

Delhivery : Feature Engineering

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.

The Data team builds intelligence and capabilities using this data that helps them to widen the gap between the quality, efficiency, and profitability of their business versus their competitors.

Column Profiling:

1. data - tells whether the data is testing or training data
2. trip_creation_time – Timestamp of trip creation
3. route_schedule_uuid – Unique ID for a particular route schedule
4. route_type – Transportation type
 - a. 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
 - b. Carting: Handling system consisting of small vehicles (carts)
5. trip_uuid - Unique ID given to a particular trip (A trip may include different source and destination centers)
6. source_center - Source ID of trip origin
7. source_name - Source Name of trip origin
8. destination_center – Destination ID
9. destination_name – Destination Name
10. od_start_time – Trip start time
11. od_end_time – Trip end time
12. start_scan_to_end_scan – Time taken to deliver from source to destination
13. is_cutoff – Unknown field
14. cutoff_factor – Unknown field
15. cutoff_timestamp – Unknown field
16. actual_distance_to_destination – Distance in kms between source and destination warehouse
17. actual_time – Actual time taken to complete the delivery (Cumulative)
18. 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)
19. 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)
20. factor – Unknown field
21. segment_actual_time – This is a segment time. Time taken by the subset of the package delivery
22. segment_osrm_time – This is the OSRM segment time. Time taken by the subset of the package delivery
23. segment_osrm_distance – This is the OSRM distance. Distance covered by subset of the package delivery
24. segment_factor – Unknown field

Objective

The company wants 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

Why this case study?

From Delhivery's Perspective:

1. Delhivery aims to establish itself as the premier player in the logistics industry. This case study is of paramount importance as it aligns with the company's core objectives and operational excellence.
2. It provides a practical framework for understanding and processing data, which is integral to their operations. By leveraging data engineering pipelines and data analysis techniques, Delhivery can achieve several critical goals.
3. First, it allows them to ensure data integrity and quality by addressing missing values and structuring the dataset appropriately.
4. Second, it enables the extraction of valuable features from raw data, which can be utilized for building accurate forecasting models.
5. Moreover, it facilitates the identification of patterns, insights, and actionable recommendations crucial for optimizing their logistics operations.
6. By conducting hypothesis testing and outlier detection, Delhivery can refine their processes and further enhance the quality of service they provide.

Business Insights:

- The data covers the period from September 12, 2018, to October 8, 2018 (26 days).
- 88% of the trips occurred in October, with the remaining in November.
- The data distribution is heavily right-skewed.
- Almost all features are strongly positively correlated, which is intuitive.
- Trips are slightly lower at the start and end of the month compared to the mid-month, though the difference is minor.
- No trips were recorded from the 4th to the 11th day of the month.
- Most orders are placed mid-month, indicating a higher volume of bookings during that period.

Route type:

The analysis shows that a larger share of shipments is routed via Full Truck Load (FTL) rather than carting. This has significant implications for both the efficiency and speed of the delivery process.

Geographical Focus:

Identifying the busiest routes and distances can aid in optimizing logistics operations, enhancing transportation efficiency, and potentially lowering costs.

State: Haryana, Maharashtra, and Karnataka are not only active source states but also stand out as the busiest, highlighting high demand or substantial business activity originating from these regions.

Source city: Gurgaon, Bangalore, and Bhiwandi have been identified as the busiest source cities, indicating their critical role in driving overall business operations and transportation activities.

Destination city: Gurgaon, Bangalore, and Hyderabad are identified as the busiest destination cities, underscoring their significance in terms of business activities or population movement.

Busiest corridor: Overall, the busiest corridor is Mumbai_Maharashtra and Bangalore_Karnataka which has the maximum trips.

- Average distance: 74.852844 kms
- Average time (in hours): 5.346577921457034

Delivery Time & Distance Accuracy:

1. OSRM Time vs. Actual Time:

- The difference between the mean values of estimated delivery time and actual delivery time suggests that there may be variations or delays in the actual delivery process compared to the initial estimates.
- The fact that the mean of OSRM time is less than the mean of actual delivery time indicates that the estimated times provided by the OSRM (Open Source Routing Machine) service tend to be optimistic.

2. OSRM Distance vs. Actual Distance:

- The mean of OSRM distance being greater than the mean of actual distance to the destination suggests that the OSRM might overestimate the distances. This could impact route planning and fuel efficiency calculations.

3. Segment-wise time Analysis:

- The equality in the mean values of actual time and segment actual time suggests that the time measurements are consistent across different segments of the delivery process

4. Segment-wise distance Analysis:

- The mean of segment OSRM distance being greater than the mean of OSRM distance implies that the OSRM might provide more conservative estimates for distance within individual segments.

5. Further look into :

- As its depicted from the analysis that there is absolutely no trip from 4th- 11th day of the month, The reason for that can be figured out and catered to receive the orders in the these dates as well.
- More ways to promote FTL route handling system can be implemented to increase this percentage

Business Recommendations:

1. Route Optimization:

- With Karnataka being the busiest state for transportation routes, optimizing the network within the state could enhance efficiency and reduce congestion. Implementing route optimization algorithms and real-time traffic monitoring would further improve the system.
- Given that Gurgaon and Bangalore are the busiest source and destination cities, respectively, city-specific strategies should be developed to effectively manage the high volume of traffic in these areas.

2. Operational Efficiency:

- Since the average OSRM time is lower than the actual delivery time, businesses can use this insight to set more realistic delivery time expectations for customers.
- As the average OSRM distance exceeds the actual distance, businesses should adjust their distance estimates to improve the accuracy of logistics planning.
- With the segment OSRM distance also greater than the overall OSRM distance, and actual distance data available, businesses can leverage this information to fine-tune route planning and optimize logistics for specific segments.
- Advanced demand forecasting techniques should be implemented to predict peak travel times, allowing businesses to adjust transportation services accordingly. This proactive approach will enable better resource allocation and help minimize congestion during busy periods.
- Overall, the analysis points to opportunities for operational improvement. Businesses can focus on refining their route planning, addressing discrepancies in time and distance estimates, and streamlining processes across various delivery stages to enhance overall efficiency.

3. Customer Satisfaction:

- Enhancing the accuracy of delivery time and distance estimates can lead to improved customer satisfaction.
- Full Truck Load (FTL) shipments contribute to faster delivery times, which positively impacts customer satisfaction. Customers value timely deliveries, and the focus on FTL shipments aligns with meeting or exceeding their expectations for delivery speed.

4. Customer Profiling:

- Profiling customers from Maharashtra, Karnataka, Haryana, Tamil Nadu, and Uttar Pradesh will provide insights into why a large volume of orders originates from these states and help businesses improve the buying and delivery experience for these customers.

5. Cost Optimization:

- Understanding the discrepancies between estimated and actual times and distances can support cost optimization efforts.
- Fine-tuning logistics planning based on more accurate data can lead to better resource allocation and potentially reduce operational costs.

6. Strategic Decision-Making:

- The preference for FTL over carting highlights a strategic decision by logistics management.
- Continuously evaluating the impact of this choice and understanding the reasons behind it can inform future decision-making and help businesses adapt to changing needs.

7. Collaboration with Stakeholders:

- Collaborating with stakeholders, including government authorities, transportation companies, and local communities, will be crucial in developing comprehensive strategies to manage and optimize transportation in the busiest corridors and cities.

Jupyter Notebook:

<https://drive.google.com/file/d/1iZOvPluHdhRnAJd-4bwHlj3vUC862mwk/view?usp=sharing>