

# Business Case: Delhivery - Feature Engineering

## About Delhivery

Delhivery is the largest and fastest-growing fully integrated player in India by revenue in Fiscal 2021. They aim to build the operating system for commerce, through a combination of world-class infrastructure, logistics operations of the highest quality, and cutting-edge engineering and technology capabilities.

```
In [129]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import norm, probplot
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import LabelEncoder
```

```
In [130]: df=pd.read_csv("delhivery_data.csv")
```

```
In [131]: df
```

```
Out[131]:
```

		data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination
0	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
1	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
2	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
3	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
4	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78- b351-4c0e-a951- fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
...	...		...	...	...	...	...	...	...
144862	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144863	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144864	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144865	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144866	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f- 4e20-4c31-8542- 67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB

144867 rows × 24 columns

Data set

- data - whether the data is testing or training data
- trip\_creation\_time – timestamp of trip creation
- route\_schedule\_uuid – unique Id for a particular route schedule route\_type – Transportation type
- FTL – Full Truck Load: FTL shipments get to the destination sooner, as the truck is making no other pickups or drop-offs along the way. Carting: handling system consisting of small vehicles (carts)
- trip\_uuid - unique ID given to a particular trip (A trip may include different source and destination centers)
- source\_center - source ID of trip origin
- source\_name - source Name of trip origin
- destination\_cente – destination ID
- destination\_name – destination Name
- od\_start\_time – trip start time
- od\_end\_time – trip end time

- `start_scan_to_end_scan` – time taken to deliver from source to destination
- `actual_distance_to_destination` – distance in Kms between source and destination warehouse
- `actual_time` – actual time taken to complete the delivery (cumulative) \* `osrm_time` – an open-source routing engine time calculator which computes the shortest path between points in a given map (includes usual traffic, distance through major and minor roads) and gives the time (cumulative)
- `osrm_distance` – an open-source routing engine which computes the shortest path between points in a given map (includes usual traffic, distance through major and minor roads) (cumulative)
- `segment_actual_time` – this is a segment time. Time taken by the subset of the package delivery
- `segment_osrm_time` – this is the OSRM segment time. Time taken by the subset of the package delivery
- `segment_osrm_distance` – this is the OSRM distance. Distance covered by subset of the package delivery

## Basic Data Exploration

In [132]: `df.head()`

		data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center
0	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620.
1	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620.
2	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620.
3	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620.
4	training		2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620.

5 rows × 24 columns

In [133]: `df.tail()`

		data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center
144862	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0000
144863	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0000
144864	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0000
144865	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0000
144866	training		2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0000

5 rows × 24 columns

In [134]: `df.shape`

Out[134]: (144867, 24)

In [135]: `df.size`

Out[135]: 3476808

In [136]: `df.dtypes`

```

Out[136]: data                object
trip_creation_time          object
route_schedule_uuid         object
route_type                  object
trip_uuid                   object
source_center               object
source_name                 object
destination_center          object
destination_name            object
od_start_time               object
od_end_time                 object
start_scan_to_end_scan     float64
is_cutoff                   bool
cutoff_factor               int64
cutoff_timestamp            object
actual_distance_to_destination float64
actual_time                 float64
osrm_time                   float64
osrm_distance               float64
factor                      float64
segment_actual_time         float64
segment_osrm_time           float64
segment_osrm_distance       float64
segment_factor              float64
dtype: object

```

data types consist object type,integer,float and bool

```
In [137]: df.isnull().sum()
```

```

Out[137]: data                0
trip_creation_time          0
route_schedule_uuid         0
route_type                  0
trip_uuid                   0
source_center               0
source_name                 293
destination_center          0
destination_name            261
od_start_time               0
od_end_time                 0
start_scan_to_end_scan     0
is_cutoff                   0
cutoff_factor               0
cutoff_timestamp            0
actual_distance_to_destination 0
actual_time                 0
osrm_time                   0
osrm_distance               0
factor                      0
segment_actual_time         0
segment_osrm_time           0
segment_osrm_distance       0
segment_factor              0
dtype: int64

```

there is null values in source\_name and destination\_name

```
In [138]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144867 entries, 0 to 144866
Data columns (total 24 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   data                                  144867 non-null object
1   trip_creation_time                   144867 non-null object
2   route_schedule_uuid                 144867 non-null object
3   route_type                           144867 non-null object
4   trip_uuid                            144867 non-null object
5   source_center                        144867 non-null object
6   source_name                          144574 non-null object
7   destination_center                  144867 non-null object
8   destination_name                     144606 non-null object
9   od_start_time                       144867 non-null object
10  od_end_time                          144867 non-null object
11  start_scan_to_end_scan               144867 non-null float64
12  is_cutoff                            144867 non-null bool
13  cutoff_factor                        144867 non-null int64
14  cutoff_timestamp                     144867 non-null object
15  actual_distance_to_destination        144867 non-null float64
16  actual_time                          144867 non-null float64
17  osrm_time                            144867 non-null float64
18  osrm_distance                        144867 non-null float64
19  factor                               144867 non-null float64
20  segment_actual_time                  144867 non-null float64
21  segment_osrm_time                   144867 non-null float64
22  segment_osrm_distance                144867 non-null float64
23  segment_factor                       144867 non-null float64
dtypes: bool(1), float64(10), int64(1), object(12)
memory usage: 25.6+ MB

```

we need to change trip\_creation\_time ,od\_start\_time,od\_end\_time , cutoff\_timestamp to date time format

```

In [139.. df['trip_creation_time']=pd.to_datetime(df['trip_creation_time'])
df['od_start_time']=pd.to_datetime(df['od_start_time'])
df['od_end_time']=pd.to_datetime(df['od_end_time'])
df['cutoff_timestamp']=pd.to_datetime(df['cutoff_timestamp'])

```

```

In [140.. 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 datetime64[ns]
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 datetime64[ns]
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](4), float64(10), int64(1), object(8)
memory usage: 25.6+ MB

```

the info showing values changed in to date time format

```

In [141.. df.describe(include="all")

```

```
C:\Users\risha\AppData\Local\Temp\ipykernel_19992\1985922364.py:1: FutureWarning: Treating datetime data as categorical rather than numeric in `.describe` is deprecated and will be removed in a future version of pandas. Specify `datetime_is_numeric=True` to silence this warning and adopt the future behavior now.
df.describe(include="all")
C:\Users\risha\AppData\Local\Temp\ipykernel_19992\1985922364.py:1: FutureWarning: Treating datetime data as categorical rather than numeric in `.describe` is deprecated and will be removed in a future version of pandas. Specify `datetime_is_numeric=True` to silence this warning and adopt the future behavior now.
df.describe(include="all")
C:\Users\risha\AppData\Local\Temp\ipykernel_19992\1985922364.py:1: FutureWarning: Treating datetime data as categorical rather than numeric in `.describe` is deprecated and will be removed in a future version of pandas. Specify `datetime_is_numeric=True` to silence this warning and adopt the future behavior now.
df.describe(include="all")
C:\Users\risha\AppData\Local\Temp\ipykernel_19992\1985922364.py:1: FutureWarning: Treating datetime data as categorical rather than numeric in `.describe` is deprecated and will be removed in a future version of pandas. Specify `datetime_is_numeric=True` to silence this warning and adopt the future behavior now.
df.describe(include="all")
```

Out[141]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center
count	144867	144867	144867	144867	144867	144867	144574	144867
unique	2	14817	1504	2	14817	1508	1498	1481
top	training	2018-09-28 05:23:15.359220	thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...	FTL	trip-153811219535896559	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB
freq	104858	101	1812	99660	101	23347	23347	15192
first	NaN	2018-09-12 00:00:16.535741	NaN	NaN	NaN	NaN	NaN	NaN
last	NaN	2018-10-03 23:59:42.701692	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

13 rows × 9 columns

In [142]: df.describe(include=['object'])

Out[142]:

	data	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination_center
count	144867	144867	144867	144867	144867	144574	144867
unique	2	1504	2	14817	1508	1498	1481
top	training	thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069f...	FTL	trip-153811219535896559	IND000000ACB	Gurgaon_Bilaspur_HB (Haryana)	IND000000ACB
freq	104858	1812	99660	101	23347	23347	15192

In [143]: cat\_cols=df.dtypes==object  
cat\_cols=list(cat\_cols[cat\_cols].index)  
cat\_cols

Out[143]: ['data',  
'route\_schedule\_uuid',  
'route\_type',  
'trip\_uuid',  
'source\_center',  
'source\_name',  
'destination\_center',  
'destination\_name']

the above showing categorical values in object types

In [144]: num\_cols=df.dtypes!=object  
num\_cols=list(num\_cols[num\_cols].index)  
num\_cols

```
Out[144]: ['trip_creation_time',
          'od_start_time',
          'od_end_time',
          'start_scan_to_end_scan',
          'is_cutoff',
          'cutoff_factor',
          'cutoff_timestamp',
          'actual_distance_to_destination',
          'actual_time',
          'osrm_time',
          'osrm_distance',
          'factor',
          'segment_actual_time',
          'segment_osrm_time',
          'segment_osrm_distance',
          'segment_factor']
```

the above showing numerical values not in object types

```
In [145]: for i in df.columns:
          print(i,':',df[i].nunique())
```

```
data : 2
trip_creation_time : 14817
route_schedule_uuid : 1504
route_type : 2
trip_uuid : 14817
source_center : 1508
source_name : 1498
destination_center : 1481
destination_name : 1468
od_start_time : 26369
od_end_time : 26369
start_scan_to_end_scan : 1915
is_cutoff : 2
cutoff_factor : 501
cutoff_timestamp : 93180
actual_distance_to_destination : 144515
actual_time : 3182
osrm_time : 1531
osrm_distance : 138046
factor : 45641
segment_actual_time : 747
segment_osrm_time : 214
segment_osrm_distance : 113799
segment_factor : 5675
```

it showing number of unique values in each columns

```
In [146]: for i in df.columns:
          print(i,':\n',df[i].unique())
```

```
data :
['training' 'test']
trip_creation_time :
['2018-09-20T02:35:36.476840000' '2018-09-23T06:42:06.021680000'
 '2018-09-14T15:42:46.437249000' ... '2018-09-22T11:30:41.399439000'
 '2018-09-17T11:35:28.838714000' '2018-09-20T16:24:28.436231000']
route_schedule_uuid :
['thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3297ef'
 'thanos::sroute:ff52ef7a-4d0d-4063-9bfe-cc211728881b'
 'thanos::sroute:a16bfa03-3462-4bce-9c82-5784c7d315e6' ...
 'thanos::sroute:72cf9feb-f4e3-4a55-b92a-0b686ee8fabc'
 'thanos::sroute:5e08be79-8a4c-4a91-a514-5350403c0e31'
 'thanos::sroute:a3c30562-87e5-471c-9646-0ed49c150996']
route_type :
['Carting' 'FTL']
trip_uuid :
['trip-153741093647649320' 'trip-153768492602129387'
 'trip-153693976643699843' ... 'trip-153761584139918815'
 'trip-153718412883843340' 'trip-15374606684355182']
source_center :
['IND388121AAA' 'IND388620AAB' 'IND421302AAG' ... 'IND361335AAA'
 'IND562132AAC' 'IND639104AAB']
source_name :
['Anand_VUNagar_DC (Gujarat)' 'Khambhat_MotvdDPP_D (Gujarat)'
 'Bhiwandi_Mankoli_HB (Maharashtra)' ... 'Dwarka_StnRoad_DC (Gujarat)'
 'Bengaluru_Nelmngla_L (Karnataka)' 'Kulithalai_AnnaNGR_D (Tamil Nadu)']
destination_center :
['IND388620AAB' 'IND388320AAA' 'IND411033AAA' ... 'IND600004AAA'
 'IND134203AAA' 'IND400701AAA']
destination_name :
['Khambhat_MotvdDPP_D (Gujarat)' 'Anand_Vaghasi_IP (Gujarat)'
 'Pune_Tathawde_H (Maharashtra)' ... 'Chennai_Mylapore (Tamil Nadu)'
 'Naraingarh_Ward2DPP_D (Haryana)' 'Mumbai_Ghansoli_DC (Maharashtra)']
od_start_time :
['2018-09-20T03:21:32.418600000' '2018-09-20T04:47:45.236797000'
 '2018-09-23T06:42:06.021680000' ... '2018-09-22T11:30:41.399439000']
```

```
'2018-09-17T11:35:28.838714000' '2018-09-20T16:24:28.436231000']
od_end_time :
['2018-09-20T04:47:45.236797000' '2018-09-20T06:36:55.627764000'
'2018-09-23T11:44:28.365845000' ... '2018-09-22T21:45:05.128533000'
'2018-09-17T13:32:21.128357000' '2018-09-20T23:32:09.618069000']
start_scan_to_end_scan :
[ 86. 109. 302. ... 2476. 1161. 2949.]
is_cutoff :
[ True False]
cutoff_factor :
[ 9 18 27 36 39 43 22 44 66 88 100 16 110 132
154 176 198 220 242 264 286 308 330 331 20 38 31 28
33 37 45 54 56 52 128 17 62 79 29 125 147 61
34 208 21 352 374 396 418 440 442 63 64 265 55 203
289 277 46 67 83 49 123 101 24 60 19 10 252 53
98 75 59 51 462 484 506 528 550 572 594 616 638 660
682 704 726 748 770 792 814 836 858 880 902 924 946 962
69 50 93 35 25 14 267 731 99 366 401 511 23 15
13 968 990 1012 1034 1056 1078 1100 1122 1144 1166 1188 1210 1232
1254 1276 1298 1320 1342 1364 1386 1408 1430 1452 1474 1496 1518 1540
1562 1584 1606 1628 1650 1672 1689 41 26 48 30 12 11 32
72 81 735 42 47 82 87 236 514 78 65 142 58 247
281 210 40 151 141 103 1236 157 223 70 85 447 68 57
71 74 120 86 448 181 102 172 320 219 113 202 1687 271
112 91 235 159 84 158 754 144 137 129 115 196 367 174
73 80 195 563 224 253 90 1120 237 764 97 108 117 126
135 146 136 361 1366 250 382 423 76 77 127 365 1200 130
161 160 1113 134 207 133 1269 856 283 296 178 175 143 1690
733 197 1300 422 139 153 164 259 273 1356 441 266 1694 1716
1722 307 1010 107 409 116 121 1201 804 119 473 122 803 145
95 89 106 104 148 105 92 270 150 140 155 510 1119 644
306 1011 152 189 829 494 674 263 173 412 170 156 383 300
149 96 562 402 182 171 871 1091 169 225 94 221 162 262
1014 268 109 209 830 1237 1148 515 138 410 282 1114 411 114
607 321 959 364 248 1367 165 1720 180 216 234 931 111 356
381 673 561 1301 1688 290 1463 1129 254 217 177 730 765 349
249 226 384 332 608 403 805 294 350 1299 118 354 1295 348
1238 1461 233 368 258 163 218 472 624 1532 168 732 1302 251
215 124 200 932 187 131 1270 347 255 167 272 1118 1015 493
222 1076 1566 355 873 1090 527 855 1692 204 1202 206 260 231
424 317 1721 276 232 188 1198 645 241 399 345 257 1629 315
806 1462 802 957 960 872 606 261 295 297 230 1738 1760 1782
1804 1826 1848 1870 1892 1914 1927 245 734 1149 179 513 346 183
278 269 690 874 753 496 201 1117 1089 305 958 360 299 1147
288 309 1092 256 184 469 228 304 362 1685 239 199 246 675
240 310 319 508 205 1691 625 509 192 831 275]
cutoff_timestamp :
['2018-09-20T04:27:55.000000000' '2018-09-20T04:17:55.000000000'
'2018-09-20T04:01:19.505586000' ... '2018-09-17T12:30:41.952774000'
'2018-09-17T12:05:21.000000000' '2018-09-20T16:24:28.436231000']
actual_distance_to_destination :
[10.43566024 18.9368423 27.63727904 ... 66.16359134 73.68066734
70.03901016]
actual_time :
[ 14. 24. 40. ... 3169. 3318. 2980.]
osrm_time :
[ 11. 20. 28. ... 1340. 1439. 1312.]
osrm_distance :
[ 11.9653 21.7243 32.5395 ... 97.0933 111.2709 88.7319]
factor :
[1.27272727 1.2 1.42857143 ... 2.01142857 1.73880597 4.48421053]
segment_actual_time :
[ 1.400e+01 1.000e+01 1.600e+01 2.100e+01 6.000e+00 1.500e+01
2.800e+01 2.600e+01 3.800e+01 3.700e+01 4.100e+01 2.300e+01
4.600e+01 3.000e+01 5.000e+01 9.300e+01 6.200e+01 4.900e+01
2.700e+01 3.500e+01 6.700e+01 2.000e+01 5.100e+01 9.400e+01
1.900e+01 1.200e+01 1.800e+01 1.800e+01 1.100e+01 2.000e+00 1.300e+01
2.400e+01 5.700e+01 1.000e+00 0.000e+00 2.200e+01 2.500e+01
4.500e+01 8.000e+00 3.600e+01 4.200e+01 4.400e+01 7.500e+01
7.800e+01 2.900e+01 3.400e+01 3.200e+01 6.000e+01 4.300e+01
7.900e+01 4.000e+01 6.900e+01 5.800e+01 5.200e+01 4.800e+01
5.500e+01 5.400e+01 4.700e+01 9.000e+00 8.700e+01 1.700e+01
6.800e+01 5.600e+01 3.300e+01 4.000e+00 5.300e+01 2.000e+02
6.500e+01 3.100e+01 8.800e+01 7.000e+00 8.400e+01 8.200e+01
3.900e+01 1.010e+02 1.900e+02 2.020e+02 9.700e+01 8.600e+01
2.910e+02 1.620e+02 4.310e+02 1.060e+02 1.530e+02 8.900e+01
7.300e+01 3.000e+00 1.360e+02 6.600e+01 6.300e+01 5.000e+00
4.220e+02 7.600e+01 8.300e+01 1.230e+02 7.200e+01 1.320e+02
6.850e+02 1.038e+03 6.100e+01 5.900e+01 1.710e+02 1.410e+02
7.000e+01 7.700e+01 1.250e+02 9.200e+01 7.100e+01 6.400e+01
1.040e+02 1.120e+02 9.000e+01 9.800e+01 3.030e+02 1.240e+02
8.100e+01 1.730e+02 9.100e+01 2.200e+02 1.750e+02 2.920e+02
1.170e+02 4.680e+02 6.940e+02 1.090e+02 1.300e+02 3.710e+02
6.110e+02 -2.600e+01 1.480e+02 1.070e+02 5.040e+02 1.150e+02
8.000e+01 1.790e+02 1.080e+02 9.600e+01 1.000e+02 8.500e+01
4.930e+02 4.440e+02 4.240e+02 7.600e+02 1.030e+02 2.320e+02
1.490e+02 2.050e+02 9.420e+02 1.270e+02 7.400e+01 1.660e+02
9.500e+01 1.440e+02 2.220e+02 1.540e+02 1.210e+02 1.840e+02
3.250e+02 1.020e+02 5.270e+02 1.110e+02 5.390e+02 1.590e+02]
```

5.860e+02	3.460e+02	1.180e+02	3.190e+02	2.690e+02	2.950e+02
6.580e+02	2.410e+02	2.960e+02	9.900e+01	1.160e+02	1.140e+02
1.520e+02	-2.100e+01	2.130e+02	1.050e+02	1.220e+02	1.670e+02
-5.000e+00	1.780e+02	1.136e+03	1.190e+02	1.820e+02	2.120e+02
2.930e+02	1.870e+02	1.350e+02	9.010e+02	1.600e+02	2.240e+02
2.790e+02	1.280e+02	6.370e+02	1.310e+02	1.340e+02	5.590e+02
1.580e+02	1.200e+02	5.580e+02	3.940e+02	2.280e+02	2.770e+02
2.040e+02	2.297e+03	1.630e+02	1.130e+02	5.700e+02	2.720e+02
1.510e+02	7.080e+02	1.380e+02	2.810e+02	8.330e+02	-1.000e+00
5.020e+02	1.100e+02	1.570e+02	1.650e+02	2.080e+02	1.910e+02
3.480e+02	1.500e+02	1.430e+02	2.430e+02	2.330e+02	1.470e+02
3.550e+02	1.370e+02	6.590e+02	2.620e+02	2.440e+02	6.200e+02
1.810e+02	1.560e+02	2.700e+02	1.420e+02	6.270e+02	2.480e+02
2.510e+02	1.770e+02	1.860e+02	1.390e+02	2.880e+02	2.530e+02
2.230e+02	1.400e+02	1.330e+02	2.150e+02	3.470e+02	3.560e+02
2.670e+02	1.720e+02	1.290e+02	6.800e+02	3.450e+02	1.450e+02
4.760e+02	3.300e+02	1.880e+02	3.950e+02	3.850e+02	7.190e+02
1.039e+03	5.510e+02	4.590e+02	4.890e+02	2.100e+02	4.740e+02
1.700e+02	9.900e+02	1.550e+02	2.170e+02	3.230e+02	4.850e+02
3.180e+02	1.690e+02	4.190e+02	6.360e+02	1.850e+02	1.260e+02
3.020e+02	6.350e+02	2.030e+02	1.980e+02	3.600e+02	2.090e+02
2.290e+02	7.430e+02	1.117e+03	3.790e+02	3.090e+02	2.500e+02
1.460e+02	2.780e+02	3.140e+02	1.760e+02	2.340e+02	6.300e+02
3.930e+02	1.890e+02	4.160e+02	4.610e+02	2.630e+02	2.380e+02
1.140e+03	4.210e+02	2.260e+02	2.760e+02	1.610e+02	2.060e+02
1.153e+03	4.270e+02	3.390e+02	8.410e+02	5.220e+02	2.140e+02
1.830e+02	4.490e+02	6.120e+02	1.017e+03	1.640e+02	2.160e+02
3.520e+02	4.140e+02	4.470e+02	6.710e+02	2.210e+02	3.800e+02
2.940e+02	2.070e+02	2.370e+02	3.900e+02	3.750e+02	2.010e+02
1.800e+02	1.847e+03	3.880e+02	9.430e+02	5.160e+02	9.930e+02
2.640e+02	5.150e+02	9.470e+02	6.700e+02	1.680e+02	4.280e+02
4.040e+02	3.980e+02	2.270e+02	2.890e+02	3.990e+02	7.320e+02
5.560e+02	5.980e+02	1.050e+03	2.350e+02	7.660e+02	4.920e+02
6.460e+02	3.120e+02	6.950e+02	2.580e+02	3.590e+02	2.360e+02
3.540e+02	6.010e+02	6.400e+02	6.410e+02	5.070e+02	3.720e+02
2.750e+02	5.180e+02	4.400e+02	2.390e+02	5.100e+02	1.716e+03
7.410e+02	9.370e+02	3.860e+02	2.180e+02	2.650e+02	2.820e+02
2.990e+02	1.990e+02	6.890e+02	6.870e+02	3.220e+02	4.460e+02
3.810e+02	4.950e+02	5.930e+02	2.300e+02	9.340e+02	1.143e+03
3.840e+02	9.940e+02	3.160e+02	8.770e+02	4.150e+02	1.086e+03
1.960e+02	8.790e+02	8.960e+02	6.600e+02	5.240e+02	3.360e+02
5.440e+02	3.380e+02	6.620e+02	7.050e+02	3.110e+02	2.680e+02
1.211e+03	5.630e+02	1.981e+03	6.670e+02	3.320e+02	2.310e+02
8.430e+02	8.220e+02	-5.800e+01	5.730e+02	3.000e+02	-2.110e+02
1.740e+02	6.440e+02	3.570e+02	2.190e+02	2.600e+02	1.940e+02
2.110e+02	2.450e+02	2.032e+03	2.860e+02	2.710e+02	4.200e+02
5.430e+02	3.200e+02	5.570e+02	4.320e+02	2.460e+02	5.010e+02
6.480e+02	3.690e+02	5.250e+02	3.920e+02	6.430e+02	7.570e+02
4.780e+02	7.670e+02	2.850e+02	2.250e+02	7.170e+02	1.970e+02
9.580e+02	5.890e+02	4.620e+02	4.050e+02	1.036e+03	6.560e+02
2.351e+03	4.520e+02	6.530e+02	2.740e+02	1.077e+03	5.210e+02
4.830e+02	3.370e+02	4.500e+02	2.610e+02	9.500e+02	6.830e+02
3.500e+02	2.400e+02	3.910e+02	9.910e+02	2.590e+02	2.281e+03
5.910e+02	1.083e+03	3.280e+02	1.124e+03	1.207e+03	3.640e+02
5.360e+02	4.330e+02	5.870e+02	4.510e+02	7.110e+02	2.550e+02
2.980e+02	-1.200e+01	8.500e+02	9.350e+02	2.660e+02	5.480e+02
4.900e+02	2.464e+03	3.650e+02	5.750e+02	6.420e+02	-3.600e+01
1.008e+03	7.800e+02	2.540e+02	1.950e+02	3.820e+02	6.250e+02
6.510e+02	4.770e+02	5.060e+02	6.180e+02	4.350e+02	6.690e+02
9.920e+02	6.470e+02	3.270e+02	9.180e+02	1.067e+03	4.720e+02
1.125e+03	7.370e+02	4.130e+02	1.046e+03	6.970e+02	6.390e+02
4.290e+02	5.850e+02	3.080e+02	3.780e+02	5.400e+02	4.670e+02
5.170e+02	8.690e+02	5.660e+02	2.520e+02	4.340e+02	3.510e+02
6.230e+02	3.060e+02	7.680e+02	4.110e+02	1.065e+03	-4.200e+01
5.110e+02	3.490e+02	5.340e+02	5.090e+02	4.560e+02	1.055e+03
2.490e+02	4.410e+02	-5.100e+01	9.020e+02	3.010e+02	3.350e+02
3.050e+02	9.530e+02	9.050e+02	6.650e+02	4.090e+02	4.980e+02
5.690e+02	5.670e+02	5.950e+02	5.740e+02	7.990e+02	3.130e+02
6.280e+02	4.600e+02	4.870e+02	1.630e+03	7.700e+02	2.120e+03
2.800e+02	4.750e+02	4.800e+02	6.020e+02	4.660e+02	3.530e+02
6.980e+02	1.131e+03	7.620e+02	7.180e+02	9.950e+02	8.090e+02
2.730e+02	3.340e+02	2.970e+02	4.640e+02	4.170e+02	-2.440e+02
6.000e+02	4.970e+02	8.140e+02	9.320e+02	8.700e+02	3.580e+02
1.097e+03	6.260e+02	3.680e+02	7.550e+02	1.061e+03	1.032e+03
4.230e+02	8.910e+02	3.240e+02	3.420e+02	6.820e+02	1.152e+03
5.880e+02	4.100e+02	9.680e+02	7.710e+02	3.290e+02	7.310e+02
-3.000e+00	9.590e+02	4.020e+02	1.166e+03	8.150e+02	4.960e+02
4.390e+02	9.160e+02	1.182e+03	8.490e+02	2.830e+02	8.940e+02
5.940e+02	5.280e+02	2.470e+02	2.900e+02	7.120e+02	1.041e+03
2.420e+02	1.128e+03	-7.400e+01	7.070e+02	6.290e+02	3.330e+02
5.600e+02	6.770e+02	6.100e+02	9.820e+02	1.920e+02	8.550e+02
3.610e+02	6.210e+02	6.740e+02	1.001e+03	9.610e+02	7.330e+02
6.880e+02	7.200e+02	3.310e+02	2.570e+02	2.408e+03	5.310e+02
1.025e+03	8.400e+02	1.584e+03	1.246e+03	7.470e+02	8.930e+02
6.310e+02	6.780e+02	4.010e+02	5.970e+02	9.270e+02	1.167e+03
1.115e+03	3.260e+02	7.150e+02	1.930e+02	1.395e+03	9.670e+02
5.370e+02	6.040e+02	3.210e+02	3.770e+02	8.510e+02	1.016e+03
1.031e+03	4.550e+02	6.220e+02	3.890e+02	5.830e+02	-4.800e+01
4.180e+02	1.047e+03	6.930e+02	9.700e+02	7.040e+02	7.850e+02



```

-2.000e+00  2.491e+03  9.460e+02  4.650e+02  2.541e+03  1.122e+03
 3.051e+03  9.740e+02  9.780e+02  9.040e+02 -1.600e+01  8.530e+02
 4.790e+02  1.148e+03  5.720e+02  4.250e+02  5.530e+02  4.060e+02
-7.000e+00  3.070e+02  1.093e+03  4.630e+02  9.390e+02  5.450e+02
 1.325e+03  9.150e+02  5.460e+02  7.530e+02  5.290e+02  4.370e+02
 5.200e+02  8.300e+02  1.677e+03  1.020e+03  7.480e+02  4.880e+02
 6.130e+02  9.510e+02  3.740e+02  7.360e+02  9.330e+02  5.790e+02
 2.625e+03  7.520e+02 -1.500e+01  1.192e+03  6.640e+02  1.320e+03
 2.870e+02  3.700e+02  1.104e+03]
segment_osrm_time :
[1.100e+01  9.000e+00  7.000e+00  1.200e+01  5.000e+00  6.000e+00  1.000e+01
 2.400e+01  2.700e+01  2.600e+01  1.400e+01  1.500e+01  3.000e+01  1.800e+01
 3.800e+01  3.700e+01  2.500e+01  1.700e+01  2.200e+01  3.600e+01  3.200e+01
 1.600e+01  7.000e+01  3.500e+01  4.500e+01  1.300e+01  0.000e+00  8.000e+00
 2.000e+00  1.900e+01  2.300e+01  2.800e+01  2.000e+01  2.100e+01  3.300e+01
 3.400e+01  8.100e+01  3.000e+00  4.400e+01  1.000e+00  4.000e+00  3.900e+01
 4.000e+01  2.900e+01  5.300e+01  3.100e+01  7.500e+01  7.900e+01  9.700e+01
 4.800e+01  4.100e+01  4.300e+01  5.000e+01  5.400e+01  6.800e+01  4.200e+01
 4.600e+01  6.000e+01  5.800e+01  7.600e+01  4.900e+01  1.300e+02  7.800e+01
 6.700e+01  6.400e+01  5.500e+01  5.100e+01  1.420e+02  7.700e+01  7.100e+01
 5.600e+01  6.600e+01  5.900e+01  9.200e+01  6.200e+01  5.200e+01  5.700e+01
 8.000e+01  4.700e+01  7.400e+01  6.500e+01  8.800e+01  7.200e+01  6.900e+01
 7.300e+01  8.400e+01  8.200e+01  8.300e+01  1.540e+02  9.100e+01  6.100e+01
 9.400e+01  1.220e+02  6.300e+01  2.180e+02  9.800e+01  8.700e+01  3.830e+02
 9.000e+01  1.080e+02  9.300e+01  8.600e+01  8.900e+01  1.020e+02  1.600e+02
 2.210e+02  1.050e+02  1.330e+02  1.000e+02  1.060e+02  4.070e+02  8.500e+01
 1.340e+02  4.690e+02  1.800e+02  2.340e+02  1.490e+02  1.010e+02  1.450e+02
 1.140e+02  1.840e+02  2.270e+02  1.740e+02  1.320e+02  9.900e+01  9.600e+01
 1.310e+02  1.110e+02  1.040e+02  1.750e+02  2.300e+02  9.500e+01  1.250e+02
 2.950e+02  1.560e+02  1.160e+02  1.460e+02  1.410e+02  1.030e+02  1.170e+02
 2.310e+02  2.540e+02  2.200e+02  2.330e+02  1.810e+02  1.210e+02  1.270e+02
 3.700e+02  3.750e+02  1.500e+02  1.070e+02  1.610e+02  2.320e+02  1.090e+02
 1.200e+02  1.100e+02  9.970e+02  1.790e+02  1.130e+02  1.660e+02  9.960e+02
 1.240e+02  2.150e+02  1.570e+02  3.620e+02  1.430e+02  1.150e+02  1.280e+02
 1.700e+02  1.440e+02  2.350e+02  1.510e+02  3.560e+02  1.180e+02  1.390e+02
 1.710e+02  1.290e+02  1.190e+02  1.690e+02  1.630e+02  2.040e+02  1.480e+02
 1.830e+02  4.810e+02  3.410e+02  3.280e+02  2.130e+02  1.890e+02  1.910e+02
 1.400e+02  1.470e+02  2.080e+02  2.860e+02  2.160e+02  1.720e+02  1.380e+02
 1.670e+02  2.940e+02  1.230e+02  1.260e+02  2.110e+02  1.611e+03  2.190e+02
 2.490e+02  1.850e+02  1.580e+02  3.240e+02  1.770e+02  4.530e+02  1.520e+02
 1.760e+02  7.370e+02  1.730e+02  1.032e+03]
segment_osrm_distance :
[11.9653  9.759  10.8152 ... 20.7053 18.8885  8.8088]
segment_factor :
[ 1.27272727  1.11111111  2.28571429 ...  2.09322034  5.72
 29.77777778]

```

## Value Counts

```

In [147]: for i in df.columns:
          print(i, '\n', df[i].value_counts())

data :
training    104858
test        40009
Name: data, dtype: int64
trip_creation_time :
2018-09-28 05:23:15.359220    101
2018-10-02 06:05:53.086094    101
2018-09-27 04:47:19.425867    101
2018-09-22 04:55:04.835022    101
2018-09-29 05:04:57.639067    101
...
2018-09-27 18:08:18.207639     1
2018-09-28 17:31:07.690205     1
2018-09-29 14:56:33.655170     1
2018-09-19 04:35:44.776558     1
2018-09-14 17:04:32.989471     1
Name: trip_creation_time, Length: 14817, dtype: int64
route_schedule_uuid :
thanos::sroute:4029a8a2-6c74-4b7e-a6d8-f9e069fbcea9    1812
thanos::sroute:0456b740-1dad-4929-bbe0-87d8843f5a10    1608
thanos::sroute:dca6268f-741a-4d1a-b1b0-aab13095a366    1605
thanos::sroute:alb25549-1e77-498f-8538-00292e5bd5a2    1285
thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e5720d    1280
...
thanos::sroute:d563d17e-2123-40a4-9eec-40018966caba     1
thanos::sroute:036f372d-28d8-4d19-877c-6277077ad09e     1
thanos::sroute:e00eb6aa-d792-4b28-81fa-fdee413ef326     1
thanos::sroute:889b9cf5-da6a-48ce-b3bd-6983c8090164     1
thanos::sroute:404cbabf-d2a5-4e46-bf79-8b3c518f082b     1
Name: route_schedule_uuid, Length: 1504, dtype: int64
route_type :
FTL    99660
Carting 45207
Name: route_type, dtype: int64
trip_uuid :
trip-153811219535896559    101

```

```

trip-153846035308581166      101
trip-153802363942560700      101
trip-153759210483476123      101
trip-153819749763881430      101
...
trip-153807169820740041       1
trip-153815586768995663       1
trip-153823299365493206       1
trip-153733174477629450       1
trip-153694467298919626       1
Name: trip_uuid, Length: 14817, dtype: int64
source_center :
  IND000000ACB      23347
  IND562132AAA      9975
  IND421302AAG      9088
  IND411033AAA      4061
  IND501359AAE      3340
...
  IND741121AAA       1
  IND207123AAA       1
  IND242001AAA       1
  IND222001AAA       1
  IND741101AAB       1
Name: source_center, Length: 1508, dtype: int64
source_name :
  Gurgaon_Bilaspur_HB (Haryana)      23347
  Bangalore_Nelmngla_H (Karnataka)    9975
  Bhiwandi_Mankoli_HB (Maharashtra)   9088
  Pune_Tathawde_H (Maharashtra)       4061
  Hyderabad_Shamshbd_H (Telangana)    3340
...
  Shahjhnpur_NavdaCln_D (Uttar Pradesh) 1
  Soro_UttarDPP_D (Orissa)              1
  Kayamkulam_Bhrnikvu_D (Kerala)        1
  Krishnanagar_AnadiDPP_D (West Bengal) 1
  Faridabad_Old (Haryana)               1
Name: source_name, Length: 1498, dtype: int64
destination_center :
  IND000000ACB      15192
  IND562132AAA      11019
  IND421302AAG      5492
  IND501359AAE      5142
  IND712311AAA      4892
...
  IND520011AAA       1
  IND741201AAC       1
  IND400705AAA       1
  IND110046AAA       1
  IND504215AAA       1
Name: destination_center, Length: 1481, dtype: int64
destination_name :
  Gurgaon_Bilaspur_HB (Haryana)      15192
  Bangalore_Nelmngla_H (Karnataka)    11019
  Bhiwandi_Mankoli_HB (Maharashtra)   5492
  Hyderabad_Shamshbd_H (Telangana)    5142
  Kolkata_Dankuni_HB (West Bengal)    4892
...
  Hyd_Trimulgherry_Dc (Telangana)      1
  Vijayawada (Andhra Pradesh)          1
  Baghpat_Barout_D (Uttar Pradesh)     1
  Mumbai_Sanpada_CP (Maharashtra)      1
  Basta_Central_DPP_1 (Orissa)         1
Name: destination_name, Length: 1468, dtype: int64
od_start_time :
  2018-09-21 18:37:09.322207      81
  2018-10-03 04:55:30.039225      79
  2018-09-30 05:56:48.299467      79
  2018-09-26 05:33:10.899941      79
  2018-09-14 07:13:24.396869      79
...
  2018-09-25 04:58:40.930230       1
  2018-09-22 05:19:27.704727       1
  2018-09-25 09:05:19.576386       1
  2018-09-20 02:16:44.645390       1
  2018-09-27 02:59:59.877566       1
Name: od_start_time, Length: 26369, dtype: int64
od_end_time :
  2018-09-24 09:59:15.691618      81
  2018-10-05 11:15:01.115906      79
  2018-10-02 10:36:25.970169      79
  2018-09-28 12:13:41.675546      79
  2018-09-16 17:00:03.263746      79
...
  2018-09-25 13:57:24.614624       1
  2018-09-22 06:11:51.820645       1
  2018-09-25 15:47:35.535055       1
  2018-09-20 04:04:45.620456       1
  2018-09-27 15:47:50.386008       1
Name: od_end_time, Length: 26369, dtype: int64

```

```

start_scan_to_end_scan :
  110.0      459
  72.0       424
  99.0       411
  95.0       405
  86.0       399
  ...
  1076.0     1
  22.0       1
  905.0      1
  3361.0     1
  2706.0     1
Name: start_scan_to_end_scan, Length: 1915, dtype: int64
is_cutoff :
  True      118749
  False     26118
Name: is_cutoff, dtype: int64
cutoff_factor :
  22      13157
  9       12378
  44      8334
  18      8263
  66      5795
  ...
  245     1
  734     1
  1149    1
  412     1
  275     1
Name: cutoff_factor, Length: 501, dtype: int64
cutoff_timestamp :
  2018-09-24 05:19:20.000000    40
  2018-09-24 05:19:21.000000    33
  2018-09-14 05:29:26.000000    19
  2018-09-24 07:21:24.000000    18
  2018-09-14 02:35:24.000000    17
  ..
  2018-09-27 01:15:55.000000     1
  2018-09-26 08:20:38.000000     1
  2018-09-26 07:32:28.000000     1
  2018-09-26 05:57:51.000000     1
  2018-09-20 16:24:28.436231     1
Name: cutoff_timestamp, Length: 93180, dtype: int64
actual_distance_to_destination :
  100.282892    2
  19.122553     2
  44.552586     2
  27.297724     2
  23.955638     2
  ..
  18.495678     1
  9.380550      1
  10.583412     1
  9.162213      1
  70.039010     1
Name: actual_distance_to_destination, Length: 144515, dtype: int64
actual_time :
  32.0      1443
  36.0      1420
  30.0      1350
  38.0      1329
  42.0      1241
  ...
  2709.0     1
  2608.0     1
  2910.0     1
  2870.0     1
  2980.0     1
Name: actual_time, Length: 3182, dtype: int64
osrm_time :
  21.0      2414
  20.0      2361
  18.0      2253
  22.0      2147
  17.0      2098
  ...
  1560.0     1
  1614.0     1
  1130.0     1
  1250.0     1
  1312.0     1
Name: osrm_time, Length: 1531, dtype: int64
osrm_distance :
  48.0394     11
  11.2300     5
  11.3233     4
  12.5416     4
  39.8109     4
  ..

```

```

23.9719      1
55.8507      1
28.7254      1
307.6147     1
88.7319      1
Name: osrm_distance, Length: 138046, dtype: int64
factor :
 2.000000    2351
1.500000    1278
1.666667     830
1.750000     667
1.333333     599
...
1.922049      1
1.710396      1
1.923513      1
1.918675      1
4.484211      1
Name: factor, Length: 45641, dtype: int64
segment_actual_time :
 24.0      6188
26.0      5479
30.0      4903
27.0      4439
23.0      4401
...
869.0      1
517.0      1
540.0      1
1847.0     1
1104.0     1
Name: segment_actual_time, Length: 747, dtype: int64
segment_osrm_time :
 16.0      11483
17.0      10856
18.0       8734
19.0       6925
15.0       6846
...
169.0      1
163.0      1
204.0      1
156.0      1
1032.0     1
Name: segment_osrm_time, Length: 214, dtype: int64
segment_osrm_distance :
 0.0000      1536
25.6081      8
22.6267      8
58.2708      7
26.6974      7
...
13.6619      1
0.7952      1
8.1647      1
10.8209      1
8.8088      1
Name: segment_osrm_distance, Length: 113799, dtype: int64
segment_factor :
 2.000000     6001
1.500000     4637
1.000000     2371
1.666667     2370
-1.000000     2347
...
1.844444      1
1.380000      1
4.103448      1
2.614458      1
29.777778      1
Name: segment_factor, Length: 5675, dtype: int64

```

Merging some rows based on the columns trip\_uuid, source ID and destination ID

```

In [148.. merged_df=df.groupby(['trip_uuid', 'source_center', 'destination_center']).agg({'data': 'first', 'trip_creation_tim
'route_schedule_uuid': 'first', 'route_type': 'first', 'source_name': 'first', 'destination_name': 'last', 'od_start_ti
'od_end_time': 'last', 'start_scan_to_end_scan': 'max', 'actual_distance_to_destination': 'max', 'actual_time': 'max',
'osrm_time': 'max', 'osrm_distance': 'max', 'segment_actual_time': 'sum', 'segment_osrm_time': 'sum', 'segment_osrm_di
merged_df

```

	trip_uuid	source_center	destination_center	data	trip_creation_time	route_schedule_uuid	route_type	
0	trip-153671041653548748	IND209304AAA	IND000000ACB	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	Kanpu (
1	trip-153671041653548748	IND462022AAA	IND209304AAA	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	Bhop (Ma
2	trip-153671042288605164	IND561203AAB	IND562101AAA	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	Doddablpu
3	trip-153671042288605164	IND572101AAA	IND561203AAB	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	Tumh
4	trip-153671043369099517	IND000000ACB	IND160002AAC	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	Gurgaon
...	...	...	...	...	...	...	...	...
26363	trip-153861115439069069	IND628204AAA	IND627657AAA	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	Tirchchnr.
26364	trip-153861115439069069	IND628613AAA	IND627005AAA	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	Peikulam
26365	trip-153861115439069069	IND628801AAA	IND628204AAA	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	Eral_Buss
26366	trip-153861118270144424	IND583119AAA	IND583101AAA	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	Sandur.
26367	trip-153861118270144424	IND583201AAA	IND583119AAA	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	Hosr

26368 rows × 9 columns

Further row merging the above dataframe based on the trip\_uuid column

```
In [149]: merged_df1=merged_df.groupby('trip_uuid').agg({'source_center':'first','destination_center':'last','data':'first','route_schedule_uuid':'first','route_type':'first','source_name':'first','destination_name':'last','od_start_time':'last','start_scan_to_end_scan':'sum','actual_distance_to_destination':'sum','actual_time':'sum','osrm_time':'sum','osrm_distance':'sum','segment_actual_time':'sum','segment_osrm_time':'sum','segment_osrm_distance':'sum'})
merged_df1.head(7)
```

	trip_uuid	source_center	destination_center	data	trip_creation_time	route_schedule_uuid	route_type	source_name
0	trip-153671041653548748	IND209304AAA	IND209304AAA	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	Kanpur_Ce (Uttar
1	trip-153671042288605164	IND561203AAB	IND561203AAB	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	Doddablpur_Chi (K
2	trip-153671043369099517	IND000000ACB	IND000000ACB	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	Gurgaon_Bil
3	trip-153671046011330457	IND400072AAB	IND401104AAA	training	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	Mu (Mal
4	trip-153671052974046625	IND583101AAA	IND583119AAA	training	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	Bellary_Dc (K
5	trip-153671055416136166	IND600056AAA	IND600056AAA	training	2018-09-12 00:02:34.161600	thanos::sroute:9b0f3170-d0a2-4a3f-aa4d-9aabb3d...	Carting	Chennai_Poc (Ta
6	trip-153671066201138152	IND600044AAD	IND600048AAA	training	2018-09-12 00:04:22.011653	thanos::sroute:a97698cc-846e-41a7-916b-88b1741...	Carting	Chennai_Chron (Ta

```
In [150]: merged_df1['trip_creation_time'] = pd.to_datetime(merged_df1['trip_creation_time'])
```

Split and extract features out of the column 'destination\_name'

```
In [151]: dr=merged_df1.copy()
x=dr['destination_name'].str.split("(", expand = True)
x[1]=x[1].str.rstrip(')')
y=x[0].str.split("_", expand = True)
y
```

```
Out[151]:
```

	0	1	2	3
0	Kanpur	Central	H	6
1	Doddablpur	ChikaDPP	D	None
2	Gurgaon	Bilaspur	HB	None
3	Mumbai	MiraRd	IP	None
4	Sandur	WrdN1DPP	D	None
...	...	...	...	...
14812	Chandigarh	Mehmdpur	H	None
14813	Faridabad	Blbgarh	DC	None
14814	Kanpur	GovndNgr	DC	None
14815	Tirchchndr	Shnmgprn	D	None
14816	Sandur	WrdN1DPP	D	None

14817 rows × 4 columns

split and extract features out of the column 'source\_name'

```
In [152]: x=dr['source_name'].str.split("(", expand = True)
x[1]=x[1].str.rstrip(')')
y=x[0].str.split("_", expand = True)
#print(x)
y
```

```
Out[152]:
```

	0	1	2	3
0	Kanpur	Central	H	6
1	Doddablpur	ChikaDPP	D	None
2	Gurgaon	Bilaspur	HB	None
3	Mumbai Hub	None	None	None
4	Bellary	Dc	None	None
...	...	...	...	...
14812	Chandigarh	Mehmdpur	H	None
14813	FBD	Balabhgarh	DPC	None
14814	Kanpur	GovndNgr	DC	None
14815	Tirunelveli	VdkkuSrt	I	None
14816	Sandur	WrdN1DPP	D	None

14817 rows × 4 columns

```
In [153]: merged_df1['dest_city'] = y[0]
merged_df1['dest_place'] = y[1]
merged_df1['dest_state'] = x[1]

merged_df1['source_city'] = y[0]
merged_df1['source_place'] = y[1]
merged_df1['source_state'] = x[1]

merged_df1.drop(columns=['source_name','destination_name'],axis=1, inplace=True)
```

Extract features like month, year and day from the column 'trip\_creation\_time'

```
In [154]: merged_df1['total_min_diff']=(merged_df1.od_end_time - merged_df1.od_start_time) / pd.Timedelta(minutes=1)

merged_df1['trip_creation_year'] = merged_df1['trip_creation_time'].dt.year
merged_df1['trip_creation_month'] = merged_df1['trip_creation_time'].dt.month
merged_df1['trip_creation_day'] = merged_df1['trip_creation_time'].dt.day

merged_df1.drop(columns=['od_end_time','od_start_time'],axis=1, inplace=True)
merged_df1.head(7)
```

Out[154]:

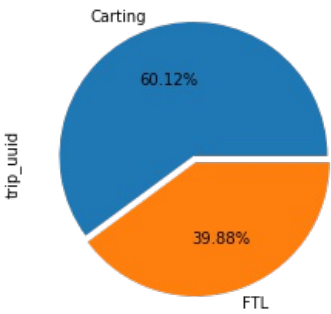
	trip_uuid	source_center	destination_center	data	trip_creation_time	route_schedule_uuid	route_type	start_scan_to_e
0	trip-153671041653548748	IND209304AAA	IND209304AAA	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	
1	trip-153671042288605164	IND561203AAB	IND561203AAB	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	
2	trip-153671043369099517	IND000000ACB	IND000000ACB	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	
3	trip-153671046011330457	IND400072AAB	IND401104AAA	training	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	
4	trip-153671052974046625	IND583101AAA	IND583119AAA	training	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	
5	trip-153671055416136166	IND600056AAA	IND600056AAA	training	2018-09-12 00:02:34.161600	thanos::sroute:9bf03170-d0a2-4a3f-aa4d-9aaab3d...	Carting	
6	trip-153671066201138152	IND600044AAD	IND600048AAA	training	2018-09-12 00:04:22.011653	thanos::sroute:a97698cc-846e-41a7-916b-88b1741...	Carting	

7 rows × 25 columns

data visualization

Route type preferred for the delivery

```
In [155] r_type = merged_df1.groupby('route_type')['trip_uuid'].count().plot(kind='pie', explode=(0.05,0), autopct='%0.2
```



more time taken for carting

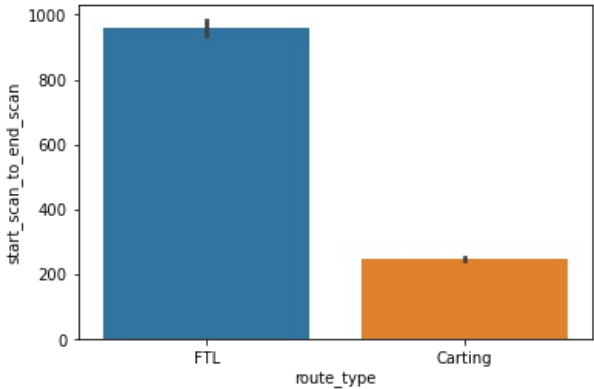
Average time taken by different route types to deliver products

```
In [156] merged_df1.groupby('route_type')['start_scan_to_end_scan'].mean().reset_index()
```

Out[156]:

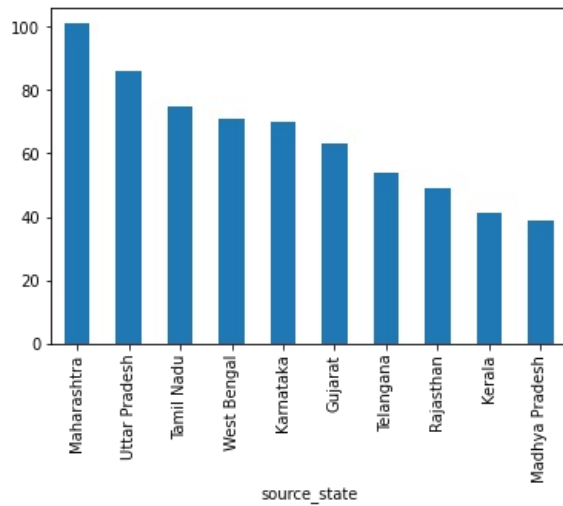
	route_type	start_scan_to_end_scan
0	Carting	246.537382
1	FTL	959.359790

```
In [157] sns.barplot(data=merged_df1, x='route_type', y='start_scan_to_end_scan')
plt.show()
```



Top Indian states with most delivery source centres

```
In [158.. top_state = merged_df1.groupby('source_state')['source_center'].nunique().sort_values(ascending=False)[:10].plo
```

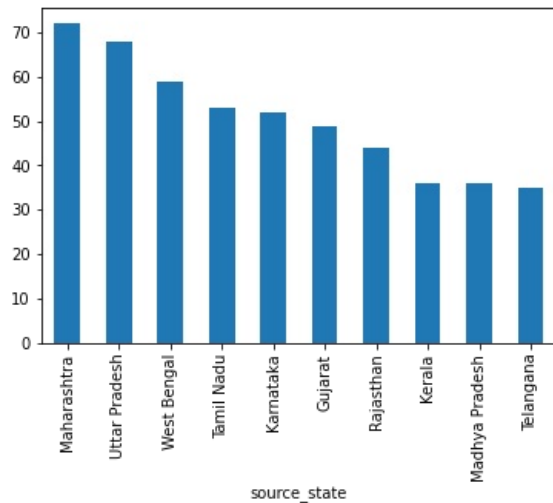


Maharashtra is the most delivery source centre ,kerala and hariyana the lowest

Top Indian states with most delivery source cities

```
In [159.. merged_df1.groupby('source_state')['source_city'].nunique().sort_values(ascending=False)[:10].plot(kind='bar')
```

```
Out[159]: <AxesSubplot:xlabel='source_state'>
```

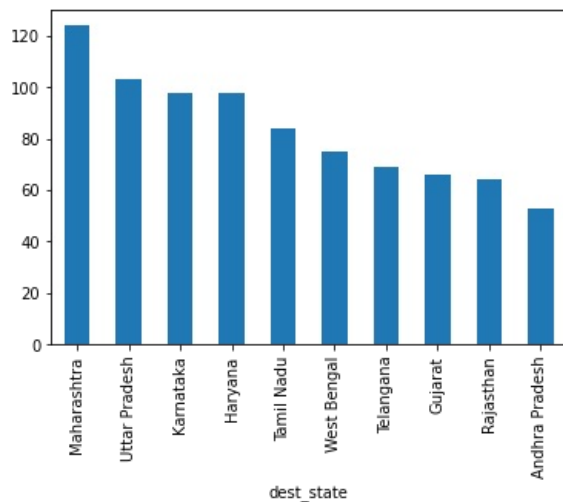


punjab is the lowest delivery source city

Top Indian states with most delivery destination centres

```
In [160.. merged_df1.groupby('dest_state')['destination_center'].nunique().sort_values(ascending=False)[:10].plot(kind='b
```

```
Out[160]: <AxesSubplot:xlabel='dest_state'>
```

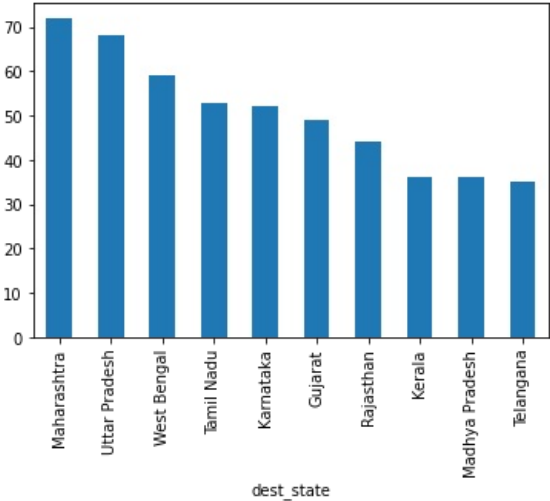


Top Indian states with most delivery destination cities



```
In [161]: merged_df1.groupby('dest_state')['dest_city'].nunique().sort_values(ascending=False)[:10].plot(kind='bar')
```

Out[161]: <AxesSubplot:xlabel='dest\_state'>



Maharashtra is the most delivery destination city and punjab is the lowest destination city

Weekday with most delivery of the products

```
In [162]: merged_df1['weekday'] = merged_df1['trip_creation_time'].dt.day_name()  
merged_df1
```

	trip_uuid	source_center	destination_center	data	trip_creation_time	route_schedule_uuid	route_type	start_scan_
0	trip-153671041653548748	IND209304AAA	IND209304AAA	training	2018-09-12 00:00:16.535741	thanos::sroute:d7c989ba-a29b-4a0b-b2f4-288cdc6...	FTL	
1	trip-153671042288605164	IND561203AAB	IND561203AAB	training	2018-09-12 00:00:22.886430	thanos::sroute:3a1b0ab2-bb0b-4c53-8c59-eb2a2c0...	Carting	
2	trip-153671043369099517	IND000000ACB	IND000000ACB	training	2018-09-12 00:00:33.691250	thanos::sroute:de5e208e-7641-45e6-8100-4d9fb1e...	FTL	
3	trip-153671046011330457	IND400072AAB	IND401104AAA	training	2018-09-12 00:01:00.113710	thanos::sroute:f0176492-a679-4597-8332-bbd1c7f...	Carting	
4	trip-153671052974046625	IND583101AAA	IND583119AAA	training	2018-09-12 00:02:09.740725	thanos::sroute:d9f07b12-65e0-4f3b-bec8-df06134...	FTL	
...	...	...	...	...	...	...	...	...
14812	trip-153861095625827784	IND160002AAC	IND160002AAC	test	2018-10-03 23:55:56.258533	thanos::sroute:8a120994-f577-4491-9e4b-b7e4a14...	Carting	
14813	trip-153861104386292051	IND121004AAB	IND121004AAA	test	2018-10-03 23:57:23.863155	thanos::sroute:b30e1ec3-3bfa-4bd2-a7fb-3b75769...	Carting	
14814	trip-153861106442901555	IND208006AAA	IND208006AAA	test	2018-10-03 23:57:44.429324	thanos::sroute:5609c268-e436-4e0a-8180-3db4a74...	Carting	
14815	trip-153861115439069069	IND627005AAA	IND628204AAA	test	2018-10-03 23:59:14.390954	thanos::sroute:c5f2ba2c-8486-4940-8af6-d1d2a6a...	Carting	
14816	trip-153861118270144424	IND583119AAA	IND583119AAA	test	2018-10-03 23:59:42.701692	thanos::sroute:412fea14-6d1f-4222-8a5f-a517042...	FTL	

14817 rows × 26 columns

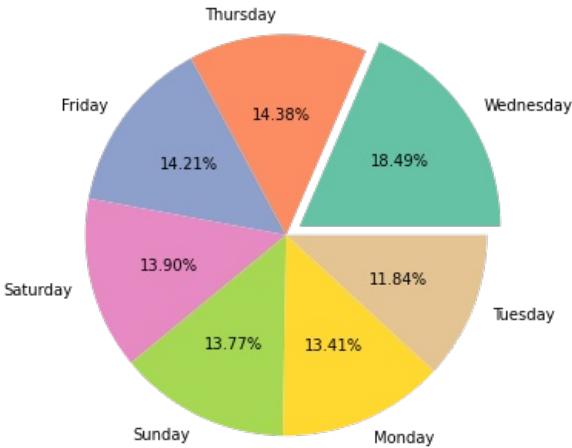
```
In [163]: x=merged_df1['weekday'].value_counts().to_frame('count').reset_index()  
x
```

Out[163]:

	index	count
0	Wednesday	2739
1	Saturday	2130
2	Thursday	2106
3	Friday	2060
4	Tuesday	2040
5	Monday	1987
6	Sunday	1755

```
In [164]: plt.figure(figsize=(8,6))
palette_color = sns.color_palette('Set2')
plt.pie(data=x, x=x['count'], colors=palette_color, labels=['Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sun
plt.plot()
```

Out[164]: []



Wednesday ,they are delivering more porducts.

Relation between actual time and estimated time for the completion of product delivery

```
In [165]: x1=merged_df1.groupby('source_state').aggregate({'actual_time':'mean','osrm_time':'mean'}).reset_index()
x1=pd.melt(x1, id_vars=['source_state'], value_vars=['actual_time', 'osrm_time'])
x1
```

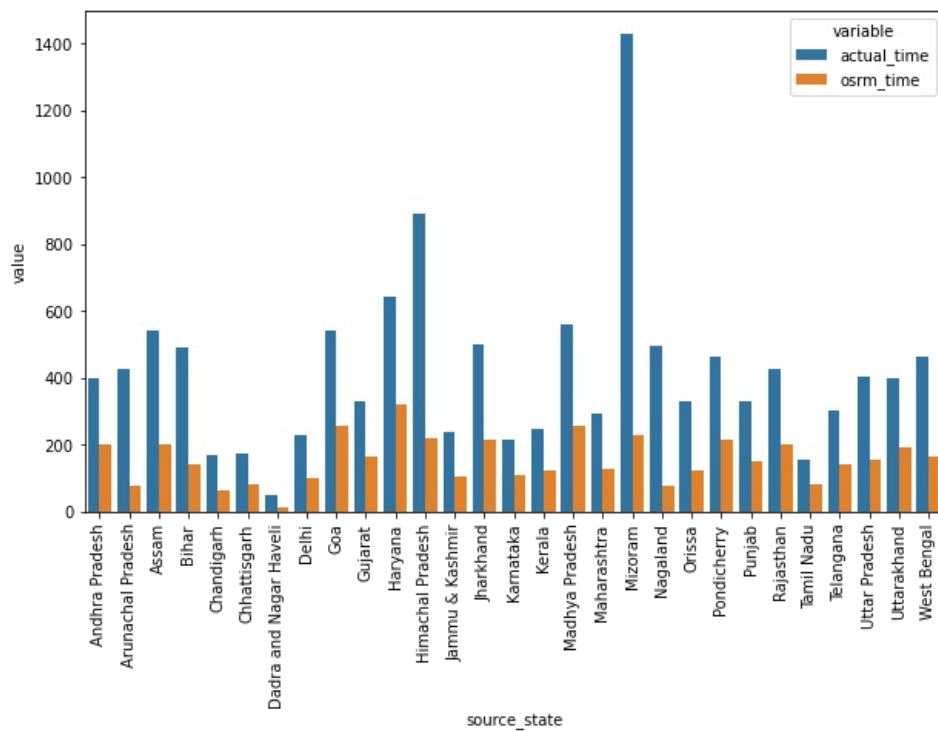
Out[165]:

	source_state	variable	value
0	Andhra Pradesh	actual_time	398.170115
1	Arunachal Pradesh	actual_time	429.250000
2	Assam	actual_time	540.171642
3	Bihar	actual_time	491.985755
4	Chandigarh	actual_time	168.741935
5	Chhattisgarh	actual_time	174.139535
6	Dadra and Nagar Haveli	actual_time	48.333333
7	Delhi	actual_time	230.550824
8	Goa	actual_time	540.138462
9	Gujarat	actual_time	331.470667
10	Haryana	actual_time	641.968988
11	Himachal Pradesh	actual_time	891.088235
12	Jammu & Kashmir	actual_time	237.235294
13	Jharkhand	actual_time	498.981250
14	Karnataka	actual_time	216.933738
15	Kerala	actual_time	247.173010
16	Madhya Pradesh	actual_time	559.406940
17	Maharashtra	actual_time	296.011054
18	Mizoram	actual_time	1427.000000
19	Nagaland	actual_time	494.600000
20	Orissa	actual_time	332.308411

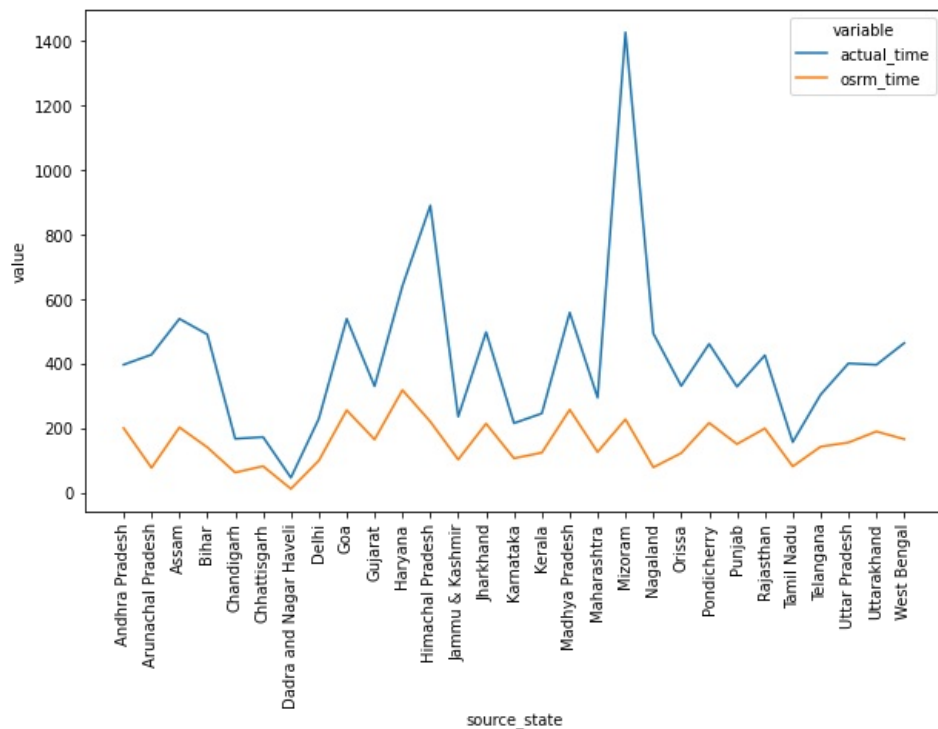
21	Pondicherry	actual_time	462.416667
22	Punjab	actual_time	329.917910
23	Rajasthan	actual_time	427.264591
24	Tamil Nadu	actual_time	158.486044
25	Telangana	actual_time	305.440764
26	Uttar Pradesh	actual_time	402.257218
27	Uttarakhand	actual_time	397.728070
28	West Bengal	actual_time	465.124812
29	Andhra Pradesh	osrm_time	201.549425
30	Arunachal Pradesh	osrm_time	78.750000
31	Assam	osrm_time	204.085821
32	Bihar	osrm_time	142.324786
33	Chandigarh	osrm_time	64.215054
34	Chhattisgarh	osrm_time	83.906977
35	Dadra and Nagar Haveli	osrm_time	13.733333
36	Delhi	osrm_time	100.934066
37	Goa	osrm_time	257.276923
38	Gujarat	osrm_time	166.368000
39	Haryana	osrm_time	319.556039
40	Himachal Pradesh	osrm_time	221.852941
41	Jammu & Kashmir	osrm_time	104.352941
42	Jharkhand	osrm_time	215.600000
43	Karnataka	osrm_time	108.295847
44	Kerala	osrm_time	125.650519
45	Madhya Pradesh	osrm_time	258.952681
46	Maharashtra	osrm_time	127.739499
47	Mizoram	osrm_time	229.000000
48	Nagaland	osrm_time	80.200000
49	Orissa	osrm_time	124.411215
50	Pondicherry	osrm_time	217.916667
51	Punjab	osrm_time	152.354478
52	Rajasthan	osrm_time	200.708171
53	Tamil Nadu	osrm_time	83.142445
54	Telangana	osrm_time	144.253503
55	Uttar Pradesh	osrm_time	157.217848
56	Uttarakhand	osrm_time	191.324561
57	West Bengal	osrm_time	167.622556

In [166...

```
plt.figure(figsize=(10,6))
sns.barplot(data=x1,x='source_state',y='value',hue='variable')
plt.xticks(rotation=90)
plt.show()
```



```
In [167]: plt.figure(figsize=(10,6))
sns.lineplot(data=x1,x='source_state',y='value',hue='variable')
plt.xticks(rotation=90)
plt.show()
```



Relation between actual distance and estimated distance for product delivery

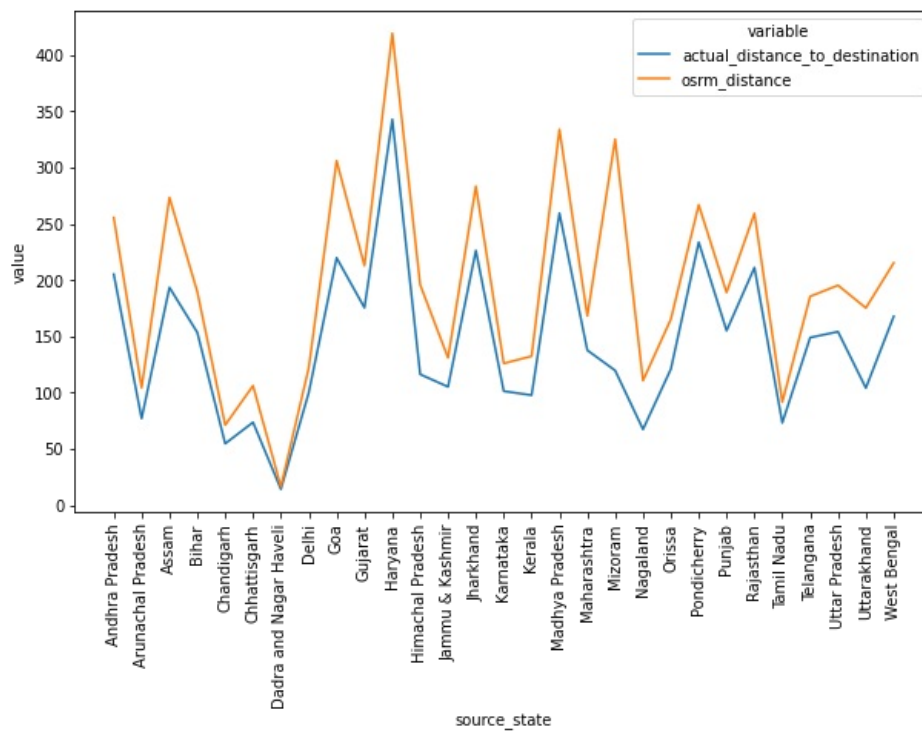
```
In [168]: x1=merged_df1.groupby('source_state').aggregate({'actual_distance_to_destination':'mean','osrm_distance':'mean'})
x1=pd.melt(x1, id_vars=['source_state'], value_vars=['actual_distance_to_destination', 'osrm_distance'])
x1
```

Out[168]:

	source_state	variable	value
0	Andhra Pradesh	actual_distance_to_destination	205.325072
1	Arunachal Pradesh	actual_distance_to_destination	77.150532
2	Assam	actual_distance_to_destination	193.532247
3	Bihar	actual_distance_to_destination	153.533454
4	Chandigarh	actual_distance_to_destination	54.811432
5	Chhattisgarh	actual_distance_to_destination	73.734984
6	Dadra and Nagar Haveli	actual_distance_to_destination	14.349976
7	Delhi	actual_distance_to_destination	100.054953

8	Goa	actual_distance_to_destination	219.883526
9	Gujarat	actual_distance_to_destination	175.549300
10	Haryana	actual_distance_to_destination	342.895103
11	Himachal Pradesh	actual_distance_to_destination	116.370463
12	Jammu & Kashmir	actual_distance_to_destination	105.300217
13	Jharkhand	actual_distance_to_destination	226.406040
14	Karnataka	actual_distance_to_destination	101.412585
15	Kerala	actual_distance_to_destination	97.875658
16	Madhya Pradesh	actual_distance_to_destination	259.562780
17	Maharashtra	actual_distance_to_destination	137.783489
18	Mizoram	actual_distance_to_destination	119.774782
19	Nagaland	actual_distance_to_destination	67.510835
20	Orissa	actual_distance_to_destination	121.063622
21	Pondicherry	actual_distance_to_destination	233.749084
22	Punjab	actual_distance_to_destination	155.165304
23	Rajasthan	actual_distance_to_destination	211.217983
24	Tamil Nadu	actual_distance_to_destination	73.311371
25	Telangana	actual_distance_to_destination	149.049494
26	Uttar Pradesh	actual_distance_to_destination	154.254277
27	Uttarakhand	actual_distance_to_destination	104.240000
28	West Bengal	actual_distance_to_destination	167.877726
29	Andhra Pradesh	osrm_distance	255.621811
30	Arunachal Pradesh	osrm_distance	104.452450
31	Assam	osrm_distance	273.490164
32	Bihar	osrm_distance	189.747497
33	Chandigarh	osrm_distance	71.311419
34	Chhattisgarh	osrm_distance	106.256274
35	Dadra and Nagar Haveli	osrm_distance	16.592587
36	Delhi	osrm_distance	122.053160
37	Goa	osrm_distance	306.198908
38	Gujarat	osrm_distance	212.907671
39	Haryana	osrm_distance	419.345857
40	Himachal Pradesh	osrm_distance	195.963321
41	Jammu & Kashmir	osrm_distance	131.068324
42	Jharkhand	osrm_distance	283.488592
43	Karnataka	osrm_distance	126.073458
44	Kerala	osrm_distance	132.494697
45	Madhya Pradesh	osrm_distance	333.969259
46	Maharashtra	osrm_distance	168.249902
47	Mizoram	osrm_distance	325.281675
48	Nagaland	osrm_distance	110.817940
49	Orissa	osrm_distance	165.384751
50	Pondicherry	osrm_distance	266.931242
51	Punjab	osrm_distance	189.070016
52	Rajasthan	osrm_distance	259.339060
53	Tamil Nadu	osrm_distance	91.682308
54	Telangana	osrm_distance	185.625192
55	Uttar Pradesh	osrm_distance	195.463270
56	Uttarakhand	osrm_distance	175.364052
57	West Bengal	osrm_distance	215.371641

```
In [169.. plt.figure(figsize=(10,6))
sns.lineplot(data=x1,x='source_state',y='value',hue='variable')
plt.xticks(rotation=90)
plt.show()
```



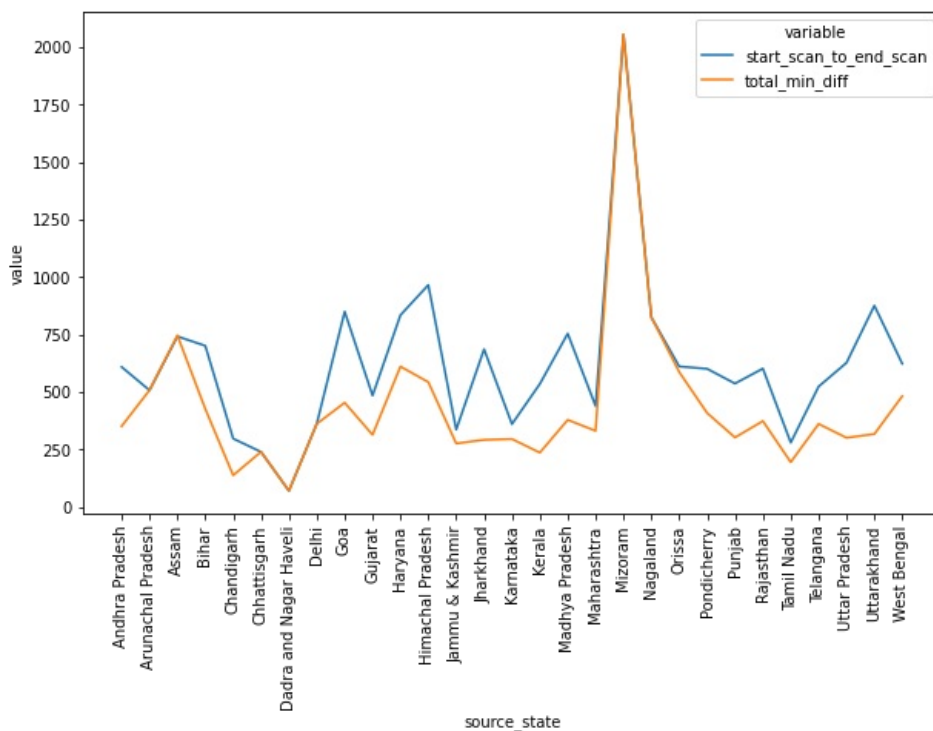
Relation between actual time and estimated time for product deliver from source to destination

```
In [170]: x1=merged_df1.groupby('source_state').aggregate({'start_scan_to_end_scan':'mean','total_min_diff':'mean'}).reset_index()
x1=pd.melt(x1, id_vars=['source_state'], value_vars=['start_scan_to_end_scan', 'total_min_diff'])
x1
```

	source_state	variable	value
0	Andhra Pradesh	start_scan_to_end_scan	609.694253
1	Arunachal Pradesh	start_scan_to_end_scan	508.500000
2	Assam	start_scan_to_end_scan	742.720149
3	Bihar	start_scan_to_end_scan	701.903134
4	Chandigarh	start_scan_to_end_scan	298.290323
5	Chhattisgarh	start_scan_to_end_scan	240.395349
6	Dadra and Nagar Haveli	start_scan_to_end_scan	71.200000
7	Delhi	start_scan_to_end_scan	363.497253
8	Goa	start_scan_to_end_scan	850.507692
9	Gujarat	start_scan_to_end_scan	485.690667
10	Haryana	start_scan_to_end_scan	835.249728
11	Himachal Pradesh	start_scan_to_end_scan	966.352941
12	Jammu & Kashmir	start_scan_to_end_scan	337.411765
13	Jharkhand	start_scan_to_end_scan	686.812500
14	Karnataka	start_scan_to_end_scan	361.357909
15	Kerala	start_scan_to_end_scan	536.048443
16	Madhya Pradesh	start_scan_to_end_scan	754.949527
17	Maharashtra	start_scan_to_end_scan	440.095063
18	Mizoram	start_scan_to_end_scan	2054.750000
19	Nagaland	start_scan_to_end_scan	825.600000
20	Orissa	start_scan_to_end_scan	612.018692
21	Pondicherry	start_scan_to_end_scan	601.916667
22	Punjab	start_scan_to_end_scan	537.755597
23	Rajasthan	start_scan_to_end_scan	602.838521
24	Tamil Nadu	start_scan_to_end_scan	280.906641
25	Telangana	start_scan_to_end_scan	524.649682
26	Uttar Pradesh	start_scan_to_end_scan	628.146982
27	Uttarakhand	start_scan_to_end_scan	876.456140
28	West Bengal	start_scan_to_end_scan	623.929323
29	Andhra Pradesh	total_min_diff	352.277578
30	Arunachal Pradesh	total_min_diff	508.420443

30	Arunachal Pradesh	total_min_diff	509.150113
31	Assam	total_min_diff	747.093302
32	Bihar	total_min_diff	427.482453
33	Chandigarh	total_min_diff	138.687010
34	Chhattisgarh	total_min_diff	240.979895
35	Dadra and Nagar Haveli	total_min_diff	71.703475
36	Delhi	total_min_diff	362.499357
37	Goa	total_min_diff	454.558358
38	Gujarat	total_min_diff	314.738026
39	Haryana	total_min_diff	611.743258
40	Himachal Pradesh	total_min_diff	543.952309
41	Jammu & Kashmir	total_min_diff	277.121679
42	Jharkhand	total_min_diff	292.998060
43	Karnataka	total_min_diff	296.529968
44	Kerala	total_min_diff	237.169945
45	Madhya Pradesh	total_min_diff	379.722736
46	Maharashtra	total_min_diff	332.091113
47	Mizoram	total_min_diff	2056.028526
48	Nagaland	total_min_diff	826.324622
49	Orissa	total_min_diff	587.600940
50	Pondicherry	total_min_diff	409.609242
51	Punjab	total_min_diff	303.230200
52	Rajasthan	total_min_diff	374.573627
53	Tamil Nadu	total_min_diff	195.835152
54	Telangana	total_min_diff	362.046573
55	Uttar Pradesh	total_min_diff	301.936576
56	Uttarakhand	total_min_diff	318.202144
57	West Bengal	total_min_diff	482.943198

```
In [171]: plt.figure(figsize=(10,6))
sns.lineplot(data=x1,x='source_state',y='value',hue='variable')
plt.xticks(rotation=90)
plt.show()
```



Destination state with the highest product delivery

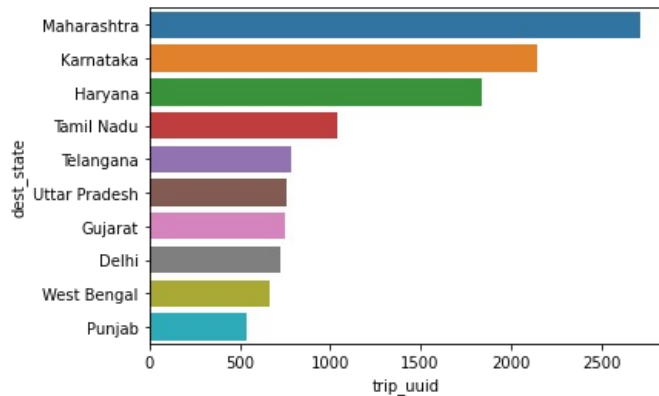
```
In [172]: d_state=merged_df1.groupby('dest_state')['trip_uuid'].count().sort_values(ascending=False).reset_index().head(1)
d_state
```

Out[172]:

	dest_state	trip_uuid
0	Maharashtra	2714
1	Karnataka	2143
2	Haryana	1838
3	Tamil Nadu	1039
4	Telangana	785
5	Uttar Pradesh	762
6	Gujarat	750
7	Delhi	728
8	West Bengal	665
9	Punjab	536

In [173]: sns.barplot(data=d\_state, y='dest\_state', x='trip\_uuid')

Out[173]: <AxesSubplot:xlabel='trip\_uuid', ylabel='dest\_state'>



Source state providing highest product delivery

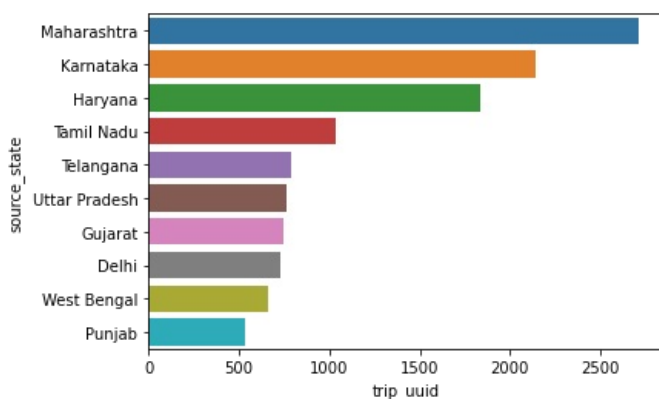
In [174]: s\_state=merged\_df1.groupby('source\_state')['trip\_uuid'].count().sort\_values(ascending=False).reset\_index().head(s\_state)

Out[174]:

	source_state	trip_uuid
0	Maharashtra	2714
1	Karnataka	2143
2	Haryana	1838
3	Tamil Nadu	1039
4	Telangana	785
5	Uttar Pradesh	762
6	Gujarat	750
7	Delhi	728
8	West Bengal	665
9	Punjab	536

In [175]: sns.barplot(data=s\_state, y='source\_state', x='trip\_uuid')

Out[175]: <AxesSubplot:xlabel='trip\_uuid', ylabel='source\_state'>

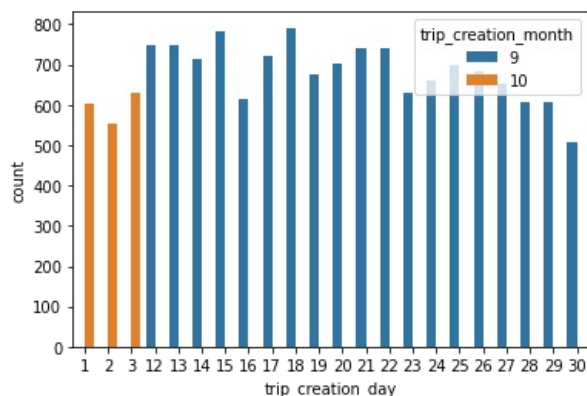


How the product delivery has changed across the given months



```
In [176]: sns.countplot(data=merged_df1,x='trip_creation_day',hue='trip_creation_month')
```

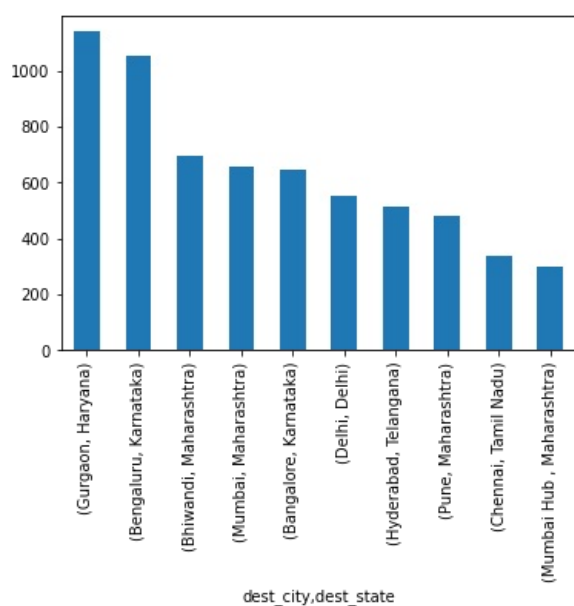
```
Out[176]: <AxesSubplot:xlabel='trip_creation_day', ylabel='count'>
```



Destination city having most product deliveries in a state

```
In [177]: merged_df1.groupby(['dest_city','dest_state'])['trip_uuid'].count().sort_values(ascending=False).head(10).plot()
```

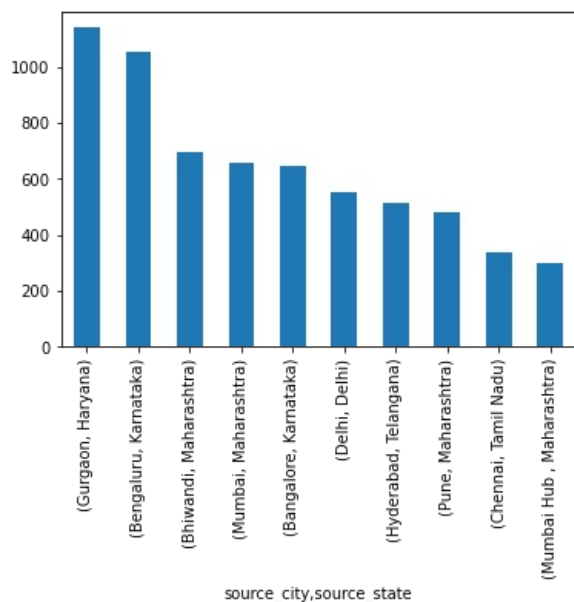
```
Out[177]: <AxesSubplot:xlabel='dest_city,dest_state'>
```



Source city providing most product deliveries in a state

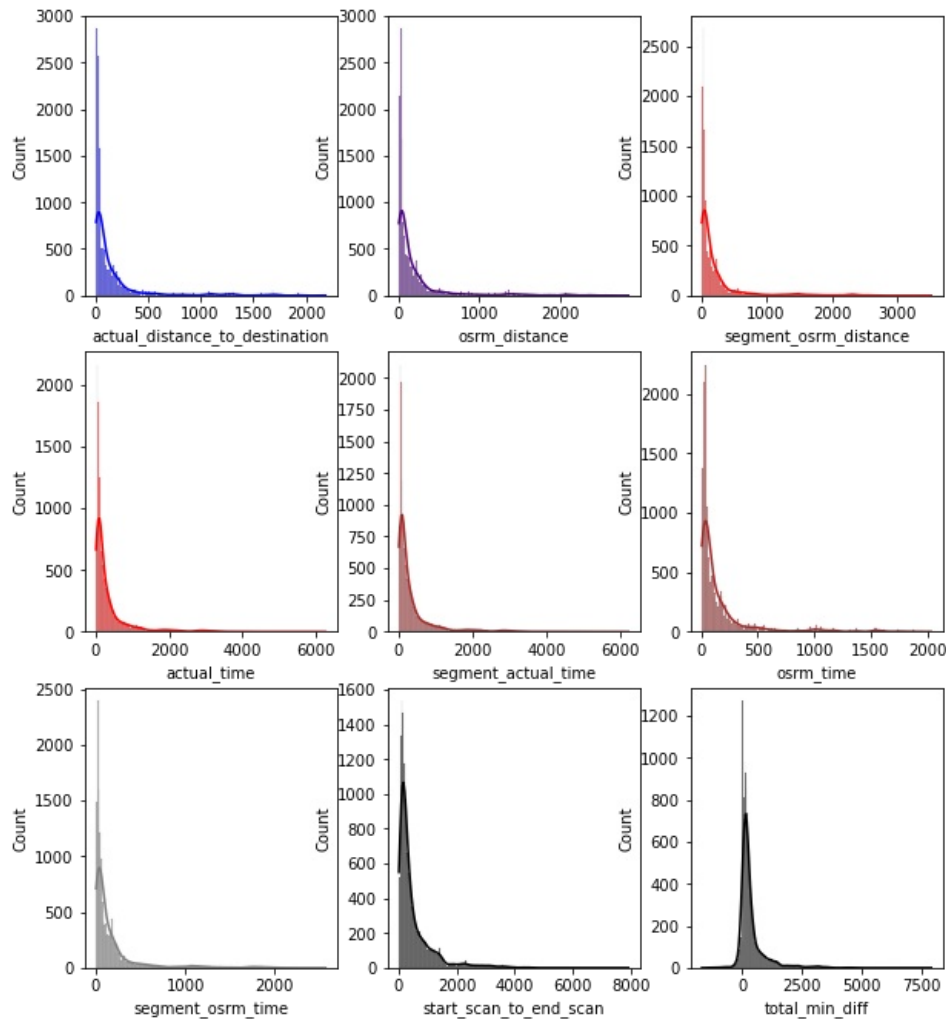
```
In [178]: merged_df1.groupby(['source_city','source_state'])['trip_uuid'].count().sort_values(ascending=False).head(10).plot()
```

```
Out[178]: <AxesSubplot:xlabel='source_city,source_state'>
```



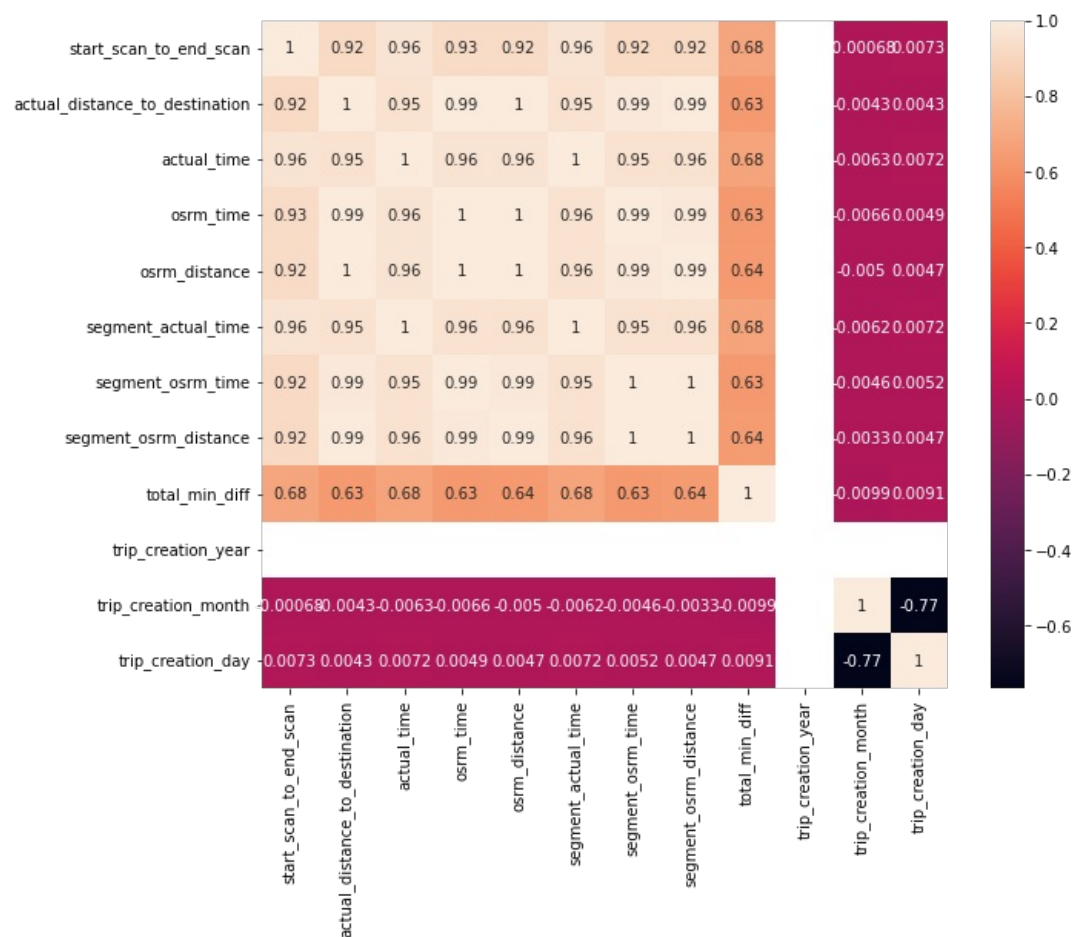
Distribution of the quantitive data

```
In [179.. fig, axis = plt.subplots(nrows=3, ncols=3, figsize=(10,8))
fig.subplots_adjust(top=1.2)
sns.histplot(data=merged_df1, x="actual_distance_to_destination", kde=True,color="blue", ax=axis[0,0])
sns.histplot(data=merged_df1, x="osrm_distance", kde=True,color="indigo", ax=axis[0,1])
sns.histplot(data=merged_df1, x="segment_osrm_distance", kde=True,color="red", ax=axis[0,2])
sns.histplot(data=merged_df1, x="actual_time", kde=True,color="red", ax=axis[1,0])
sns.histplot(data=merged_df1, x="segment_actual_time", kde=True,color="brown", ax=axis[1,1])
sns.histplot(data=merged_df1, x="osrm_time", kde=True,color="brown", ax=axis[1,2])
sns.histplot(data=merged_df1, x="segment_osrm_time", kde=True,color="grey", ax=axis[2,0])
sns.histplot(data=merged_df1, x="start_scan_to_end_scan", kde=True,color="black", ax=axis[2,1])
sns.histplot(data=merged_df1, x="total_min_diff", kde=True,color="black", ax=axis[2,2])
plt.show()
```



Correlation of the quantitative data

```
In [180.. plt.figure(figsize=(10,8))
sns.heatmap(merged_df1.corr(), annot=True)
plt.show()
```



```
In [181]: sns.pairplot(merged_df1)
```

```
Out[181]: <seaborn.axisgrid.PairGrid at 0x22021377df0>
```



## FEATURE ENGINEERING

```
In [182]: def missing_to_df(df):
total_missing_df=df.isnull().sum().sort_values(ascending=False)
percent_missing_df=(df.isnull().sum()/df.isnull().count()*100).sort_values(ascending=False)
missing_data_df = pd.concat([total_missing_df,percent_missing_df],axis=1,keys=['Total', 'Percent'])
return missing_data_df
```

```
In [183]: missing_df = missing_to_df(df)
missing_df[missing_df['Total']>0]
```

```
Out[183]:
```

	Total	Percent
source_name	293	0.202254
destination_name	261	0.180165

```
In [184]: cat_missing=['source_name','destination_name']
freq_imputer=SimpleImputer(strategy='most_frequent')
for col in cat_missing:
df[col]=pd.DataFrame(freq_imputer.fit_transform(pd.DataFrame(df[col])))
```

```
In [185]: missing_to_df(df)
```

Out[185]:

	Total	Percent
data	0	0.0
trip_creation_time	0	0.0
segment_osrm_distance	0	0.0
segment_osrm_time	0	0.0
segment_actual_time	0	0.0
factor	0	0.0
osrm_distance	0	0.0
osrm_time	0	0.0
actual_time	0	0.0
actual_distance_to_destination	0	0.0
cutoff_timestamp	0	0.0
cutoff_factor	0	0.0
is_cutoff	0	0.0
start_scan_to_end_scan	0	0.0
od_end_time	0	0.0
od_start_time	0	0.0
destination_name	0	0.0
destination_center	0	0.0
source_name	0	0.0
source_center	0	0.0
trip_uuid	0	0.0
route_type	0	0.0
route_schedule_uuid	0	0.0
segment_factor	0	0.0

## In-depth analysis and feature engineering

In [186.. `data=df.copy()`

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

In [187.. `data['total_min_diff']=(df['od_end_time']-df['od_start_time'])/pd.Timedelta(minutes=1)`  
`data.drop(columns=['od_end_time','od_start_time'],axis=1,inplace=True)`  
`data.head(10)`

Out[187]:	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388
5	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388
6	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388
7	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388
8	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388
9	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388620AAB	Khambhat_MotvdDPP_D (Gujarat)	IND388

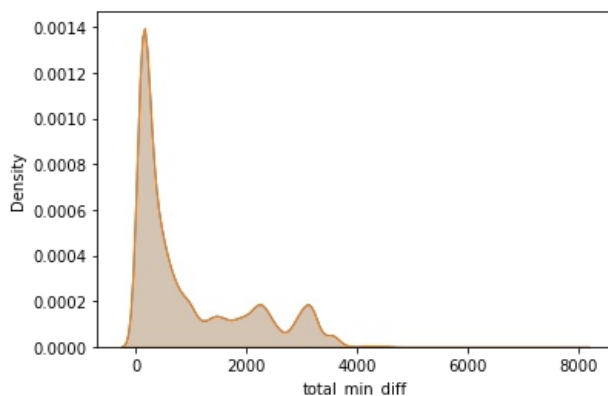
10 rows × 23 columns

## Hypothesis Testing

- Ho : Mean of start\_scan\_to\_end\_scan and total\_min\_diff are same
- H1 : Mean of start\_scan\_to\_end\_scan and total\_min\_diff are not same

```
In [188]: sns.kdeplot(data['total_min_diff'],fill=True)
sns.kdeplot(data['start_scan_to_end_scan'],fill=True)

Out[188]: <AxesSubplot:xlabel='total_min_diff', ylabel='Density'>
```



```
In [189]: observed_mean_diff = np.mean(data['total_min_diff']) - np.mean(data['start_scan_to_end_scan'])
observed_mean_diff
```

```
Out[189]: 0.49601665918658
```

```
In [190]: p = np.concatenate((data['total_min_diff'], data['start_scan_to_end_scan']))
p
```

```
Out[190]: array([ 86.21363662,  86.21363662,  86.21363662, ..., 427.
        427.          , 427.          ])
```

```
In [191]: test_sample=[]
itr=10000
for i in range(itr):
    np.random.shuffle(p)
    new_c1=p[:144867]
    new_c2=p[144867:]
    mean_diff=np.mean(new_c1)-np.mean(new_c2)
    test_sample.append(mean_diff)
```

In [192]: test\_sample

```
Out[192]: [2.9130080569036636,
1.714433824299931,
3.9213194724794675,
-1.6862885632442612,
-1.3524307801213808,
-0.17139391092518963,
-2.3011318053513605,
-6.139588956577654,
-0.4286654923705555,
-2.069839623111875,
0.4090150083932258,
2.2971077637489543,
0.5498343903736895,
2.6186824782269014,
1.1060234917126763,
-5.571143576403415,
-0.40461470640639163,
5.533397186922571,
1.3239129960597893,
-6.194217919992752,
-1.905102003845741,
-4.476058707635161,
1.5256382198344909,
-3.4550447993616444,
-0.7821487274678702,
-2.154121548497983,
3.5820572202242147,
-0.02129475667970837,
4.260938877646595,
1.181967727356323,
-4.725656846494985,
0.8930090859425945,
0.27368186708986286,
-1.2974378494996017,
-3.388293571186182,
-6.4854015086635854,
-0.7799943828715641,
3.610209735335502,
-2.1087137376214287,
1.5641794578416466,
8.114973556873338,
-2.1699182727606967,
-3.8153456802942856,
0.8442289837041699,
0.21719960725909004,
-6.1807152834774115,
1.8095805449518139,
-2.938937051590983,
3.349469376238062,
0.5166871431136997,
6.336790944645827,
-2.052584999120654,
-0.5867406011871026,
-2.5391679689344073,
1.3559667207939583,
-1.9526283668160431,
1.0635847498363091,
-7.041314819798117,
1.1802134071873525,
-0.761689116035086,
5.334086963637901,
3.4034679274892596,
0.6921267529476154,
2.1544379805025073,
-4.01684008881773,
-2.6438837820114713,
2.014542400901746,
-3.3262903138843285,
2.0441953745232695,
1.4945097922227433,
-1.6081439037689051,
-5.2291763608197925,
0.8871502164664662,
7.704646753354382,
-3.404909932537066,
7.674666418839479,
-2.0156960232991423,
2.2059067150289593,
-3.4409415670193084,
-5.849088269294953,
-4.537532788033104,
0.6995884155026033,
3.7582260641240737,
1.394691759025136,
0.8024635666306494,
-2.0756556512517363,
```

-2.7021988990147747,  
5.706051828675072,  
-3.225110907532553,  
-5.013663002319845,  
-0.0845979989363741,  
-2.7749909249582743,  
-1.53578247076166,  
2.603038555064586,  
7.615835991629069,  
-2.171815086168408,  
-2.3597669452828995,  
3.702772993563485,  
0.32043355381074434,  
0.14579520580196004,  
-5.08696837109494,  
1.8464344956919376,  
5.921743253278009,  
-3.635841924689771,  
-2.4338649402772035,  
-3.390780422240482,  
2.9070538245722446,  
-2.711607305310622,  
0.2521339114214243,  
4.230364147222872,  
4.145871781428696,  
3.753902083361254,  
0.6749530599322497,  
-1.2841014397745312,  
1.4860587836692503,  
-1.652130000756415,  
-6.567528288075323,  
-1.6117124407805932,  
0.8318680844570281,  
8.136596424457139,  
3.61399295051433,  
0.24556658655842512,  
-4.335042652502807,  
-0.02213261085148588,  
-2.851427247909214,  
1.5707206634835984,  
0.40601987146101237,  
-0.10572337066116688,  
2.20628032008608,  
-3.677764708734003,  
6.744528284318449,  
3.3147120622511466,  
1.5878182468018167,  
2.591305388549472,  
2.2336000195953147,  
-2.9834354882161733,  
1.5686885633333532,  
1.8232641011707074,  
-3.2973479290969863,  
-4.67308611341582,  
-4.757354918906003,  
6.656041848914356,  
0.9385788041549858,  
2.9532570887913607,  
6.657346065685715,  
5.764247763778485,  
4.670242851410876,  
-2.832425715767613,  
9.508659290892183,  
1.8912286199522441,  
-1.1644820395379156,  
3.1883936076228565,  
-0.7256268362730225,  
-4.038979647679525,  
5.227940735743459,  
-0.4206664089933838,  
3.478523103456382,  
3.683989084928271,  
-5.090461157660911,  
2.040039942767862,  
4.535881119701344,  
-8.052449087341984,  
0.49126285611714593,  
1.8747036107131407,  
0.25851262654123275,  
-5.651294768749608,  
1.9858490357894425,  
1.907511417094156,  
-5.394941094904311,  
1.664838553658683,  
2.552326693044506,  
-11.83951107052053,  
-2.6884617467784437,  
-6.524392552402787,  
-1.3877054109636902,



-4.887569645552276,  
0.16594860396730837,  
-3.9294487352626675,  
0.9261160552800902,  
-2.3018783904192333,  
-4.3322265211811555,  
5.020299440473764,  
-1.8623771856080111,  
4.244492332175923,  
-1.7679618143167772,  
4.477583647983124,  
-1.0212029263298064,  
-3.1835770065392808,  
0.5316371664323469,  
-1.7866886124290886,  
-0.7797894676905344,  
-3.969419553613079,  
1.247772384043401,  
0.6337495493615961,  
0.8131728766005608,  
-1.7445558442211677,  
-2.2184799200488214,  
-1.661054371141745,  
7.020553709620344,  
6.023437774926151,  
0.8753491384734389,  
-1.3700553542860234,  
1.9246586944782393,  
-2.8707436674435485,  
2.720458462625402,  
6.924925118982628,  
2.15560180365037,  
10.963008158033063,  
2.572591473811144,  
6.666344693311771,  
2.4880712637827855,  
5.282157167206606,  
-3.633577081595604,  
4.6034096155888165,  
1.2491247876530451,  
-4.916889301239962,  
-1.4305411833395283,  
-0.48562687197852483,  
3.5889174875106846,  
-2.586725283514852,  
1.903279829224175,  
2.8147297492212147,  
2.1022855846216544,  
-0.4532556516421664,  
-7.238096550754108,  
-2.59869649699192,  
-4.58517113111111,  
3.368276153315378,  
0.38481649138043394,  
-3.3182957931219335,  
1.702269219278378,  
-5.693996431996311,  
0.4610073812769997,  
-0.17152028695318222,  
0.4397596429441819,  
8.983759388786666,  
2.2152946608073307,  
-3.874501874404814,  
3.0904510661704307,  
3.46173643888892,  
6.725111581005649,  
-1.6169308999144505,  
3.3259934718786326,  
0.34156024628293835,  
3.271553161339625,  
-4.2597607910097395,  
-0.7353143731560294,  
-1.9498173614138068,  
4.701399884261036,  
3.5079554746644135,  
-0.32302442866694037,  
-3.277492873741039,  
-2.8280816076992323,  
2.380585538875721,  
4.605355230846953,  
-6.020477924388274,  
8.826100112059976,  
-2.7663424466336437,  
1.9313580850537164,  
10.499901756739519,  
0.8122142041311236,  
0.36797940566316356,  
0.714907015208837,  
-3.708689546321466,

-0.670276618584694,  
-6.348432839220891,  
-3.8106756989227506,  
-2.011586859345016,  
4.622691683735297,  
-2.5107222725296197,  
0.8076048792959227,  
3.762397496961057,  
-1.625795044594156,  
-0.33660544904000744,  
-0.6005668427025057,  
4.986423944590115,  
-0.7459547327068776,  
-6.066029294906684,  
0.34122579444442636,  
-7.740055491731027,  
2.6940534540888166,  
-0.46724620417603546,  
-1.8827127010169988,  
3.4590852693819443,  
-0.022724785718537532,  
-6.0403352908257375,  
1.739222520118119,  
-3.0946852064860195,  
0.0032569212648922985,  
2.760980807189526,  
4.24615075992233,  
2.811446632866364,  
-2.2895702924022316,  
3.630695185452396,  
-0.3884919554645876,  
4.478010043731729,  
7.161684733854827,  
2.4156696712720986,  
-6.676779653266522,  
-6.677101251536442,  
-4.703519875396751,  
3.177866145844064,  
0.22791063337581363,  
2.7627524888538346,  
0.8068233227143082,  
3.4865367184081606,  
5.192731821980715,  
3.535614965156583,  
-5.739357529628705,  
-1.4758747899966238,  
-0.6848293317193566,  
8.71111534040665,  
2.798460445481396,  
5.148828264971598,  
3.8019700183519944,  
7.599612282767566,  
-1.6571329576377138,  
9.335663627018675,  
-6.669926296834774,  
3.7405814879773516,  
5.059714601678252,  
-3.185738605212009,  
-6.3423637458136,  
-1.0169237344950943,  
-1.5810025420396414,  
5.2353192777572986,  
6.366106491042501,  
0.7445877790976283,  
-0.5889894777426434,  
3.9310594902781304,  
-0.45005710380837627,  
-4.231004239895128,  
-1.8976101608807312,  
-1.2790705597763008,  
4.029628508738369,  
0.03503627020154454,  
-5.980297290606359,  
6.5626871512489515,  
2.1182413503527187,  
-4.166528179294346,  
2.9714789386305256,  
2.910042154019493,  
1.8157583018048626,  
1.1846245809169886,  
-5.12369740246595,  
5.8946670949037525,  
0.19616333940905406,  
-6.781510253781448,  
-0.9747181319071387,  
-7.79146184021738,  
4.5439482139017855,  
6.784422789689188,  
-6.493836226725989,

-3.0833120203882345,  
1.953087895458566,  
2.4500210844682897,  
3.5915241134290454,  
-2.659158660629373,  
1.9310646267215361,  
-0.8761109016963928,  
0.6479424800659217,  
4.465687741571969,  
-6.948839682952212,  
3.5723341130216113,  
-1.9481430330629337,  
1.0412235908218008,  
-4.209041530305058,  
0.8918524699511181,  
5.654249467288992,  
-4.7100893913420805,  
1.568633004679441,  
-9.426913694622726,  
3.748532923581479,  
3.093673480988855,  
2.2985057003531892,  
4.048710165335706,  
1.5253398194099645,  
-0.24916890395456903,  
-6.703598098685916,  
-3.525247817882814,  
1.814633492091616,  
5.522750445272322,  
-1.288523648436012,  
3.584932701602611,  
-5.252655015778373,  
-6.766005142767085,  
4.808535013184837,  
-1.3736700733454654,  
-5.647287140089702,  
-0.534572065116663,  
9.215093498926421,  
0.01968472882572314,  
0.2543632057133891,  
-2.5544727338022994,  
0.6106329535139139,  
6.033913128652102,  
0.09902397118185036,  
6.20988053592032,  
2.07667652373857,  
-0.9840891454712164,  
2.6813141345085114,  
-0.650031187525201,  
-2.1818144092053444,  
-4.488831398855837,  
5.585913073578695,  
3.16786039234114,  
1.8081561897840857,  
2.854065255360865,  
-3.128982925002447,  
0.19284632418316505,  
-1.5201624311555406,  
-4.929670113406473,  
3.2578611927499423,  
-4.256401768655451,  
-2.6905833786562425,  
2.1069156489720626,  
-1.5841657715233168,  
0.2823711269352316,  
-3.144510167631779,  
-0.7859434653903463,  
-3.7419738187729763,  
0.510306784741033,  
2.1254685311203048,  
5.002939359209904,  
1.8363915000807083,  
-2.7133300812554353,  
5.973562014706658,  
2.469051378249219,  
3.8691400805832927,  
-1.126269088600111,  
-2.161853556447795,  
1.8206593756201528,  
0.3850892379398374,  
-2.897745110169126,  
-9.354278109591405,  
-1.528870517240648,  
3.006400897072126,  
-7.238944685395836,  
-2.5114611961939772,  
1.825885976404379,  
-1.4898681192398726,  
8.461647841036893,

4.333530327799281,  
6.259993455439144,  
5.138156652648831,  
-2.807077633221297,  
-5.389482139285292,  
-0.755135163070463,  
-0.665057390897573,  
3.704965943763682,  
0.9177028611995866,  
-8.054712198756988,  
-1.5073359811246974,  
0.33716811049316675,  
2.251372238106569,  
-0.5140161829144745,  
-7.344871525237295,  
0.7512184945751414,  
10.890383797218192,  
1.068694183906473,  
-1.886488063236584,  
3.8672218073409113,  
0.6007615549724505,  
-0.6056763449747677,  
1.715223473016522,  
-10.472710042807421,  
-9.273744781455662,  
5.580534935564174,  
0.5593186809003328,  
2.5990752097379755,  
-0.40849362012318124,  
2.6714489556738954,  
3.664941360593616,  
-0.012953315555705558,  
-6.57271866585927,  
0.2846496991166987,  
-1.2037613101526858,  
4.249307803224838,  
-1.8813419918276395,  
2.7512978154890106,  
-1.4545230274759433,  
4.626849297021863,  
-3.4680794510538817,  
-0.5851711342077124,  
-2.5366815664419846,  
-6.269192390212311,  
7.322893484622682,  
0.5174338787883244,  
-0.8636586214377076,  
0.3807778330351539,  
4.02105943924073,  
4.682000105844963,  
-3.1853968322323,  
-0.4191197065140386,  
0.923035087699418,  
-7.596820580690633,  
0.35012503793291216,  
-1.3786335816157589,  
2.893290496414238,  
3.0982520174093224,  
-1.3739098157495846,  
-3.1197751554698243,  
2.3627613644487155,  
4.003699216031009,  
-0.21568717967943485,  
3.751378447075467,  
-2.672726117220236,  
2.2546428808311703,  
2.929918680728406,  
7.283201714866891,  
-3.349452145369696,  
-2.1022623079718414,  
4.822057334632063,  
3.9463524798429717,  
-1.2372145753888617,  
-4.06780685558067,  
3.9126232804126175,  
-1.363134255293744,  
4.527348092904731,  
-2.2676220194327925,  
3.7890218211609863,  
1.1423908852987097,  
-6.085767829298106,  
-8.940523184849098,  
-1.1145000222491035,  
-1.8235068696992585,  
4.322752010383056,  
-1.3351350943075886,  
-1.1368548997559174,  
-0.6052232411921068,  
1.9460332057365122,

-4.419907841492659,  
5.241100518500957,  
2.9009182447366584,  
-6.698863723121121,  
-0.003460389033421052,  
1.0302259301809045,  
1.899852918644342,  
-2.685968562972448,  
-0.7444035086234635,  
-2.867992197601552,  
10.699451007847301,  
2.6321395996678802,  
3.895371689670583,  
-3.8250131035279082,  
1.0254539275215393,  
-0.749831947239727,  
1.786856969782093,  
-4.092616749994022,  
3.7665959619075693,  
-1.126769575723074,  
2.387107549439861,  
-7.294196893176604,  
-1.4946306365623059,  
2.09845798005324,  
1.5564625720246568,  
2.9133306068338243,  
-5.997410285518754,  
-3.99790044980773,  
-1.6042110473085813,  
-2.8383369044747724,  
-3.536151024828996,  
7.68458437994218,  
-6.098004904090658,  
2.5416037103102553,  
1.6569220298276832,  
-0.944055913324064,  
1.2689600500527831,  
-0.353834849109262,  
1.2426147677718973,  
7.718544432355429,  
5.497028811108066,  
-5.245786502800229,  
-0.07025282001313826,  
3.519185798968465,  
8.06359933117119,  
0.7994456991548304,  
2.362869719177297,  
-2.5237173598991376,  
-0.3000414631289914,  
-2.2993693834515625,  
1.9945354200986003,  
2.9563393363991963,  
2.1492157152210893,  
-6.054523604679389,  
2.5428225473448265,  
8.621289163176357,  
-0.8945993009861013,  
6.194680487375763,  
4.704673631149717,  
4.445719064202876,  
-2.4536153682931854,  
0.7667587077819462,  
-1.4628613974668951,  
-1.4633769985248364,  
-2.6884281318748435,  
-2.246063426683577,  
1.5917137337639815,  
4.452585944198745,  
-1.4590235503046642,  
3.0790182131415804,  
3.394030240624943,  
-0.7088257488485397,  
4.651157264879203,  
1.0486944992806002,  
-0.6770623655720556,  
-3.71985524211334,  
1.462002971478114,  
2.1825514722233947,  
-2.7334725451539725,  
-5.526916595067178,  
2.7612764568589228,  
0.04170356481927229,  
-4.392268630241233,  
-13.729703281656157,  
1.320220029764414,  
0.24103194313534004,  
5.710207585960006,  
5.004288417314001,  
-0.5062698093156541,

3.8093378349497016,  
0.9781105026338537,  
-0.2926997497982029,  
-2.1031690822449036,  
0.10073227217822023,  
0.1493336805714307,  
-0.34224780553927303,  
7.748570703689666,  
-2.460448371475195,  
-0.1557225396334161,  
-5.809941269327737,  
-4.332872392052423,  
-1.0712143199907587,  
-4.389817903930634,  
-1.2383050717928654,  
2.7121612435630595,  
3.4137065099500887,  
5.302720487690749,  
-4.754517426518646,  
-3.4260518545086143,  
1.5219821342939213,  
-4.624716776561058,  
0.5309984272748807,  
3.5857721171985304,  
-6.647931565624276,  
-0.24503945209858102,  
-2.3578807725887145,  
2.804942758891684,  
3.06529319493211,  
1.9500994127984086,  
-2.8353646980419853,  
0.2662486181965278,  
4.200447030727901,  
4.033079432219324,  
-8.535457040978372,  
3.4177730546251723,  
-2.425114077692797,  
-0.2844042187132345,  
1.6495325230107483,  
3.613199651862601,  
-2.985698771610487,  
4.136647222682313,  
-5.180406093138572,  
6.836814134155475,  
-4.38215357505851,  
-4.893635385637253,  
-0.1439321779362217,  
4.981229932800261,  
-2.361285728094117,  
3.847614110703489,  
-15.50119938125772,  
-3.3178779399872838,  
-3.204698806150418,  
6.6028011871328545,  
-3.4375195411668074,  
1.9265436778046023,  
6.148354130158623,  
-0.7386106129516747,  
0.020316453302257287,  
-4.4049370568145605,  
-1.0153334840665593,  
-0.0016113544352265308,  
-1.8927591910501178,  
3.040652577471633,  
0.10507121388809537,  
-2.696904100932443,  
-2.934075470959442,  
-2.9478159749944552,  
3.252090540348263,  
7.433187303219256,  
4.518033921146866,  
-0.23384120690752752,  
3.192137399346848,  
-6.056385518566458,  
-4.954154310590866,  
-3.350915754871039,  
-1.9739365659333998,  
-11.201762857393533,  
-3.2690660405531844,  
0.10425240483709786,  
5.729776798242256,  
-1.0364749947673317,  
4.12039413591333,  
-2.5111229031458606,  
-5.513136934360205,  
1.012155326656739,  
-2.212568426628877,  
-2.806941116207099,  
4.986433702823888,

-4.985470607158959,  
3.5492934027329284,  
9.056059015120013,  
-1.952600621170859,  
-4.27035274092259,  
2.4638270992884372,  
2.461209698685252,  
5.161937504443813,  
-0.923294961137799,  
2.6825834679277705,  
-5.102600282818003,  
5.536359512520448,  
1.6556952175915285,  
3.589031958283158,  
-2.2195011413393786,  
4.245409414980145,  
-2.3820090361045914,  
-0.9214526795799429,  
0.8134804212295421,  
-2.0629596162229973,  
-1.542882438515221,  
0.527284305684816,  
-4.752159949960287,  
4.140443315045559,  
4.849616417993957,  
3.5682790560388185,  
5.935620792279224,  
1.9865536632041767,  
-3.1696687715716507,  
-0.22780074730519573,  
3.0099882595905,  
-0.0601377694747498,  
1.3727265187230842,  
-2.1171770978990025,  
1.0013361998828714,  
0.3526524682545187,  
2.0663727527448827,  
-2.779872207862809,  
3.269536077376756,  
-0.7092547126927684,  
-2.3613956560122915,  
1.834739145708454,  
1.9948847709588335,  
3.2622987420857044,  
4.368542488138701,  
-0.3563549422513006,  
-3.943127822980614,  
0.7543905341026402,  
-1.2846265450389183,  
4.613104952243475,  
0.8870729591110376,  
3.481077640932199,  
5.221145142149908,  
-2.2481760401310567,  
8.608561321792536,  
2.726629114513571,  
0.6358296306482316,  
1.5376261880333004,  
-0.9997493067610321,  
-6.343786655782083,  
7.184583674909163,  
0.3717645590254506,  
2.828609083145693,  
2.7737039436129862,  
-0.4918013418632654,  
6.058285193508482,  
-1.8250138834691825,  
-2.974165574838935,  
-6.895867578921184,  
-2.4050928331098476,  
-1.0781625989448003,  
4.285823049818191,  
-4.601728882586372,  
0.3018860517072426,  
-0.35967497799833836,  
2.7596036754717943,  
-0.37555989165809933,  
4.552290341269099,  
-7.5906236601942965,  
3.6723309232877455,  
2.2560758724437164,  
-1.0953577387888345,  
-2.8414998459211347,  
-2.741073176007262,  
1.6659298104455047,  
-6.4668886897652555,  
1.674203422538426,  
-0.8922640683375676,  
1.936633994585236,

0.05052188797412782,  
-8.146706647050905,  
1.7571394548078842,  
-5.216075345408626,  
-2.475281148914519,  
1.2289530205134724,  
0.07069874889862149,  
-5.8217073671648905,  
5.204849214987576,  
-0.9172749045376349,  
-1.1119719446101044,  
7.099730445662544,  
4.636323107330327,  
-2.6434787233806674,  
6.005674032722482,  
-4.3636612984485055,  
-1.3020101292490835,  
-3.0441184342262204,  
-6.018000383160825,  
-5.6913462128202355,  
7.094593195433163,  
-8.031182514790999,  
2.8940623144835627,  
4.014157514456315,  
1.376052925478234,  
1.3626219154132286,  
0.8384247509054603,  
-1.1785951828264842,  
0.14275504310262477,  
3.4180494060001365,  
-0.03751337596918347,  
4.6993216611867865,  
0.39428691720013376,  
1.1361791221640942,  
-2.0869083399984447,  
-2.6053360362157036,  
9.184567129891661,  
2.0938388439892606,  
0.3025817196612479,  
10.007709927933774,  
5.163616300482772,  
2.870427091561851,  
7.206922545931548,  
-11.819791714191979,  
1.0377735532351835,  
-4.767458529055375,  
-11.862874460431158,  
5.876498048245594,  
-0.7729257265298202,  
-0.1572138255627351,  
-5.480253954873206,  
9.492338743560595,  
1.1644496545989114,  
-2.407762724131885,  
-3.355622493448209,  
-6.901989353039085,  
1.3091830809767089,  
-0.43461884734574596,  
-4.525605162259808,  
4.239854777905975,  
-0.7498453042690016,  
2.101716515586986,  
6.874458372927393,  
7.628138344947047,  
7.570413338274761,  
2.8636824469208477,  
-2.014845326371187,  
6.665323580426957,  
-2.015618399159848,  
4.33261498470597,  
-2.7369034318193144,  
-4.330131994296494,  
-5.191568203911402,  
2.7969626272083588,  
0.5814778833803302,  
-0.7646034639645904,  
-1.905641629325828,  
-0.9725677104036095,  
-7.044342371426637,  
3.0016557761457534,  
-0.6298471193767909,  
-2.1048046381115455,  
-1.655667518477685,  
-3.6694294672299748,  
0.7138898908053761,  
2.1600506093325293,  
4.1399383342322835,  
0.1223770895587677,  
0.7888242423869087,

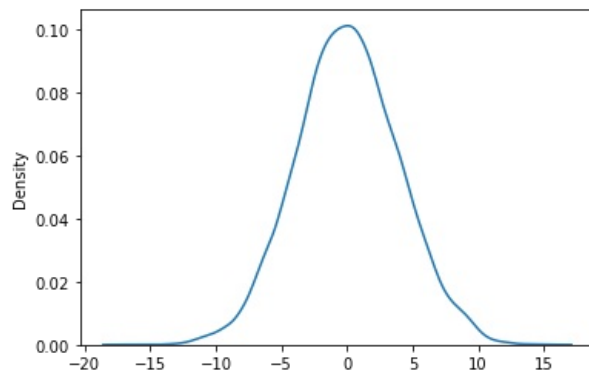


-3.243540983346975,  
-0.40193669960513034,  
-2.403719244263357,  
-0.3444946200827417,  
0.28231488839935537,  
0.03017337672258691,  
-0.6072133460831992,  
-6.780378507181922,  
-6.017695001126299,  
1.9090601174486892,  
-3.3151912281680325,  
-0.8404130153915048,  
0.38820937174568826,  
-6.4339836555229795,  
5.016859387518252,  
-4.483555752723078,  
3.022909685945592,  
-3.153001031973872,  
-2.546946830060506,  
-0.5833842693079987,  
1.598910259144077,  
0.29605282674015143,  
-9.325371605186774,  
2.266656690783975,  
-0.5787799880539524,  
0.07983446482239742,  
0.22721858257659733,  
2.0582679337569516,  
1.8103543754552902,  
-1.5565678803495757,  
-1.1422166751915483,  
4.481637506713923,  
-5.462916726031381,  
4.425744773891665,  
-2.7333022482457636,  
-0.21804221702814175,  
0.8287920472089354,  
9.170594991365988,  
-1.732012058537066,  
2.0824395077688678,  
3.258072323050328,  
-6.80601201714785,  
0.9757903736610842,  
4.816368086600846,  
-5.799945417248182,  
-1.4957415471814102,  
-3.2671644939348425,  
-1.005828575301848,  
-4.958562437070896,  
-3.2158627107538678,  
-2.122409113623007,  
5.579184808864738,  
6.751758992653663,  
-1.55606067846702,  
1.8405807701362846,  
-2.327837431672606,  
-3.4499572630907096,  
-0.617906997379805,  
-1.6298085443324908,  
7.351332839156953,  
5.290420571528784,  
-2.482388494410543,  
-5.728743066330026,  
-5.14374758488384,  
-4.766847094613013,  
-1.0582090092207181,  
-1.843388539396301,  
1.5296116933657231,  
-3.702644029125395,  
-0.8548564558182079,  
-0.4141909643155941,  
3.6749908813001184,  
-1.5935844893169815,  
3.2690802736485693,  
-1.525667808459957,  
-3.8372129261372265,  
8.609409860328014,  
5.442056155428872,  
3.6255647097129895,  
-0.5216191363074358,  
-1.7741769568172003,  
4.266365911686535,  
1.3113039692152597,  
2.9271911055545843,  
-9.55418464150057,  
2.939713998454181,  
3.835877955664955,  
1.795050718139919,  
-1.5078953672409625,

```
-3.321112564118266,
-1.2768307331105007,
6.786317686119446,
6.519074595177244,
-2.2676867570660306,
-7.4977323188611535,
-2.6801704085519304,
-0.9141147479870142,
-2.7177831473800325,
1.2572932175523874,
-1.946285818334104,
5.30793337884063,
-2.508332300471693,
3.2754081281925664,
-0.4231416788836668,
-1.592305845644887,
-3.254578253211548,
4.663643966288191,
-2.5118992856416753,
1.2787643614508397,
-5.437355736724157,
0.23835160978160275,
2.7610976050143563,
-1.4846023788859384,
...]
```

```
In [193]: sns.kdeplot(test_sample)
```

```
Out[193]: <AxesSubplot:ylabel='Density'>
```



```
In [194]: np.mean(test_sample)
```

```
Out[194]: 0.04919490209908773
```

```
In [195]: np.std(test_sample)
```

```
Out[195]: 3.9153546977783518
```

significant level =  $0.05 \times 1 = (z \times \sigma) + \text{mean}$

```
In [196]: from scipy.stats import norm
ucv=(norm.ppf(0.975)*np.std(test_sample)+np.mean(test_sample))
lcv=(norm.ppf(0.025)*np.std(test_sample)+np.mean(test_sample))
```

```
In [197]: print('lower critical value',lcv)
print('upper critical value',ucv)
print('observed_mean_diff',observed_mean_diff)
```

```
lower critical value -7.624759292246192
upper critical value 7.723149096444366
observed_mean_diff 0.49601665918658
```

observed mean difference 0.49 is within the range of acceptance region [lcv, ucv]. So Null hypothesis is accepted

```
In [198]: x=data.groupby(['trip_uuid', 'source_center']).agg({'osrm_time':'max', 'segment_osrm_time':'sum'})
x
```

Out[198]:

	trip_uid	source_center	osrm_time	segment_osrm_time
	trip-153671041653548748	IND209304AAA	349.0	534.0
		IND462022AAA	394.0	474.0
	trip-153671042288605164	IND561203AAB	26.0	26.0
		IND572101AAA	42.0	39.0
	trip-153671043369099517	IND000000ACB	212.0	231.0
	...	...	...	...
	trip-153861115439069069	IND628204AAA	41.0	42.0
		IND628613AAA	48.0	77.0
		IND628801AAA	14.0	14.0
	trip-153861118270144424	IND583119AAA	42.0	42.0
		IND583201AAA	26.0	25.0

26309 rows × 2 columns

In [199..

```
x=data.groupby(['trip_uid','source_center']).agg({'actual_time':'max','osrm_time':'max'})
x = x.groupby('trip_uid').agg({'actual_time':'sum','osrm_time':'sum'}).reset_index()
x.rename(columns={'actual_time':'actual_time_aggregated','osrm_time':'osrm_time_aggregated'},inplace=True)
x
```

Out[199]:

	trip_uid	actual_time_aggregated	osrm_time_aggregated
0	trip-153671041653548748	1562.0	743.0
1	trip-153671042288605164	143.0	68.0
2	trip-153671043369099517	3347.0	1741.0
3	trip-153671046011330457	59.0	15.0
4	trip-153671052974046625	341.0	117.0
...	...	...	...
14812	trip-153861095625827784	83.0	62.0
14813	trip-153861104386292051	21.0	12.0
14814	trip-153861106442901555	282.0	54.0
14815	trip-153861115439069069	264.0	184.0
14816	trip-153861118270144424	275.0	68.0

14817 rows × 3 columns

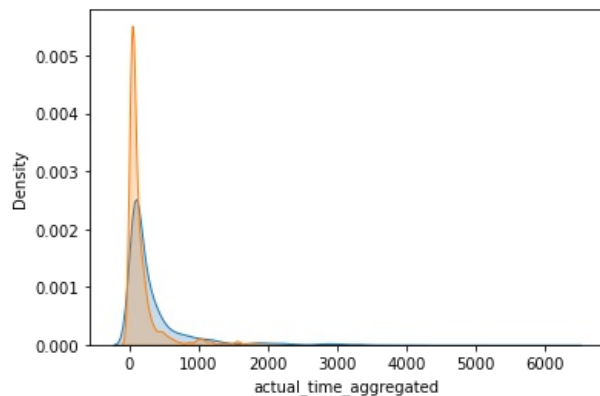
## Hypothesis Testing

In [200..

```
sns.kdeplot(x['actual_time_aggregated'],fill=True)
sns.kdeplot(x['osrm_time_aggregated'],fill=True)
```

Out[200]:

<AxesSubplot:xlabel='actual\_time\_aggregated', ylabel='Density'>



In [201..

```
observed_mean_diff = np.mean(x['actual_time_aggregated']) - np.mean(x['osrm_time_aggregated'])
observed_mean_diff
```

Out[201]:

194.9047040561517

In [202..

```
p = np.concatenate((x['actual_time_aggregated'], x['osrm_time_aggregated']))
p
```

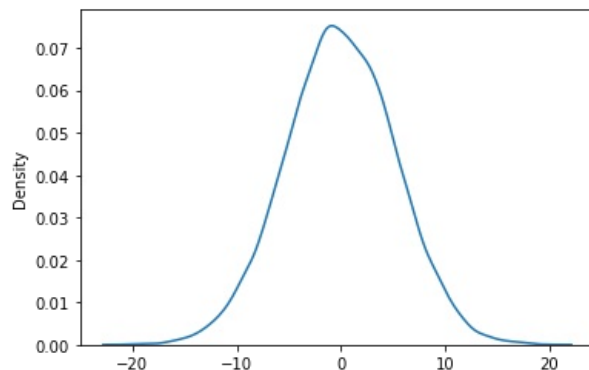
Out[202]:

array([1562., 143., 3347., ..., 54., 184., 68.]

```
In [203] test_sample = []
itr = 10000
for i in range(itr):
    np.random.shuffle(p)
    new_c1 = p[:14817]
    new_c2 = p[14817:]
    mean_diff = np.mean(new_c1) - np.mean(new_c2)
    test_sample.append(mean_diff)
```

```
In [204] sns.kdeplot(test_sample)
```

```
Out[204]: <AxesSubplot:ylabel='Density'>
```



significant level = 0.05  $x_1 = (z \cdot \sigma) + \text{mean}$

```
In [205] np.mean(test_sample)
```

```
Out[205]: -0.03988224336910284
```

```
In [206] np.std(test_sample)
```

```
Out[206]: 5.258359105296313
```

```
In [207] ucv = (norm.ppf(0.975)*np.std(test_sample) + np.mean(test_sample))
lcv = (norm.ppf(0.025)*np.std(test_sample) + np.mean(test_sample))
```

```
In [208] print('lower critical value', lcv)
print('upper critical value', ucv)
print('observed_mean_diff', observed_mean_diff)
```

```
lower critical value -10.34607670752814
upper critical value 10.266312220789931
observed_mean_diff 194.9047040561517
```

### 3.4

Hypothesis testing between actual time aggregated value and segment actual time aggregated value

```
In [209] x=data.groupby(['trip_uuid','source_center']).agg({'actual_time':'max','segment_actual_time':'sum'})
x
```

```
Out[209]:
```

	trip_uuid	source_center	actual_time	segment_actual_time
trip-153671041653548748	IND209304AAA		732.0	728.0
		IND462022AAA	830.0	820.0
trip-153671042288605164	IND561203AAB		47.0	46.0
		IND572101AAA	96.0	95.0
trip-153671043369099517	IND000000ACB		611.0	608.0
trip-153861115439069069	IND628204AAA		51.0	49.0
		IND628613AAA	90.0	89.0
		IND628801AAA	30.0	29.0
trip-153861118270144424	IND583119AAA		233.0	233.0
		IND583201AAA	42.0	41.0

26309 rows × 5 columns

```
In [210] x=data.groupby(['trip_uuid','source_center']).agg({'actual_time':'max','segment_actual_time':'sum'})
x=x.groupby(['trip_uuid']).agg({'actual_time':'sum','segment_actual_time':'sum'}).reset_index()
```

```
x.rename(columns={'actual_time':'actual_time_aggregated','segment_actual_time':'segment_actual_time_aggregated'})
x
```

```
Out[210]:
```

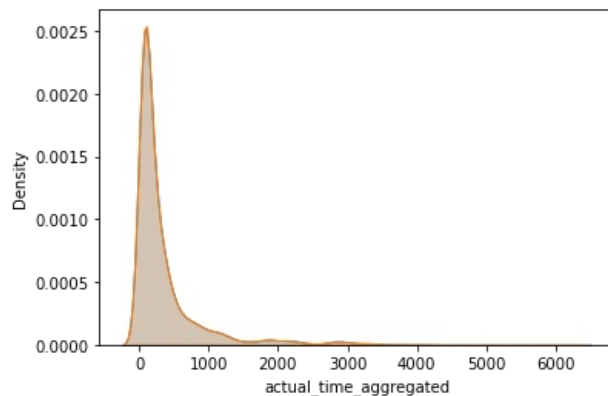
	trip_uid	actual_time_aggregated	segment_actual_time_aggregated
0	trip-153671041653548748	1562.0	1548.0
1	trip-153671042288605164	143.0	141.0
2	trip-153671043369099517	3347.0	3308.0
3	trip-153671046011330457	59.0	59.0
4	trip-153671052974046625	341.0	340.0
...	...	...	...
14812	trip-153861095625827784	83.0	82.0
14813	trip-153861104386292051	21.0	21.0
14814	trip-153861106442901555	282.0	281.0
14815	trip-153861115439069069	264.0	258.0
14816	trip-153861118270144424	275.0	274.0

14817 rows × 3 columns

## Hypothesis Testing

```
In [211]: sns.kdeplot(x['actual_time_aggregated'],fill=True)
sns.kdeplot(x['segment_actual_time_aggregated'],fill=True)
```

```
Out[211]: <AxesSubplot:xlabel='actual_time_aggregated', ylabel='Density'>
```



```
In [212]: observed_mean_diff = np.mean(x['actual_time_aggregated']) - np.mean(x['segment_actual_time_aggregated'])
observed_mean_diff
```

```
Out[212]: 2.922588918134579
```

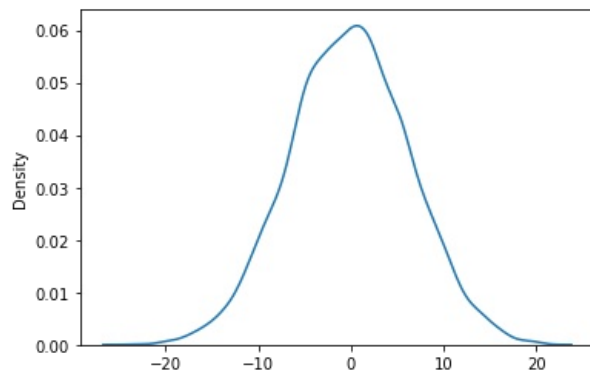
```
In [213]: p = np.concatenate((x['actual_time_aggregated'], x['segment_actual_time_aggregated']))
p
```

```
Out[213]: array([1562., 143., 3347., ..., 281., 258., 274.])
```

```
In [214]: test_sample = []
itr = 10000
for i in range(itr):
    np.random.shuffle(p)
    new_c1 = p[:14817]
    new_c2 = p[14817:]
    mean_diff = np.mean(new_c1) - np.mean(new_c2)
    test_sample.append(mean_diff)
```

```
In [215]: sns.kdeplot(test_sample)
```

```
Out[215]: <AxesSubplot:ylabel='Density'>
```



significant level = 0.05

$x1 = (z * \sigma) + \text{mean}$

```
In [216.. np.mean(test_sample)
```

```
Out[216]: -0.1490087197138418
```

```
In [217.. np.std(test_sample)
```

```
Out[217]: 6.44271832289436
```

```
In [218.. ucv = (norm.ppf(0.975)*np.std(test_sample) + np.mean(test_sample))
lcv = (norm.ppf(0.025)*np.std(test_sample) + np.mean(test_sample))
```

```
In [219.. print('lower critical value', lcv)
print('upper critical value', ucv)
print('observed_mean_diff', observed_mean_diff)
```

```
lower critical value -12.77650459512309
upper critical value 12.478487155695403
observed_mean_diff 2.922588918134579
```

observed mean difference 2.92 is in the range of acceptance region [lcv, ucv]. So Null hypothesis is accepted

3.5

Hypothesis testing between osrm time aggregated value and segment osrm time aggregated value

```
In [220.. x=data.groupby(['trip_uuid', 'source_center']).agg({'osrm_time': 'max', 'segment_osrm_time': 'sum'})
x
```

Out[220]:

		osrm_time	segment_osrm_time
trip_uuid		source_center	
trip-153671041653548748	IND209304AAA	349.0	534.0
	IND462022AAA	394.0	474.0
trip-153671042288605164	IND561203AAB	26.0	26.0
	IND572101AAA	42.0	39.0
trip-153671043369099517	IND000000ACB	212.0	231.0
...		...	...
trip-153861115439069069	IND628204AAA	41.0	42.0
	IND628613AAA	48.0	77.0
	IND628801AAA	14.0	14.0
trip-153861118270144424	IND583119AAA	42.0	42.0
	IND583201AAA	26.0	25.0

26309 rows × 2 columns

In [221..

```
x=data.groupby(['trip_uuid','source_center']).agg({'osrm_time':'max','segment_osrm_time':'sum'})
x=data.groupby(['trip_uuid']).agg({'osrm_time':'sum','segment_osrm_time':'sum'}).reset_index()
x.rename(columns={'osrm_time':'osrm_time_aggregated','segment_osrm_time':'segment_osrm_time_aggregated'},inplace=True)
```

Out[221]:

	trip_uuid	osrm_time_aggregated	segment_osrm_time_aggregated
0	trip-153671041653548748	7787.0	1008.0
1	trip-153671042288605164	210.0	65.0
2	trip-153671043369099517	65768.0	1941.0
3	trip-153671046011330457	24.0	16.0
4	trip-153671052974046625	207.0	115.0
...		...	...
14812	trip-153861095625827784	148.0	62.0
14813	trip-153861104386292051	19.0	11.0
14814	trip-153861106442901555	134.0	88.0
14815	trip-153861115439069069	446.0	221.0
14816	trip-153861118270144424	106.0	67.0

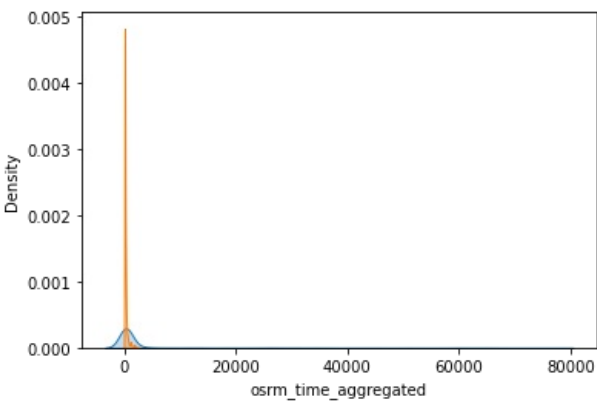
14817 rows × 3 columns

Hypothesis Testing

In [222..

```
sns.kdeplot(x['osrm_time_aggregated'],fill=True)
sns.kdeplot(x['segment_osrm_time_aggregated'],fill=True)
```

Out[222]:

<AxesSubplot:xlabel='osrm\_time\_aggregated', ylabel='Density'>

In [223..

```
observed_mean_diff = np.mean(x['osrm_time_aggregated']) - np.mean(x['segment_osrm_time_aggregated'])
observed_mean_diff
```

Out[223]:

1910.0575015185261

In [224..

```
p = np.concatenate((x['osrm_time_aggregated'], x['segment_osrm_time_aggregated']))
p
```

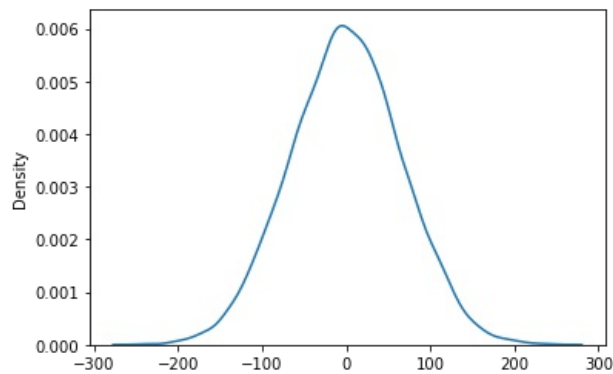
Out[224]:

array([ 7787., 210., 65768., ..., 88., 221., 67.])

```
In [225... test_sample = []
itr = 10000
for i in range(itr):
    np.random.shuffle(p)
    new_c1 = p[:14817]
    new_c2 = p[14817:]
    mean_diff = np.mean(new_c1) - np.mean(new_c2)
    test_sample.append(mean_diff)
```

```
In [226... sns.kdeplot(test_sample)
```

```
Out[226]: <AxesSubplot:ylabel='Density'>
```



significant level = 0.05

$x1 = (z * \sigma) + \text{mean}$

```
In [227... np.mean(test_sample)
```

```
Out[227]: -0.1372627927380704
```

```
In [228... np.std(test_sample)
```

```
Out[228]: 66.30969076065354
```

```
In [229... ucv = (norm.ppf(0.975)*np.std(test_sample) + np.mean(test_sample))
lcv = (norm.ppf(0.025)*np.std(test_sample) + np.mean(test_sample))
```

```
In [230... print('lower critical value', lcv)
print('upper critical value', ucv)
print('observed_mean_diff', observed_mean_diff)
```

```
lower critical value -130.1018685096074
upper critical value 129.82734292413127
observed_mean_diff 1910.0575015185261
```

observed mean difference not in the range of acceptance region [lcv, ucv]. So Null hypothesis is rejected

Hypothesis testing between osrm distance aggregated value and segment osrm distance aggregated value

```
In [231... x=data.groupby(['trip_uuid', 'source_center']).agg({'osrm_distance': 'max', 'segment_osrm_distance': 'sum'})
x
```

```
Out[231]:
```

	trip_uuid	source_center	osrm_distance	segment_osrm_distance
trip-153671041653548748	IND209304AAA		446.5496	670.6205
		IND462022AAA	544.8027	649.8528
trip-153671042288605164	IND561203AAB		28.1994	28.1995
		IND572101AAA	56.9116	55.9899
trip-153671043369099517	IND000000ACB		281.2109	317.7408
trip-153861115439069069	IND628204AAA		42.5213	42.1431
		IND628613AAA	40.6080	78.5869
		IND628801AAA	16.0185	16.0184
trip-153861118270144424	IND583119AAA		52.5303	52.5303
		IND583201AAA	28.0484	28.0484

26309 rows × 4 columns

```
In [232... x=data.groupby(['trip_uuid', 'source_center']).agg({'osrm_distance': 'max', 'segment_osrm_distance': 'sum'})
```



```
In [232]: x=data.groupby(['trip_uid', 'source_center']).agg({'osrm_distance': 'max', 'segment_osrm_distance': 'sum'})
x = x.groupby('trip_uid').agg({'osrm_distance': 'sum', 'segment_osrm_distance': 'sum'}).reset_index()
x.rename(columns={'osrm_distance': 'osrm_distance_aggregated', 'segment_osrm_distance': 'segment_osrm_distance_agg'}, inplace=True)
```

```
Out[232]:
```

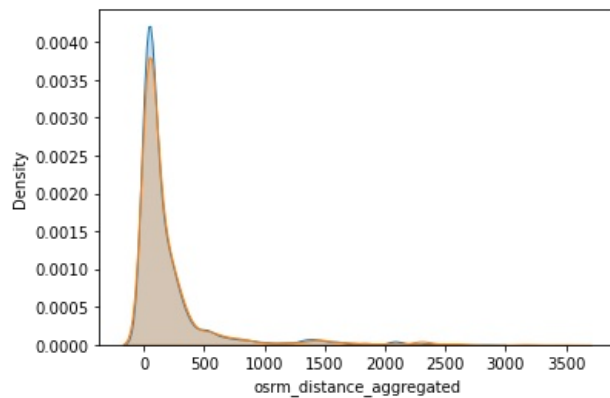
	trip_uid	osrm_distance_aggregated	segment_osrm_distance_aggregated
0	trip-153671041653548748	991.3523	1320.4733
1	trip-153671042288605164	85.1110	84.1894
2	trip-153671043369099517	2372.0852	2545.2678
3	trip-153671046011330457	19.6800	19.8766
4	trip-153671052974046625	146.7918	146.7919
...	...	...	...
14812	trip-153861095625827784	73.4630	64.8551
14813	trip-153861104386292051	16.0882	16.0883
14814	trip-153861106442901555	63.2841	104.8866
14815	trip-153861115439069069	177.6635	223.5324
14816	trip-153861118270144424	80.5787	80.5787

14817 rows × 3 columns

## Hypothesis Testing

```
In [233]: sns.kdeplot(x['osrm_distance_aggregated'], fill=True)
sns.kdeplot(x['segment_osrm_distance_aggregated'], fill=True)
```

```
Out[233]: <AxesSubplot:xlabel='osrm_distance_aggregated', ylabel='Density'>
```



```
In [234]: observed_mean_diff = np.mean(x['osrm_distance_aggregated']) - np.mean(x['segment_osrm_distance_aggregated'])
observed_mean_diff
```

```
Out[234]: -18.289453742322962
```

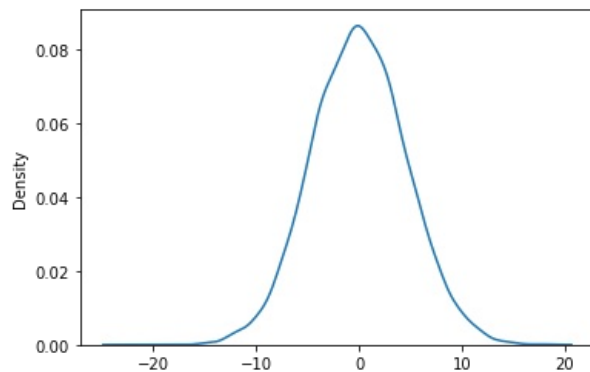
```
In [235]: p=np.concatenate((x['osrm_distance_aggregated'],x['segment_osrm_distance_aggregated']))
p
```

```
Out[235]: array([ 991.3523,  85.111 , 2372.0852, ..., 104.8866, 223.5324,
        80.5787])
```

```
In [236]: test_sample = []
itr = 10000
for i in range(itr):
    np.random.shuffle(p)
    new_c1 = p[:14817]
    new_c2 = p[14817:]
    mean_diff = np.mean(new_c1) - np.mean(new_c2)
    test_sample.append(mean_diff)
```

```
In [237]: sns.kdeplot(test_sample)
```

```
Out[237]: <AxesSubplot:ylabel='Density'>
```



significant level = 0.05

$x1 = (z \cdot \sigma) + \text{mean}$

```
In [238... np.mean(test_sample)
```

```
Out[238]: 0.0017765004467840497
```

```
In [239... np.std(test_sample)
```

```
Out[239]: 4.59056439778655
```

```
In [240... ucv = (norm.ppf(0.975)*np.std(test_sample) + np.mean(test_sample))
lcv = (norm.ppf(0.025)*np.std(test_sample) + np.mean(test_sample))
```

```
In [241... print('lower critical value', lcv)
print('upper critical value', ucv)
print('observed_mean_diff', observed_mean_diff)
```

```
lower critical value -8.995564387926658
upper critical value 8.999117388820224
observed_mean_diff -18.289453742322962
```

observed mean difference not in the range of acceptance region [lcv, ucv]. So Null hypothesis is rejected

3.7

Find outliers in the numerical variables and check it using visual analysis

```
In [242... df
```

Out[242]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destinati
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND3i
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND3i
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND3i
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND3i
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND3i
...	...	...	...	...	...	...	...	...
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0i
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0i
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0i
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0i
144866	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND0i

144867 rows × 24 columns

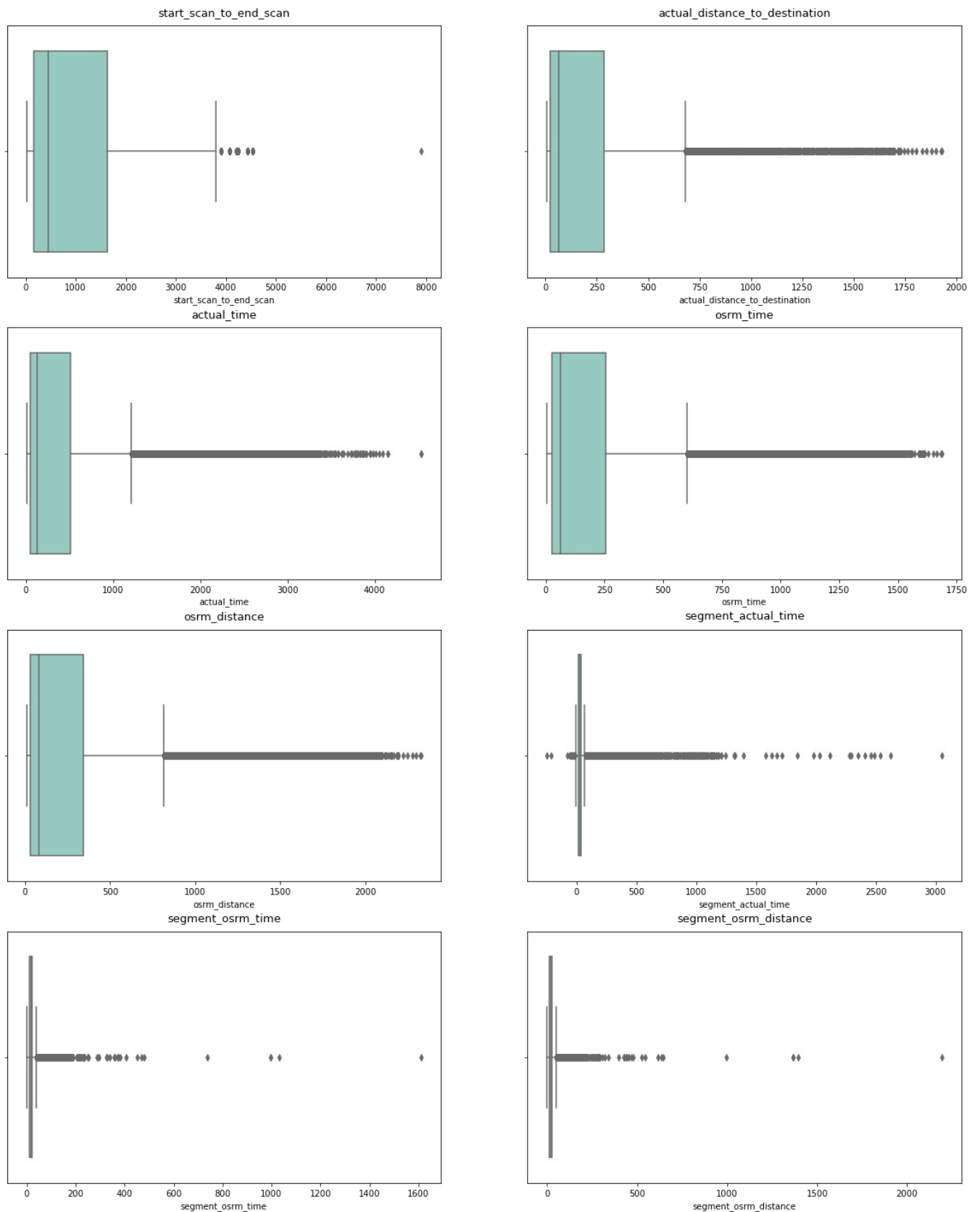
In [243... df.describe()

Out[243]:

	start_scan_to_end_scan	cutoff_factor	actual_distance_to_destination	actual_time	osrm_time	osrm_distance	factor
count	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000	144867.000000
mean	961.262986	232.926567	234.073372	416.927527	213.868272	284.771297	2.120107
std	1037.012769	344.755577	344.990009	598.103621	308.011085	421.119294	1.715421
min	20.000000	9.000000	9.000045	9.000000	6.000000	9.008200	0.144000
25%	161.000000	22.000000	23.355874	51.000000	27.000000	29.914700	1.604264
50%	449.000000	66.000000	66.126571	132.000000	64.000000	78.525800	1.857143
75%	1634.000000	286.000000	286.708875	513.000000	257.000000	343.193250	2.213483
max	7898.000000	1927.000000	1927.447705	4532.000000	1686.000000	2326.199100	77.387097

In [244...

```
attrs = ['start_scan_to_end_scan', 'actual_distance_to_destination', 'actual_time', 'osrm_time', 'osrm_distance']
fig, axs = plt.subplots(nrows=4, ncols=2, figsize=(20, 16))
fig.subplots_adjust(top=1.3)
count = 0
for row in range(4):
    for col in range(2):
        sns.boxplot(data=df, x=attrs[count], ax=axs[row, col], palette='Set3')
        axs[row,col].set_title(f"{attrs[count]}", pad=12, fontsize=13)
        count += 1
plt.show()
```



In [ ]:

```
In [245.. outlier=data.copy()
attrs = ['start_scan_to_end_scan', 'actual_distance_to_destination', 'actual_time', 'osrm_time', 'osrm_distance']
for i in attrs:
    q1=df[i].quantile(0.25)
    q3=df[i].quantile(0.75)
    iqr=q3 - q1
    lower = q1-(1.5)*iqr
    upper = q3+(1.5)*iqr
    print(f'The lower limit of outliers for {i} = {lower}')
    print('The upper limit of outliers for',i,'=',upper)
    print('-----')
    outlier=outlier[~((outlier[i]<lower)|(outlier[i]>upper))]
outlier
```

```

The lower limit of outliers for start_scan_to_end_scan = -2048.5
The upper limit of outliers for start_scan_to_end_scan = 3843.5
-----
The lower limit of outliers for actual_distance_to_destination = -371.6736259929169
The upper limit of outliers for actual_distance_to_destination = 681.7383749520162
-----
The lower limit of outliers for actual_time = -642.0
The upper limit of outliers for actual_time = 1206.0
-----
The lower limit of outliers for osrm_time = -318.0
The upper limit of outliers for osrm_time = 602.0
-----
The lower limit of outliers for osrm_distance = -440.0031250000001
The upper limit of outliers for osrm_distance = 813.1110750000001
-----
The lower limit of outliers for segment_actual_time = -10.0
The upper limit of outliers for segment_actual_time = 70.0
-----
The lower limit of outliers for segment_osrm_time = -5.5
The upper limit of outliers for segment_osrm_time = 38.5
-----
The lower limit of outliers for segment_osrm_distance = -11.544625
The upper limit of outliers for segment_osrm_distance = 51.427975
-----

```

Out[245]:

	data	trip_creation_time	route_schedule_uuid	route_type	trip_uuid	source_center	source_name	destination
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	Carting	153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388121AAA
...	...	...	...	...	...	...	...	...
144861	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	Carting	153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND131028AAB

114085 rows × 23 columns

## Categorical to Numerical encoding

One-hot encoding of categorical variables (route\_type)

One Hot Encoding

```
In [246]: one_hot_encoded_data = pd.get_dummies(outlier, columns=['route_type'])
one_hot_encoded_data
```

Out[246]:

	data	trip_creation_time	route_schedule_uuid	trip_uuid	source_center	source_name	destination_center
0	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB
1	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB
2	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB
3	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB
4	training	2018-09-20 02:35:36.476840	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	Anand_VUNagar_DC (Gujarat)	IND388620AAB
...	...	...	...	...	...	...	...
144861	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB
144862	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB
144863	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB
144864	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB
144865	training	2018-09-20 16:24:28.436231	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	Sonipat_Kundli_H (Haryana)	IND000000ACB

114085 rows × 24 columns

In [247]:

```
df_new=one_hot_encoded_data.drop(['data','trip_creation_time','source_name','destination_name','is_cutoff','cut
df_new
```

Out[247]:

	route_schedule_uuid	trip_uuid	source_center	destination_center	start_scan_to_end_scan	actual_distance_to_destina
0	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	IND388620AAB	86.0	10.43
1	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	IND388620AAB	86.0	18.93
2	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	IND388620AAB	86.0	27.63
3	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	IND388620AAB	86.0	36.11
4	thanos::sroute:eb7bfc78-b351-4c0e-a951-fa3d5c3...	trip-153741093647649320	IND388121AAA	IND388620AAB	86.0	39.38
...	...	...	...	...	...	...
144861	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	IND000000ACB	427.0	37.40
144862	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	IND000000ACB	427.0	45.25
144863	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	IND000000ACB	427.0	54.09
144864	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	IND000000ACB	427.0	66.16
144865	thanos::sroute:f0569d2f-4e20-4c31-8542-67b86d5...	trip-153746066843555182	IND131028AAB	IND000000ACB	427.0	73.68

114085 rows × 15 columns

In [249]:

```
from category_encoders import TargetEncoder
te = TargetEncoder()
```

```
cols=['route_schedule_uuid','trip_uuid','source_center','destination_center']
for col in cols:
    df_new[col] = te.fit_transform(df_new[col], df_new['segment_actual_time'])
df_new
```

```
Out[249]:
```

	route_schedule_uuid	trip_uuid	source_center	destination_center	start_scan_to_end_scan	actual_distance_to_destination	actual_tin
0	15.016514	24.705609	14.602140	14.602140	86.0	10.435660	14
1	15.016514	24.705609	14.602140	14.602140	86.0	18.936842	24
2	15.016514	24.705609	14.602140	14.602140	86.0	27.637279	40
3	15.016514	24.705609	14.602140	14.602140	86.0	36.118028	62
4	15.016514	24.705609	14.602140	14.602140	86.0	39.386040	68
...	...	...	...	...	...	...	...
144861	20.109914	25.735087	26.855453	28.812624	427.0	37.406091	81
144862	20.109914	25.735087	26.855453	28.812624	427.0	45.258278	94
144863	20.109914	25.735087	26.855453	28.812624	427.0	54.092531	120
144864	20.109914	25.735087	26.855453	28.812624	427.0	66.163591	140
144865	20.109914	25.735087	26.855453	28.812624	427.0	73.680667	158

114085 rows × 15 columns

Normalize/ Standardize the numerical features using MinMaxScaler or StandardScaler.

```
In [250]: from sklearn.preprocessing import StandardScaler, MinMaxScaler

scaler = StandardScaler()
std_data = scaler.fit_transform(df_new)
std_data = pd.DataFrame(std_data, columns=df_new.columns)
std_data
```

```
Out[250]:
```

	route_schedule_uuid	trip_uuid	source_center	destination_center	start_scan_to_end_scan	actual_distance_to_destination	actual_tin
0	-1.652946	-1.049382	-2.082120	-1.999254	-0.733058	-0.706326	-0.76448
1	-1.652946	-1.049382	-2.082120	-1.999254	-0.733058	-0.649305	-0.72454
2	-1.652946	-1.049382	-2.082120	-1.999254	-0.733058	-0.590947	-0.66063
3	-1.652946	-1.049382	-2.082120	-1.999254	-0.733058	-0.534063	-0.57275
4	-1.652946	-1.049382	-2.082120	-1.999254	-0.733058	-0.512143	-0.54878
...	...	...	...	...	...	...	...
114080	-0.995287	-0.750225	-0.121535	0.182547	-0.315850	-0.525424	-0.49686
114081	-0.995287	-0.750225	-0.121535	0.182547	-0.315850	-0.472756	-0.44493
114082	-0.995287	-0.750225	-0.121535	0.182547	-0.315850	-0.413500	-0.34107
114083	-0.995287	-0.750225	-0.121535	0.182547	-0.315850	-0.332535	-0.26118
114084	-0.995287	-0.750225	-0.121535	0.182547	-0.315850	-0.282115	-0.18928

114085 rows × 15 columns

## Recommendations and Insights

- There is null values in source\_name and destination\_name
- Here ,change trip\_creation\_time ,od\_start\_time,od\_end\_time , cutoff\_timestamp to date time format
- There is numerical and categorical values in the dataset.
- merging and split dataframe based on the trip\_uuid column
- Merging some rows based on the columns trip\_uuid, source ID and destination ID.
- More time taken for carting as compared to FTL
- Maharashtra is the most delivery source centre ,kerala and hariyana the lowest.
- punjab is the lowest delivery source city.
- Wednesday ,they are delivering more products.
- observed mean difference od od\_start \_time and od\_end\_time is 0.49 is within the range of acceptance region [lcv, ucv]. So Null hypothesis is accepted
- observed mean difference of actual\_time and segment\_time is 2.92 , it is in the range of acceptance region [lcv, ucv]. So Null hypothesis is accepted
- One-hot encoding of categorical variables is possible in route\_type.
- To reduce total delivering time ,we can reduce carting time.

