Automobile Price Analysis and Prediction

Exploring the relationship between vehicle specifications and pricing through data-driven analysis and machine learning models.

Project Overview

Understanding Automobile Pricing

The automobile industry is highly competitive, where pricing plays a critical role in customer decision-making and business profitability. This project performs end-to-end analysis using 26 automobile attributes.

We explore relationships between vehicle features and price, conduct detailed exploratory analysis, and build regression models for accurate price prediction.

Project Objectives

Data Understanding

Analyze structure and characteristics of automobile data

Data Preprocessing

Clean and prepare raw data for analysis and modeling

Visual Analysis

Perform EDA using comprehensive visualizations

Key Factors

Identify features influencing automobile prices

Model Building

Build and evaluate regression prediction models

Insights

Interpret results for data-driven decisions

Dataset Overview

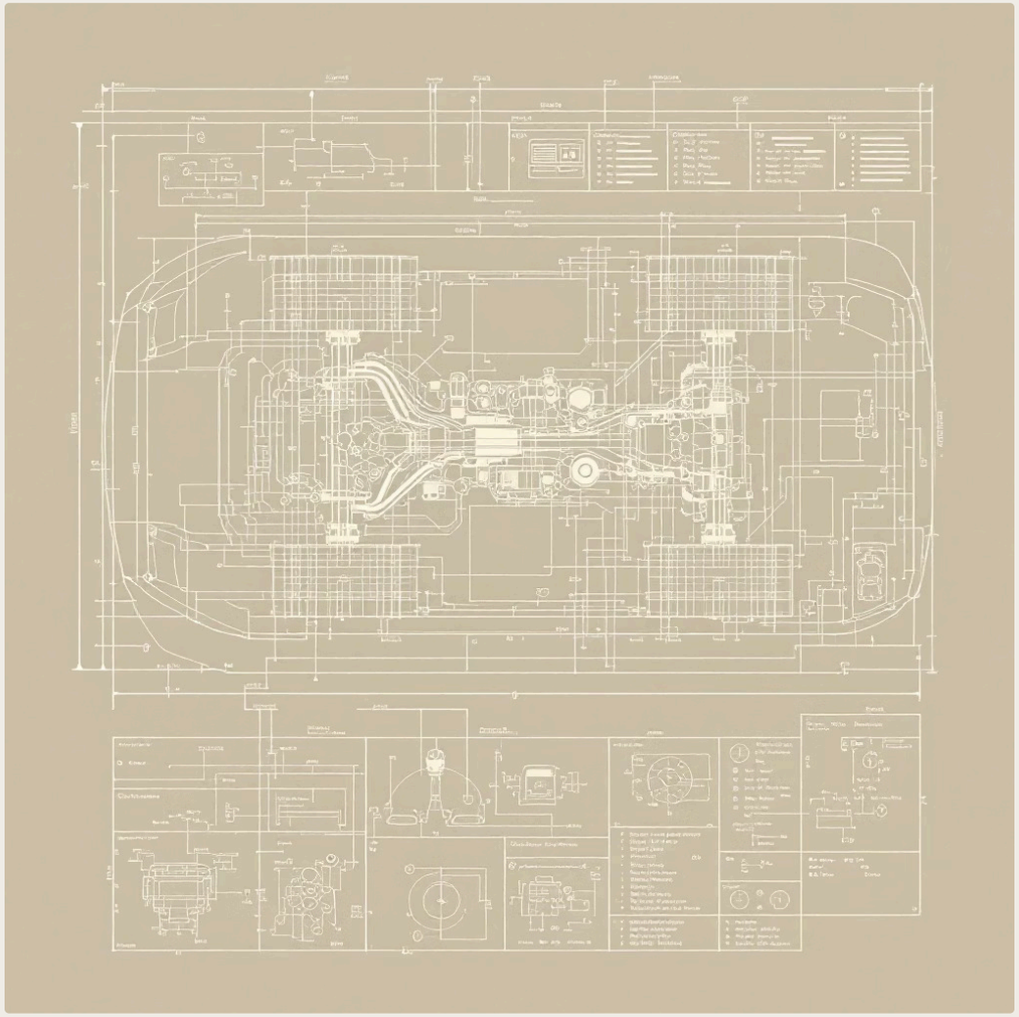
The dataset consists of **26 attributes** representing technical specifications and efficiency metrics of automobiles.

Categorical Attributes

- Make, fuel-type, aspiration
- Body-style, drive-wheels
- Engine-type, fuel-system

Numerical Attributes

- Engine-size, horsepower
- Curb-weight, city-mpg, highway-mpg
- Compression-ratio, **price**



Data Cleaning and Preprocessing

01

Handling Missing Values

Identified and replaced missing values using mean, median, or mode techniques

02

Data Type Conversion

Converted numerical attributes from object/string types to proper numeric formats

03

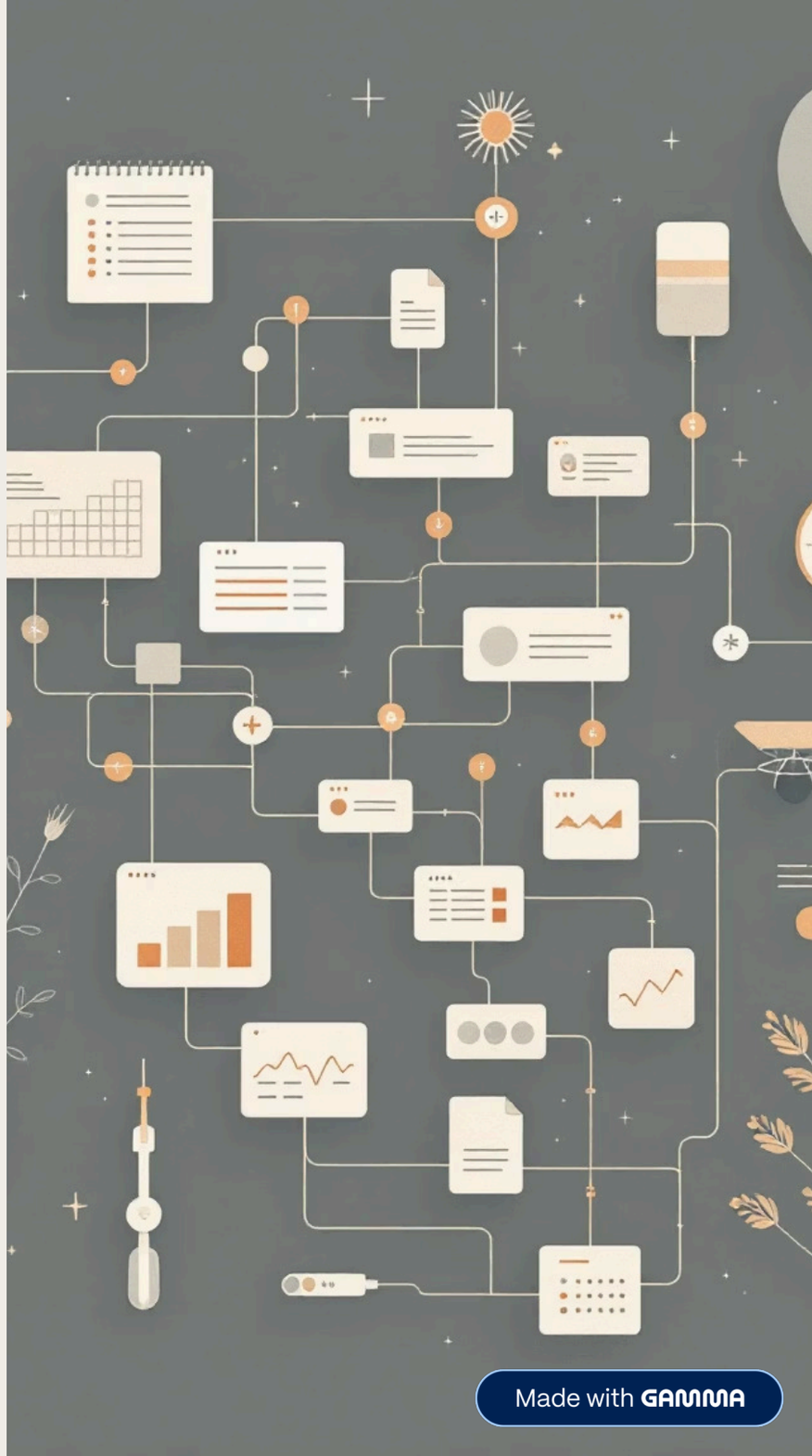
Feature Encoding

Transformed categorical variables into numerical representations for modeling

04

Feature Scaling

Normalized numerical features to ensure uniform scale for regression models



Exploratory Data Analysis

Key Findings

Price Distribution

Right-skewed distribution indicating presence of high-end premium vehicles

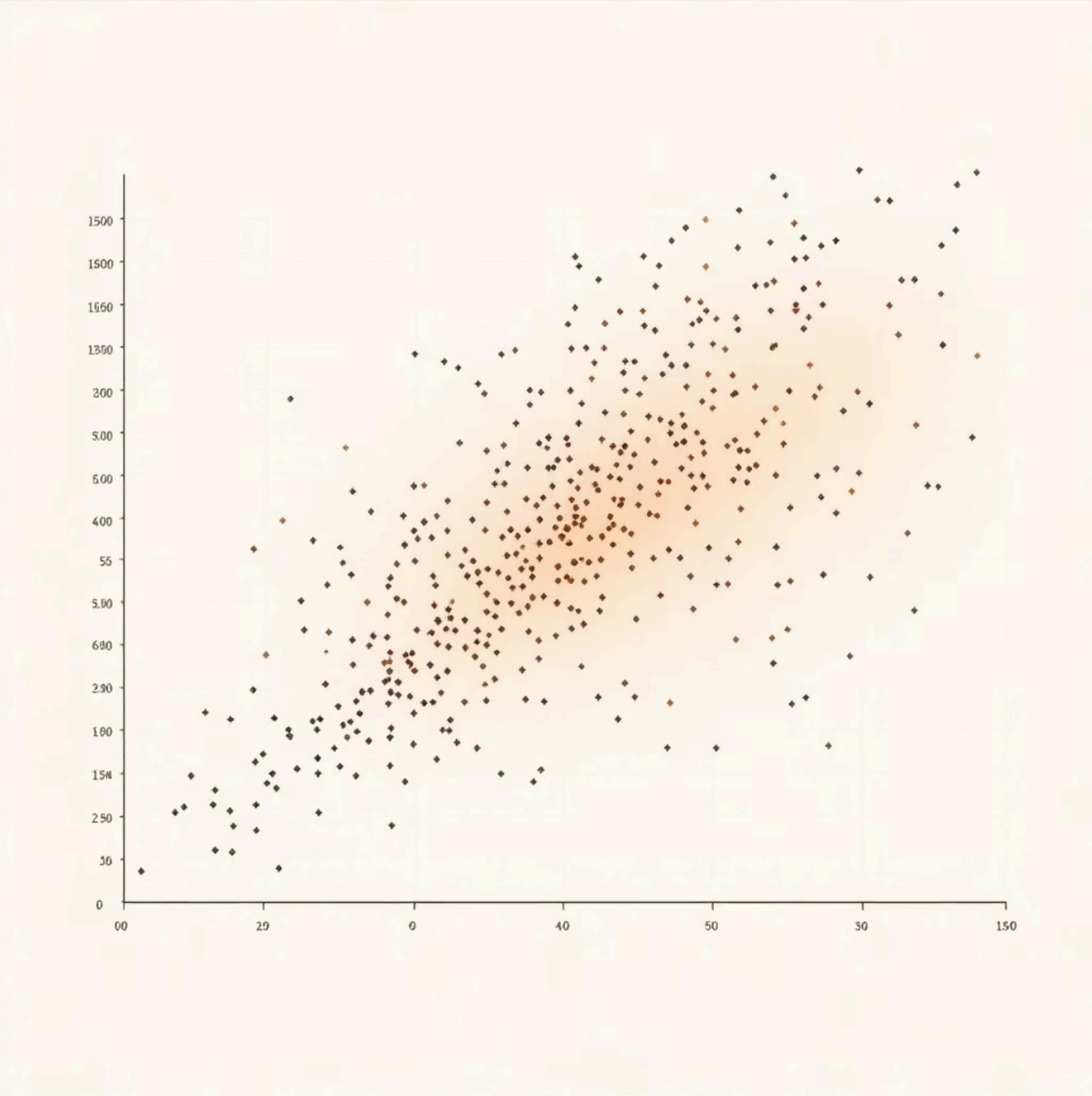
Engine Specifications

Strong positive correlation: higher engine capacity and horsepower result in higher prices

Fuel Efficiency

Inverse relationship: vehicles with higher MPG tend to have lower prices

Feature Correlation Analysis



Key Correlations

Strong Positive

Price correlates with engine-size, horsepower, and curb-weight

Negative

Price inversely related to fuel efficiency metrics (MPG)

Multicollinearity

Observed among engine-related features

Model Development and Evaluation

Model Selection

Linear and Multiple Linear Regression models chosen for continuous target variable

Train-Test Split

Dataset divided to evaluate model generalization performance

Model Training

Models trained using preprocessed features with importance analysis

Evaluation

Performance measured using R^2 Score, MSE, and RMSE metrics

Key Insights

1

Top Predictors

Engine size, horsepower, and curb weight are strongest price predictors

2

Efficiency Trade-off

Fuel efficiency shows inverse relationship with automobile price

3

Preprocessing Impact

Data preprocessing and feature scaling significantly improve model performance

4

EDA Importance

Visualization-driven analysis essential for understanding complex relationships

Conclusion and Future Work

Project Summary

This project demonstrates a complete data analytics and machine learning workflow, from raw automobile data to meaningful insights and predictive modeling.

Showcases strong fundamentals in data analysis suitable for Data Analyst, Business Analyst, and Entry-Level ML roles.

Tools Used: Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Jupyter Notebook

Future Enhancements

- Experiment with Ridge, Lasso, and Random Forest Regression
- Perform feature selection to reduce multicollinearity
- Hyperparameter tuning for improved accuracy
- Deploy model as web application