

Big Data - Deliverable 1

Group 21

1. Team:

a. Members:

1. Lakshmi Tulasi Bhavanam (801275438)
2. Ram Deepak Vinjam (801275255)
3. Jaswanth Tumati (801274978)
4. Rohith Sri Sai Mukkamala (801275202)
5. Jaswanth Yadla (801275432)

b. Communication plan to include project artifact repository

- (1) We will be having Zoom calls every week.
- (2) Github repository- https://github.com/developer-rohith/bigdata_project for discussing the project progress.

2. Business Problem:

Analysis of data is a potent modern technology that could greatly assist businesses in concentrating on the most crucial data from the information they have gathered about the shipping information of their items and associated product specifics. It unearths data-related information that reports, and queries are unable to fully expose.

This dataset contains information about a variety of products and their shipments, including the origin, days, carriers, and distance of the shipments. They intend to employ the data analysis to develop a solution to lessen the shipment arrival delay time and want to enhance the experience of customers based on the information.

This is the function of a business consultant employed by a logistics-related company. Predicting the shipment's expected arrival time in days is the objective.

3. Dataset:

Shipment dataset: The shipping-logs.csv dataset is one of the sample datasets from the Amazon SageMaker Canvas. This dataset includes detailed shipment

information for all delivered goods, such as expected shipping times, carriers, and origins.

This dataset includes all delivery-related shipment information, such as expected delivery time, shipping priority, carrier, and origin. It has approximately 10000 rows and 12 characteristic columns.

Data schema:

1. ActualShippingDays: Days required to deliver the shipment.
2. Carrier: Transporter for shipments
3. YShippingDistance: The shipment's distance along the Y axis
4. XShippingDistance: measured along the X-axis
5. ExpectedShippingDays: Anticipated shipping days
6. Origin: Origin of the shipment
7. OrderDate: Order placement date
8. OrderID: Order ID
9. ShippingPriority: Priority of On-Time Delivery
10. Delivery: revealed whether the package was delivered on schedule.
11. Product ID: Item Number
12. ComputerBrand: designates a brand of computer that is being shipped.
13. ComputerModel: A computer model's shipping status is indicated by the code.
14. ScreenSize: provided the screen size.
15. PackageWeight: The weight of the shipped package.

4. Research Objectives and Questions:

1. Predictions:
 - a. the time of arrival of the product.
 - b. the estimated delay of the shipment based on the location.
2. Analyze the shipment delays
 - a. How much is the average delay of each carrier?
 - b. The shipments from which area are mostly delayed?