

Data Analysis With Python Case Study – 2

Case Scenario:

A logistics business based in Delhi specialises in transporting items between different regions. The business interacts with other contractors for extra transportation services and has a fleet of delivery trucks. The company chooses to use the data on available delivery truck trips to optimize their trucking operations as part of their continual efforts to increase operational effectiveness and offer better customer service.

Data provided is pertaining to logistics and supply chain operations:

<https://www.kaggle.com/datasets/ramakrishnanthiyagu/delivery-truck-trips-data>

Variable description:

- GpsProvider - Vendor who provides GPS
- BookingID - Unique Identification for a trip
- Market/Regular - Type of trip.
 - Regular - Vendors with whom we will have contract.
 - Market - Vendor with whom we will not have contract
- BookingID_Date - Date when booking was created
- vehicle_no - Truck Number
- Origin_Location - Trip start place
- Destination_Location - Trip end place
- Org_lat_lon - Latitude/Longitude of start place
- Des_lat_lon - Latitude/Longitude of end place
- Data_Ping_time - Time when we receive GPS ping
- Planned_ETA - Planned Estimated Time of Arrival
- Current_Location - Live location
- DestinationLocation - Repeat of destination location
- actual_eta - Time when the truck arrived
- Curr_lat - current latitude - changes each time when we receive GPS ping
- Curr_lon - current longitude - changes each time when we receive GPS ping
- ontime - If the truck arrived on time - calculated based on Planned and Actual ETA
- delay - If the truck arrived with a delay - calculated based on Planned and Actual ETA
- OriginLocation_Code - Origin code
- DestinationLocation_Code - Destination code
- trip_start_date - Date/Time when trip started
- trip_end_date - Date/Time when trip ended - based on documentation (can't be considered for calculating delay)
- TRANSPORTATION_DISTANCE_IN_KM - Total KM of travel
- vehicleType - Type of Truck
- Minimum_kms_to_be_covered_in_a_day - Minimum KM the driver needs to cover in a day
- Driver_Name - Driver details
- Driver_MobileNo - Driver details
- customerID - Customer details

- customerNameCode - Customer details
- supplierID - Supplier who provides the vehicle
- supplierNameCode - Supplier who provides the vehicle
- Material Shipped – the product shipped to the customer

Problems:

- 1) Transform the booking data more effectively by extracting specific dates and times associated with each booking.
- 2) What is the frequency distribution of trips based on the destination location code?
- 3) Which vendors offer GPS information for more than 100 routes?
- 4) How many different materials are supplied by the supplier with ID 68018?
- 5) Who is the most common customer and supplier?
- 6) Calculate the range of covered distances that encompasses the majority of trucks.
- 7) What are the top 5 materials most frequently shipped, and how many shipments does each material category represent?
- 8) What type of vehicle was used to ship the highest number of materials?
- 9) How many users bought at least 50 different products?
- 10) Calculate the number of bookings made on 26th August 2020.
- 11) How many products were dispatched to the state of Karnataka?
- 12) How does the number of trips vary across different months?
- 13) What are the top 10 states with the highest number of vehicles registered, and what are their respective vehicle counts?
- 14) Who is the most common customer among the regular bookings in the dataset?
- 15) Does the choice of GPS provider significantly vary between regular and market bookings?
- 16) How does the total transportation distance vary across different vehicle types?
- 17) What are the top 10 GPS providers used, and what are their respective shares among the whole of GPS providers recorded?
- 18) What is the distribution of booking type?
- 19) How does the number of trips vary between regular and market bookings over time?