Exploratory Analysis of Automative Supply Chain Data: Insights into Supplier Performance and Customer Trends

Mendoza, John Andrei D Jose Rizal University, 3rd Year College

Abstract— is an analytical workbook centered on car supply chain management data. The main data sheet, "CarSupplyChainManagement," houses detailed records, including supplier names, car models, pricing, shipping details, and customer feedback, likely gathered to provide insights into the operational metrics of the supply chain. Supplementary sheets include filtered and sorted views, which may be tailored to highlight specific segments or trends within the dataset. Conditional formatting is also applied to bring attention to particular records, possibly based on critical thresholds or notable values.

Additional sheets provide summary tables and statistical measures. Sheets labeled "Measures of Location" contain calculated averages, medians, and other descriptive statistics that likely serve as reference points for performance analysis. Quarterly sheets (Q1, Q2, Q3) focus on specific data breakdowns, such as car models by location, price ratings, and monthly sales distribution. Overall, the workbook appears designed to deliver both a detailed dataset and interpretive views, useful for understanding and optimizing supply chain processes for car manufacturers or suppliers.

Keywords—car supply chain management, supplier information, car models, pricing, shipping details, customer feedback, filtered data, sorted views, conditional formatting, measures of location, statistical measures, sales distribution, quarterly data, operational metrics, performance analysis, and dataset analysis.

Introduction

The Supply Chain Management for Car dataset, compiled by Prashant Kumar on Kaggle in 2019, provides a detailed examination of supply chain dynamics in the automotive sector. It contains data points related to supplier names, product categories, order quantities, and delivery schedules, offering a comprehensive view of inventory and logistics management. This dataset was designed to address key operational challenges, such as production delays and supplier inefficiencies, and aims to help businesses optimize their supply chain processes.

The dataset holds particular relevance due to the global disruptions that have impacted automotive supply chains in recent years, including shortages of essential components and logistical delays. With its detailed records, this dataset allows users to perform in-depth analyses of supplier performance, inventory control, and overall production planning. The data also provides a valuable foundation for predictive modeling,

allowing businesses to forecast potential disruptions and strategize solutions for enhancing efficiency and resilience.

Available in CSV format, the dataset is accessible to data scientists, analysts, and industry professionals who seek to derive insights from supply chain data. Its intended users can leverage the data for academic research, business process optimization, and decision-making purposes.

I. DOMAIN DESCRIPTION

The domain is about automotive supply chain management. It focuses on the logistics and data tracking processes involved in the distribution of car products from suppliers to end customers. This includes tracking suppliers, managing inventories, analyzing pricing strategies, monitoring shipping logistics, and gathering customer feedback to assess satisfaction and quality of service. Key metrics and statistical analyses are also part of the domain, serving to optimize various stages of the supply chain—like procurement, storage, and distribution.

The dataset aids decision-makers within the automotive industry by providing structured, data-driven insights into product flow, customer demand, cost management, and performance monitoring, all aimed at enhancing operational efficiency, improving customer satisfaction, and reducing costs across the supply chain.

II. PROBLEM STATEMENT (3 QUESTIONS)

- A. Are the patterns in purchasing behavior based on customer location?
- B. Are there correlations between customer feedback and factors like price?
 - C. How does the customer location(e.g State) impact delivery times and order fulfillment?

III. DATA COLLECTION

A. Dataset Description

The dataset in this file appears to capture extensive information relevant to car supply chain management. Key data attributes include:

- **Supplier Information**: Identifies sources for cars or car components, potentially tracking supplier names, locations, and service levels.
- Car Model Details: Provides specifics about each car model, including its make, model, year, and

- unique identifiers that distinguish it within the dataset.
- Pricing Data: Tracks costs associated with different car models, which may include base prices, supplier costs, and any additional shipping or handling charges.
- Shipping and Delivery Details: Contains information on shipment dates, delivery times, and other logistics details that are critical for tracking product movement from supplier to customer.
- Customer Feedback: Likely records customer satisfaction ratings, comments, or complaints, helping assess product quality and service effectiveness.

Additional sheets in the dataset offer filtered and sorted views, conditional formatting to highlight significant data points, and statistical measures (mean, median, etc.) to analyze and summarize trends. Quarterly summaries (Q1, Q2, Q3) focus on specific analyses, such as car models by location, price distribution, and sales trends over time. This dataset provides a comprehensive view of the car supply chain, allowing for in-depth analysis of logistics, customer satisfaction, and financial metrics.

B. Data Exploration

Data Dictionary

Attributes: SupplierID

Description: A unique numeric identifier assigned to each supplier. This helps differentiate between suppliers in the dataset.

Sample data: 1, 2, 3, 4, 5, 6, 7, 8, 9,

Attributes: SupplierAddress

Description: The full address of the supplier, including details like street name, city, and region. This is useful for logistical and shipping purposes.

Sample data: 542 Dayton Center, 0674 Springview Circle, 70 Autumn Leaf Center, 649 Corben Lane, 46347 Dunning Drive

Attributes: SupplierName

Description: The name of the supplier company or organization providing the products (cars).

Sample data: Bubbletube, Tagopia, Zoomdog, Oozz, Rhynyx, Roombo, Wordify

Attributes: SupplierContactDetails

Description: The contact information for the supplier, typically a phone number, used for communication regarding orders or shipments.

Sample data: 871-57-6028, 337-64-4060, 337-64-4060, 218-19-1802, 635-15-3112, 849-23-6788, 378-57-0118

Attributes: ProductID

Description: A unique identifier assigned to each product (in this case, a car). It helps in tracking and managing product inventory.

Sample data: 8993, 9444, 253, 1283, 8905, 8877, 101, 8854, 9043, 492

Attributes: CarMaker

Description: The manufacturer or brand of the car (e.g., Toyota, Ford, Dodge). This specifies which company produced the car.

Sample data: Dodge, Toyota, GMC, Volkswagen, Subaru, BMW, Mitsubishi, Nissan

Attributes: CarModel

Description: The specific model name of the car (e.g., Tundra, Ram 2500), which identifies the version and design of the car.

Sample data: Ram 2500, Tundra, Savana 1500, Cabriolet, Mariner, Land Cruiser, Imperza, X6

Attributes: CarColor

Description: The exterior color of the car (e.g., Goldenrod, Crimson), specifying the paint color.

Sample data: Goldenrod, Crimson, Fuscia, Teal, Indigo, Purple, Orange, Red, Yellow

Attributes: CarModelYear

Description: The year the car model was produced or released. It helps to identify the age of the car.

Sample data: 2007, 2010, 2011, 1990, 2009, 2002, 2008, 2013, 1996

Attributes: CarPrice

Description: CThe price of the car, usually listed before any discounts or promotions.

5ample data: 521965.45, 672222.04, 504465.72, 646077.11, 699890.24, 546977.55,

Attributes:CustomerID

Description: A unique identifier assigned to each customer, used to track their purchases and order history.

Sample data: 60760-224, 67457-594, 58411-135, 0591-5307, 51655-189, 65811-0001, 31722-328

Attributes: CustomerName

Description: The full name of the customer who made the purchase.

Sample data: Louie Hinsche, Alexandros Manuel, Alvie Weighell, Flint Gunston, Alyssa Filpi, Wiatt Piche,

Attributes: Gender

Description: The gender of the customer (e.g., Male, Female, Non-binary). This could be used for demographic analysis.

Sample data: Male, Female

Attributes: JobTitle

Description: The job title or occupation of the customer, which can give insight into their purchasing behavior

Sample data: Mechanical Systems Engineer, Structural Engineer, Systems Administrator III, Operator, Software Engineer III

Attributes: PhoneNumber

Description: The customer's phone number, which is used for communication regarding orders or deliveries.

Sample data: 907-966-3323, 320-943-9301, 312-561-8337, 407-830-3175, 313-870-7247, 404-590-7068

Α

Attributes: EmailAddress

Description: IThe customer's email address, used for order confirmation, customer service, and other communications.

amanuel1@usnews.com,
aweighell2@nature.com,
fgunston3@mozilla.org, owittman7@nps.gov,

Attributes: City

Description: The city where the customer resides.
This can help in understanding regional demand and logistics

Sample data: Chicago, Orlando, Detroit, Saint Cloud, San Jose, Washington, New York City

Attributes: Country

Description: The country in which the customer resides.

Sample data: United States,

Attributes: CountryCode

Description: A standardized code representing the country (e.g., US for the United States). This helps in international shipping and data management.

Sample data: US

Attributes: State

Description: The state or province in which the customer resides, helping to further pinpoint the customer's location.

Sample data: Alaska, Minnesota, Illinois, Florida, Michigan, Georgia, Arizona, Indiana

Attributes: CustomerAddress

Description: The full address of the customer, including street, city, and region, used for delivery purposes

purposes. Sample data: Tindependence Circle, 209 Lerdahl Alley, 1803 Evergreen Trail, 9682 Warbler Terrace, 5751 Tony Avenue, 26066 Valley Edge Parkway, 60797 Rowland Pass, 0

Attributes: OrderDate

Description: The date on which the customer placed the order. It provides insight into purchasing trends over time.

Sample data: 12/16/2018, 2/13/2019, 8/16/2018, 2/22/2019, 9/6/2018, 8/16/2018, 3/16/2019, 11/4/2018

Attributes: OrderID

Description: A unique identifier for each order placed by a customer. This is essential for tracking and referencing specific transactions.

Sample data: 0268-6694, 0699-5728, 42043-251, 57520-1044, 36987-3308, 11822-0580, 24909-111, 49527-013

Attributes: ShipDate

Description: The date when the product (car) was shipped to the customer or intermediary. It helps track the fulfillment process.

Sample data: 3/14/2019, 3/6/2019, 1/20/2019, 3/16/2019, 1/29/2019, 3/24/2019, 1/24/2019, 1/1/2019, 1/16/2019

Attributes: ShipMode

Description: The shipping class or method used for delivery (e.g., Standard Class, Express Delivery). This affects delivery speed and cost.

Sample data: Standard Class, First Class, Same Day, Second Class

Attributes: CreditCard

Description: The actual credit card number used for payment, often masked or partially hidden in records.

Sample data: 30408016042565,

3549221112237760, 3557159608180900, 3529909223663920, 5602235978541510, 6706247841149200, 30528106894581,

3589144247894730

Attributes: CustomerFeedback

Description: Feedback provided by the customer regarding their purchase experience, which can help in analyzing customer satisfaction.

Sample data: Good, Bad, Okay, Very Good, Very Bad

Attributes: Shipping

Description: The mode of transportation used to ship the product (e.g., Truck, Air), which influences the delivery time and cost.

Sample data: Truck, Air

Attributes: PostalCode

Description: The postal or ZIP code of the customer's delivery address. It helps with shipping logistics and ensuring accurate delivery.

Sample data: 99522, 56398, 60674, 32885, 48232, 31136, 2208, 25336, 85035, 95113

Attributes: Sales

Description: The total sales amount from the order, including all cars and possibly other fees or taxes.

Sample data: 744796.41, 794773.17, 968244.9, 942213.82, 879519.57, 947785.4, 883836.81, 779626.29

Attributes: Quantity

Description: The number of units of the product (cars) purchased in the transaction.

Sample data: 1, 2

Attributes: Discount

Description: The percentage of discount applied to the order. A value like 0.20 represents a 20% discount on the total price.

Sample data: 0.83, 0.79, 0.28, 0.76, 0.5, 0.39, 0.79

Attributes: CreditCardType

Description: The type of credit card used for the transaction (e.g., Visa, MasterCard, American Express).

Sample data: diners-club-carte-blanche, jcb, china-unionpay, laser, maestro, bankcard

C. Data preprocessing

Removing insignificant attributes



I first remove the attributes that are insignificant for the questions that I made.



Then I filtered the attribute($Car\ Maker)$ to just Honda $car\ makers$.

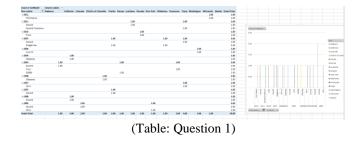


After that I sorted the attribute (Sales) from largest to smallest in order for me to identify what type of card model the customers often buy.



After that I implemented the conditional formatting in order to visualize the data easier. I then appointed the condition that when its greater than 857630.6 the text will be highlighted into a green color and if its less than 857630.6 it will be highlighted into the color red. I also implemented data bars to visualize the data way simpler.

IV. EXPERIMENTAL DESIGN PIVOT TABLE





(Table: Question 2)



(Table: Question 3)

V. RESULTS AND DISCUSSION

A. Are the patterns in purchasing behavior based on customer location?

The pivot table in Q1 organizes data by year and U.S. states, counting the availability of various car models. For instance, in 2011, models like the Accord and Accord Cross tour are noted in states such as Louisiana and Texas. This table could help determine where certain car models are concentrated or more popular, potentially revealing regional preferences or distribution strategies. Notably, the limited distribution of specific models in certain states (e.g., Accord only in Louisiana and Texas) suggests a focus on those areas, possibly influenced by factors like market demand or logistical considerations.

B. Are there correlations between customer feedback and factors like price?

In Q2, car models are rated as Good, Bad, Okay, Very Bad, and Very Good across various price points. For instance, models within the \$573,485 to \$641,469 price range are predominantly rated as Good and Very Bad, with some models rated as Very Good. This table indicates a possible relationship between price range and quality perception. Generally, there is no strict correlation, as some high-priced models still receive lower ratings, suggesting that price does not guarantee quality. However, there is a notable concentration of Good ratings in mid-range prices, which might imply that buyers see mid-range models as offering a balanced value for quality.

C. How does the customer location(e.g State) impact delivery times and order fulfillment?

In Q3, the distribution of car models is broken down by state and month, showing when models are released or available in certain areas. For instance, California shows models in February and April, while Alabama and Florida show entries in January.

This table reveals seasonal trends or timed availability of specific models, which could be strategically aligned with local demand peaks or marketing schedules. The availability concentrated in early months could point to a launch period strategy, targeting consumer demand right after the New Year or at the start of other seasonal sales cycles.

REFERENCES

Prashant K., "Supply Chain Management for Car Data," *Kaggle*, 2023. [Online]. Available: https://www.kaggle.com/datasets/prashantk93/supply-chain-management-for-car/data. [Accessed: Nov. 15, 2024].