Feature Engineering

Jupyter Notebook link:

https://drive.google.com/file/d/1TjaIKOyQjeS_f01vofgsDzUPVyucCTYh/view?usp=sharing

Basic data exploration:

1. The shape of the data is (144867, 24)
2. Data type of the columns
   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

3. Unique values in the columns:
   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

4. Null values in the columns

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

Insights based on the Analysis:

- Mean of actual time is different from mean of segment osrm time.
- Mean of osrm distance is similar to mean of segment osrm distance.
- Mean of actual time is higher than mean of segment osrm time.
- Carting Transportation : 31%
- FTL Transportation : 69%
- Most orders are coming and going to same state ie. Maharashtra.
- Most orders are going to Mumbai.

- Most orders are coming from Bhindwandi_Mankoli_HB .
- Trip between Angamaly to Chalakudy saw the least avg time for completion.
- Trip between Hyderabad to Shamshabad saw the highest avg time for completion.
- The busiest route is between Bhindwandi_Mankoli_HB to Mumbai.

Recommendations:

- Since most orders are coming and going to Maharashtra , company have to expand the strategy used in Maharashtra to other states.
- Since the busiest route is between Bhindwandi_Mankoli_HB to Mumbai ,company can use more transportation in this route.
- Company have to analyse the transportation setup between Angamaly and Chalakudy which is the fastest route and setup the same producere to all other routes.
- Company have to analyse Hyderabad to Shamshabad route and make necessary changes to make the route fast.
- Proper route type for each route should be implemented.