



Spinny's Revenue Growth Through Demand and Pricing Optimisation

Introduction:

Spinny is an online used-car retailing platform that aims to bring about transparency and convenience in the car buying and selling process in India. Spinny operates across the entire value chain of pre-owned cars, embedding superior technology and processes to deliver a premium experience to customers. It has 36 car hubs that operate across 15 cities.

Spinny wants to increase revenue by boosting the number of completed transactions. To achieve this, it is essential to understand the demand for cars across different regions and identify the most in-demand cars and their attributes. Additionally, determining the right pricing for each car, based on its attributes and market demand, is crucial for staying competitive.

Objectives:

In this project our goal is to analyse used car transaction data to increase revenue by boosting the number of completed transactions. Some objectives of the project are:

Demand Analysis:

1. Assess the demand for pre-owned cars in each region and identify geographic patterns to target high-demand markets more effectively.
2. Identify the demand for specific car models and their key attributes (e.g., condition, fuel type, mileage).

Price Analysis:

1. Analyse how prices are distributed across different key attributes (e.g., condition, transmission etc.)
2. Analyse how price varies by manufacturer and model.

Revenue Analysis:

1. Identify how factors like car age, condition, odometer reading, and title status influence revenue.
2. Identify how revenue varies on basis of region and time.
3. Investigate revenue fluctuations over time and seasonal trends.

Business Impact:

This analysis will help **Spinny** to:

Identify regional demand:

1. Knowing which regions have maximum and minimum sale can help to formulate sales strategy.
2. It will also help Spinny in inventory management in such regions where demand is high.

Optimize price:

1. As per demand patterns Spinny can finalise competitive yet profitable prices for each car.
2. Avoiding under-pricing will help to prevent revenue loss, while controlling overpricing will help to reduce unsold inventory.

3. Offering fair and transparent price will help to build trust with customers.

Formulate marketing strategies:

1. Targeted promotions of less sold cars by offering discounts and cashbacks.
2. Focused marketing of highly sold cars.
3. Long-term market strategies can be formed on basis of car trends and pricing.

Dataset Overview:

Dataset Name: Spinny Analytics Dataset

Number of Rows: 426880

Number of Columns: 21

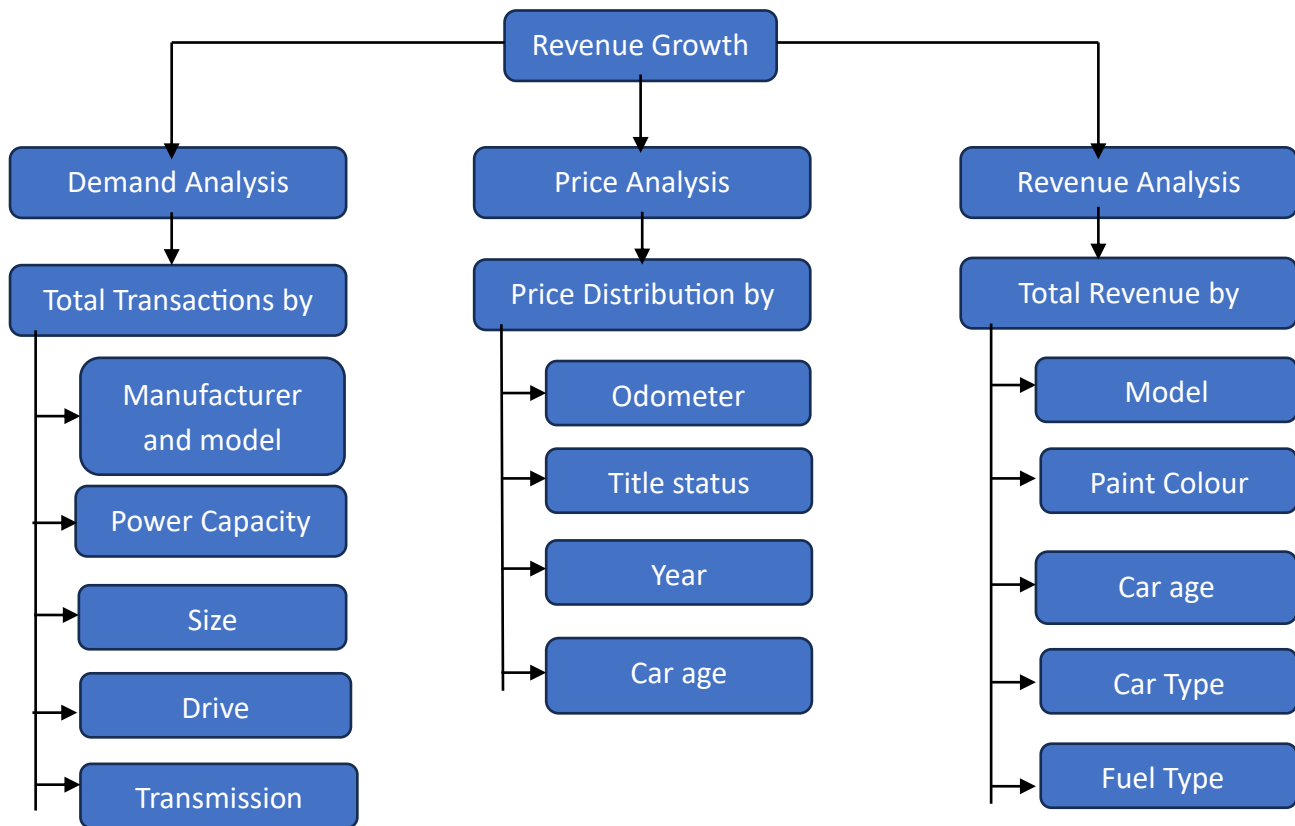
Description: The dataset contains key metrics about sales of pre-owned cars at Spinny, providing insights about sales trend and transaction patterns.

Data Columns Information:

1. **id (Transaction ID):** A unique identifier for each car transaction in the dataset.
2. **Price:** The final selling price of the vehicle in Indian Rupees (INR).
3. **Year:** The year of manufacture of the vehicle, indicating its age.
4. **Manufacturer:** The brand or company that produced the vehicle.
5. **Model:** The specific model's name of the vehicle produced by the manufacturer.
6. **Condition:** The physical and operational state of the vehicle at the time of sale.
7. **Cylinders:** The number of cylinders in the car's engine, influencing its power and performance.
8. **Fuel:** The type of fuel the vehicle uses, such as petrol, diesel, CNG, or electric.
9. **Odometer:** The total distance the vehicle has travelled, measured in kilometres.
10. **Title_status:** The condition of the car's, indicating if it is clean, or has missing parts or has liens.
11. **Transmission:** The transmission states whether the car is automatic, manual or of another category.
12. **VIN (Vehicle Identification Number):** A unique serial number assigned to each vehicle for identification.
13. **Drive:** The drivetrain configuration of the vehicle (e.g., front-wheel drive, rear-wheel drive).
14. **Size:** The size classification of the vehicle, such as compact, midsize, or full-size.
15. **Type:** The type of vehicle, such as sedan, SUV, hatchback, or truck.
16. **Paint_color:** The exterior colour of the vehicle.
17. **Image_url:** A link to an image of the vehicle for reference.
18. **State:** The Indian state where the car transaction took place.
19. **Posting_date:** The date when the car was listed for sale.
20. **Latitude:** The geographical latitude coordinate of the transaction location.
21. **Longitude:** The geographical longitude coordinate of the transaction location.

Key Metrics:

The dataset contains key metrics about sales of pre-owned cars at Spinny, providing insights about sales trend and transaction patterns.



Data Cleaning and Preparation:

Data cleaning part is very important for accurate analysis. We have 21 columns in the dataset, we will go through all the important columns and check datatype inconsistency, missing values, duplicates and outliers in the dataset and will take appropriate actions. All steps of data cleaning are listed below:

- **Data Importing:** Imported the data from google drive into google colab for analysis. Also imported important libraries such as Pandas, Numpy, Seaborn, Matplotlib.pyplot, Plotly.express.
- **Duplicates:** Checked for duplicate rows in the dataset using `duplicated().sum()` function. As there were no duplicate rows in the dataset no need to drop them.
- **Datatype:** Checked datatypes of all columns using `info()` function. Found inconsistency of data type in 'year' and 'posting_date' column.
 - ❖ Converted 'year' column data type from float to integer using `pd.to_numeric()` function.
 - ❖ Converted 'posting_date' column data type from object to timestamp using `pd.to_datetime()` function.
- **Outliers in Price column:**
 - ❖ Checked outliers in price column using IQR method. There were 34868 values as 0 price, that could mislead the analysis hence we replaced them with NaN using `replace()` function.

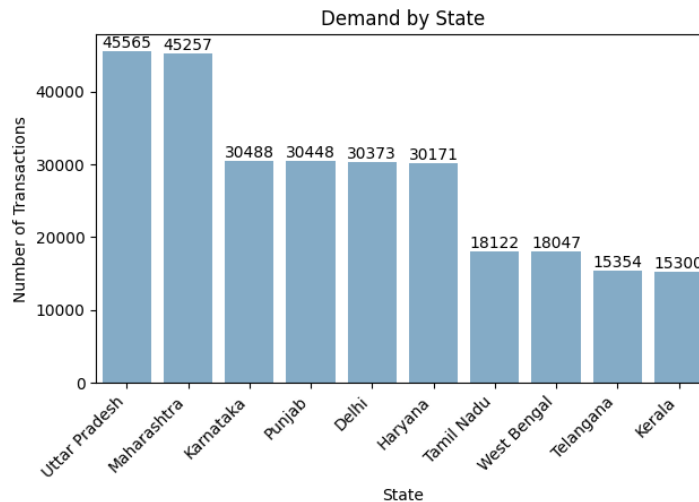
- ❖ We checked outliers in price using IQR method considering 90th - 10th percentile, Since 90-10 percentile consideration includes many values that are important for analysis hence instead of 90-10 we will consider 99-01 percentile values for imputation.
 - ❖ We imputed price values above 5 cr with 99th percentile value (considering 5 cr as upper cap for luxury vehicles). Also, we imputed price values below 01th percentile with 01th percentile values.
 - **Outliers in Odometer column:**
 - ❖ We checked outliers using IQR method considering 99-01 percentile.
 - ❖ Since many values above 99th percentile was important for analysis, we imputed values above 20,00,000 km as upper cap (maximum limit of car odometer reading) with upper bound value.
 - **Missing Values:** Used `isna().sum()` function to check missing values in each column. There were many columns with missing values which we handled accordingly.
 - ❖ There were very less missing values in `posting_date` column so we dropped them using `dropna()` function.
 - ❖ 'Year' and 'odometer' column had 1205 and 4400 missing values respectively we replaced them with median value of that particular column.
 - ❖ 'Fuel', 'title_status' and 'Transmission' column had few missing values that we replaced with most frequent column of those columns only.
 - ❖ 'Manufacturer', 'model', 'condition', 'cylinders', 'drive', 'size', 'type' and 'paint_color' column had large number of missing values imputing them could have screwed the data, so to handle them we created a different category as 'unknown' and replaced missing values with it.
 - **Drop column:** As 'image_url' and 'VIN' column is not being used in the analysis we are dropping this column using `drop()` function.
 - **Creating New Column:** Created two new columns for better analysis.
 - ❖ 'Car_power_category' created using 'cylinders' column. On basis of number of cylinders, we categorised vehicles as Ultra high power, high power, moderate power, low power category.
 - ❖ 'car_age' column is created using 'year' column. On basis of manufacturing year of vehicle, we categorised their age as New, Relatively new, Medium age, Old, Very old, and Vintage.
 - ❖ 'posting_day' column created using 'posting_date' column extracting date from timestamp using date function.
 - **Copy dataset:** Copied dataset for further analysis using `copy()` function.
-

Exploratory Data Analysis:

Exploratory Data Analysis (EDA) is essential for understanding the dynamics of the Spinny analytics dataset. The goal of EDA here is to uncover insight demand, price distribution and revenue with respect to various attributes given in the data set. Detailed analysis of the dataset is given below:

Demand Analysis:

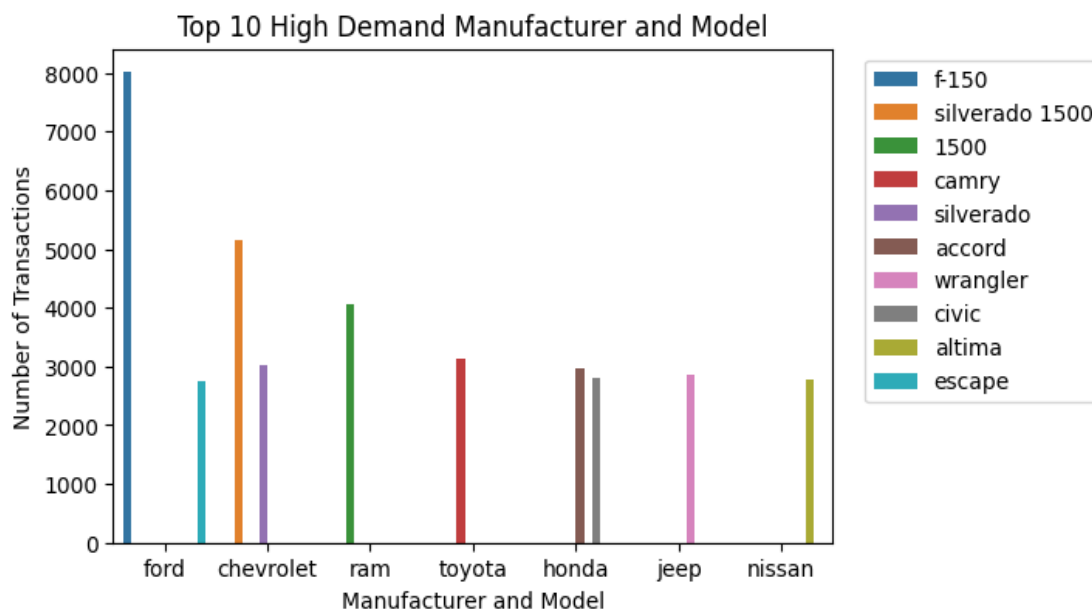
Hypothesis 1: Certain regions show higher demand for pre-owned vehicles compared to others.



Observation: Uttar Pradesh, Maharashtra, Karnataka, Punjab, Delhi and Haryana are the states where the demand for pre-owned vehicles is too high as compared to other states and UTs.

Recommendations: Prioritize stocking a larger inventory of popular pre-owned vehicles in Uttar Pradesh, Maharashtra, Karnataka, Punjab, Delhi, and Haryana. Set up additional sales hubs in these states to enhance customer access. For states with low demand launch targeted marketing campaign (e.g., SUVs in Punjab or compact cars etc).

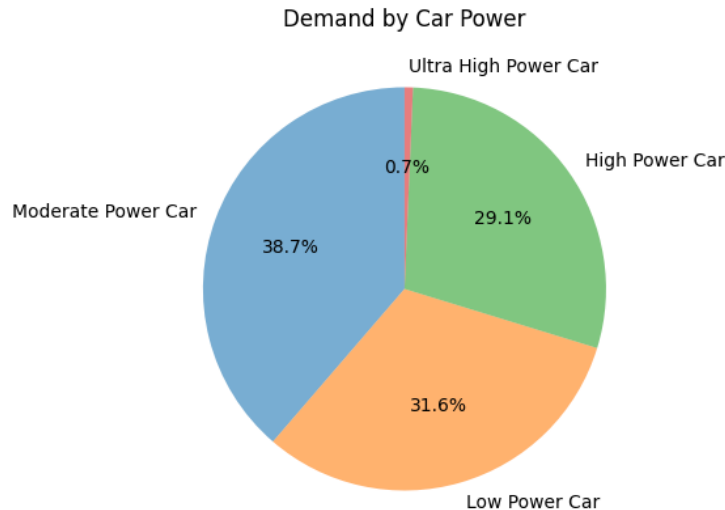
Hypothesis 2: Certain 'manufacturer' and 'model' are high in demand than others.



Observation: Top models and manufacturer are Ford (f-150, escape), Chevrolet (Silverado 1500, Silverado), Ram (1500) etc.

Recommendations: Seeing the high demand of the Ford (f-150, escape), Chevrolet (Silverado 1500, Silverado), Ram (1500) etc. Spinny should focus on sustaining sales of these cars by fulfilling demand. Also, cars with low demand should be promoted by providing extra offers.

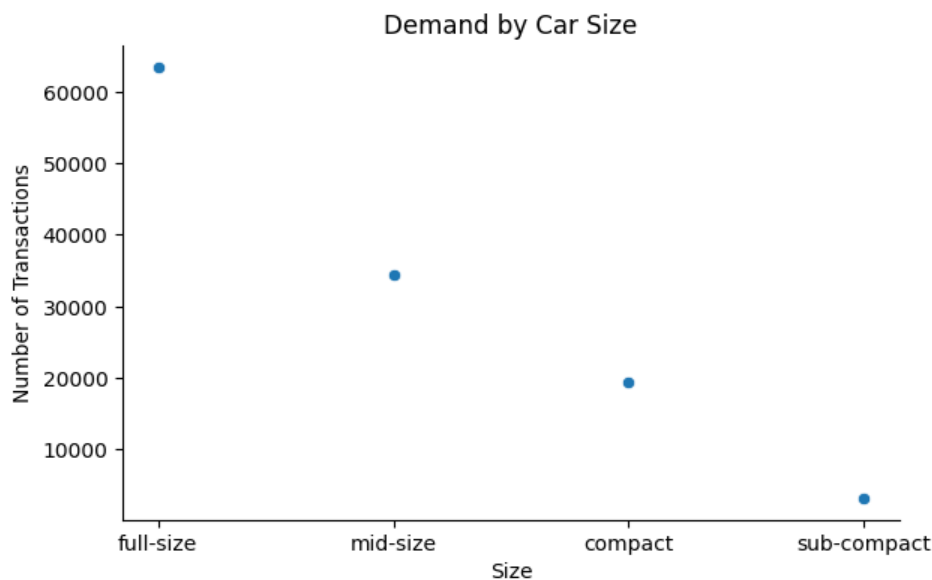
Hypothesis 3: Power capacity of vehicles influences the demand. People usually prefer moderate power cars.



Observation: Moderate Power Cars (38.7%) are high in demand followed by Low Power Cars (31.6%), may be because of all time uses and cost effectiveness. Ultra-High Power Cars have very low demand (0.7%).

Recommendations: Spinny should assure the supply of Moderate and Low Power Cars as per demand. Ultra-high-power cars should be promoted among the specific customers.

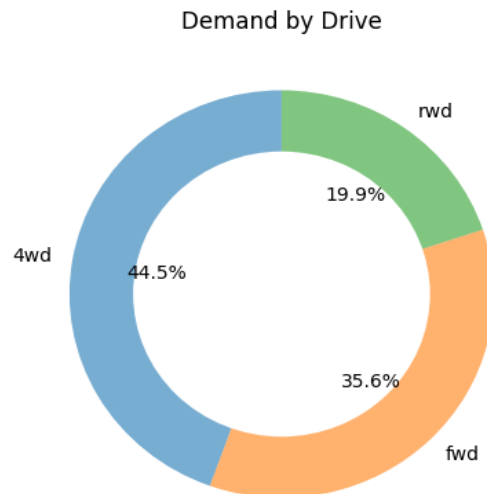
Hypothesis 4: Customers are more likely to buy large size vehicles rather than small vehicles.



Observation: Full size and Mid-size vehicles are highly in demand whereas sub-compact vehicles are the least preferred by the customers.

Recommendations: Focus should be on Full and Mid-sized vehicles for procurement, advertisement, marketing and supply. Compact and sub-compact vehicles can be promoted among the target customers such as individual users, students etc.

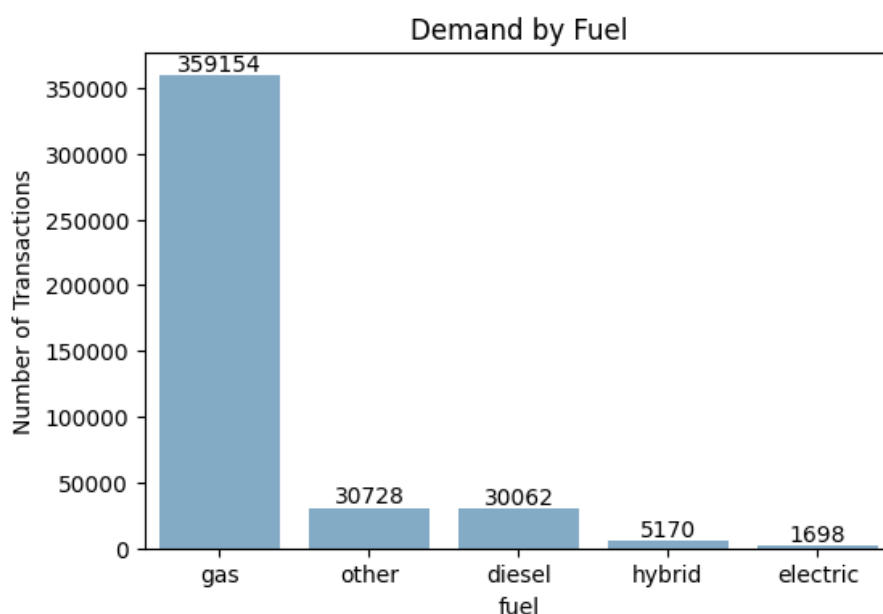
Hypothesis 5: Four-wheel drive (4WD) vehicles are more in demand because of suitability in extreme road conditions.



Observation: 4wd (four-wheel drive) and fwd (front wheel drive) vehicles are more in demand than rwd (real wheel drive).

Recommendations: Prioritize listing more 4wd and fwd vehicles on sales platforms. Implement competitive pricing strategies for rwd vehicles to improve their marketability.

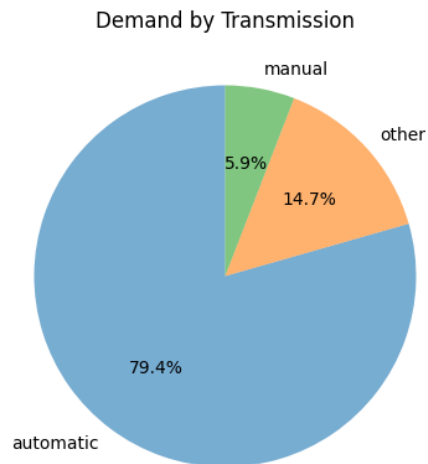
Hypothesis 6: Demand is correlated with fuel cost. People prefer buying vehicle which is cheaper to manage.



Observation: Gas based vehicles are highly in demand because it is cheap to maintain and the electric vehicles are very less preferred may be because of high cost.

Recommendations: Increase the inventory of gasoline vehicles to meet higher demand. Offer attractive financing options for electric vehicles to make them more accessible. Educate customers about available subsidies for electric vehicles.

Hypothesis 7: Automatic vehicles are more likely to be preferred by customers hence more in demand than manual.

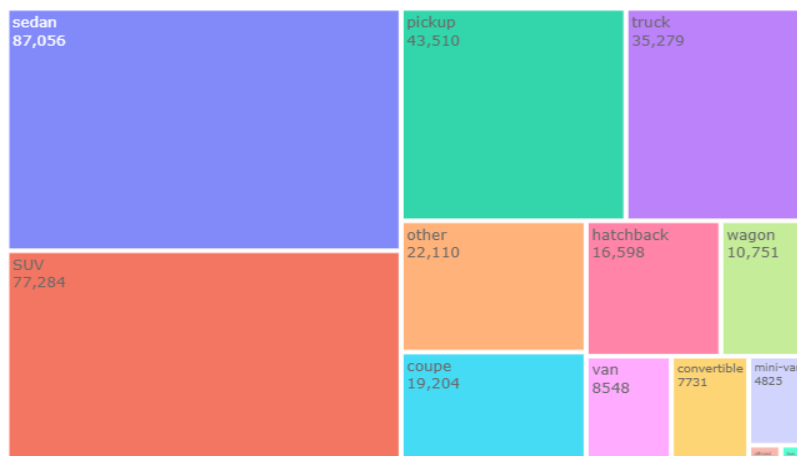


Observation: 79.4% of total transactions are of automatic vehicles which shows the incomparable demand than manual and other transmission types.

Recommendations: Align supply of automatic vehicles with respect to demand. Focus marketing on the convenience, ease of driving of modern automatic vehicles. Provide discounted packages for manual transmission vehicles to make them more attractive to buyers.

Hypothesis 8: Some specific vehicle types are more in demand than others.

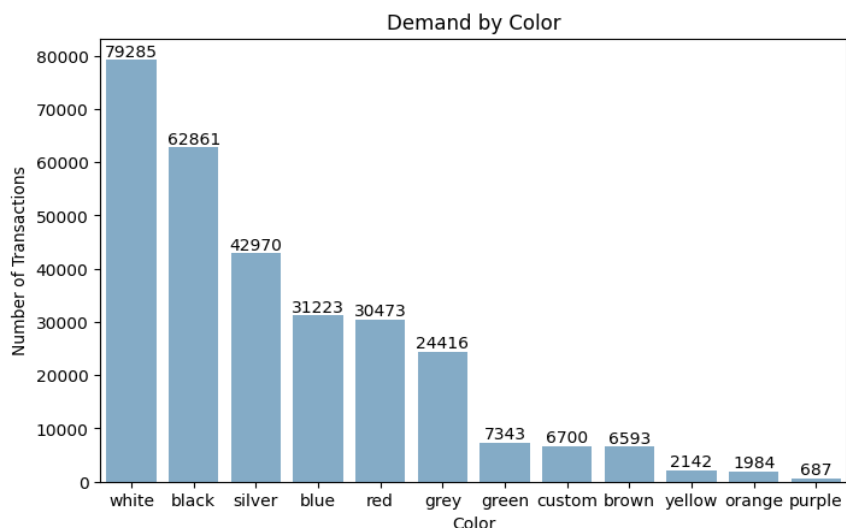
Demand by Car Type



Observation: Sedan, SUV and Pickup are highly preferred vehicles by customers and bus are the least preferred.

Recommendations: Ensure adequate supply of Sedan, SUVs, Pickup, coupe etc vehicle types to fasten sales. Reduce or limit the inventory of buses and introduce trade-in offers, to make buses more attractive to niche markets.

Hypothesis 9: Colour of the vehicle is an important attribute to enhance demand.
People normally prefer white and black colours.

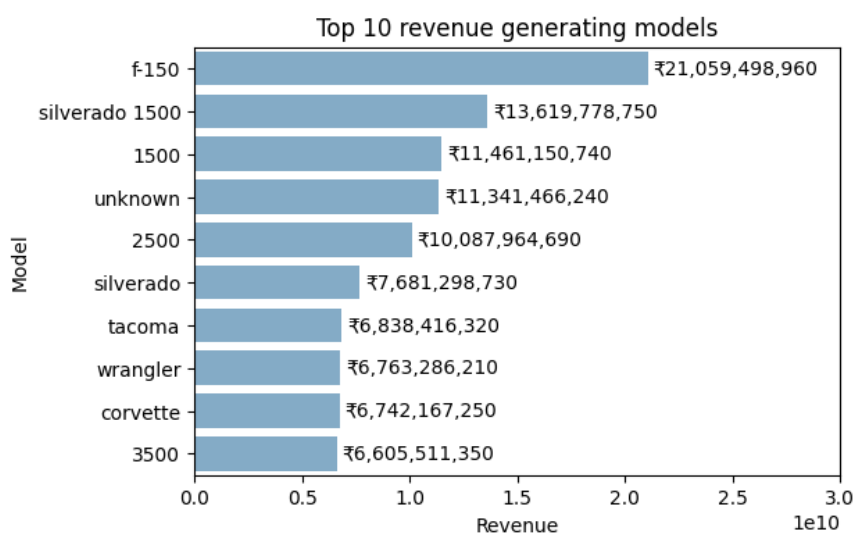


Observation: White, Black, Silver, Blue, and Red colour are high in demand whereas Yellow, Orange, Purple are least preferred.

Recommendations: Increase stock of vehicles in high demand colours (white, black, silver, blue, red) to maximize sales. Offer discounts, and incentives for vehicles in less preferred colours to clear inventory faster.

Revenue Analysis:

Hypothesis 10: Some specific models lead to higher revenue generation.

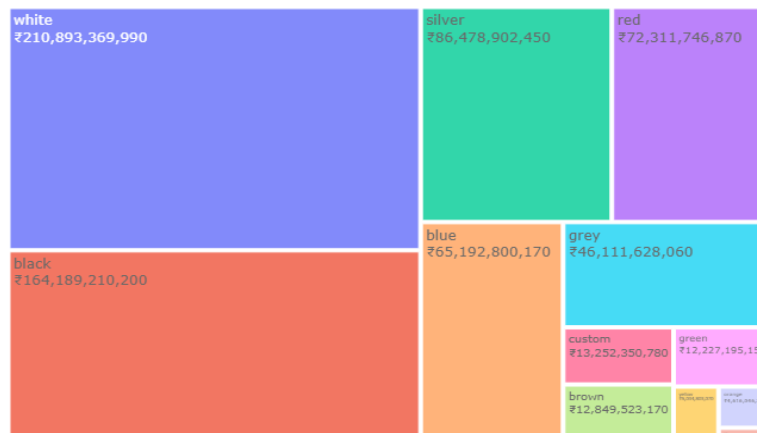


Observation: F-150, Silverado 1500, 2500, Tacoma, wrangler etc are the cars which are generating high revenue.

Recommendations: Stock of F-150, Silverado 1500, 2500, Tacoma, wrangler etc cars should be enough to sell at high numbers. Marketing for the same should be focused to retain and increase the sales. Target relevant customer segments such as businesses, outdoor enthusiasts, and families.

Hypothesis 11: Vehicles in certain colour categories (white, black) have higher share in revenue.

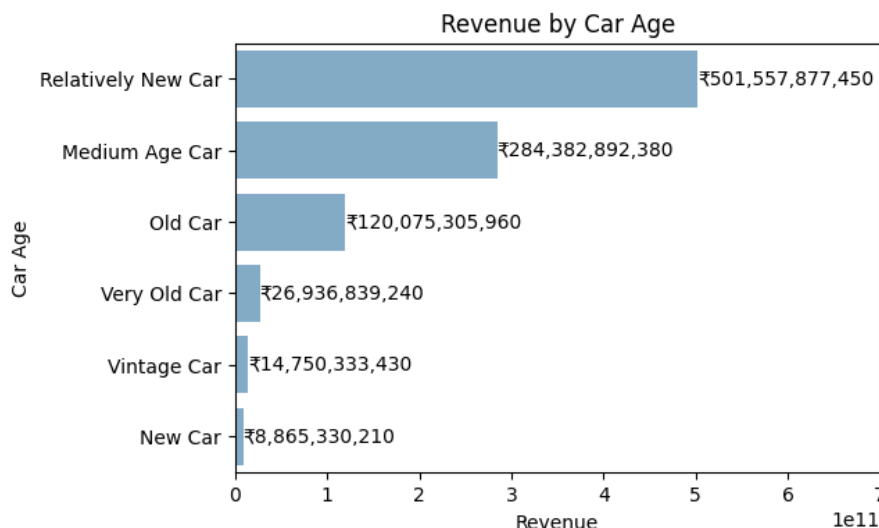
Revenue by Color



Observation: White, Black, Silver, Blue and Red colour are high in demand and also generate high revenue. Purple, Orange, Orange etc colours are very low demand and revenue share.

Recommendations: Sourcing of the vehicles in White, Black, Silver, Blue and Red colour should be at large scale to maintain and increase revenue. Least revenue generating colour vehicle should be sold at offer prices to clear the stock. Marketing for low revenue generating vehicles should focus on other feature, engine, status, type etc.

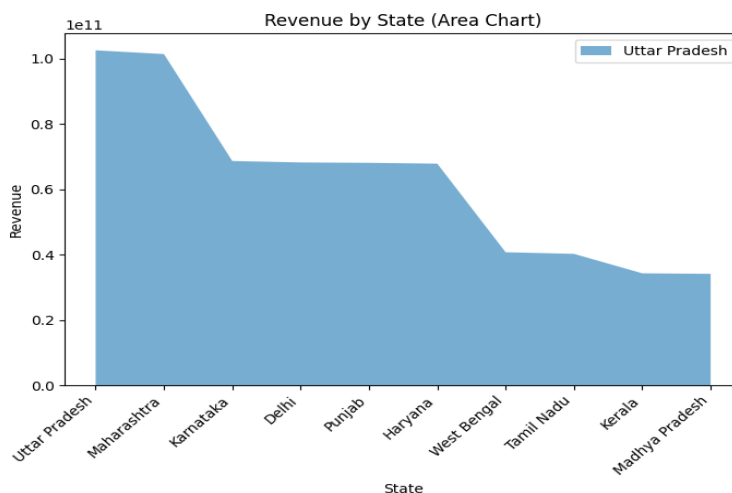
Hypothesis 12: Age of the vehicle influences the revenue.



Observation: Relatively new, Medium age, and Old age vehicles are highly important for revenue purpose. New and vintage vehicles contribute very less in revenue.

Recommendations: Relatively new (2016-2020) and medium age (2011-2015) vehicles should be in large number in inventory. Introduce tailored financing solutions for old, very old and vintage vehicles with lower monthly payments to attract budget-conscious buyers.

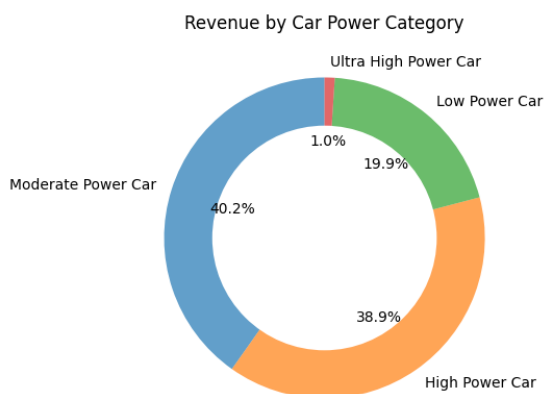
Hypothesis 13: Sates with high demand have high revenue contribution.



Observation: Uttar Pradesh, Maharashtra, Karnataka, Delhi and Punjab have high demand and high revenue also. May be because of large population living in these states.

Recommendations: Ensure a greater stock of popular models and high-revenue vehicles (such as F-150, Tacoma, Wrangler, etc.) is available in Uttar Pradesh, Maharashtra, Karnataka, Delhi, and Punjab to meet the demand. Offer festive season sales and state-specific promotions, especially during Diwali, Ganesh Chaturthi (Maharashtra), and Baisakhi (Punjab). Develop logistics and distribution centres near major urban hubs in Delhi NCR, Mumbai, Bangalore, and Lucknow to reduce delivery times.

Hypothesis 14: Certain power category of vehicles generating more revenue than other power categories.



Observation: Moderate power vehicles (40.2%) have the largest share in revenue followed by High power vehicles (38.9%). Whereas ultra-high-power vehicle have least revenue. May be because of the day to day uses and cost.

Recommendations: Include models across multiple segments (SUVs, sedans, and pickups) with moderate power output to cater to the broadest audience. Highlight the fuel efficiency, reliability, and low maintenance costs of moderate-power cars in marketing campaigns to attract to cost-conscious buyers. Organize performance test drives, racing events, or car shows to attract attention and promote ultra-high-power models to a specialized market segment.

Hypothesis 15: Luxury car manufacturers are more likely to generate high revenue per unit.

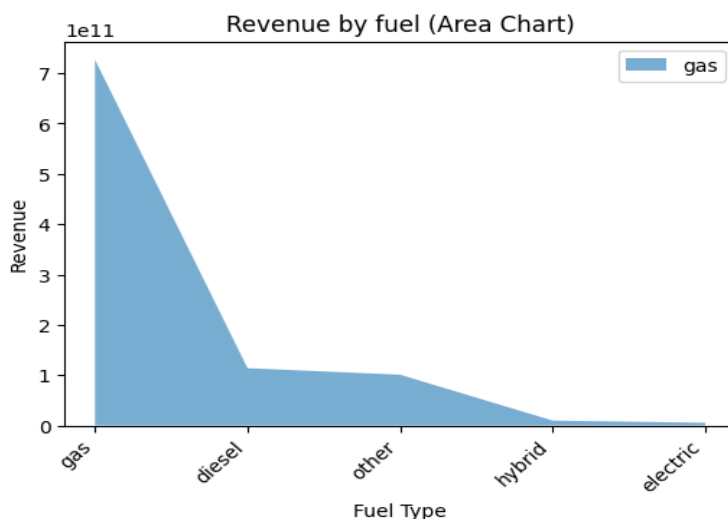
'Revenue Per Manufacturer:'

	manufacturer	total_revenue	units_sold	avg_revenue_per_unit
11	ferrari	1261836050.00	95	15772950.62
2	aston-martin	164873950.00	24	7851140.48
38	tesla	4275460330.00	868	5071720.44
33	porsche	5677993830.00	1384	4394732.07
34	ram	63140357440.00	18342	3852840.95

Observation: Ferrari, Aston-Martin, Tesla, Porche, Ram are some luxury vehicle manufacturers which have huge revenue per unit.

Recommendations: Allocate resources to regions or cities where luxury vehicles are in greater demand (e.g., metro cities like Delhi, Mumbai, Bangalore). Provide a personalized buying experience, including at-home test drives, exclusive previews, and dedicated sales advisors for affluent clients. Market Tesla's features such as Autopilot, long-range battery performance etc alongside luxury.

Hypothesis 16: Fuel type influences the revenue. Gasoline being the most sold fuel type have the highest share in revenue.

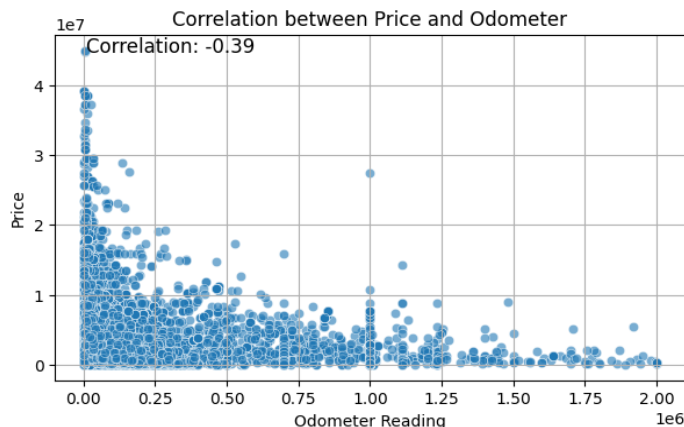


Observation: Gasoline based vehicles are earning high revenue followed by diesel. Electric vehicle contributing very low revenue may be because of cost of purchase and maintenance.

Recommendations: Prioritize maintaining a strong inventory of gasoline-powered SUVs, sedans, and pickups since they generate the most revenue. Partner with charging network providers to offer free or subsidized charging plans for Electric vehicle buyers. Use sustainability-focused marketing campaigns to appeal to buyers prioritizing environmental impact.

Price Analysis:

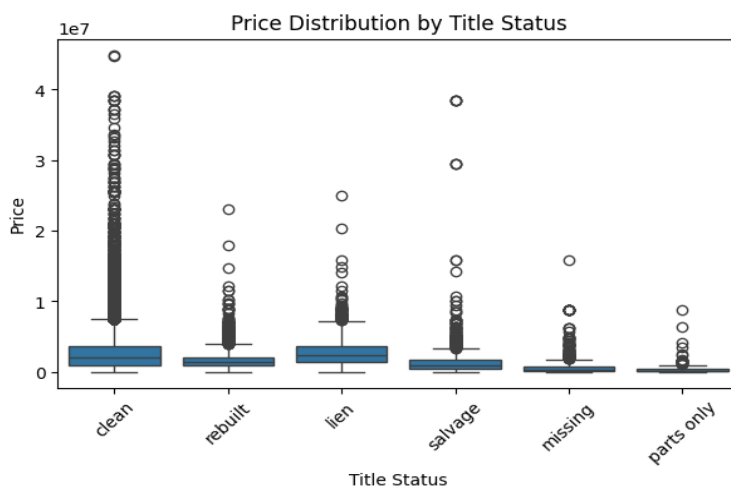
Hypothesis 17: Price of the vehicle is correlated to odometer reading. Vehicle with high odometer reading has low price.



Observation: There is a moderate negative correlation between price and odometer reading. This means that as the odometer reading increases, the price of the vehicle decreases. Vehicles with lower odometer readings show a wide range of prices, indicating variation based on vehicle attributes like manufacture, model etc. Vehicles with higher odometer readings tend to have lower prices, with fewer exceptions.

Recommendations: Focus on sourcing vehicles with lower odometer readings, to maintain good revenue. Implement dynamic pricing model for vehicles with low odometer readings. Offer extended warranties on vehicles with high odometer reading to build customer trust.

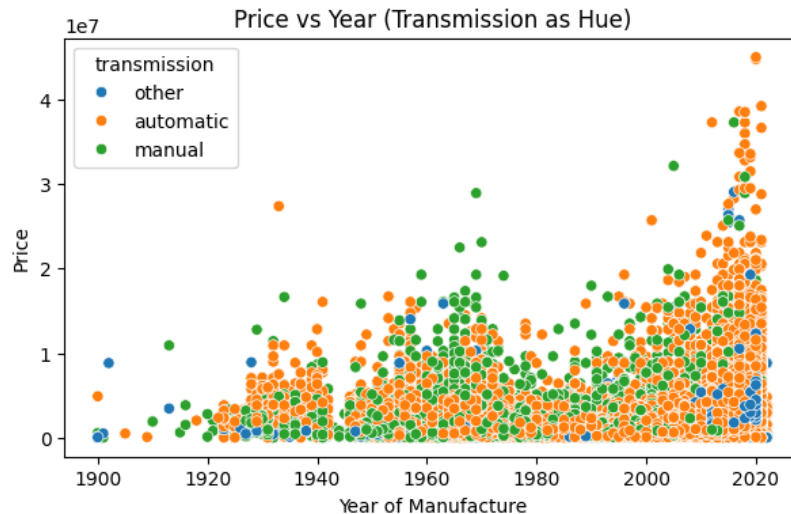
Hypothesis 18: Price of the vehicles vary by title status of the vehicle. People tend to buy clean vehicle.



Observation: Vehicles with a "clean" title status show the widest price distribution, indicating a strong market preference for these vehicles. Rebuilt and "lien" title vehicles have lower price ranges and "salvage," "missing," and "parts Only" have the lowest Prices.

Recommendations: Prioritize acquiring and selling vehicles with "clean" title status as they have high demand. Highlight models that fall into lower title status categories and work on them.

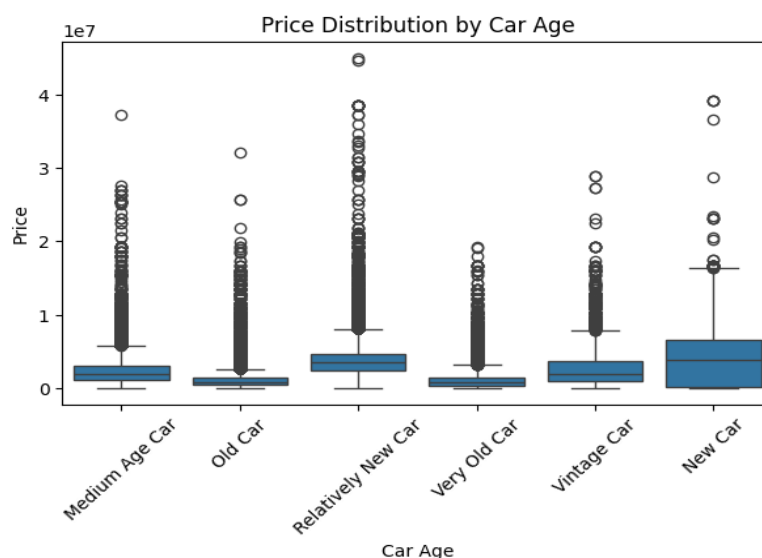
Hypothesis 19: Price is correlated with the manufacturing year of the vehicle. New vehicle has high prices whereas old have low prices.



Observation: Price and year of vehicle manufacture have positive correlation (0.32). As the year of manufacturing increase prices also increase. Maximum distribution and increase of price are for automatic vehicles.

Recommendations: Highlight automatic transmission benefits such as ease of driving, comfort in traffic, and higher resale value in marketing campaigns to keep demand high. Offer model clearance sales for older vehicles to reduce inventory faster.

Hypothesis 20: Price distribution of vehicles is correlated with car age; old vehicles have low price distribution whereas new vehicle has wide price distribution.

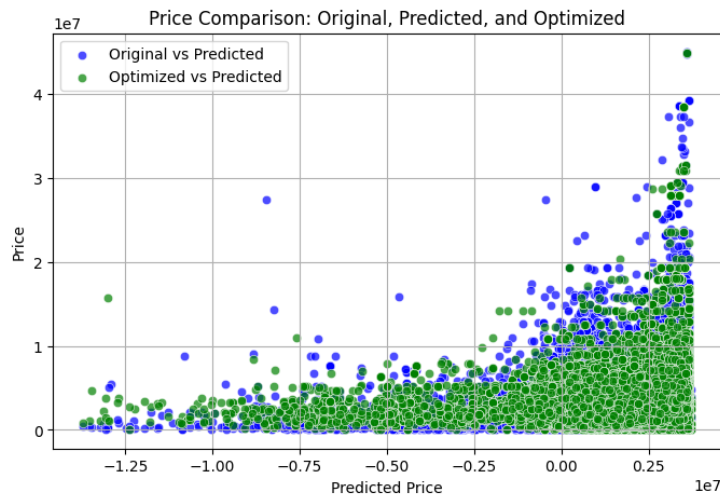


Observation: Relatively new car (2016-2020) and medium age car (2011-2015) have largest price distribution. Old, very old, new and vintage cars have low price distribution.

Recommendations: Optimize the prices of Relatively new, medium age cars with respect to the demand. Review and correct the prices of old, very old and vintage cars.

Price Optimization:

To enhance the revenue, price prediction and price optimization of each car is very important. Based on the features like 'year' of manufacture and 'odometer' reading, we can optimize the prices of cars. We are used statistical methods like IQR for price optimization and linear regression for price prediction.



Observation: There is a spread in the blue points, especially for higher prices, which suggests that the original prices deviate more significantly from the predicted values as the predicted price increases. The green points are more tightly aligned with the predicted values, indicating that the optimization process adjusted the prices to be closer to the predicted range. The green points appear to have less scatter around the predicted values than the blue points. This suggests that the optimization has reduced the variance and brought prices closer to a statistically reasonable range.

Recommendations: Since the scatter plot shows that original prices deviate significantly from the optimized prices, standardizing pricing based on statistical optimization can reduce discrepancies. Use the optimized prices as a baseline to ensure uniform pricing for cars within similar categories or features. Develop a pricing strategy aligned with optimized prices.

Summary:

- **Top Models and Manufacturers:** High-revenue models include Ford F-150, Chevrolet Silverado, and Ram 1500. Premium vehicles like Silverado 2500 lead in price, while economy models like Altima and Civic have lower prices.
 - **Vehicle Demand:** Moderate-power, full-size/mid-size vehicles (SUVs, pickups, sedans) dominate sales. Gasoline, 4WD/FWD, and automatic vehicles are highly preferred.
 - **Regional and Consumer Trends:** Major demand comes from populous states (e.g., Uttar Pradesh, Maharashtra). Popular colours include white, black, and silver. Clean titles, good condition, and low odometer readings drive higher prices.
 - **Age and Luxury:** Newer (2016-2020) and mid-age (2011-2015) cars generate most revenue, while luxury brands like Ferrari and Tesla lead in per-unit revenue.
 - **Price Optimization:** Optimized prices reduce variance and align closely with predicted values, ensuring competitiveness and fairness.
-

Thank You