	3
2	Gujarat	2018	9	20	2	35	36	3
3	Gujarat	2018	9	20	2	35	36	3
4	Gujarat	2018	9	20	2	35	36	3
...
144862	Haryana	2018	9	20	16	24	28	3
144863	Haryana	2018	9	20	16	24	28	3
144864	Haryana	2018	9	20	16	24	28	3
144865	Haryana	2018	9	20	16	24	28	3
144866	Haryana	2018	9	20	16	24	28	3

[144867 rows x 37 columns]

Aggregated DataFrame at Trip-level:

	trip_uuid	segment_actual_time	segment_osrm_distance \
0	trip-153671041653548748	1548.0	1320.4733
1	trip-153671042288605164	141.0	84.1894
2	trip-153671043369099517	3308.0	2545.2678
3	trip-153671046011330457	59.0	19.8766
4	trip-153671052974046625	340.0	146.7919
...
14812	trip-153861095625827784	82.0	64.8551
14813	trip-153861104386292051	21.0	16.0883
14814	trip-153861106442901555	281.0	104.8866
14815	trip-153861115439069069	258.0	223.5324
14816	trip-153861118270144424	274.0	80.5787

	segment_osrm_time
0	1008.0
1	65.0
2	1941.0
3	16.0
4	115.0
...	...
14812	62.0
14813	11.0
14814	88.0
14815	221.0
14816	67.0

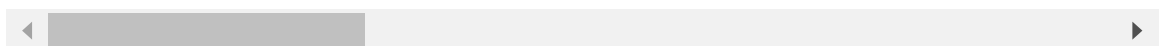
[14817 rows x 4 columns]

In [73]: df

Out[73]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355

144867 rows × 37 columns



b. Apply suitable aggregation functions like first, last, and sum specified in

the create_trip_dict dictionary to calculate summary statistics for each trip.

```
In [74]: df = pd.DataFrame(df)

print("Original DataFrame:")
print(df)

create_trip_dict = {
```

```
        'segment_actual_time': ['first', 'last', 'sum'],      # First, last, and total
        'segment_osrm_distance': ['first', 'last', 'sum'],   # First, last, and total
        'segment_osrm_time': ['first', 'last', 'sum']        # First, last, and total
    }

trip_summary = df.groupby('trip_uuid').agg(create_trip_dict).reset_index()

print("\nAggregated DataFrame with summary statistics for each trip:")
print(trip_summary)
```

Original DataFrame:

	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)	

	start_scan_to_end_scan	...	source_city	source_place_code	\
0	86.0	...	Anand	VUNagar_DC	
1	86.0	...	Anand	VUNagar_DC	
2	86.0	...	Anand	VUNagar_DC	
3	86.0	...	Anand	VUNagar_DC	
4	86.0	...	Anand	VUNagar_DC	
...	

144862	427.0	...	Sonipat	Kundli_H
144863	427.0	...	Sonipat	Kundli_H
144864	427.0	...	Sonipat	Kundli_H
144865	427.0	...	Sonipat	Kundli_H
144866	427.0	...	Sonipat	Kundli_H

	source_state	year	month	day	hour	minute	second	weekday
0	Gujarat	2018	9	20	2	35	36	3
1	Gujarat	2018	9	20	2	35	36	3
2	Gujarat	2018	9	20	2	35	36	3
3	Gujarat	2018	9	20	2	35	36	3
4	Gujarat	2018	9	20	2	35	36	3
...
144862	Haryana	2018	9	20	16	24	28	3
144863	Haryana	2018	9	20	16	24	28	3
144864	Haryana	2018	9	20	16	24	28	3
144865	Haryana	2018	9	20	16	24	28	3
144866	Haryana	2018	9	20	16	24	28	3

[144867 rows x 37 columns]

Aggregated DataFrame with summary statistics for each trip:

	trip_uuid	segment_actual_time		
		first	last	sum
0	trip-153671041653548748	43.0	20.0	1548.0
1	trip-153671042288605164	14.0	15.0	141.0
2	trip-153671043369099517	60.0	275.0	3308.0
3	trip-153671046011330457	23.0	36.0	59.0
4	trip-153671052974046625	42.0	79.0	340.0
...
14812	trip-153861095625827784	12.0	14.0	82.0
14813	trip-153861104386292051	12.0	9.0	21.0
14814	trip-153861106442901555	76.0	102.0	281.0
14815	trip-153861115439069069	18.0	51.0	258.0
14816	trip-153861118270144424	30.0	188.0	274.0

	segment_osrm_distance		segment_osrm_time		
	first	last	sum	first	last
0	55.2597	15.0693	1320.4733	39.0	10.0
1	10.3544	6.9464	84.1894	8.0	7.0
2	30.0119	32.8506	2545.2678	27.0	28.0
3	11.9675	7.9091	19.8766	9.0	7.0
4	23.8920	29.0538	146.7919	17.0	26.0
...
14812	9.3059	5.6953	64.8551	6.0	5.0
14813	10.4451	5.6432	16.0883	7.0	4.0
14814	31.0564	3.2092	104.8866	25.0	7.0
14815	11.9050	38.3058	223.5324	13.0	47.0
14816	23.2367	28.5437	80.5787	21.0	25.0

	sum
0	1008.0
1	65.0
2	1941.0
3	16.0
4	115.0
...	...
14812	62.0
14813	11.0

```
14814    88.0
14815   221.0
14816    67.0
```

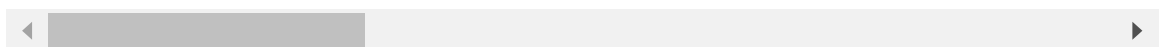
```
[14817 rows x 10 columns]
```

```
In [75]: df
```

```
Out[75]:
```

	data	trip_creation_time	route_schedule_uuid	route_type	trip_
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355

```
144867 rows x 37 columns
```



```
In [76]: trip_summary
```

Out[76]:

	trip_uuid	segment_actual_time			segment_osrm_distance			segment	
		first	last	sum	first	last	sum	first	last
0	trip-153671041653548748	43.0	20.0	1548.0	55.2597	15.0693	1320.4733	39.0	10
1	trip-153671042288605164	14.0	15.0	141.0	10.3544	6.9464	84.1894	8.0	
2	trip-153671043369099517	60.0	275.0	3308.0	30.0119	32.8506	2545.2678	27.0	20
3	trip-153671046011330457	23.0	36.0	59.0	11.9675	7.9091	19.8766	9.0	
4	trip-153671052974046625	42.0	79.0	340.0	23.8920	29.0538	146.7919	17.0	20
...	
14812	trip-153861095625827784	12.0	14.0	82.0	9.3059	5.6953	64.8551	6.0	
14813	trip-153861104386292051	12.0	9.0	21.0	10.4451	5.6432	16.0883	7.0	
14814	trip-153861106442901555	76.0	102.0	281.0	31.0564	3.2092	104.8866	25.0	
14815	trip-153861115439069069	18.0	51.0	258.0	11.9050	38.3058	223.5324	13.0	40
14816	trip-153861118270144424	30.0	188.0	274.0	23.2367	28.5437	80.5787	21.0	20

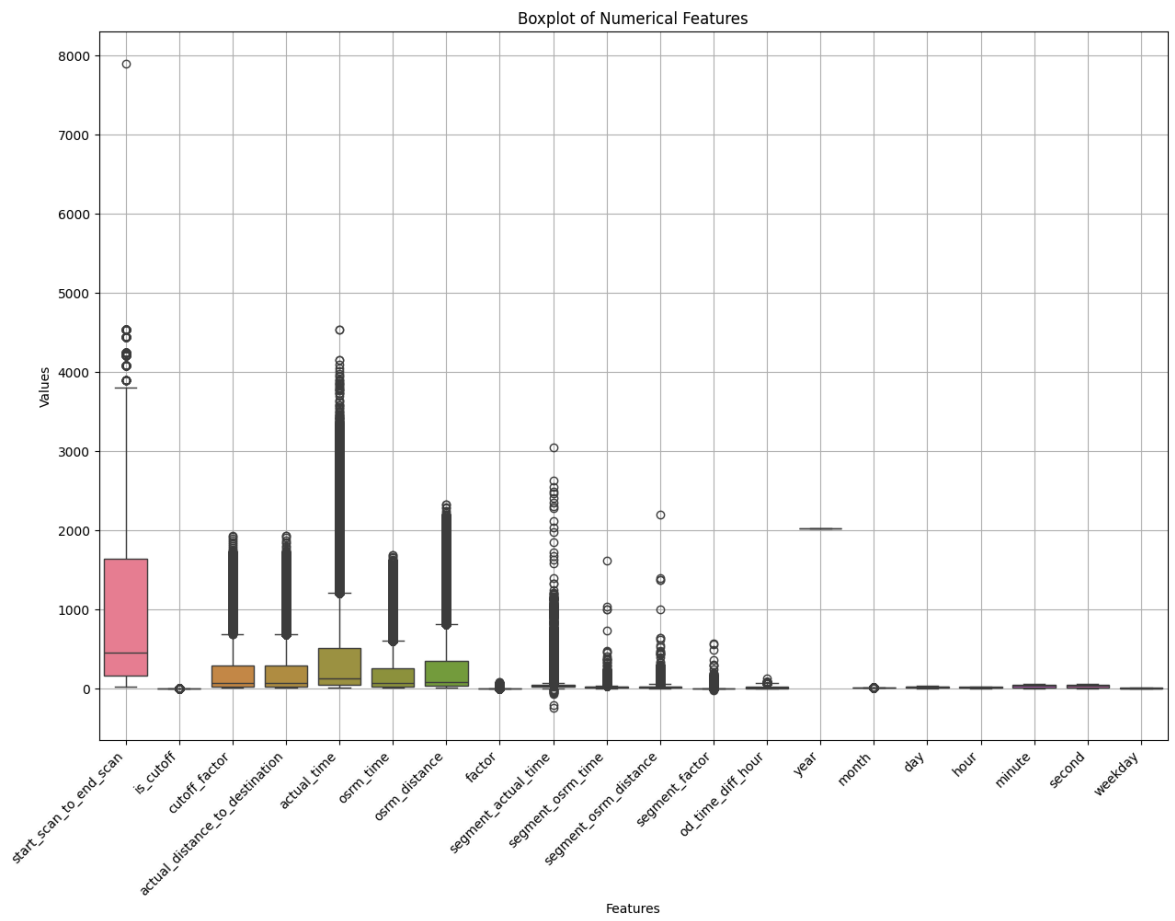
14817 rows × 10 columns



2. Outlier Detection & Treatment

a. Find any existing outliers in numerical features.

```
In [85]: plt.figure(figsize=(15, 10))
sns.boxplot(data=df)
plt.title('Boxplot of Numerical Features')
plt.xlabel('Features')
plt.ylabel('Values')
plt.xticks(rotation=45, ha='right')
plt.grid(True)
plt.show()
```

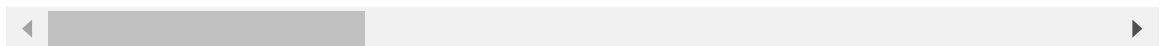



In [86]: df

Out[86]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355

144867 rows × 37 columns



In [92]:

```
df = pd.DataFrame(df)

print("Original DataFrame:")
print(df)

df_encoded = pd.get_dummies(df, columns=['route_type', 'data' ])

print("\nDataFrame after one-hot encoding:")
print(df_encoded)
```

Original DataFrame:

	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)	

	start_scan_to_end_scan	...	source_city	source_place_code	\
0	86.0	...	Anand	VUNagar_DC	
1	86.0	...	Anand	VUNagar_DC	
2	86.0	...	Anand	VUNagar_DC	
3	86.0	...	Anand	VUNagar_DC	
4	86.0	...	Anand	VUNagar_DC	
...	

144862	427.0	...	Sonipat	Kundli_H
144863	427.0	...	Sonipat	Kundli_H
144864	427.0	...	Sonipat	Kundli_H
144865	427.0	...	Sonipat	Kundli_H
144866	427.0	...	Sonipat	Kundli_H

	source_state	year	month	day	hour	minute	second	weekday
0	Gujarat	2018	9	20	2	35	36	3
1	Gujarat	2018	9	20	2	35	36	3
2	Gujarat	2018	9	20	2	35	36	3
3	Gujarat	2018	9	20	2	35	36	3
4	Gujarat	2018	9	20	2	35	36	3
...
144862	Haryana	2018	9	20	16	24	28	3
144863	Haryana	2018	9	20	16	24	28	3
144864	Haryana	2018	9	20	16	24	28	3
144865	Haryana	2018	9	20	16	24	28	3
144866	Haryana	2018	9	20	16	24	28	3

[144867 rows x 37 columns]

DataFrame after one-hot encoding:

	trip_creation_time	\
0	2018-09-20 02:35:36.476840	
1	2018-09-20 02:35:36.476840	
2	2018-09-20 02:35:36.476840	
3	2018-09-20 02:35:36.476840	
4	2018-09-20 02:35:36.476840	
...
144862	2018-09-20 16:24:28.436231	
144863	2018-09-20 16:24:28.436231	
144864	2018-09-20 16:24:28.436231	
144865	2018-09-20 16:24:28.436231	
144866	2018-09-20 16:24:28.436231	

	route_schedule_uuid	\
0	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	
1	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	
2	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	
3	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	
4	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	
...
144862	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	
144863	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	
144864	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	
144865	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	
144866	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	

	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)	

	start_scan_to_end_scan	is_cutoff	cutoff_factor	...	month	day	\
0	86.0	True	9	...	9	20	
1	86.0	True	18	...	9	20	
2	86.0	True	27	...	9	20	
3	86.0	True	36	...	9	20	
4	86.0	False	39	...	9	20	
...	
144862	427.0	True	45	...	9	20	
144863	427.0	True	54	...	9	20	
144864	427.0	True	63	...	9	20	
144865	427.0	True	72	...	9	20	
144866	427.0	False	70	...	9	20	

	hour	minute	second	weekday	route_type_Carting	route_type_FTL	\
0	2	35	36	3	True	False	
1	2	35	36	3	True	False	
2	2	35	36	3	True	False	
3	2	35	36	3	True	False	
4	2	35	36	3	True	False	
...	
144862	16	24	28	3	True	False	
144863	16	24	28	3	True	False	
144864	16	24	28	3	True	False	
144865	16	24	28	3	True	False	
144866	16	24	28	3	True	False	

	data_test	data_training
0	False	True
1	False	True
2	False	True
3	False	True
4	False	True
...
144862	False	True
144863	False	True
144864	False	True
144865	False	True
144866	False	True

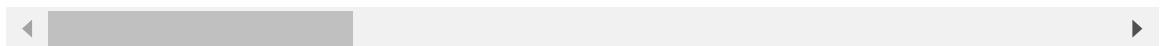
[144867 rows x 39 columns]

In [93]: df_encoded

Out[93]:

	trip_creation_time	route_schedule_uuid	trip_uuid	source_center
0	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	trip- 153741093647649320	IND388121AAA
1	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	trip- 153741093647649320	IND388121AAA
2	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	trip- 153741093647649320	IND388121AAA
3	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	trip- 153741093647649320	IND388121AAA
4	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	trip- 153741093647649320	IND388121AAA
...
144862	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	trip- 153746066843555182	IND131028AAB
144863	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	trip- 153746066843555182	IND131028AAB
144864	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	trip- 153746066843555182	IND131028AAB
144865	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	trip- 153746066843555182	IND131028AAB
144866	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	trip- 153746066843555182	IND131028AAB

144867 rows × 39 columns



5. Hypothesis Testing:

In [105...]

```
!pip install scipy
```

Requirement already satisfied: scipy in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (1.13.0)

Requirement already satisfied: numpy<2.3,>=1.22.4 in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (from scipy) (1.26.3)

[notice] A new release of pip is available: 23.2.1 -> 24.1.1

[notice] To update, run: python.exe -m pip install --upgrade pip

In [106...]

```
!pip install stats
```

Requirement already satisfied: stats in c:\users\rahul\appdata\local\programs\python\python312\lib\site-packages (0.1.2a0)

[notice] A new release of pip is available: 23.2.1 -> 24.1.1

[notice] To update, run: python.exe -m pip install --upgrade pip

In [110... `from scipy.stats import ttest_rel`

1. Perform hypothesis testing / visual analysis between :

.

a. actual_time aggregated value and OSRM time aggregated value.

```
In [111... df = pd.DataFrame(df)

# Perform paired t-test
statistic, p_value = ttest_rel(df['actual_time'], df['osrm_time'])

print("Paired t-test results:")
print("Statistic:", statistic)
print("P-value:", p_value)

alpha = 0.05 # significance level

if p_value < alpha:
    print("Reject null hypothesis: There is a significant difference between act
else:
    print("Fail to reject null hypothesis: There is no significant difference be
```

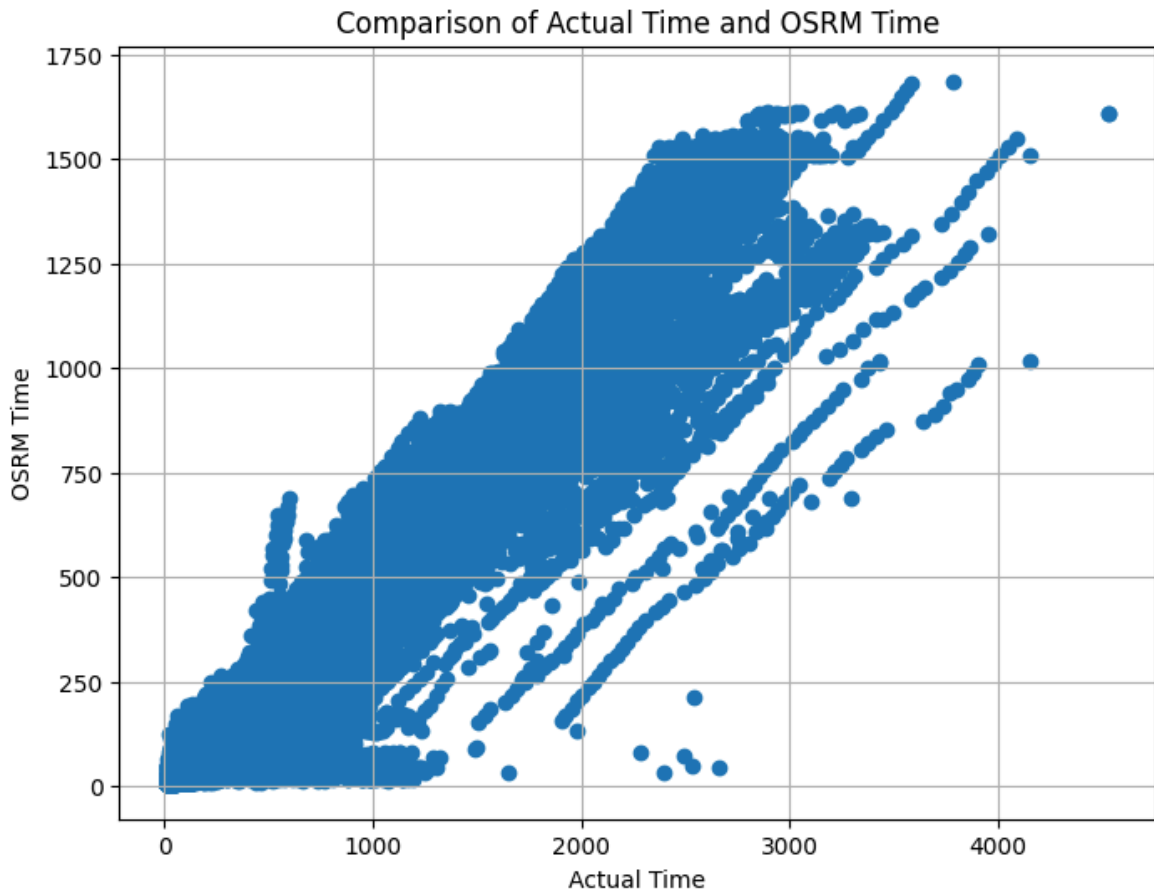
Paired t-test results:

Statistic: 254.44873336004324

P-value: 0.0

Reject null hypothesis: There is a significant difference between actual_time and OSRM time.

```
In [112... plt.figure(figsize=(8, 6))
plt.scatter(df['actual_time'], df['osrm_time'])
plt.title('Comparison of Actual Time and OSRM Time')
plt.xlabel('Actual Time')
plt.ylabel('OSRM Time')
plt.grid(True)
plt.show()
```



b. actual_time aggregated value and segment actual time aggregated value.

```
In [113...] df = pd.DataFrame(df)

statistic, p_value = ttest_rel(df['actual_time'], df['segment_actual_time'])

print("Paired t-test results:")
print("Statistic:", statistic)
print("P-value:", p_value)

alpha = 0.05

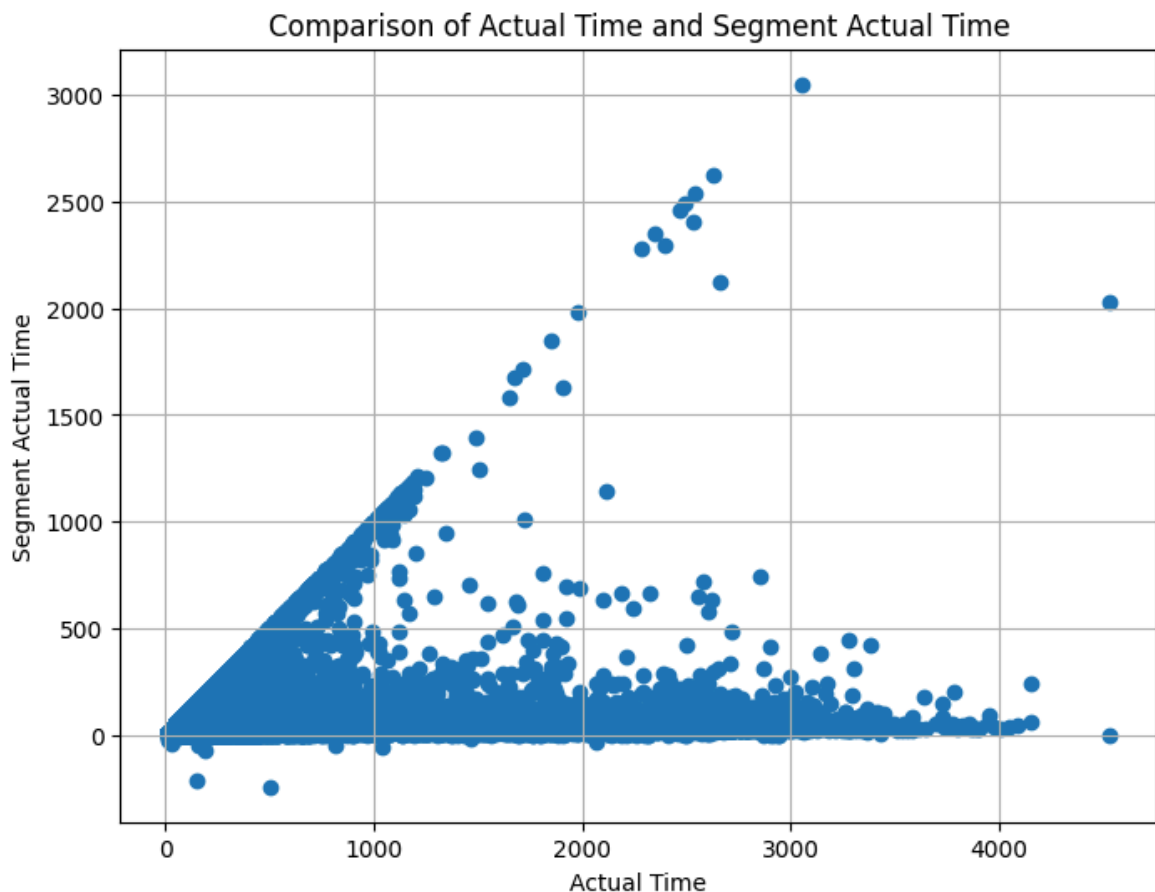
if p_value < alpha:
    print("Reject null hypothesis: There is a significant difference between act
else:
    print("Fail to reject null hypothesis: There is no significant difference be
```

```
Paired t-test results:
Statistic: 244.03197807537543
P-value: 0.0
Reject null hypothesis: There is a significant difference between actual_time and
segment_actual_time.
```

```
In [114...] plt.figure(figsize=(8, 6))
plt.scatter(df['actual_time'], df['segment_actual_time'])
plt.title('Comparison of Actual Time and Segment Actual Time')
plt.xlabel('Actual Time')
plt.ylabel('Segment Actual Time')
```



```
plt.grid(True)
plt.show()
```



c. OSRM distance aggregated value and segment OSRM distance aggregated value.

```
In [115... df = pd.DataFrame(df)

statistic, p_value = ttest_rel(df['osrm_distance'], df['segment_osrm_distance'])

print("Paired t-test results:")
print("Statistic:", statistic)
print("P-value:", p_value)

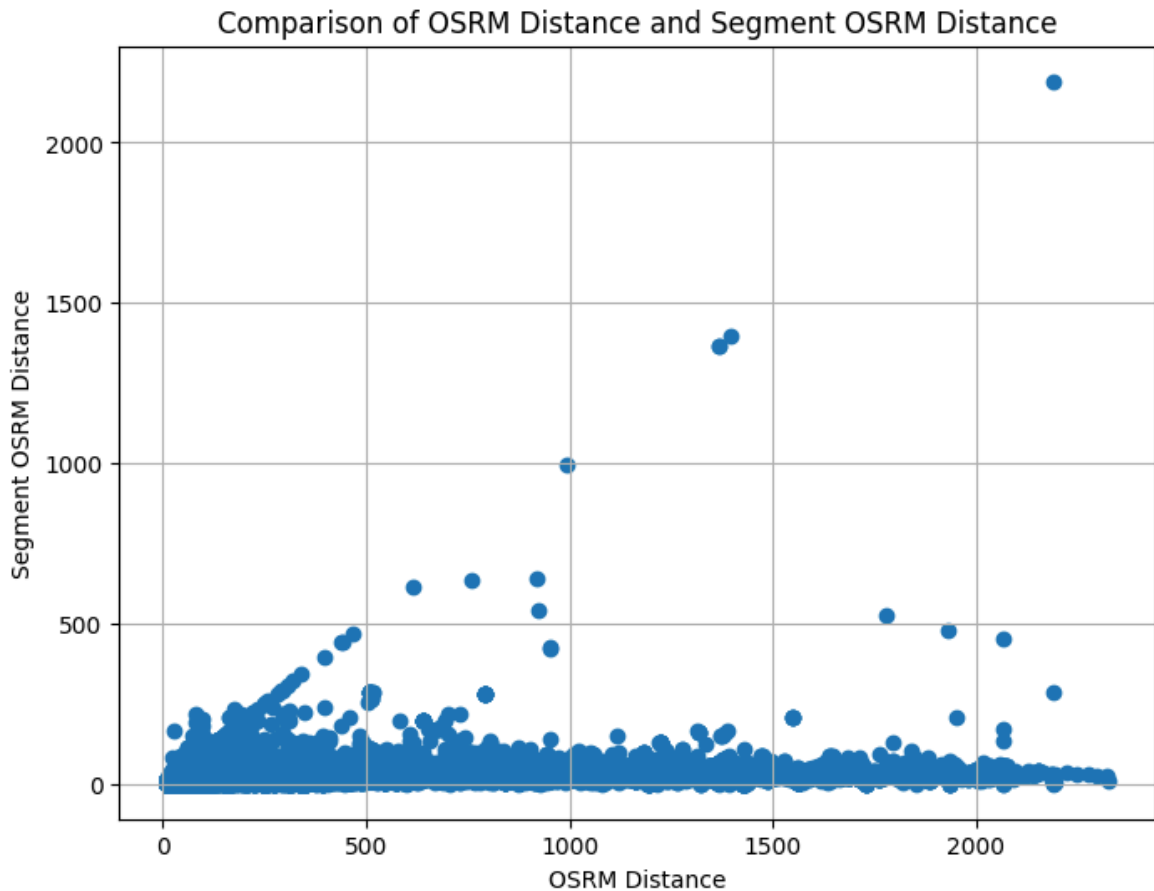
alpha = 0.05

if p_value < alpha:
    print("Reject null hypothesis: There is a significant difference between OSRM distance and segment_osrm_distance.")
else:
    print("Fail to reject null hypothesis: There is no significant difference between OSRM distance and segment_osrm_distance.")
```

```
Paired t-test results:
Statistic: 238.97194176199704
P-value: 0.0
Reject null hypothesis: There is a significant difference between OSRM distance and segment_osrm_distance.
```

```
In [116... plt.figure(figsize=(8, 6))
plt.scatter(df['osrm_distance'], df['segment_osrm_distance'])
```

```
plt.title('Comparison of OSRM Distance and Segment OSRM Distance')
plt.xlabel('OSRM Distance')
plt.ylabel('Segment OSRM Distance')
plt.grid(True)
plt.show()
```



d. OSRM time aggregated value and segment OSRM time aggregated value.

```
In [117... df = pd.DataFrame(df)

statistic, p_value = ttest_rel(df['osrm_time'], df['segment_osrm_time'])

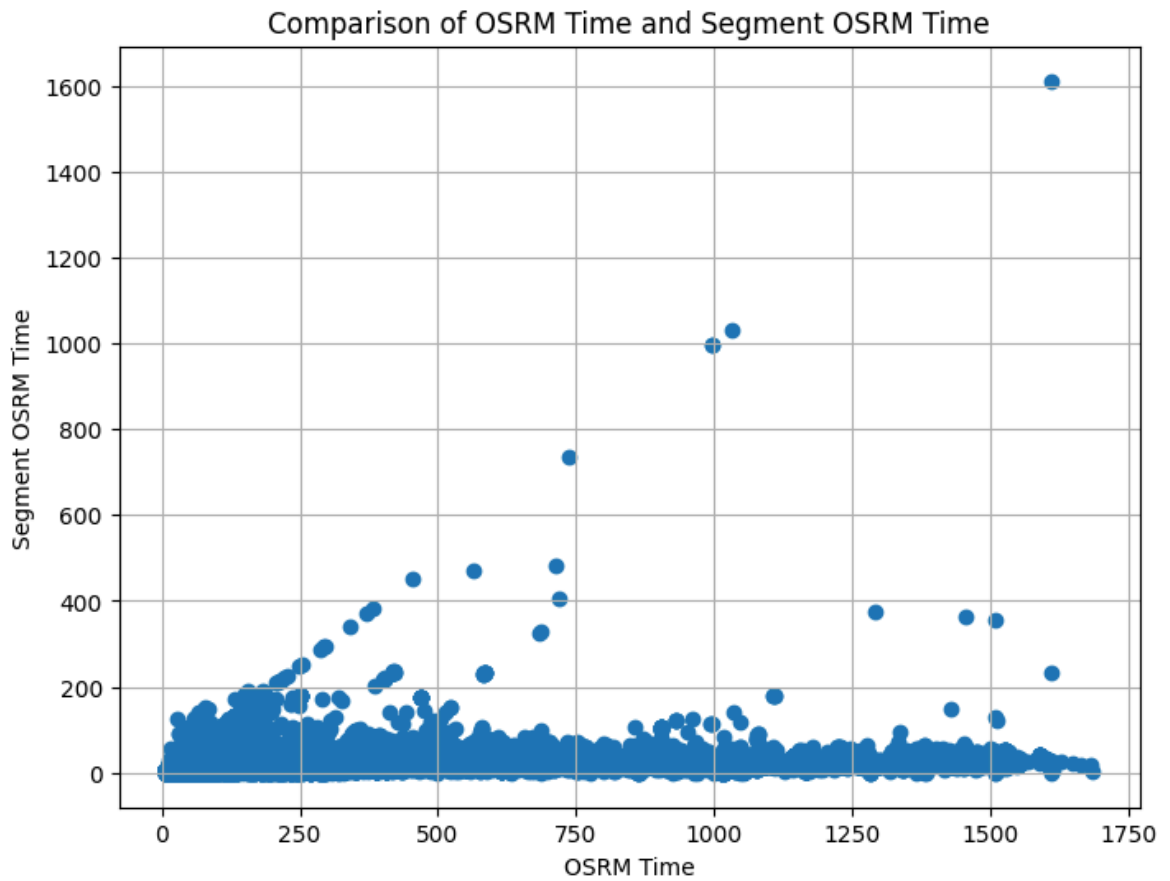
# Print results
print("Paired t-test results:")
print("Statistic:", statistic)
print("P-value:", p_value)

alpha = 0.05

if p_value < alpha:
    print("Reject null hypothesis: There is a significant difference between OSR
else:
    print("Fail to reject null hypothesis: There is no significant difference be
```

```
Paired t-test results:
Statistic: 243.2027973184123
P-value: 0.0
Reject null hypothesis: There is a significant difference between OSRM time and s
egment_osrm_time.
```

```
In [118... plt.figure(figsize=(8, 6))
plt.scatter(df['osrm_time'], df['segment_osrm_time'])
plt.title('Comparison of OSRM Time and Segment OSRM Time')
plt.xlabel('OSRM Time')
plt.ylabel('Segment OSRM Time')
plt.grid(True)
plt.show()
```

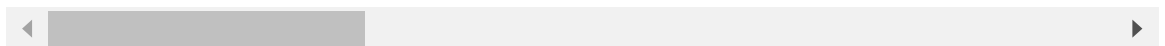


```
In [124... df
```

Out[124...

	data	trip_creation_time	route_schedule_uuid	route_type	trip_
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	15374109364764
...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	15374606684355

144867 rows × 37 columns



6. Business Insights & Recommendations

top source states

In [125...

```
df = pd.DataFrame(df)

df['state'] = df['source_name'].str.extract(r'\(((.*?)\)')

state_counts = df['state'].value_counts().reset_index()
```

```

state_counts.columns = ['state', 'order_count']

top_states = state_counts.sort_values(by='order_count', ascending=False)

print("Top Source States:")
print(top_states)

```

Top Source States:

	state	order_count
0	Haryana	27499
1	Maharashtra	21401
2	Karnataka	19578
3	Tamil Nadu	7494
4	Gujarat	7202
5	Uttar Pradesh	7137
6	Telangana	6496
7	West Bengal	5963
8	Andhra Pradesh	5539
9	Rajasthan	5267
10	Punjab	4704
11	Delhi	4398
12	Bihar	4190
13	Madhya Pradesh	4021
14	Assam	2875
15	Jharkhand	2597
16	Kerala	2413
17	Orissa	2094
18	Uttarakhand	1162
19	Himachal Pradesh	587
20	Goa	514
21	Chandigarh	507
22	Arunachal Pradesh	245
23	Chhattisgarh	229
24	Jammu & Kashmir	226
25	Meghalaya	86
26	Pondicherry	49
27	Nagaland	40
28	Dadra and Nagar Haveli	30
29	Mizoram	26
30	Tripura	5

top source cities

```

In [126... df = pd.DataFrame(df)

df['city'] = df['source_city']

city_counts = df['city'].value_counts().reset_index()
city_counts.columns = ['city', 'order_count']

top_cities = city_counts.sort_values(by='order_count', ascending=False)

```

```
print("Top Source Cities:")
print(top_cities)
```

Top Source Cities:

	city	order_count
0	Gurgaon	23646
1	Bangalore	10104
2	Bhiwandi	9088
3	Bengaluru	4237
4	Pune	4230
..
652	Sumerpur	1
651	Kothanallloor	1
650	Pachora	1
648	Bhadra	1
657	Hajipur	1

[658 rows x 2 columns]

Busiest corridor, avg distance between them, avg time taken, etc.

In [127...

```
df = pd.DataFrame(df)

df['corridor'] = df['source_center'] + ' to ' + df['destination_center']

corridor_counts = df['corridor'].value_counts().reset_index()
corridor_counts.columns = ['corridor', 'trip_count']

corridor_stats = df.groupby('corridor').agg({
    'osrm_distance': 'mean',
    'osrm_time': 'mean',
    'segment_osrm_distance': 'mean',
    'segment_osrm_time': 'mean'
}).reset_index()

corridor_summary = pd.merge(corridor_counts, corridor_stats, on='corridor')

busiest_corridors = corridor_summary.sort_values(by='trip_count', ascending=False)

print("Busiest Corridors and Their Statistics:")
print(busiest_corridors)
```

Busiest Corridors and Their Statistics:

	corridor	trip_count	osrm_distance	osrm_time	\
0	IND000000ACB to IND562132AAA	4976	1033.952561	786.834606	
1	IND562132AAA to IND000000ACB	3316	1143.918081	841.702051	
2	IND000000ACB to IND712311AAA	2862	725.691583	512.914396	
3	IND000000ACB to IND501359AAE	1639	799.937295	592.336181	
4	IND000000ACB to IND421302AAG	1617	674.050558	505.349412	
...	
2729	IND110077AAA to IND110043AAA	1	11.778600	15.000000	
2728	IND531001AAA to IND530012AAA	1	24.742100	22.000000	
2727	IND509103AAB to IND509103AAC	1	18.951200	17.000000	
2726	IND530018AAA to IND531162AAA	1	12.213500	8.000000	
2782	IND788713AAA to IND788026AAC	1	43.217000	30.000000	

	segment_osrm_distance	segment_osrm_time
0	30.352581	23.066519
1	30.685749	23.709891
2	26.755446	18.998602
3	30.298966	22.664430
4	28.582550	21.630179
...
2729	11.778600	15.000000
2728	24.742100	22.000000
2727	18.951200	17.000000
2726	12.213500	8.000000
2782	43.217000	30.000000

[2783 rows x 6 columns]

In [128... busiest_corridors.reset_index()

Out[128...

	index	corridor	trip_count	osrm_distance	osrm_time	segment_osrm_dista
0	0	IND000000ACB to IND562132AAA	4976	1033.952561	786.834606	30.352
1	1	IND562132AAA to IND000000ACB	3316	1143.918081	841.702051	30.685
2	2	IND000000ACB to IND712311AAA	2862	725.691583	512.914396	26.755
3	3	IND000000ACB to IND501359AAE	1639	799.937295	592.336181	30.298
4	4	IND000000ACB to IND421302AAG	1617	674.050558	505.349412	28.582
...
2778	2729	IND110077AAA to IND110043AAA	1	11.778600	15.000000	11.778
2779	2728	IND531001AAA to IND530012AAA	1	24.742100	22.000000	24.742
2780	2727	IND509103AAB to IND509103AAC	1	18.951200	17.000000	18.951
2781	2726	IND530018AAA to IND531162AAA	1	12.213500	8.000000	12.213
2782	2782	IND788713AAA to IND788026AAC	1	43.217000	30.000000	43.217

2783 rows × 7 columns



In []: