SPINNY'S REVENUE GROWTH THROUGH DEMAND AND PRICING OPTIMISATION

Problem Statement

Spinny, a leading Indian startup in the pre-owned car market, 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.

Objective: -

→ Understanding demand for cars across different regions.
→ Identifying the most in-demand car models and their attributes.
→ Determining competitive pricing strategies for each car.

Business Goal

➢ **Increase revenue by boosting the number of completed transactions**

WHAT DID WE UNDERSTAND FROM THE PROBLEM STATEMENT?

- **Regional Demand:** Which cars are popular in specific regions?
- **Key Attributes:** What features (e.g., fuel type, condition, mileage) make a car desirable?
- **Optimal Pricing:** How to price each car competitively to maximise sales.

Increasing Spinny's revenue by ensuring more cars are successfully sold. This involves understanding customer preferences, pricing cars competitively, and focusing on high-demand regions and car models to encourage more transactions.

Dataset Overview

Source: Next leap provided it as part of a Data Analyst course.

Dataset Name: Spinny Revenue Growth.

Rows: 426,880

Columns: 21

Description: Contains attributes like price, model, fuel type, and location to analyze demand and optimise pricing.

Columns Description

ID - Unique transaction ID.

Price - price: Final selling price (INR).

Year- Manufacturing year of the car.

Manufacturer- Brand of the car.

Model - Specific model name.

Condition- Physical state of the car (e.g., excellent, good).

Cylinders- Number of engine cylinders.

Fuel- Type of fuel (e.g., petrol, diesel).

Odometer- Distance traveled by the car (km).

Title_status- Status of the car's title (e.g., clean).

Transmission- Gearbox type (manual/automatic).

VIN - Vehicle Identification Number.

Drive- Drivetrain configuration (e.g., front-wheel drive).

Size - Vehicle size classification (e.g., compact).

Type- Type of vehicle (e.g., sedan, SUV).

Paint_color- Exterior color.

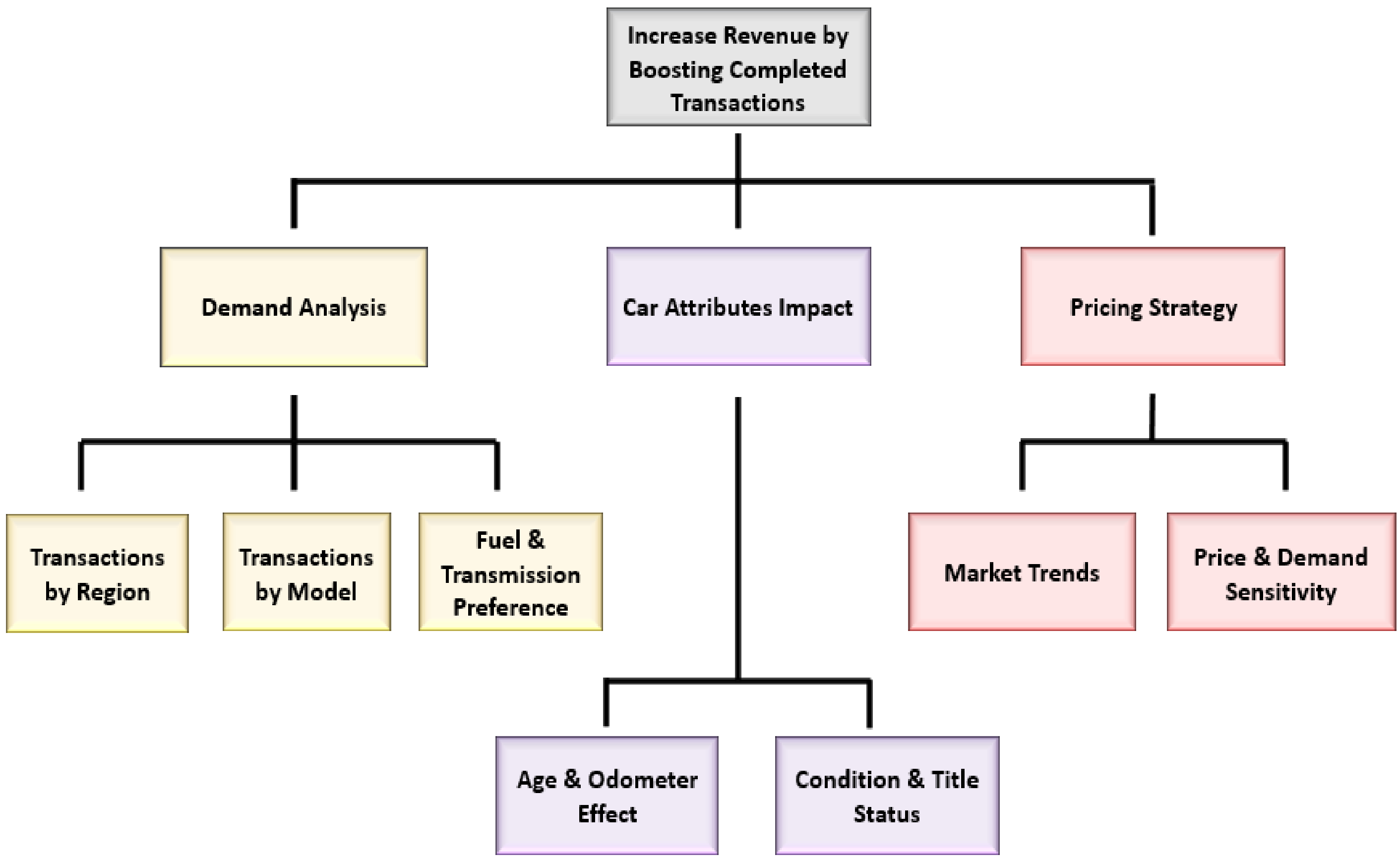
Image_url- Link to an image of the vehicle for reference.

State- Indian state where the transaction occurred

Posting_date- Date when the car was listed for sale.

Latitude &Longitude- Geographical coordinates of the transaction location.

KPI TREE:-



Loading, Importing, and Initial Data Exploration

Importing Libraries

The necessary libraries for data manipulation, visualization, and statistical analysis were imported
Pandas: For data handling.
NumPy: For numerical computations.
Matplotlib & Seaborn: For visualizations.
SciPy: For statistical analysis.

Loading the Dataset

The dataset was downloaded from Google Drive using gdown and loaded into a Pandas DataFrame. It contains **426,880 rows** and **21 columns**, detailing pre-owned car transactions.

Checking Dataset Shape

The dataset contains **426,880 rows** and **21 columns**, detailing pre-owned car transactions.

Viewing All Columns

Columns were categorized as:
Most Relevant: *price*, *year*, *manufacturer*, *model*, *condition*, *fuel*, *odometer*, *state*, *transmission*
Relevant: *type*, *drive*, *title_status*, *paint_color*, *cylinders*, *size*, *posting_date*
Less Relevant: *id*, *VIN*, *image_url*, *latitude*, *longitude*

Understanding the Dataset

A quick look at the first few rows showed that most columns are populated, but some (e.g., condition and cylinders) have missing values.
Random sampling confirmed that missing values are not uniformly distributed.

Displaying Dataset Information

Data types:
Numerical: Columns like price and odometer.
Categorical: Columns like manufacturer and fuel.

Missing Values Data Summary

A summary of missing values revealed:
Columns with high missing percentages: e.g., size (71.77%), making it unreliable for analysis.
Columns with low missing percentages: e.g., year (0.28%), which can be easily addressed.

Unique Value Counts

Unique identifiers: Columns like id and VIN.
Categorical diversity: Columns like fuel and type have limited unique values, suitable for grouping or segmentation.

Data Preparation

Creating a Backup of the Dataset

Purpose: A backup of the original dataset (df) was created to ensure the raw data remains intact. This allows for recovery in case of errors during cleaning or transformation.

Steps:

- A copy of the dataset was made using. copy().
- This ensures that all subsequent operations are performed on the duplicate, leaving the original untouched.

Structuring the Dataset

Purpose: To standardise and organise the dataset for easier analysis.

Steps:

- Converted all string columns to lowercase for consistency.
- Renamed columns (e.g., year to mfg_year) to improve clarity.
- Removed special characters from relevant columns to ensure clean, uniform data.

Datatype Correction

Purpose: Ensuring each column has an appropriate datatype for accurate analysis.

Steps:

- Converted mfg_year from float to integer for precision.
- Transformed posting_date from object to datetime format to enable time-based analysis.

Removed Duplicate Rows

Purpose: To eliminate redundancy and ensure data integrity.

Steps:

- Checked for duplicate rows using. duplicated (). sum ().
- Removed duplicates with. drop duplicates ().

Dropping Unnecessary Columns

Purpose: To streamline the dataset by removing irrelevant or redundant columns.

Steps:

○ Dropped columns like image_url and size due to high missing values or irrelevance to the analysis objectives.

○ Assessed columns like VIN (Vehicle Identification Number) and removed it after identifying issues like invalid entries and high repetition.

Outliers Handling

Purpose: To remove extreme values that could distort analysis results.

Steps:

○ Used the Interquartile Range (IQR) method to detect outliers in price and odometer.

○ Removed rows with values outside acceptable ranges.

Dropping Rows Based on Conditions

Purpose: To refine the dataset by removing irrelevant or inconsistent entries.

Steps:

○ Removed rows where mfg_year was unrealistic (e.g., future years or very old cars).

○ Eliminated rows with zero price as they do not contribute to revenue analysis.

Missing Values Imputation

Purpose: To address missing data without introducing bias.

Steps:

○ Filled missing values in categorical columns with "unknown".

○ Used hierarchical imputation for columns like condition, filling gaps based on related attributes (e.g., mfg_year, odometer).

Data Preparation for ML Models

Purpose: To prepare the dataset for machine learning algorithms.

Steps:

○ Normalised numeric columns like price and odometer using Min-Max scaling.

○ Converted object-type columns to categorical data types for memory efficiency and faster processing.

Feature Engineering

Purpose: To create new features that provide additional insights or simplify analysis.

Steps:

○ Created cars_age by subtracting mfg_year from the year in posting_date.

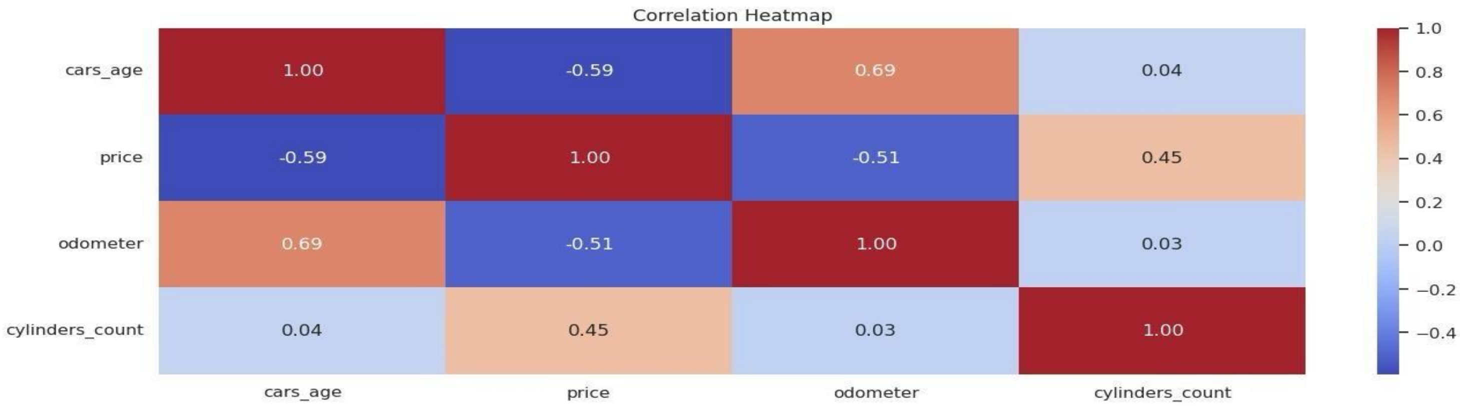
○ Extracted numeric values from cylinders into a new column, cylinders_count.

○ Categorised states into income levels (state_income) based on economic data.

Exploratory Data Analysis (EDA)

Correlation Heatmap of cars age, price, odometer, cylinders count & condition

Content: This section analyses the relationships between key numerical variables using a correlation heatmap.



Key Insights:

- Strong negative correlation (−0.59−0.59) between cars_age and price: Older cars tend to have lower prices.
- Moderate negative correlation (−0.51−0.51) between odometer and price: Higher mileage reduces resale value.
- Strong positive correlation (0.690.69) between cars_age and odometer: Older cars generally have higher mileage.
- Moderate positive correlation (0.450.45) between cylinders_count and price: Cars with more cylinders are priced higher due to better performance.

Majority of the Cars Priced Below ₹3 Million -

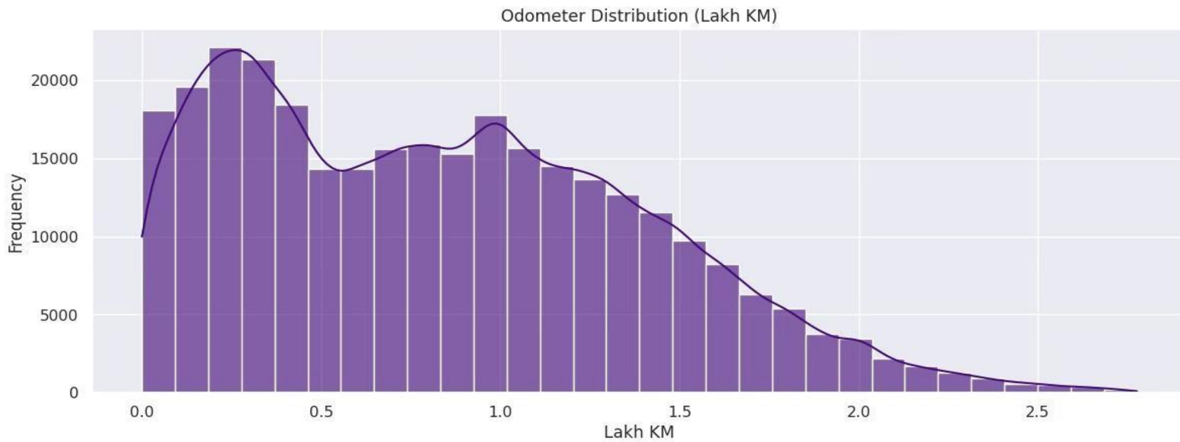
Content: This section explores the price distribution of cars.

Key Insights:

- Most cars are priced between ₹0.5M–₹3M, reflecting affordability in a price-sensitive market.
- Cars priced above ₹4M cater to a niche audience, with significantly lower demand.
- The price distribution is right-skewed, highlighting limited demand for high-priced vehicles.

Cars with Odometer Readings Between 0 to 0.5 Lakh KM Dominate the Dataset

- **Content:** Analyses the odometer readings of cars in the dataset.

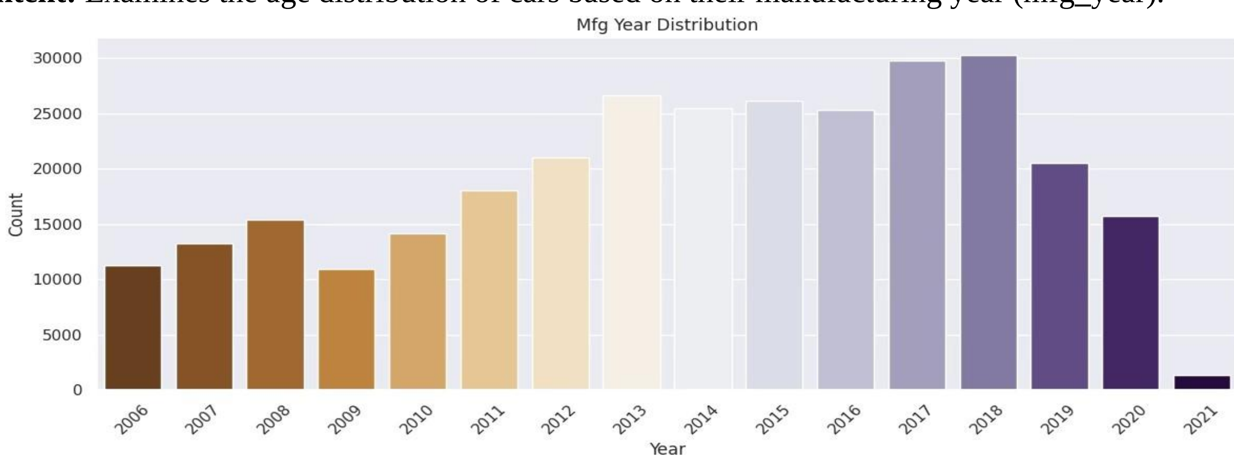


Key Insights:

- Most cars have odometer readings between 0–1 lakh KM, indicating buyer preference for low-mileage vehicles.
- The distribution is right-skewed, with fewer cars having readings above 2 lakh KM.

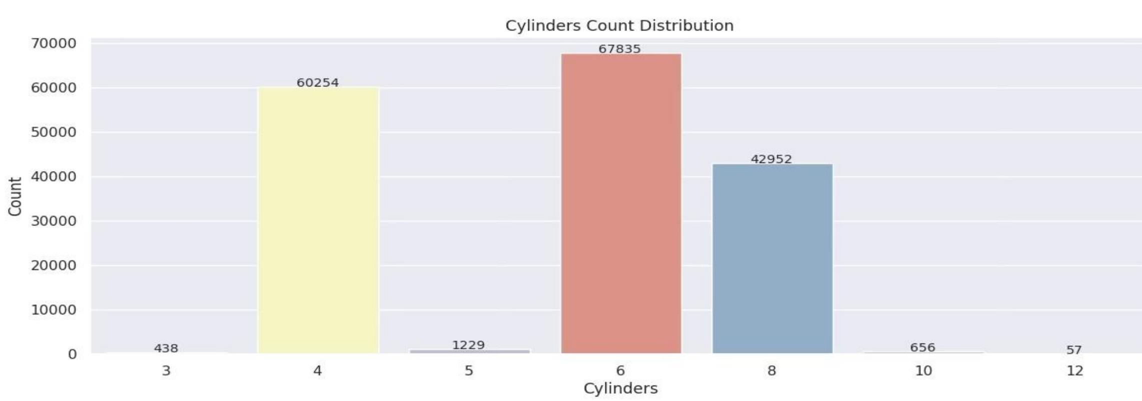
Cars Aged Between 3 to 8 Years Dominate the Dataset

Content: Examines the age distribution of cars based on their manufacturing year (mfg_year).



Key Insights:

- Cars aged 3–8 years form the largest segment, balancing affordability and reliability.
- Newer models (2020–2021) are limited, possibly due to supply constraints or lower resale availability.
- Older models (pre-2010) have reduced representation, reflecting lower demand.

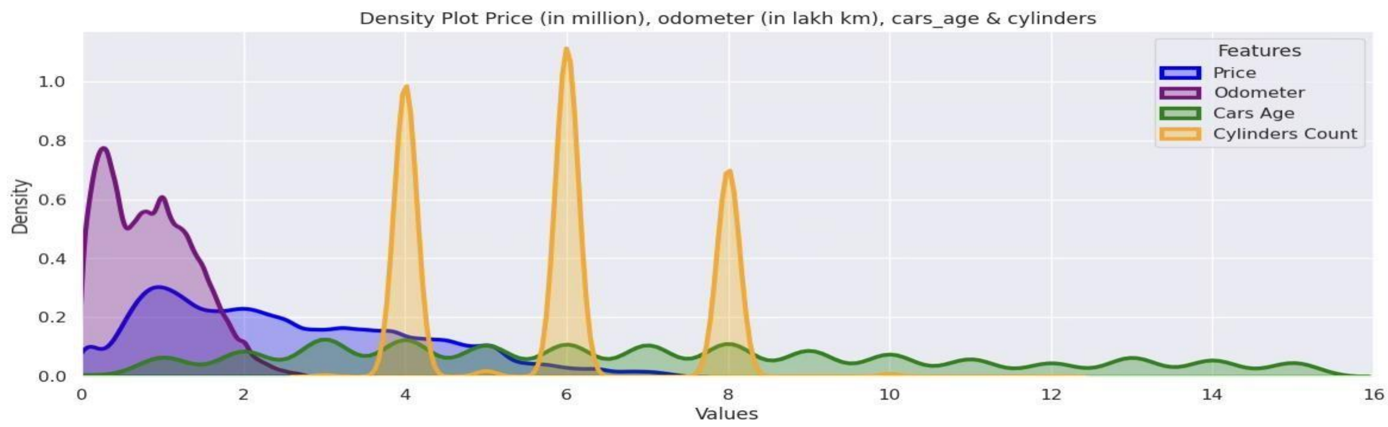


Key Insights:

- Cars with 4 cylinders (60K+) and 6 cylinders (67K+) dominate due to their efficiency and performance balance.
- Higher-cylinder configurations (e.g., 8, 10) cater to niche markets like luxury or performance vehicles.

Distributions Overview of Price, Odometer, Car Age, & Cylinders Count

Content: Provides a combined visualisation of key numerical features.

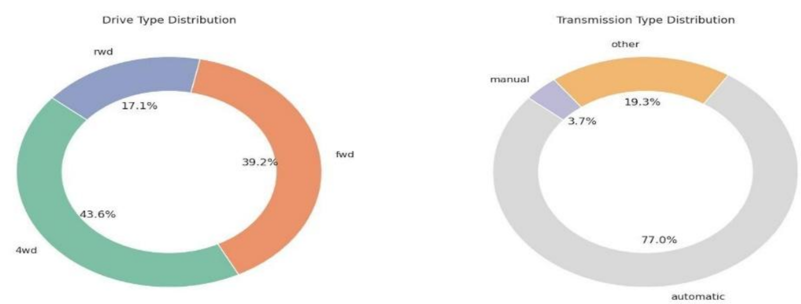


Key Insights:

- Price: Most demand lies between ₹0.5M–₹3M; higher prices see limited demand.
- Odometer: Low-mileage cars (<1 lakh KM) dominate buyer preference.
- Car Age: Cars aged 3–9 years are most popular; very new or very old cars have lower demand.
- Cylinders Count: Majority of cars have either 4 or 6 cylinders.

Cars with 4-wheel drive (4WD) and front-wheel drive (FWD) dominate the market.

Automatic transmissions are the most preferred type due to their convenience.



Content:

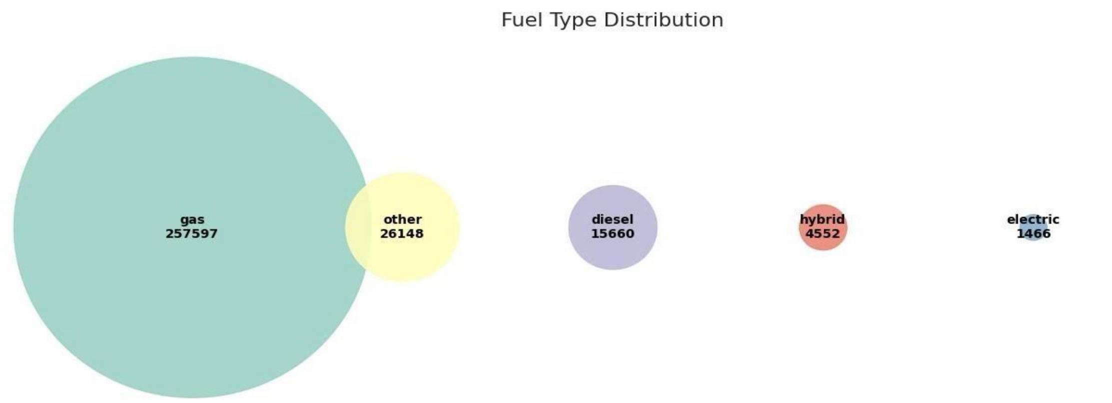
- **4WD vehicles** account for **43.6%** of the inventory, preferred for off-road and rugged conditions.
- **FWD vehicles** make up **39.2%**, valued for affordability and fuel efficiency in urban settings.
- **RWD vehicles** represent only **17.1%**, catering to niche markets like performance or luxury cars.

Insight: Focus on inventory and marketing strategies for 4WD and FWD vehicles to meet dominant demand while targeting niche buyers for RWD cars.

Content:

- **Automatic transmissions** dominate with a **77% share**, reflecting ease of use in urban areas.
- **Manual transmissions** account for **19.3%**, appealing to budget-conscious buyers or driving enthusiasts.
- Other transmission types have minimal presence (**3.7%**) due to specialised use cases or higher costs.
- **Insight:** Prioritise automatic vehicles while using targeted promotions for manual transmissions to appeal to specific buyer demographics.

Gasoline (gas) vehicles dominate the market.

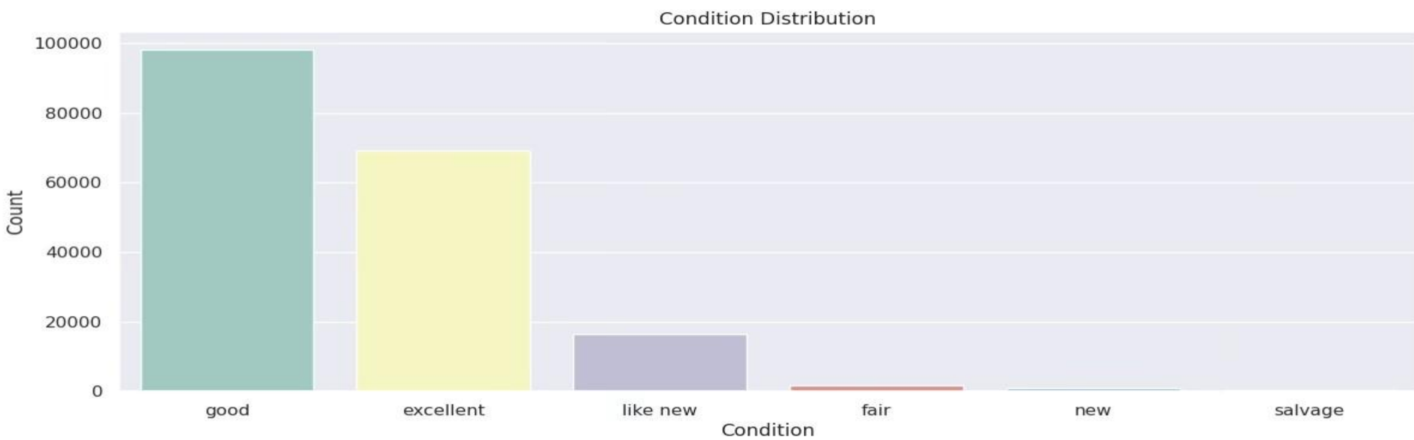


Content:

- Gasoline-powered cars are the most common due to affordability and availability.
- Diesel vehicles cater to specific needs, such as high-mileage users or rural areas.
- Electric and hybrid cars have limited representation, reflecting low adoption rates due to high costs and infrastructure challenges.

Insight: Focus inventory management on gasoline vehicles while monitoring trends in electric/hybrid markets for future opportunities.

Cars in good and excellent condition dominate the market.

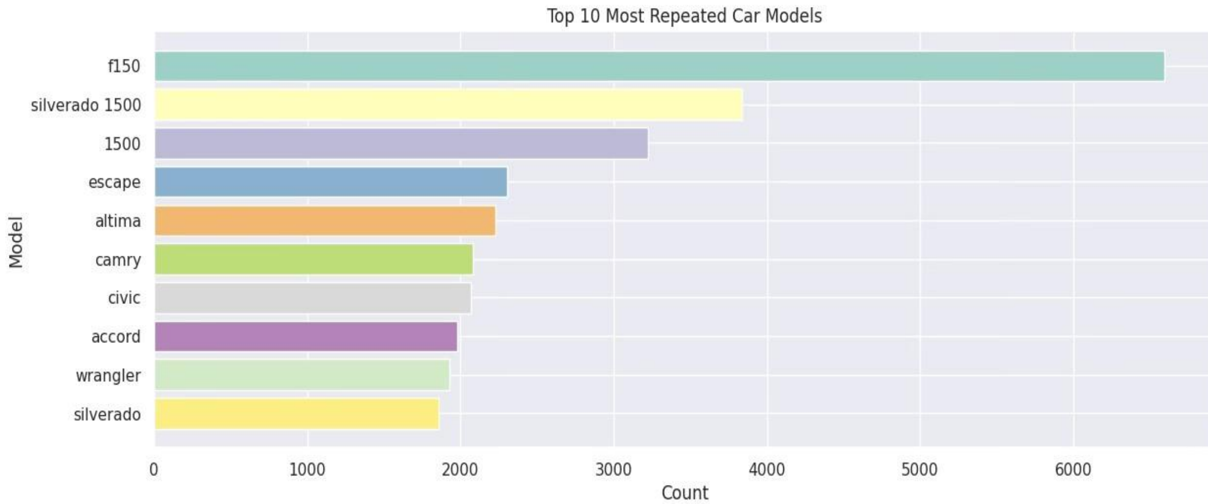


Content:

- Cars in "good" condition dominate, followed by those in "excellent" condition, as they offer a balance between quality and affordability.
- "Like new" cars cater to a smaller segment, while "fair" or "salvage" condition cars have limited demand.
- **Insight:** Focus inventory and pricing strategies on cars in good and excellent condition while exploring niche opportunities for other conditions.

Mid-range to premium models dominate -

the dataset, reflecting their popularity due to brand reliability and affordability in the second-hand car market.



Content:

- Popular models like Ford F-150 and Chevrolet Silverado 1500 dominate due to their reliability, affordability, and steady supply from affluent sellers upgrading their vehicles.

Insight: Prioritise acquiring branded second-hand cars like Ford F-150 and Chevrolet Silverado 1500 as they present significant revenue opportunities.

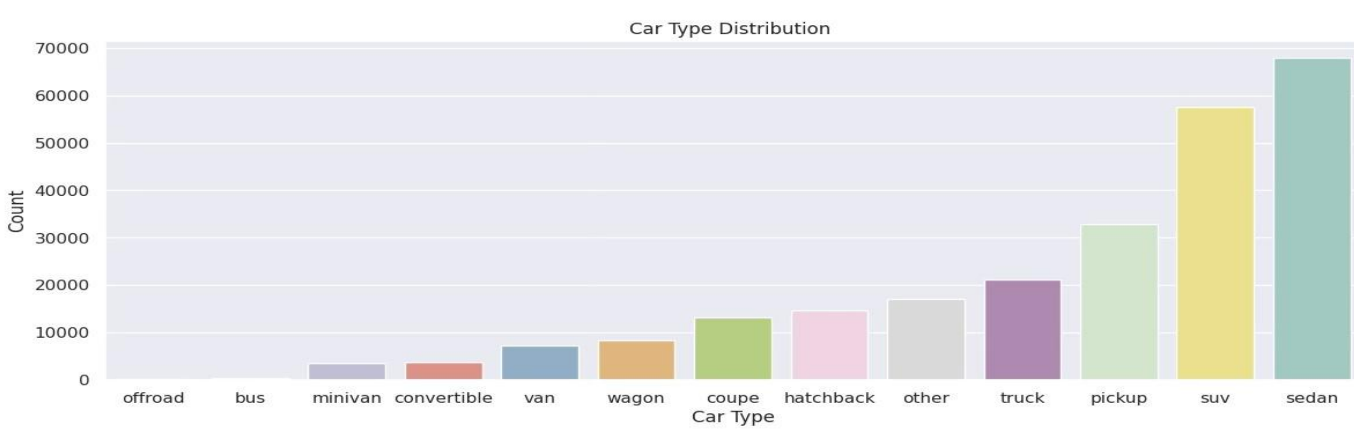
Mid-range and premium manufacturers dominate the dataset.

Content:

- Ford has the highest representation, followed by Chevrolet and Toyota, which are known for offering reliable models at competitive prices.
- Premium brands like BMW cater to niche markets but still hold notable value in the second-hand car market.

Insight: - Acquire models from Ford, Chevrolet, and Toyota to capture mainstream demand while exploring premium brands like BMW for niche segments.

Sedans and SUVs dominate the market due to their widespread appeal.

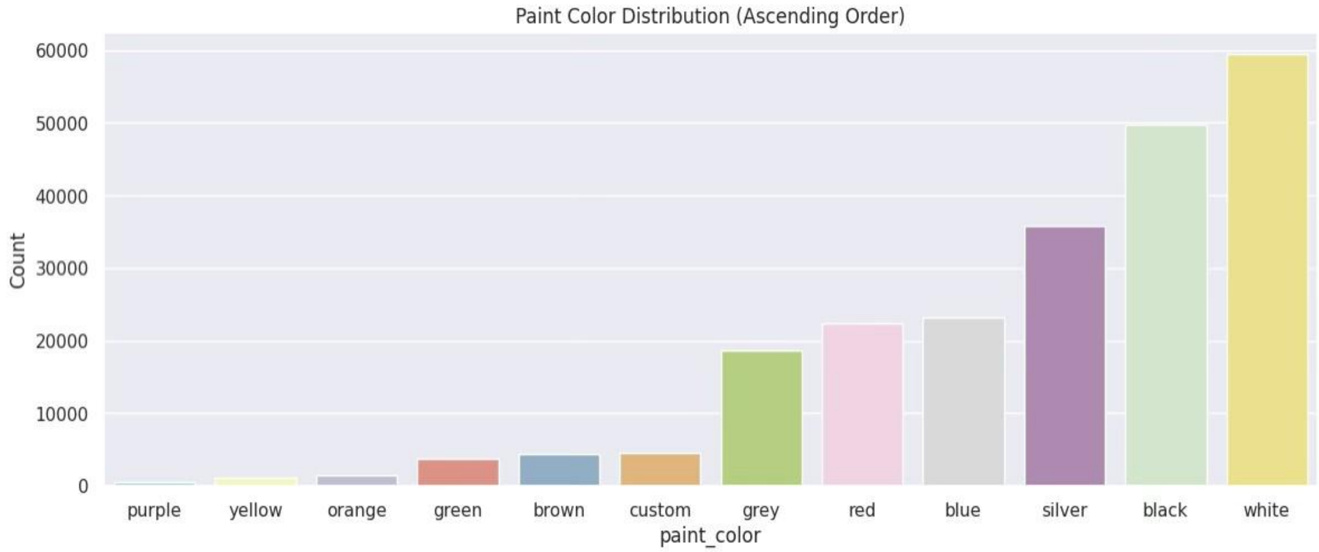


Content:

- Sedans are preferred for affordability and practicality in urban settings.
- SUVs attract buyers looking for spaciousness, versatility, and off-road capabilities.

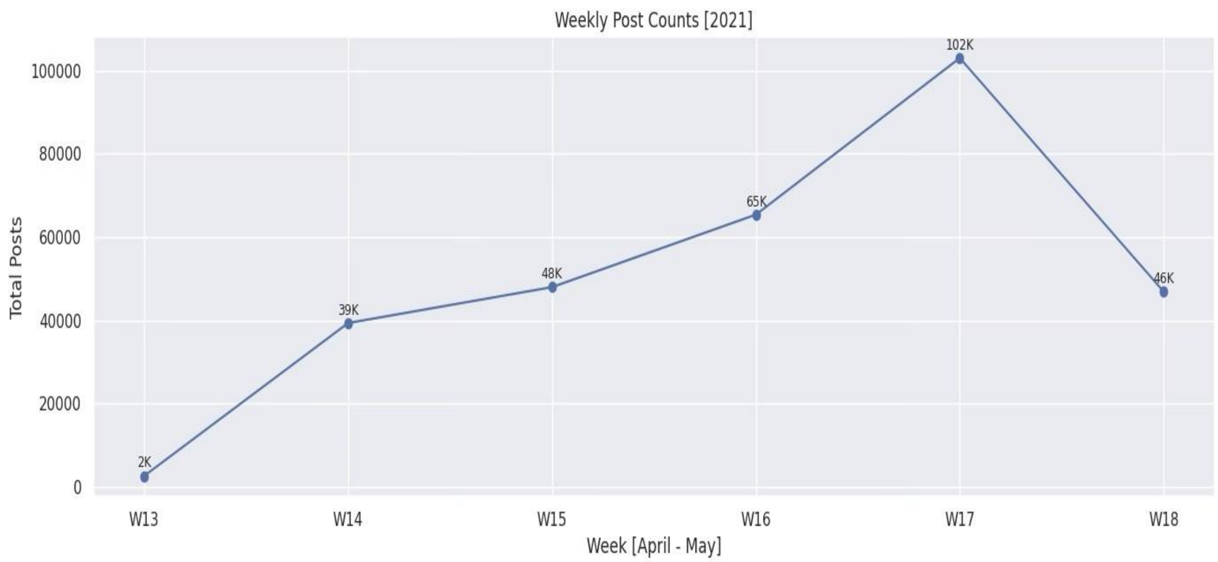
Insight: Prioritise inventory for sedans and SUVs as they capture majority demand while exploring niche segments like pickups or hatchbacks for targeted growth opportunities.

White and Black Cars Dominate the Market



- **Insight:** White and black cars are the most popular due to their neutral appeal and high resale value.
- **Actionable Point:** Focus inventory on these colours to meet majority demand, while leveraging niche colours like red or blue for targeted campaigns.

Weekly Post Counts Are Growing Rapidly

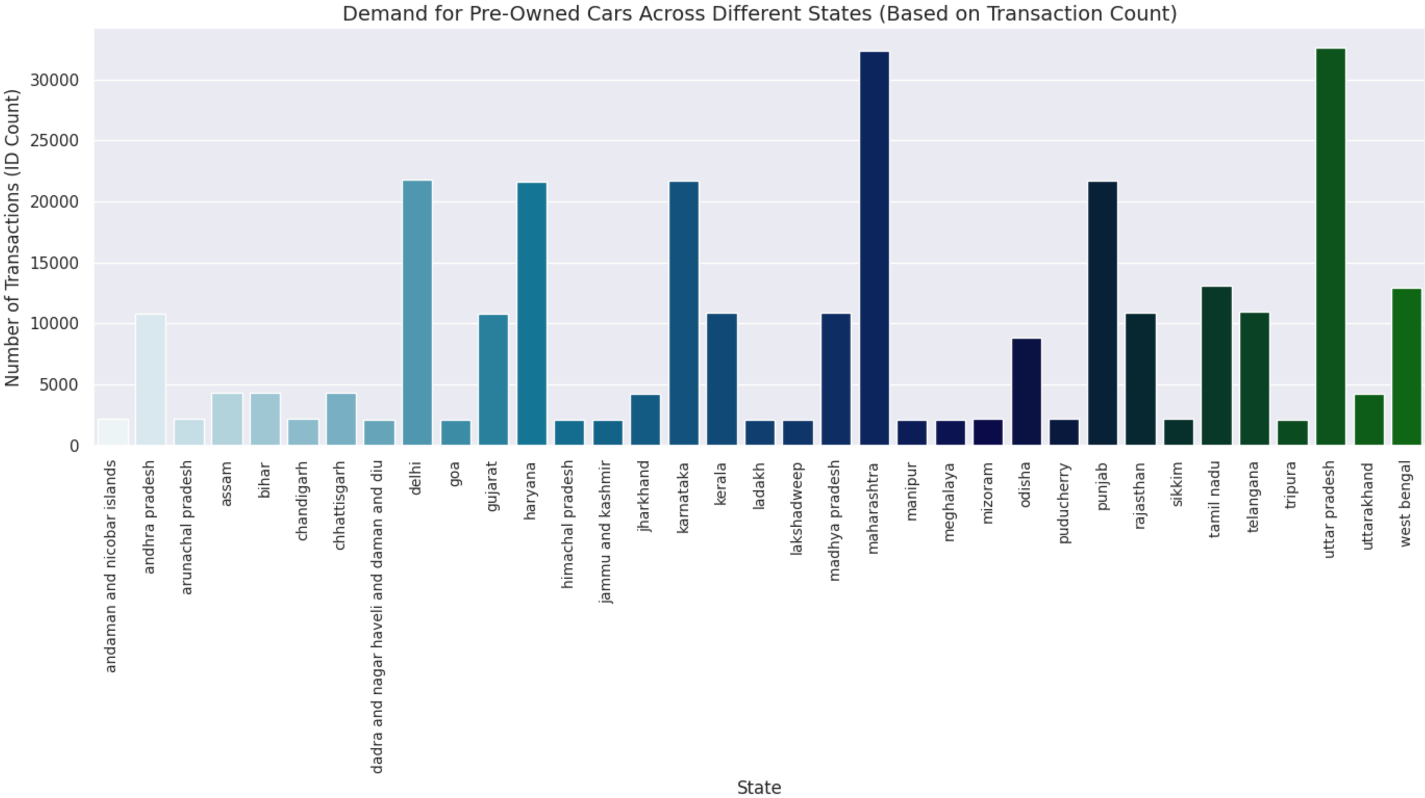


- **Insight:** Weekly post counts show consistent growth, reflecting Spinny's expanding presence in the preowned car market.
- **Actionable Point:** Scale operations and optimise marketing strategies to sustain this growth. Target regions with high post growth to capture demand hotspots effectively.
- These concise insights highlight key trends in car colour preferences and Spinny's market growth, making them suitable for HR to understand the project's impact.

Hypotheses Testing

Hypothesis 1: Regional Demand Variation

- **Null Hypothesis (H_0):** Demand for pre-owned cars does not significantly differ across states.
- **Alternate Hypothesis (H_1):** Demand for pre-owned cars significantly varies across states.

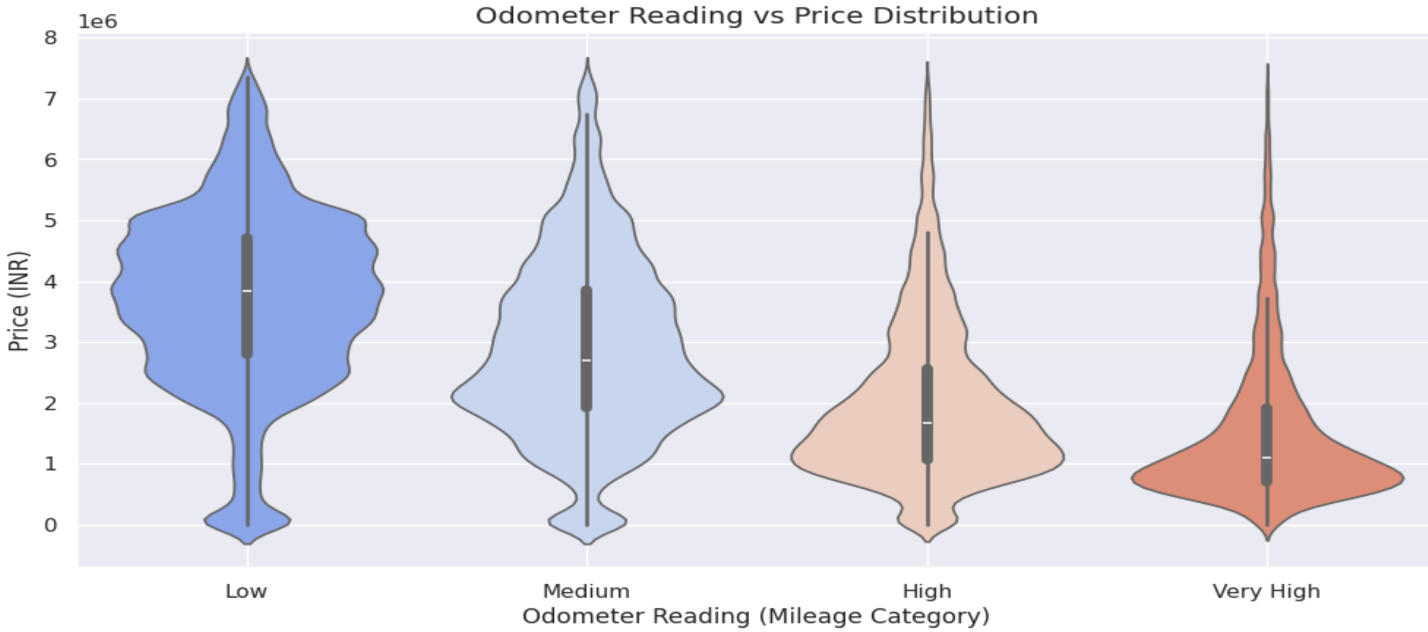


- **Methodology:** Analysed the number of car posts by state using visualisations and counts.
- **Conclusion:** States with Tier 1 cities (e.g., Maharashtra, Delhi) show higher demand due to larger populations and higher disposable incomes.
- **Decision:** Reject the null hypothesis (H_0). Demand for pre-owned cars significantly varies across states.
- **Insight:** Focus marketing efforts in urban areas while exploring rural markets for growth
- **Business Implications:**
  - Prioritize inventory and marketing in high-demand states.
  - Improve visibility and promotions in low-demand regions.
  - Adjust regional pricing to optimize sales and revenue.

.

Hypothesis 2: Impact of Odometer Reading on Demand and Price

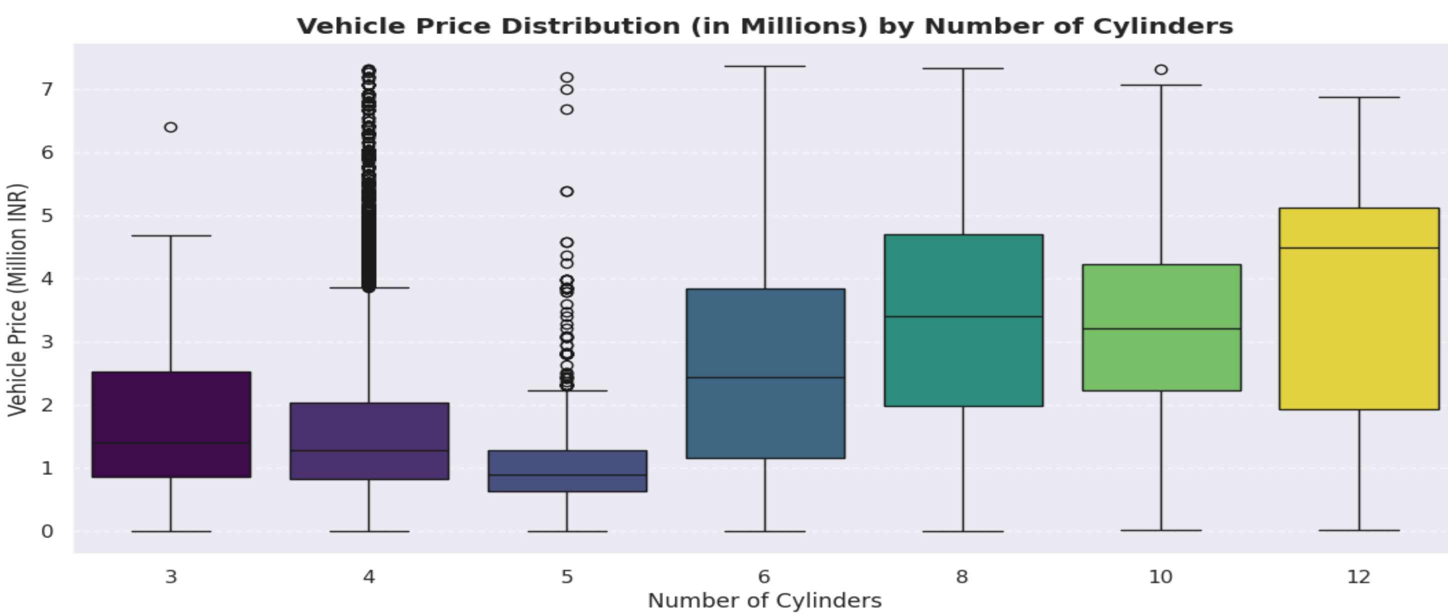
- **Null Hypothesis (H_0):** Odometer readings have no impact on car demand or pricing.
- **Alternate Hypothesis (H_1):** Odometer readings significantly influence car demand and pricing.

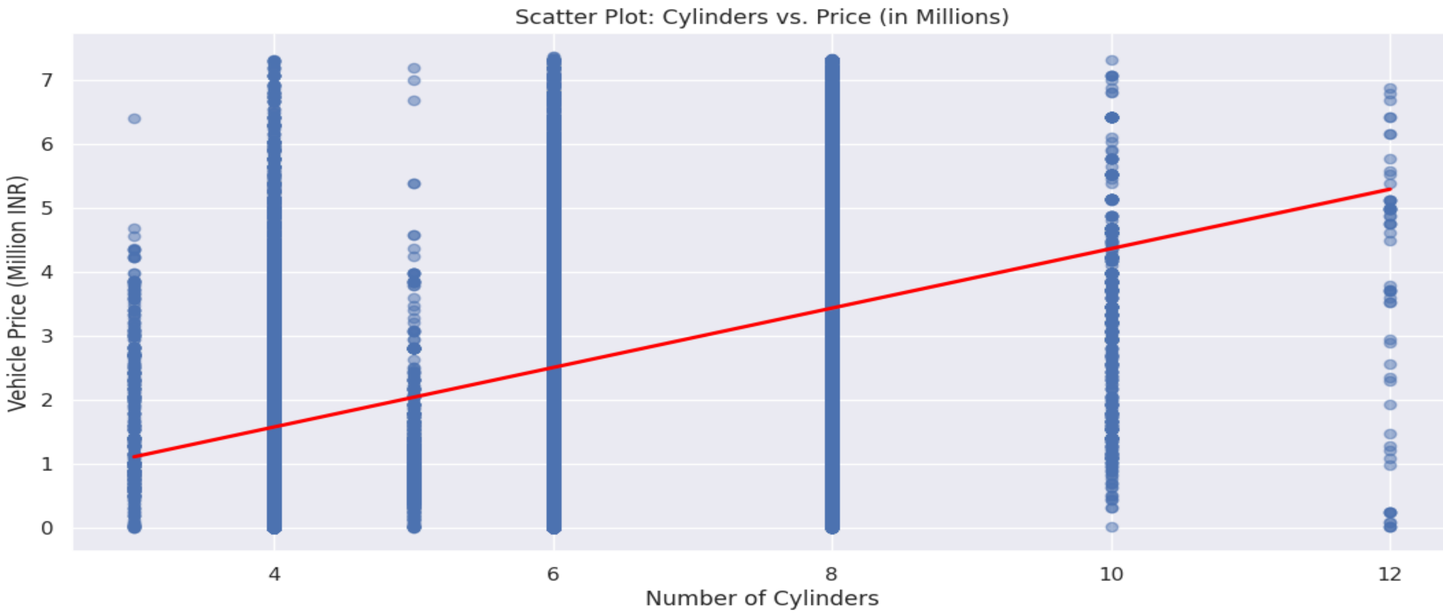


- **Methodology:** Examined odometer distributions and their correlation with price.
- **Conclusion:** Cars with odometer readings between 0–0.5 lakh km dominate the dataset, reflecting buyer preference for low-mileage vehicles.
- **Decision:** Reject the null hypothesis (H_0). Odometer readings significantly impact vehicle demand and price.
- **Key Findings:**
  - Violin plot shows higher prices for low-mileage cars, confirming strong demand.
  - Correlation heatmap indicates a moderate negative relationship (-0.51) between odometer readings and price.
- **Business Implications:**
  - **Increase Revenue:** Price low-mileage cars at a premium due to high demand.
  - **Boost Sales:** Offer value-added deals for high-mileage cars to attract buyers.
  - **Customer Satisfaction:** Maintain a well-balanced inventory, prioritizing low-mileage vehicles.

Hypothesis 3: Impact of Cylinders on Vehicle Price

- **Null Hypothesis (H_0):** The number of cylinders in a car's engine does not affect its price.
- **Alternate Hypothesis (H_1):** The number of cylinders significantly impacts vehicle price.





- **Methodology:** Analysed cylinder count distributions and their relationship with price.
- **Conclusion:** Cars with 4-cylinder and 6-cylinder engines dominate due to their balance of performance and affordability. Higher-cylinder cars cater to niche buyers.
- **Decision:** Reject the null hypothesis (H_0). The number of cylinders significantly impacts vehicle prices.

- **Key Findings:**
  - **Regression Analysis:** Each additional cylinder increases the price by ₹464,400 on average.
  - **Positive Correlation:** Vehicles with more cylinders tend to have higher prices.
  - **Market Trend:** High-cylinder cars (6, 8, 12) show greater price variability, indicating premium market demand.
- **Business Implications:**

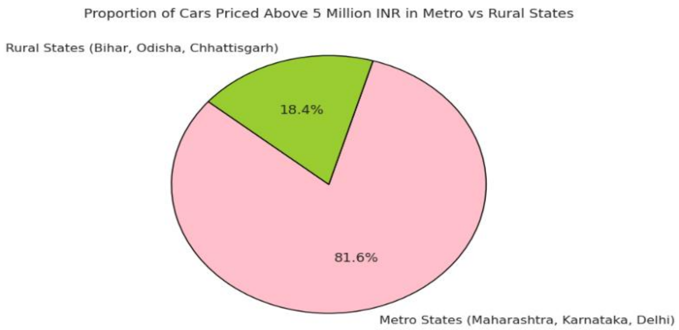
Premium Pricing: Optimize pricing strategies for high-cylinder vehicles.
Target Market: Focus on high-cylinder cars in premium and performance-driven segments.
Inventory Strategy: Stock more high-cylinder vehicles in regions with demand for luxury and performance cars.

Hypothesis 4: Relationship Between Income Levels and Average Car Prices

- **Null Hypothesis (H_0):** State income levels do not influence average car prices.
- **Alternate Hypothesis (H_1):** State income levels significantly affect average car prices.
- **Methodology:** Categorised states by income levels and compared average car prices across these categories.
- **Conclusion:** High-income states show higher average car prices due to greater purchasing power, while low-income states prefer budget-friendly options.
- **Decision:** Failed to reject the null hypothesis (H_0). There is no significant difference in average car prices between high-income and low-income states.
- **Key Findings:**
  Minimal Variation: The overall average car prices across different income levels are very close.
  No Strong Correlation: State income levels do not significantly influence car pricing trends.
- **Business Implications:**
  Focus on Other Factors: Prioritize vehicle attributes like mileage, fuel type, and brand over state income when setting prices.
  Explore Demand Drivers: Investigate other factors (e.g., buyer preferences, financing options) that might influence regional pricing trends.

Hypothesis 5: Luxury Cars Have Higher Demand in Tier One Cities Compared to Rural States

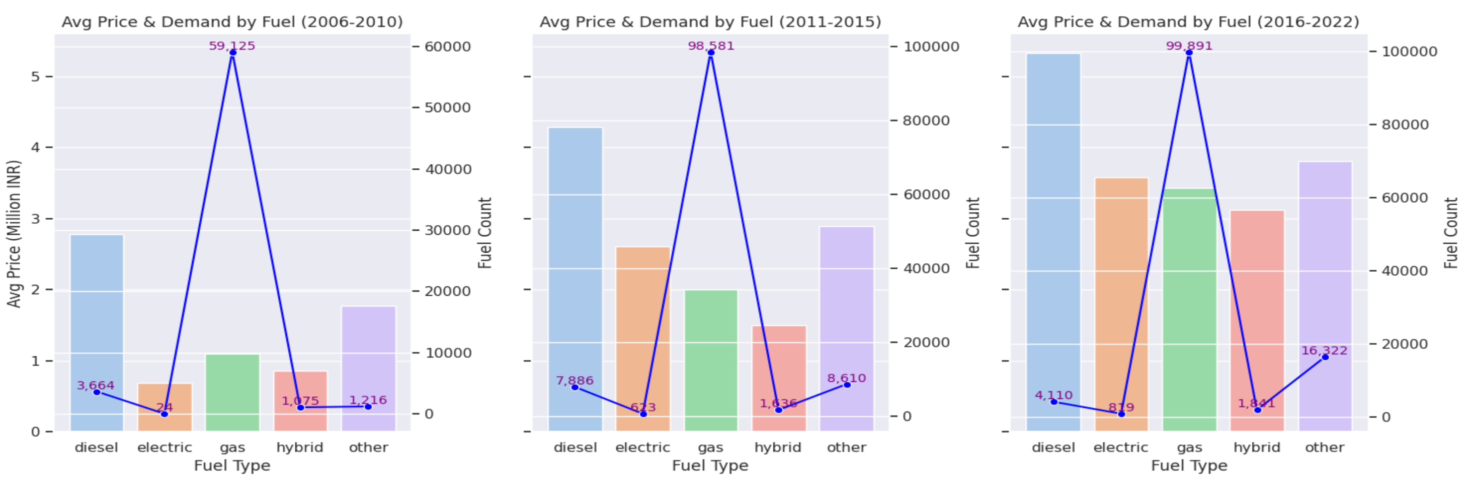
- **Null Hypothesis (H_0):** Luxury cars do not have higher demand in Tier 1 cities compared to rural states.
- **Alternate Hypothesis (H_1):** Luxury cars have significantly higher demand in Tier 1 cities compared to rural states.



- **Methodology:** Analysed the distribution of luxury cars across urban and rural regions using sales data.
- **Conclusion:** Tier 1 cities exhibit higher demand for luxury vehicles due to affluent buyers, while rural areas focus more on affordable options.
- **Decision:** Reject the null hypothesis (H_0). Luxury car demand is significantly higher in metro states than in rural states.
- **Key Findings:**
  - **Luxury Car Distribution:** 81.6% of cars priced above ₹5 million are in metro states, while only 18.4% are in rural states.
  - **Higher Disposable Income & Better Infrastructure:** Metro areas support luxury car ownership with better purchasing power and road conditions.
- **Business Implications:**
  - **Target Market:** Expand luxury offerings in urban markets while maintaining budget-friendly options for rural regions.
  - **Marketing Strategy:** Focus advertising, financing, and dealership expansion in metro cities to capture high-end buyers.
  - **Regional Expansion:** Explore premium offerings in emerging urban centers with rising affluence.

Hypothesis 6: Fuel Price and Demand Analysis for Different Manufactured Year Cars

- **Null Hypothesis (H_0):** The average price of vehicles does not significantly vary across different fuel types and manufacturing year brackets, and fuel type count does not influence pricing trends.
- **Alternate Hypothesis (H_1):** The average price of vehicles significantly varies across different fuel types and manufacturing year brackets, and fuel type count influences pricing trends.



- **Methodology:** Analyzed fuel type distribution and average car prices across three manufacturing year brackets (2006-2010, 2011-2015, 2016-2022) using visualizations.
- **Conclusion:**
    - Diesel vehicles maintain higher resale value despite moderate availability.
    - Gas vehicles dominate the market but show price fluctuations.
    - Electric & Hybrid cars see rising adoption in 2016-2022, indicating shifting consumer preferences.
- **Insight:**
    - **Diesel:** Price as premium listings due to sustained demand.
    - **Gas:** Adjust pricing dynamically based on supply-demand trends.
    - **Electric & Hybrid:** Expand inventory to meet growing market interest.
    - Optimize stock based on fuel type demand to maximize revenue.
- **Decision:** Reject the null hypothesis (H_0). Fuel type and manufacturing year significantly impact car prices and demand, requiring strategic pricing and inventory management.

Recommendations Based on EDA and Hypotheses

1. Optimizing Pricing Strategy

- Set pricing within ₹0.5M–₹3M, as this range has the highest transaction volume.
- Adjust gas vehicle pricing dynamically, given their highest count but fluctuating demand.
- Price diesel cars at a premium, as they retain strong resale value and demand stability.
- Introduce tiered pricing for electric & hybrid vehicles, as their demand has risen post-2016.
- Reduce prices for high-mileage (>1 lakh km) and older (>10 years) cars to improve sales.

---

2. Enhancing Inventory Management for Market Demand

- Stock more low-mileage cars (<1 lakh km), as they dominate buyer preference.
- Prioritize mid-age (3–8 years old) vehicles, as they form the largest segment.
- Expand gasoline vehicle inventory, as they are the most available and in demand.
- Maintain a strong mix of 4-cylinder and 6-cylinder cars, as they dominate registrations.
- Balance stock for automatic transmissions, as they are the most preferred option.

---

3. Expanding Market Reach & Regional Growth Strategy

- Focus on Tier-1 cities for premium cars, as metro areas show higher luxury car demand.
- Increase budget-friendly vehicle availability in rural regions, where affordability is key.
- Expand listings for Ford, Chevrolet, and Toyota models, as they dominate demand.
- Leverage the rising popularity of SUVs & sedans, which are the most sought-after body types.
- Introduce region-specific pricing strategies, based on historical demand trends and buyer preferences.

---

4. Improving Customer Experience & Service Offerings

- Prioritize vehicles in "Good" and "Excellent" condition, as they have the highest demand.
- Use targeted promotions for high-mileage and older vehicles to clear low-demand stock.
- Focus on popular models like the Ford F-150 and Chevrolet Silverado 1500, as they dominate transactions.
- Monitor EV and hybrid adoption trends, gradually increasing their inventory.
- Leverage AI-based customer engagement tools, such as personalized recommendations and predictive buying trends.