Project 7:
Used Bike Price Analysis

By: Umesh Uday Salvi
UNID: 2025045

Project Overview

- Objective: Predict used bike prices using data analysis and ML

- Domain: Financial Analysis

- Tools: Python, Jupyter Notebook, Pandas, Matplotlib, Seaborn, Sklearn

- Target Variable: Price (in INR)

Data Cleaning

- Handled missing values in mileage, power, location, etc.

- Corrected inconsistent brand/ model names.

- Extracted CC and brand from model_name column.

- Created new features: cc, brand, age of bike

- Processed mileage, power, and kms_driven columns

Exploratory Data Analysis (EDA)

- Analysed price distribution and trends by year, brand, owner type

- Prices varies significantly by cc, power, and mileage

- Most bikes are 150cc or 350cc, 1st Owner, from brands like Bajaj and Yamaha
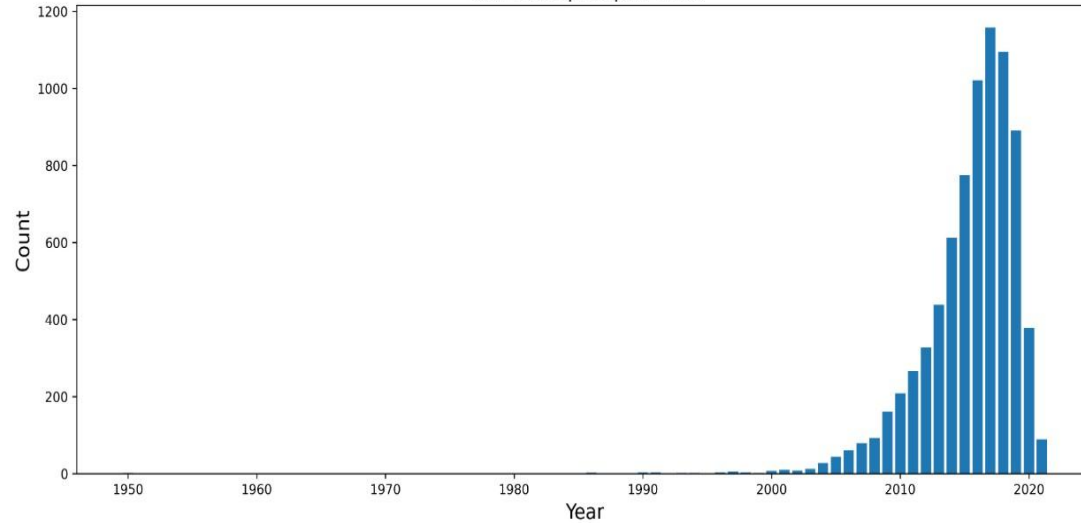
Data Visualization

- Bar Charts: Year, Brand, Owner, CC

- Histograms: Mileage, Power, Price
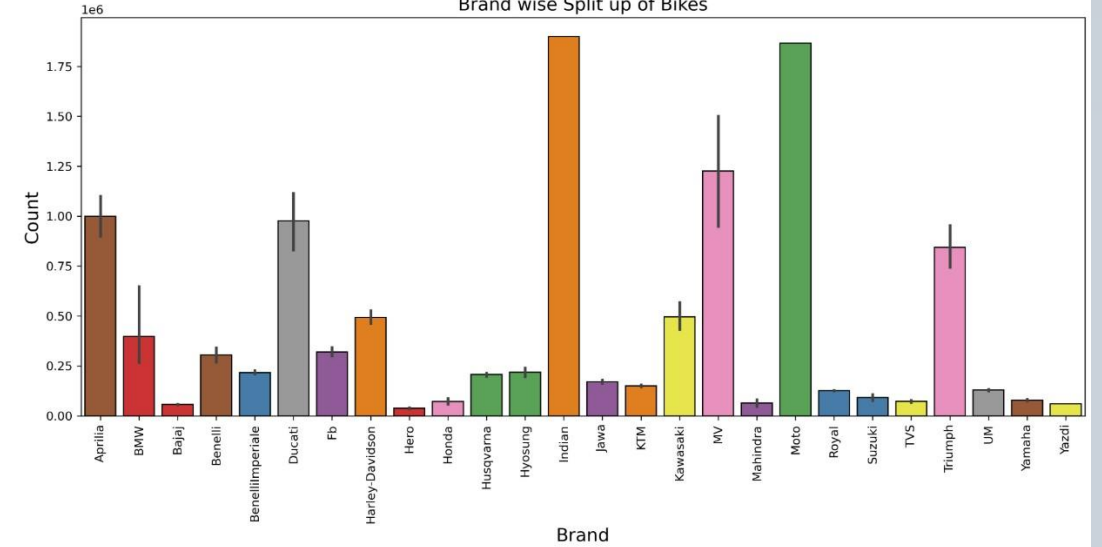
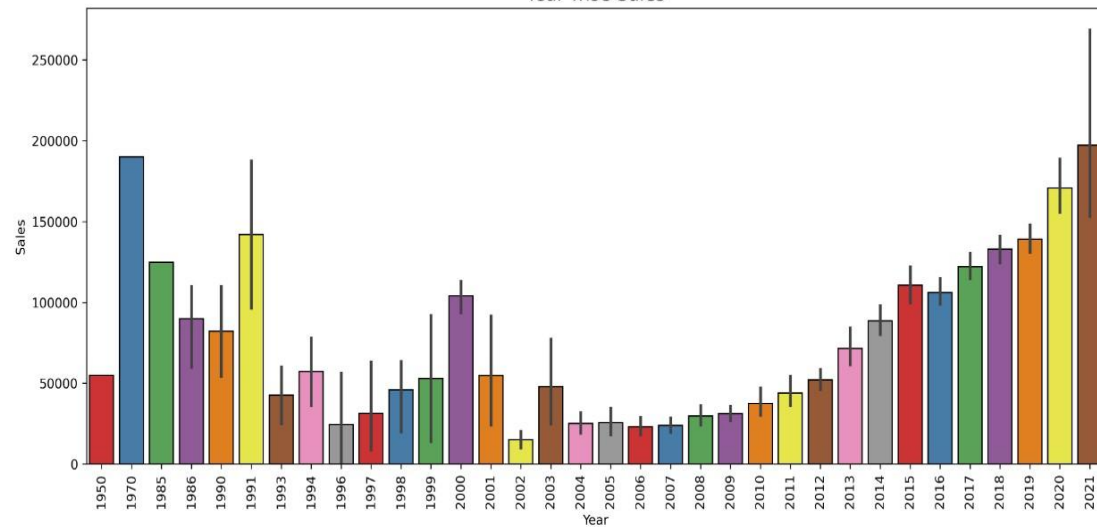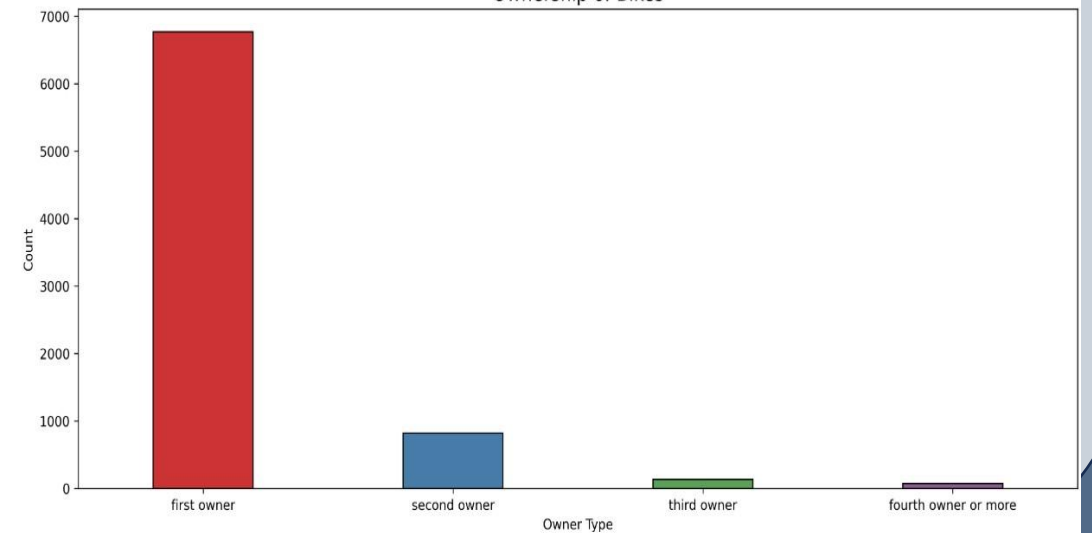- Boxplots: Price by Owner, Brand, CC

Data Visualization

Model Evaluation

- Model used: Linear Regression, Random Forest

- Metrics: Mean Squared Error (MSE), R^2 Score

- Top price influencers: model_year, mileage, power, cc

- Predictions support pricing decisions for buyers and dealers

Conclusion

- Effective use of feature engineering and EDA for price prediction

- Future Improvements: Time-based price trends

- Ensure robust preprocessing and documentation for deployment