

UCMI - USED CAR MARKET INTELLIGENCE

Data cleaning, exploratory analysis & interactive
visualization of used-car trends

PRESENTATION

Team P01

By:

Alaa Gabr

Ahmed Amin

Motivation & Goals

Why this project?

- The global used-car market is large and growing: used-car purchases remain a popular alternative to new cars.
- Without data, buyers/sellers often make decisions based on intuition — leading to overpaying, undervaluing, or missing important patterns (e.g. brand depreciation, fuel-type popularity, age vs price).
- A data-driven approach helps bring transparency: define realistic pricing, detect market trends (fuel preferences, brand demand), help stakeholders make informed decisions (when to buy/sell, what to expect).

How the Project Works

- **Data Cleaning & Preprocessing:** Fixing missing data, removing duplicates, standardizing formats. This step ensures data quality before analysis. In data-analysis workflows this is essential to avoid misleading results.
- **Exploratory Data Analysis (EDA):** Use statistical summaries and visualizations (histograms, boxplots, scatter plots, correlations) to explore distributions (price, age, mileage, fuel type, brand), detect outliers, and understand relationships.
- **Feature Engineering / Segmentation:** Create relevant variables (e.g. price-per-year-of-age, depreciation rate, fuel type categories, brand grouping) to enable meaningful comparisons across vehicles.
- **Visualization & Dashboard / Reporting:** Translate findings into clear charts, graphs, and possibly interactive dashboards — making insights accessible to non-technical audiences. Effective visualization helps communicate complex data clearly.

About the Dataset:

- **Dataset name :** Car data
- **Number of rows :** 4345 rows
- **Number of columns (Features) :** 8 columns

Brief about each feature:

- **name** : Model and variant of the car
- **year** : Manufacturing year
- **selling_price** : Listed selling price
- **km_driven** : Total distance driven
- **fuel** : Fuel type
- **seller_type** : Type of seller (First / Second)
- **transmission** : Manual or automatic
- **owner** : Ownership history (first owner, second owner, etc.)

Data Cleaning & Preprocessing

Objective: Ensure the dataset is accurate, consistent, and ready for analysis.

Key Steps Performed:

- **Handling Missing Values:**

- Identified incomplete rows (missing price, brand, fuel type, mileage).
- Applied imputation strategies or removed unrecoverable entries.

- **Removing Duplicates & Invalid Records:**

- Eliminated repeated car listings and entries with unrealistic values (e.g., negative mileage, zero price).

- **Standardizing Data Formats:**

- Unified brand names, fuel categories, date formats, and currency values.

- **Outlier Detection & Treatment**

- **Data Type Conversion**

Feature Engineering

Objective: Transform raw data into meaningful variables that capture hidden patterns.

Features Created / Enhanced:

- Brand & Model Extraction
 - Brand extracted from the first token in the car name
 - Model extracted from the remaining portion of the name

Purpose: Enables brand-level comparisons and model-specific pricing trends.

- Luxury Brand Indicator
- Flag identifies premium manufacturers

Purpose: Helps analyze premium-market pricing behavior and depreciation differences.

- Car Age
- Age computed based on manufacturing year

Purpose: Age is a primary predictor of price and depreciation.

Feature Engineering

- Old / New Car Classification

Binary flag based on age threshold

Purpose: Allows segmentation of listings into mature vs. newer vehicles.

- km_per_year

Usage intensity indicator

Purpose: Normalizes mileage relative to age — distinguishes lightly used vs. overused cars.

- High / Low Usage Flag

Compares car mileage to overall market median

Purpose: Supports pricing analysis influenced by wear-and-tear.

Exploratory Data Analysis (EDA)

Objective: Objective: Discover patterns, relationships, and market insights.

Analyses Conducted:

- **Distribution Analysis:**

Histograms of price, mileage, and age reveal typical ranges and outliers.

- **Brand Comparison:**

Boxplots comparing prices across brands show which retain value better.

- **Fuel Type Trends:**

Bar charts illustrate the market share of petrol, diesel, and hybrid models.

- **Correlation Analysis:**

Heatmaps highlight relationships (e.g., price ↓ as mileage ↑).

- **Segmentation by Age & Mileage:**

Scatterplots identify ideal combinations of affordable price and reliability.

Fuel Type Distribution Insights

Diesel Dominates the Used-Car Market

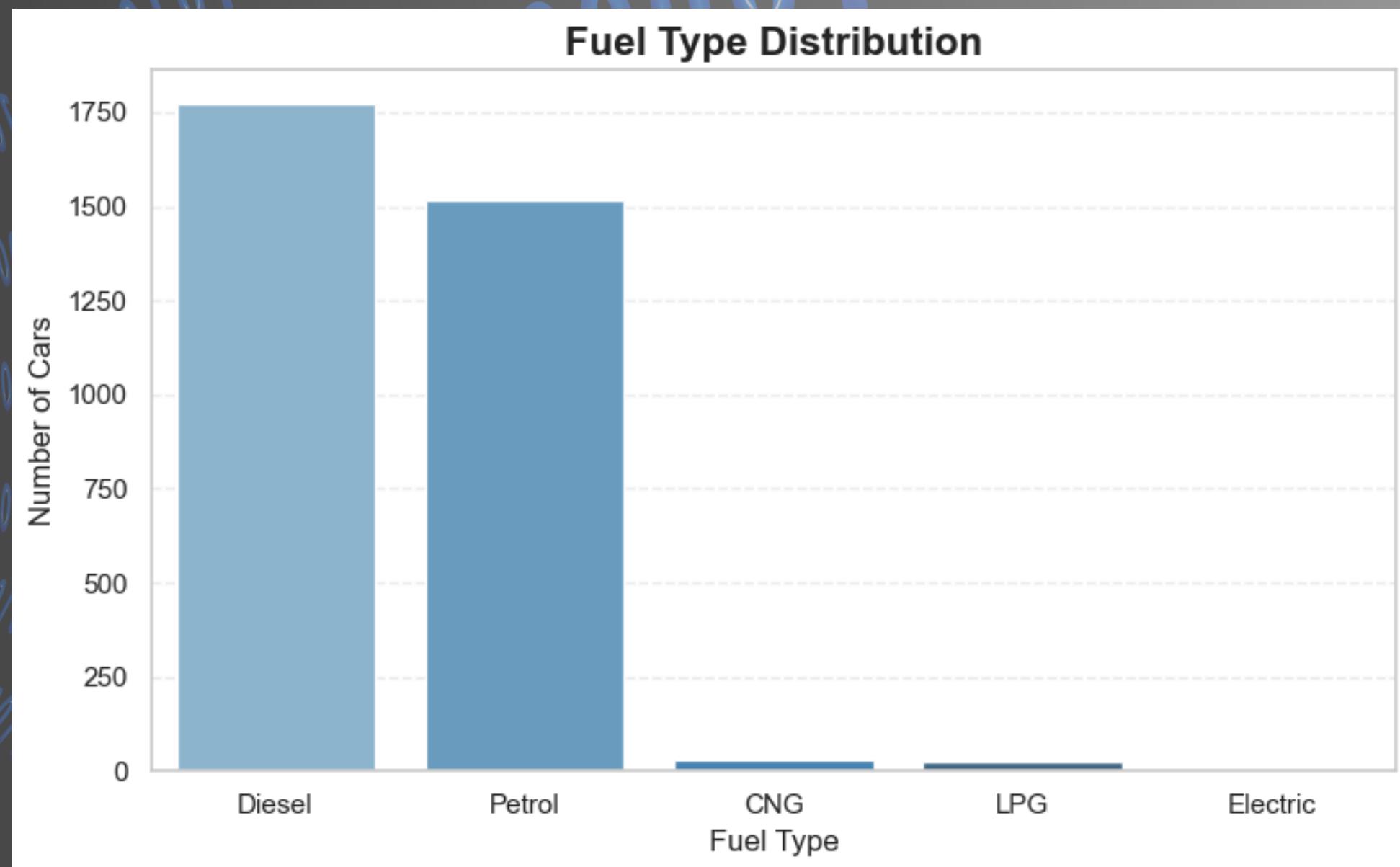
Diesel cars represent the **largest share** of available listings.

Preference for diesel engines due to:

- Better long-distance efficiency
- Higher torque for larger vehicles
- Lower fuel cost during earlier years of purchase

Petrol Close Behind with Strong Adoption

Petrol vehicles form the **second-largest** segment, not far behind diesel.



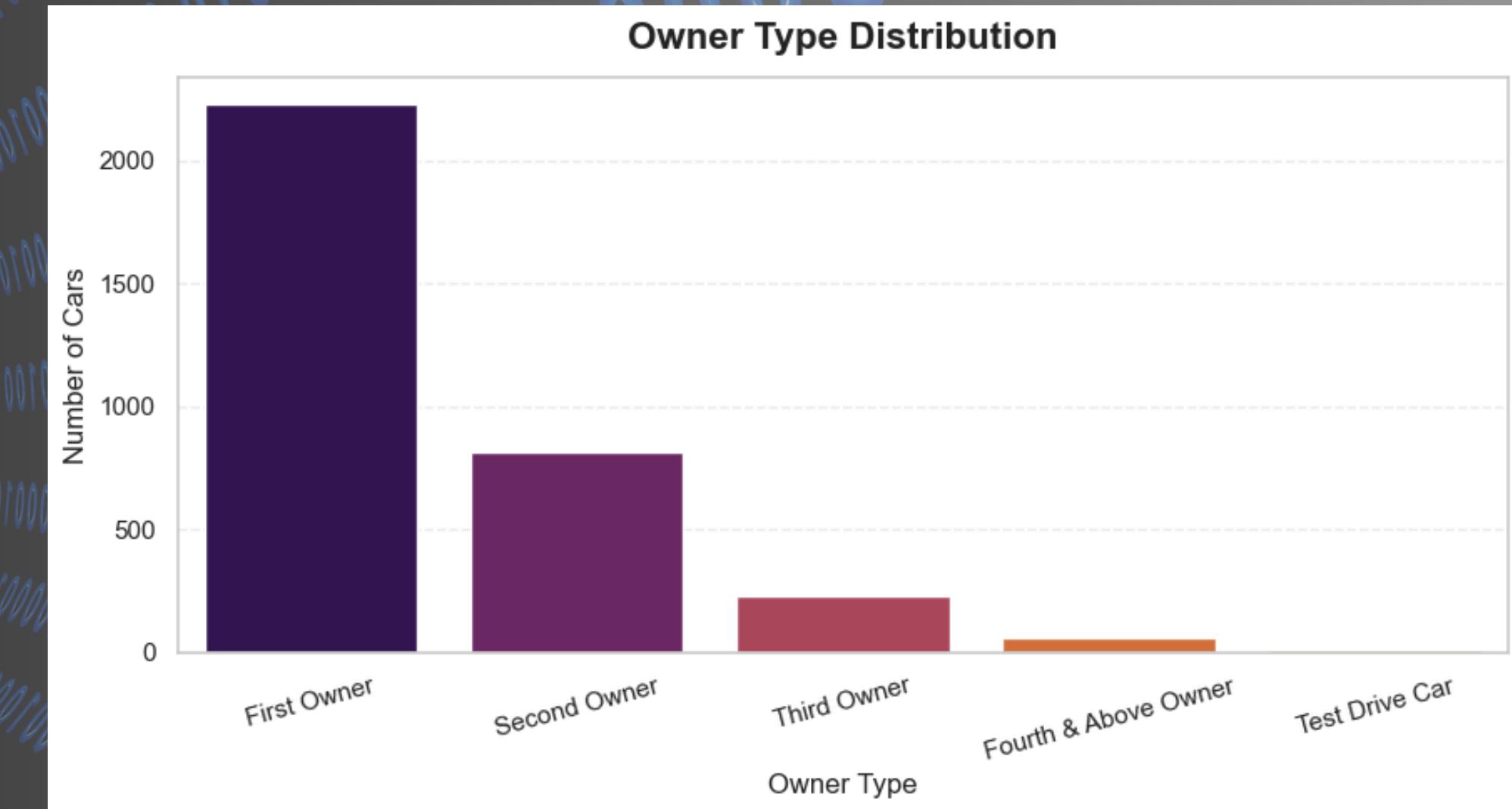
Ownership Pattern Insights

First-Owner Cars Dominate the Market

- Most listings are **first-owner** vehicles, indicating:
 - Higher trust in original owners
 - Easier verification of service history and mileage
 - Better resale value due to perceived reliability

Significant Drop for Second-Owner Cars

- These cars usually:
 - Offer a lower price point
 - Attract budget-focused buyers
 - Still retain reasonable documentation and condition

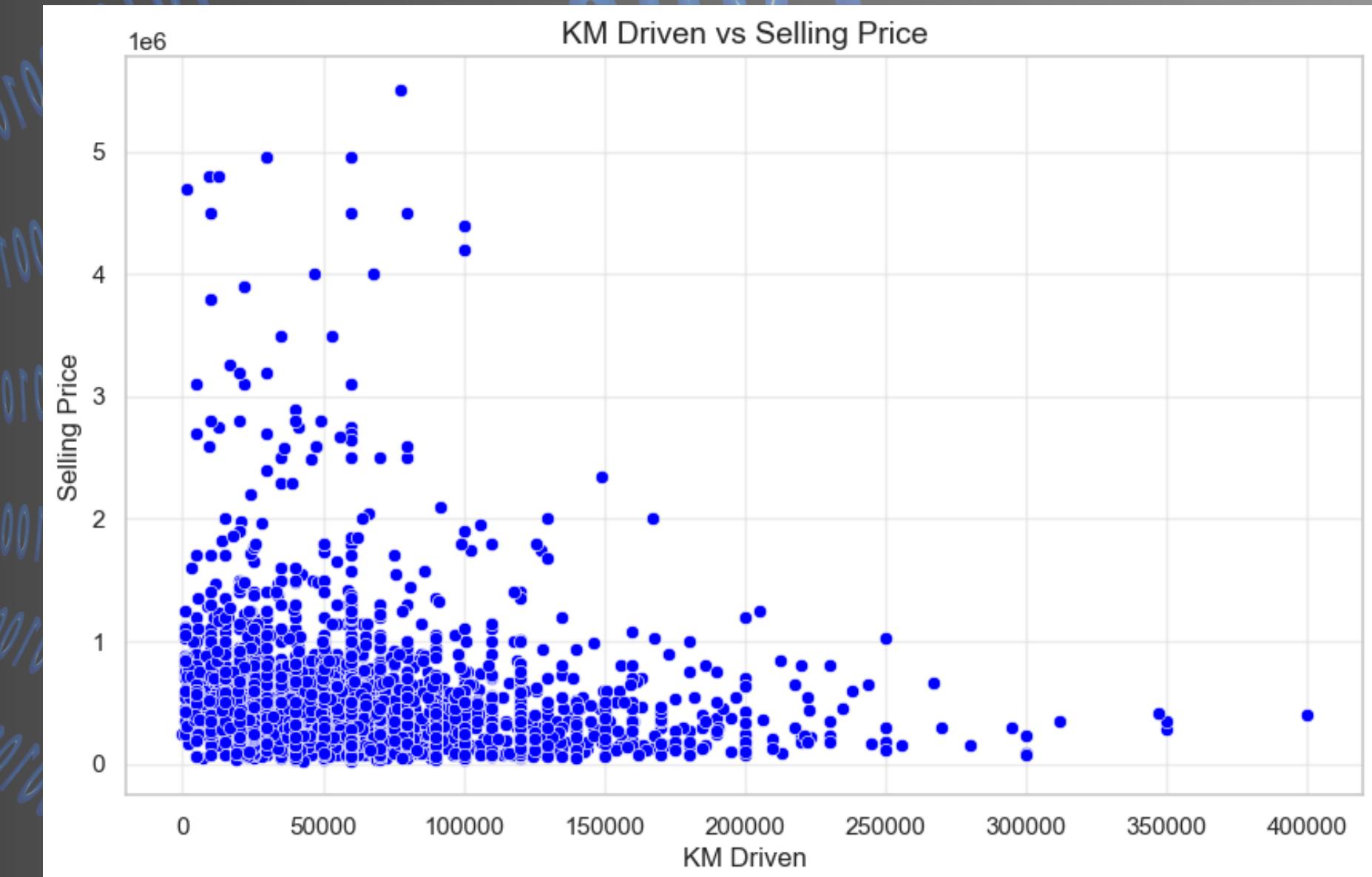


Impact of Mileage on Selling Price

Strong Negative Relationship

- As kilometers driven increase, the selling price sharply declines.
- Mileage acts as a proxy for:
 - Wear and tear
 - Maintenance costs
 - Remaining vehicle lifespan

This confirms mileage is one of the strongest determinants of resale value.



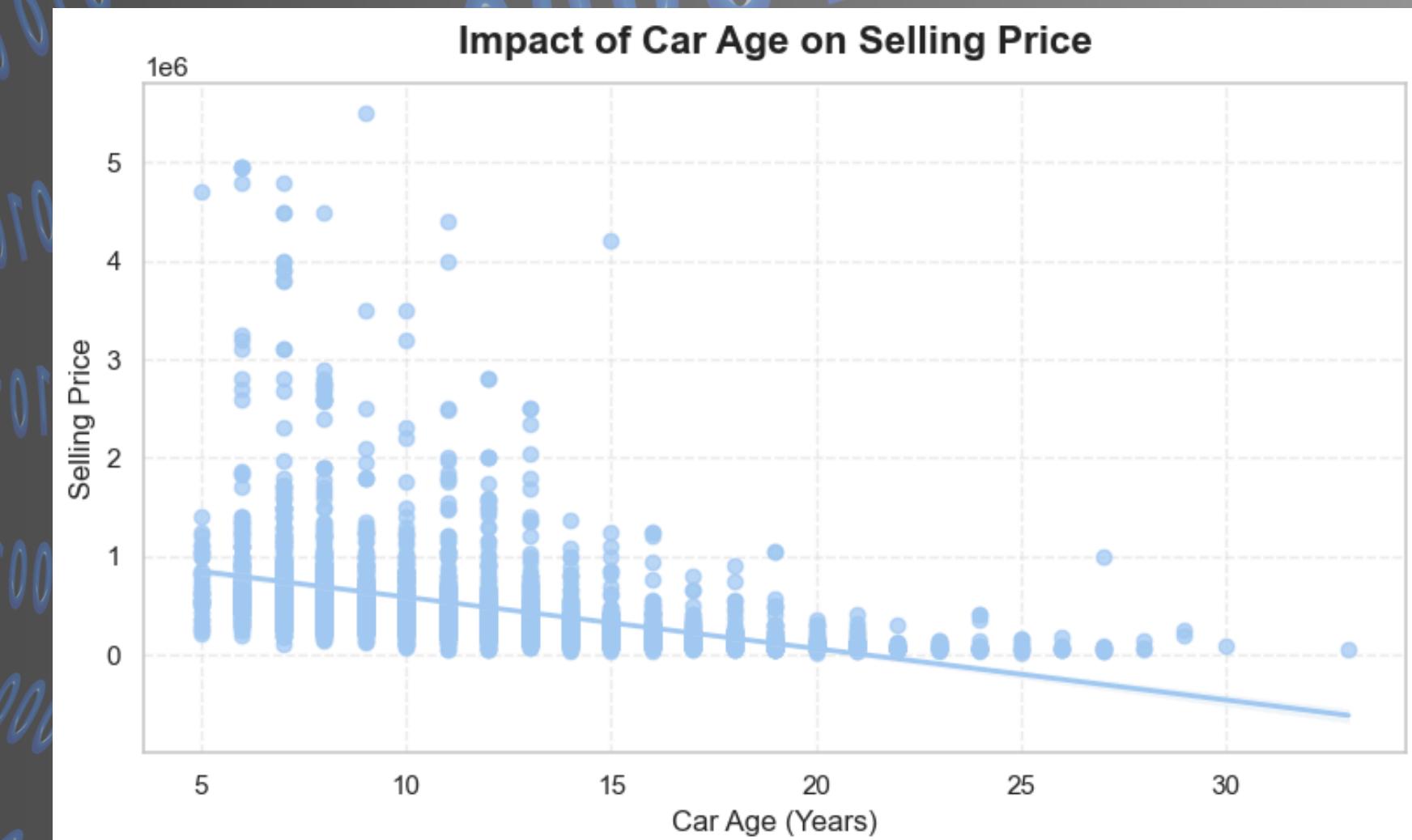
Influence of Car Age on Selling Price

Clear Depreciation Trend

The regression line shows a strong downward slope, revealing that car price decreases with age.

- Buyers associate older cars with:
 - Reduced efficiency
 - Higher probability of mechanical issues
 - Expired warranties and outdated features

Sharp Price Decline After 10 Years



Key Insights & Market Patterns (What the Data Reveals)

Depending on the data used, the analysis can surface insights such as:

Price & Value Dynamics

- Most cars are priced below ₹10 lakh, forming the market's core value band.
- Very expensive listings are rare outliers, mainly belonging to luxury brands.
- Resale prices follow rational depreciation patterns, with mileage and brand prestige acting as primary price drivers.

Fuel-Type Preferences

- Diesel is the dominant fuel type, followed closely by petrol.
- CNG and LPG cars are nearly absent, showing niche use.
- Electric vehicles are statistically negligible, indicating minimal EV penetration in the resale market.

Brand Retention & Depreciation

- Luxury brands(BMW, Mercedes, Audi, Volvo, etc.) retain higher value even at older ages and higher mileage, defying normal depreciation curves.
- Non-luxury brands experience steeper price drops, making them budget-friendly but with lower long-term value retention.

Key Insights & Market Patterns (What the Data Reveals)

Depending on the data used, the analysis can surface insights such as:

Ownership, Usage & Condition

- First-owner cars overwhelmingly dominate the dataset—buyers clearly trust cars with fewer ownership cycles.
- Cars with more than 100,000 km see sharp price reductions, confirming that usage intensity matters more than age alone.
- Cars under 10 years old retain stronger resale value, validating the effectiveness of the is_old threshold.

Overall Conclusion

The used-car market is shaped by brand perception, mileage intensity, and fuel preference. Buyers prioritize condition, trust, and long-term value, not just age or sticker price—making this dataset a powerful foundation for pricing strategy, buyer advisory, and future predictive modeling.

Market Demand Patterns

- The market gravitates toward mid-price, moderately used vehicles—the sweet spot of affordability and reliability.
- Low-kilometer, well-maintained cars command premiums, regardless of age.
- SUVs and luxury models appear in higher price tiers, reflecting status-driven buyer segments.

Visualization & Dashboard / Reporting

Objective: Communicate results clearly and enable interactive exploration.

Dashboard Components:

- Price Distribution Graphs
 - Helps users understand typical pricing bands.
- Brand Comparison Charts
 - Visualizes resale strength among brands.
- Fuel Type Breakdown
 - Shows market share and evolving preferences.
- Age vs. Price Scatterplot
 - Highlights depreciation and value-for-money zones.

Benefits:

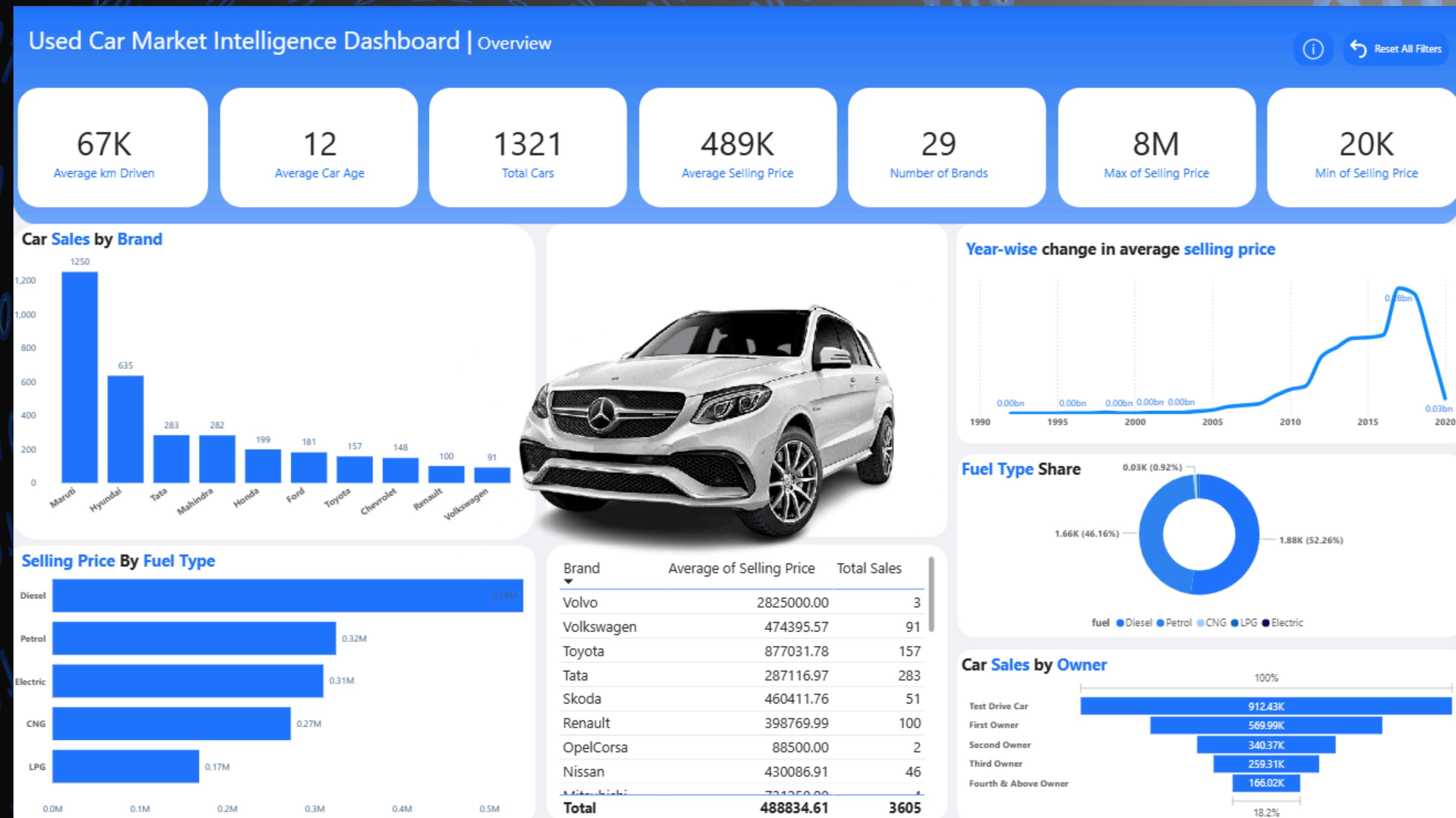
- Translates complex data into actionable insights.
- Enables non-technical users to filter and explore results.
- Supports informed decisions for buyers, sellers, and analysts.

Outcome:

A user-friendly reporting layer that transforms raw data and analysis into useful market intelligence.

-

Dashboard Preview



THANK YOU!

FOR YOUR ATTENTION