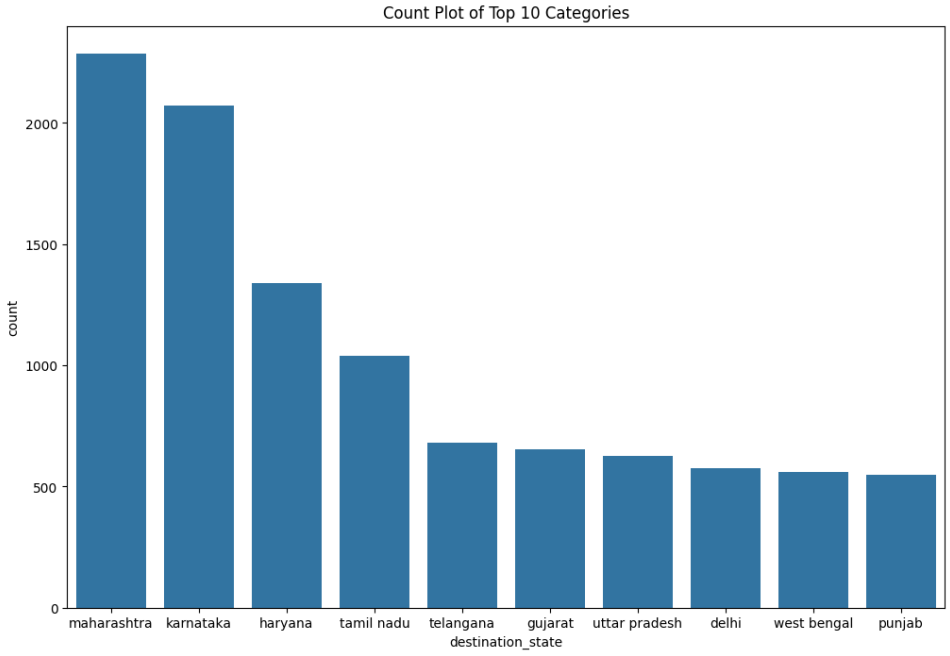
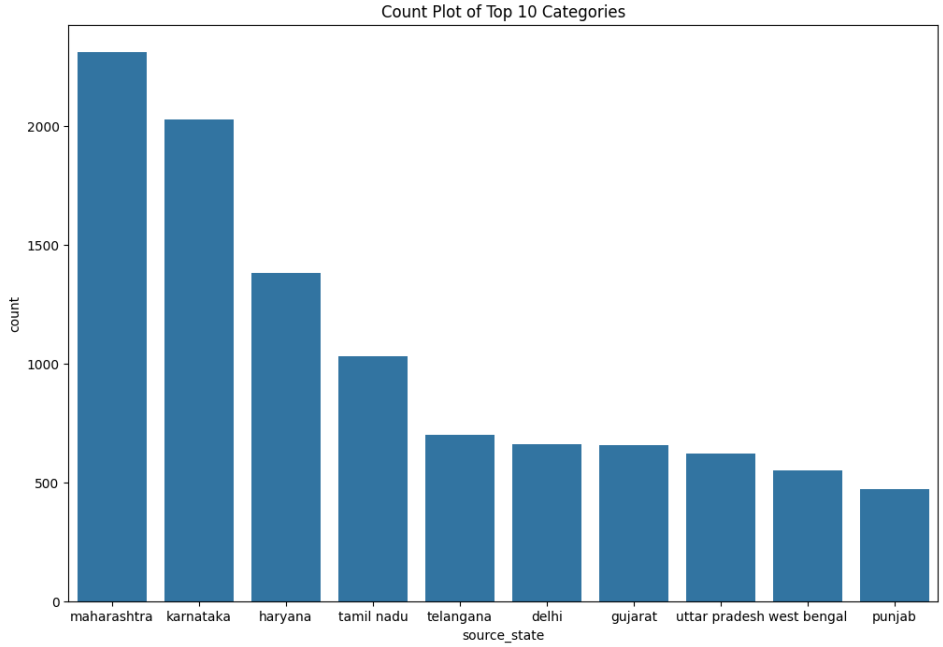
Problem Statement

- **Provide data** 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.
- To understand and **process the data** coming out of data engineering pipelines.
- **Clean, sanitize and manipulate data** to get useful features out of raw fields.
- **Make sense out of the raw data** and help the data science team to build forecasting models on it.

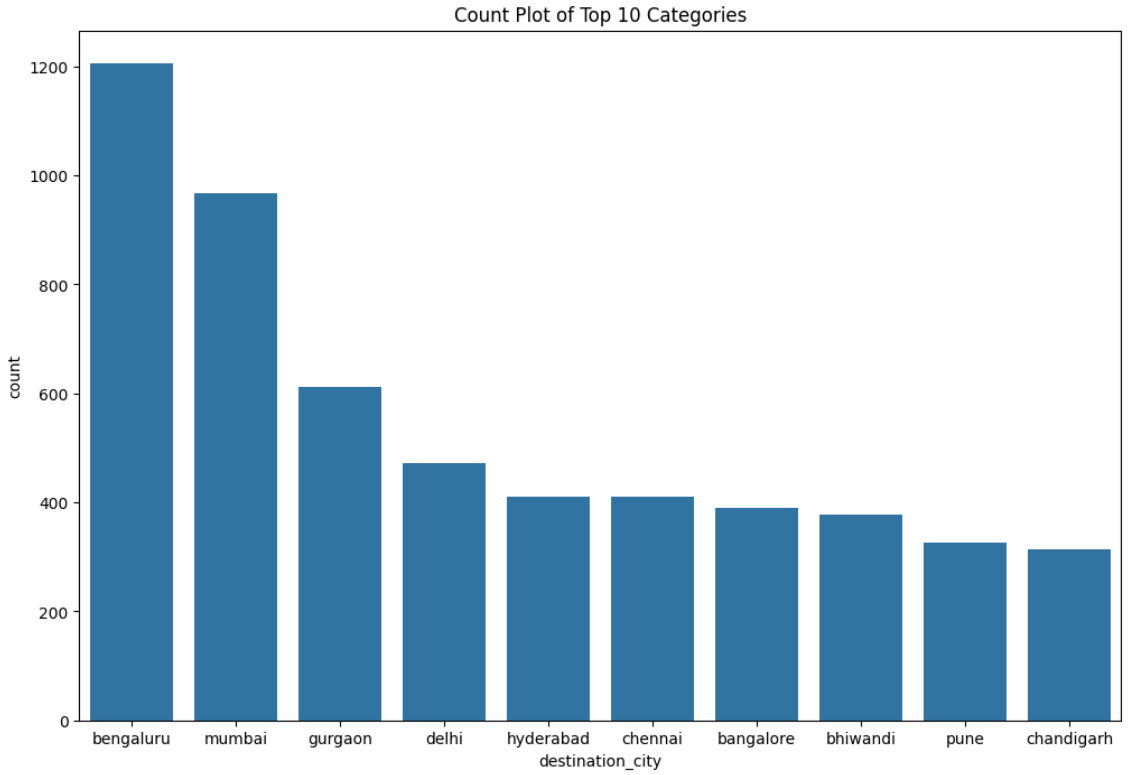
Business Insights & Recommendations

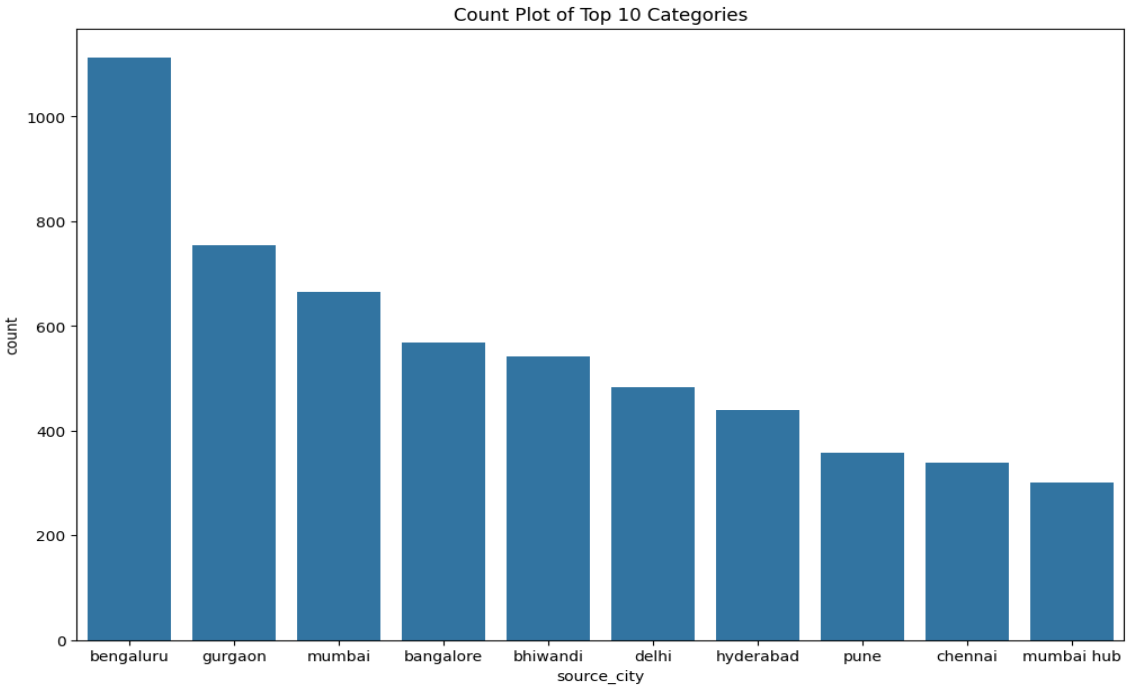
1. Maharastra seems to be the most active Source and Delivery State in India, below attached the top Source and Destination State in India :





2. Bengaluru seems to be the most active Source and Delivery City in India, below attached the top Source and Destination City in India :





3. Route Type: 8817 are Carting-1 and 33942 are of FTL -0 type.

```
route_type
1    8817
0    3942
Name: count, dtype: int64
```

4. Busiest Route:

Lowerparel is the busiest source city and
Mumbai being the busiest destination city

5. Average Time and Distance:

Average delivery time seems to be 42.67289719626168 hr,
Average distance seems to be 13.179664875398535 km

Recommendation:

1. Carting route type seems to have a lot more data as compare to FTL, so while building the ML model we have to consider the biasness in data.

2. Bengaluru seems to be our prime city and any updates in functional operations should be done considering the location.

3. After performing Hypothesis Testing, there is no significant difference and hence we can say that out model is predicting the output correctly for below mentioned parameters.

a) Actual time aggregated value Vs OSRM time aggregated value.
b) Actual time aggregated value Vs Segment actual time aggregated value.
c) OSRM distance aggregated value Vs Segment OSRM distance aggregated value.
d) OSRM time aggregated value Vs Segment OSRM time aggregated value.

<END>