BIKE SALES ANALYSIS

BY MRITUNJOY DAS
KIRLOSKAR INSTITUTE OF MANAGEMENT

Executive Summary

This report analyses a comprehensive dataset of 10,000 motorcycle records from India's 10 states to identify the primary drivers of resale value, customer behaviour, and market distribution. The analysis reveals a highly balanced market across fuel types (Petrol, Hybrid, and Electric) and geographic tiers.

I used Microsoft Excel and Power Query to visualise and analyse the data set. The most significant finding is that **the original purchase price is the dominant predictor of resale value**, explaining approximately **82% of the variance** ($R^2 = 0.819$). Interestingly, secondary factors such as mileage, engine capacity, and vehicle age showed negligible individual predictive power in this specific dataset, suggesting that the market primarily prices used bikes based on their original segment rather than operational wear. Bike sales in four types of cities are nearly the same. Bajaj emerged as the brand with the highest average resale retention, while Hybrid bikes lead slightly in market volume.

Introduction & Objectives

The objective of this analysis is to provide a data-driven overview of the motorcycle(including Scooty) market to assist stakeholders, dealers, individual sellers, and buyers in making informed valuation and inventory decisions. Dataset contain bike sales data from 10 states of India. I mainly focused on Descriptive and Predictive analysis to find insight from this dataset.

Primary Objectives:

- **Price Determinant Identification:** To quantify the relationship between original price and resale value.
- **Market Segmentation:** To analyse the distribution of sales across States, Brands, Fuel Types, and City Tiers.
- **Operational Profiling:** To understand usage patterns through "Average Daily Distance" and "Mileage" across different Indian states.
- **Predictive Modelling:** To establish a reliable formula for forecasting future resale prices based on historical trends.

Data Source:

https://www.kaggle.com/datasets/ak0212/indian-bike-sales-dataset

GitHub Portfolio Link:

https://github.com/MritunjoyDas1997

Dataset Details:

This dataset contains 10000 observations and 16 variables.

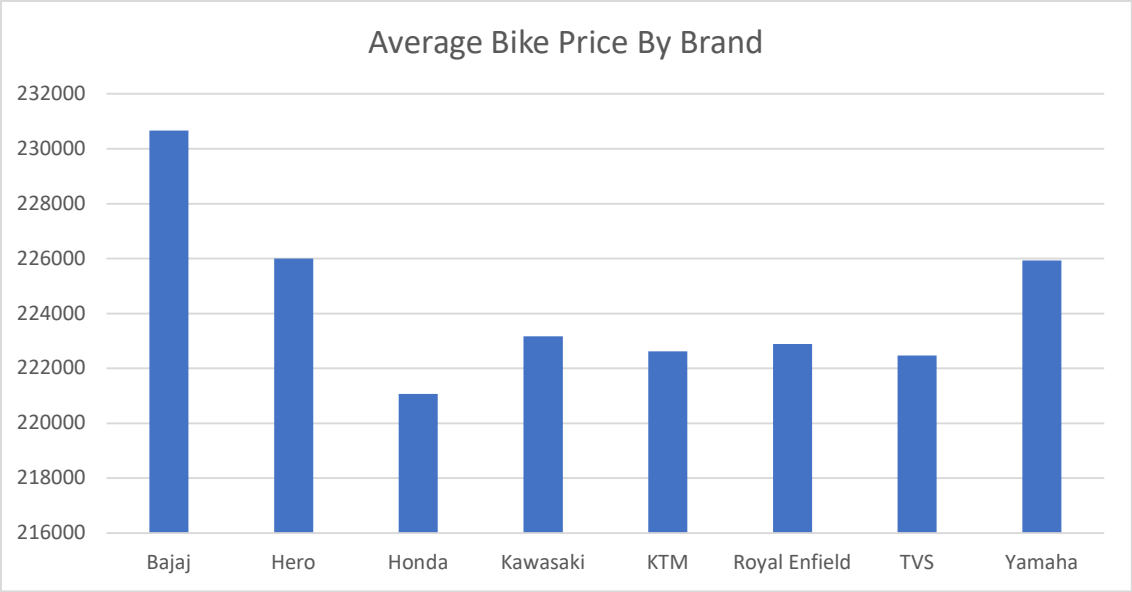
It is a Cross-Sectional Data.

This dataset contains both quantitative and qualitative data.

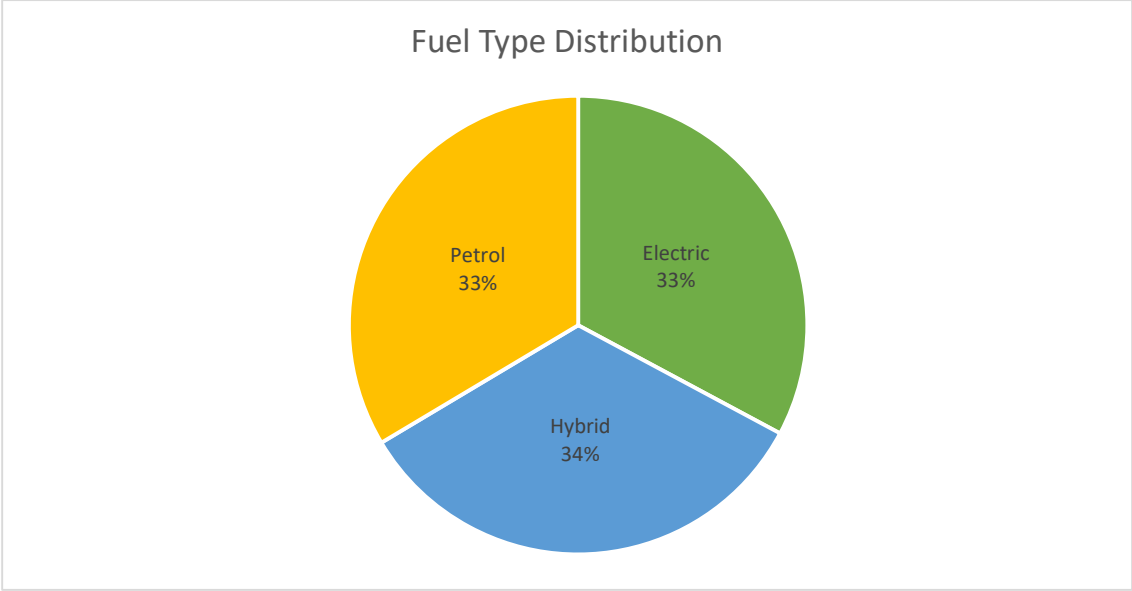
The details of the variables are given below:

Variable Name	**Data Type**	**Definition (What it represents)**
State	Nominal	The geographical location (Indian state) of the bike.
Avg. Daily Distance (km)	Ratio	Numeric distance travelled daily; has a true zero.
Brand	Nominal	The manufacturer name (e.g., Bajaj, Honda).
Model	Nominal	The specific name/version of the motorcycle.
Price (INR)	Ratio	Original cost in currency; supports all math operations.
Year of Manufacture	Interval	The calendar year of production, order and distance matter.
Engine Capacity (cc)	Ratio	Size of the engine: 0 cc would mean no engine.
Fuel Type	Nominal	Categories of power (Petrol, Electric, Hybrid).
Mileage (km/l)	Ratio	Efficiency of fuel/energy consumption.
Owner Type	Ordinal	Ranking of ownership (First > Second > Third).
Registration Year	Interval	The year the vehicle was officially recorded.
Insurance Status	Nominal	Status labels (Active, Expired, Not Available).
Seller Type	Nominal	Categorisation of the seller (Individual vs. Dealer).
Resale Price (INR)	Ratio	The used market value has a true zero.
City Tier	Ordinal	Hierarchical ranking of cities (Metro > Tier 1 > Tier 2).
Vehicle Age	Ratio	Age of the vehicle from the manufacturing year

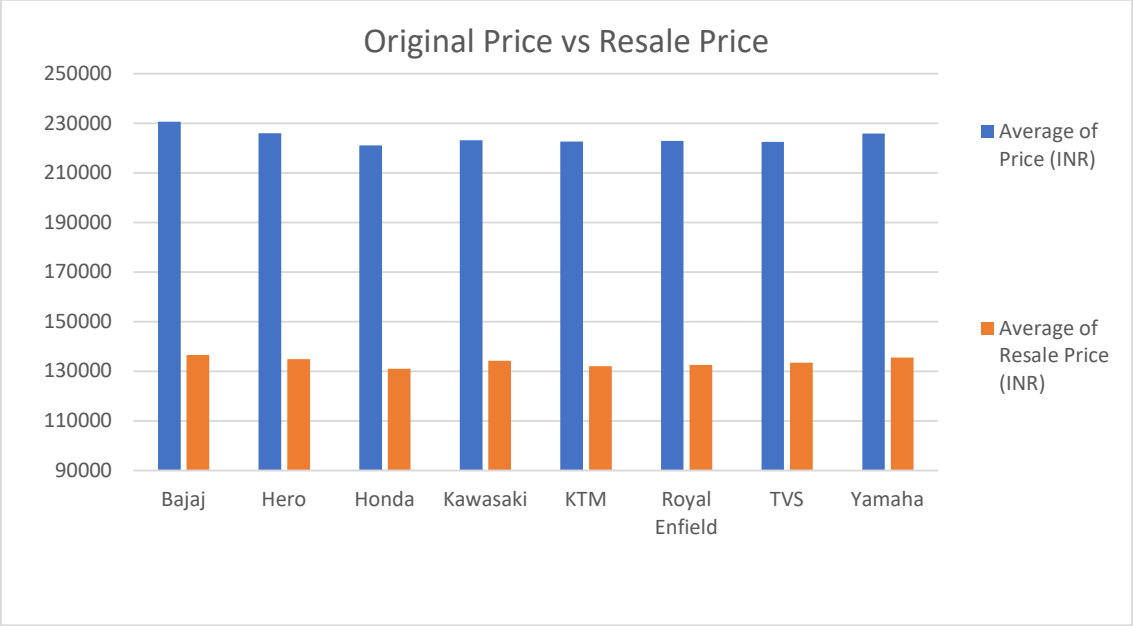
Key Findings



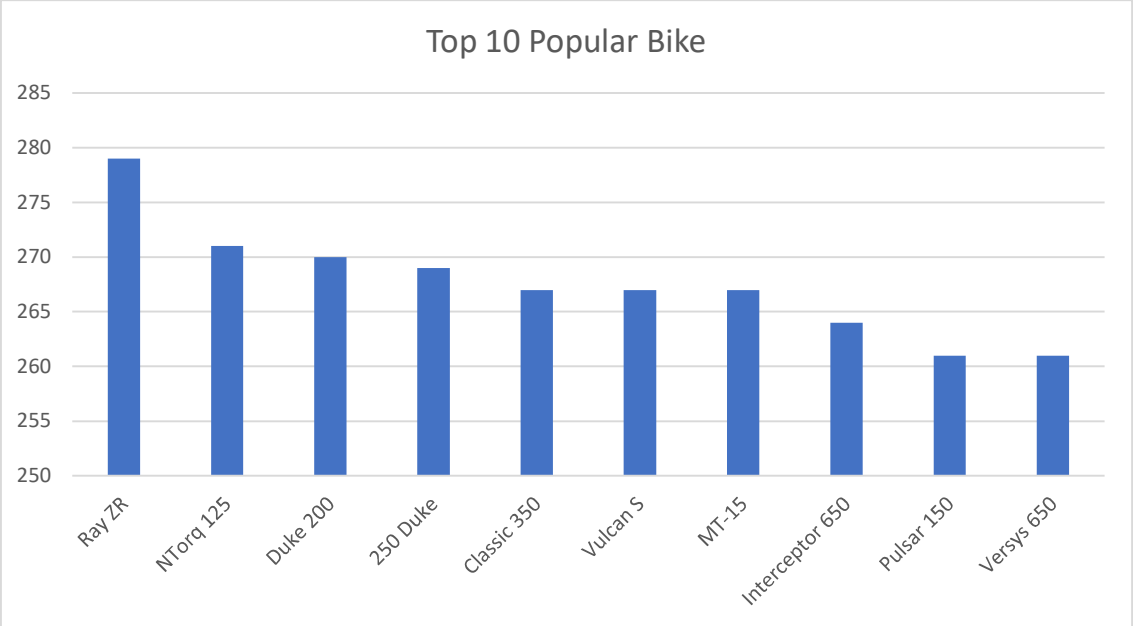
1. Bajaj has the highest average bike price, which is ₹2,30,663. The 2nd-highest average bike price is ₹2,26,001, which is Hero. Honda has the lowest average bike price, which is ₹2,21,076. It indicates Bajaj's bikes are more expensive than its competitors. While Honda offers bikes at a low price in the Indian market.



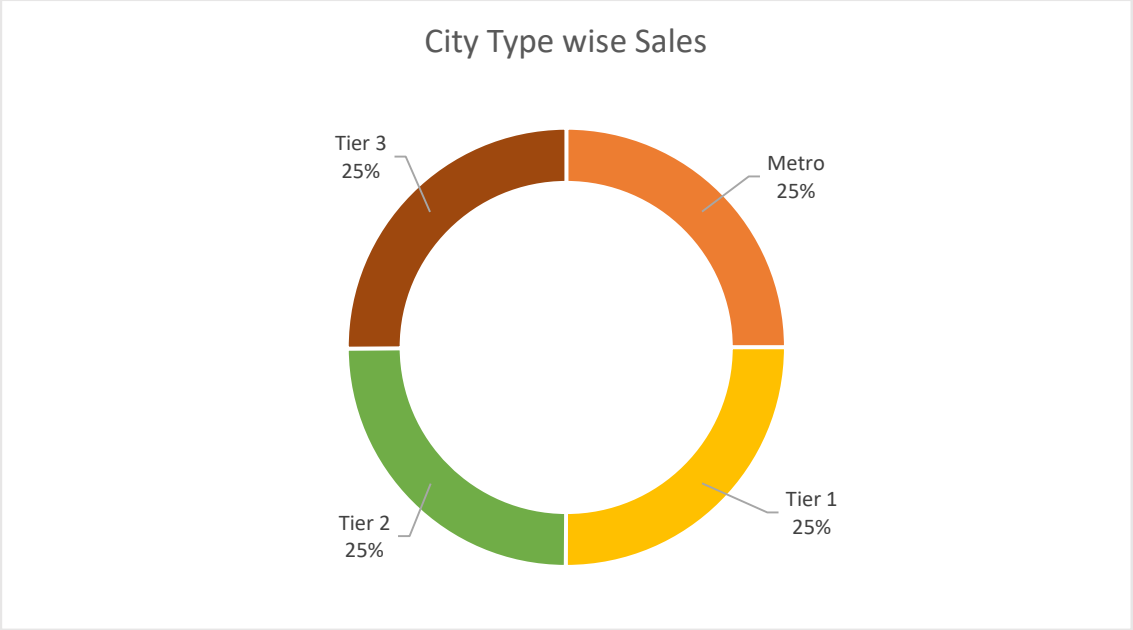
2. When it comes to fuel type, I found that all the fuel type has nearly the same market share, i.e. all Petrol, Electric, and Hybrid fuel types have 33% market share. Which means people are not preferring any special type of fuel.



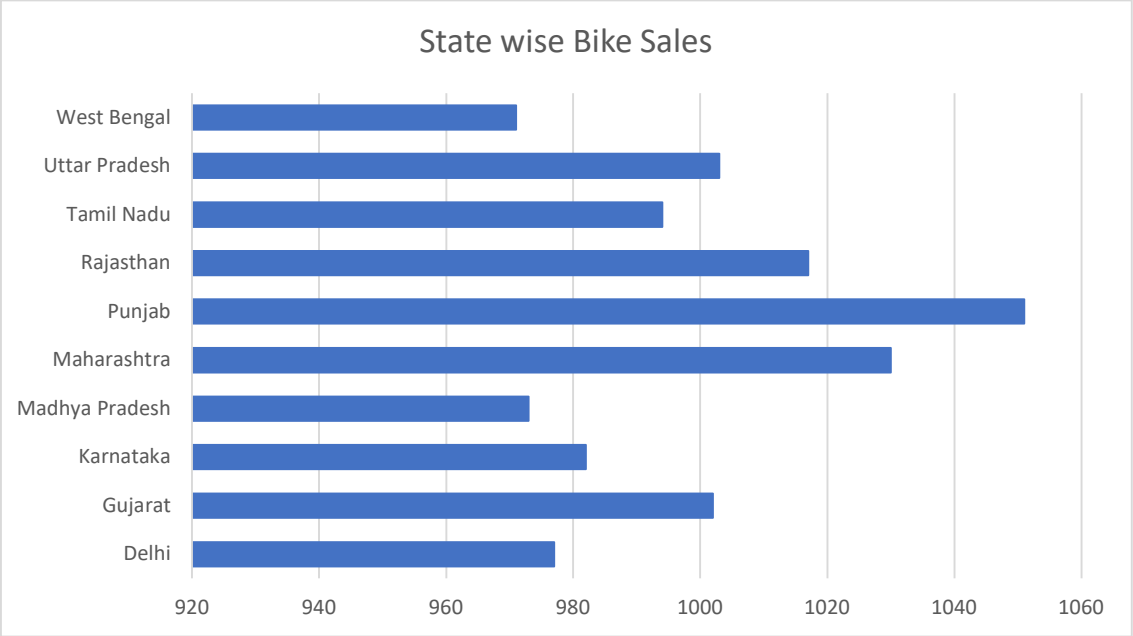
3. From the graph, we can see that the Original Price is directly proportional to the Resale Price. Bajaj's bikes have a high original price, so their resale price is also high. But for Honda, Who offer bikes in less price, has less resale value.



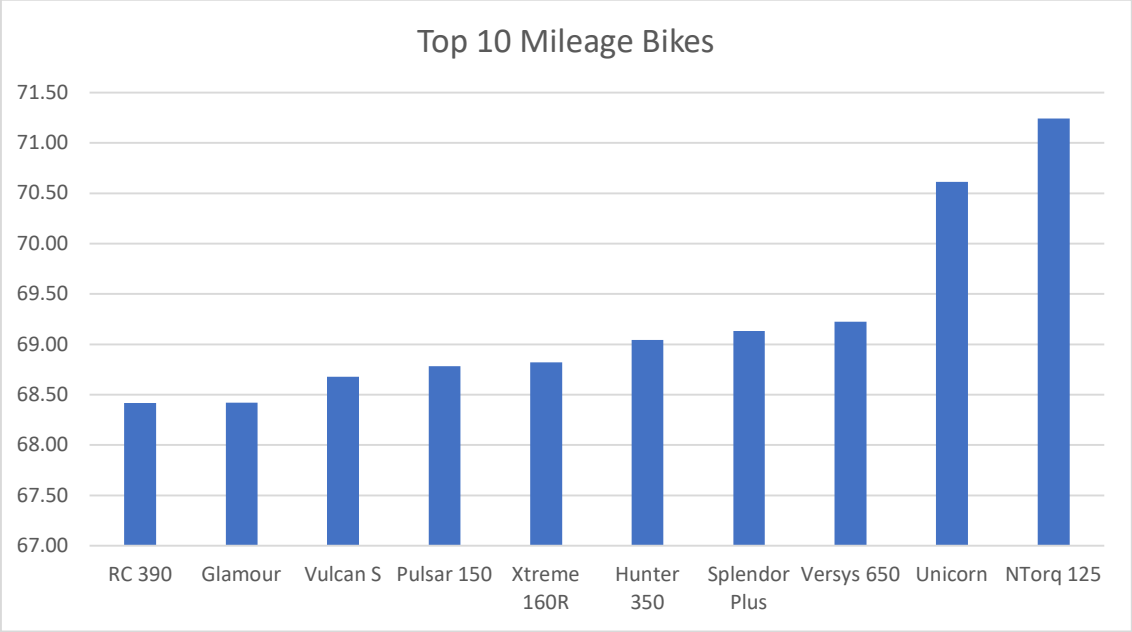
4. The Ray ZR from Yamaha is the highest-selling bike; the number of units sold is 279, followed by the NTorq 125 from TVS; the number of units sold is 271, then the Duke 200 from KTM; the number of units sold is 270. Whereas the Meteor 350 from Royal Enfield is the lowest-selling bike, the number of units sold is 214, followed by the Jupiter from TVS, which is the second lowest-selling bike, the number of units sold is 226.



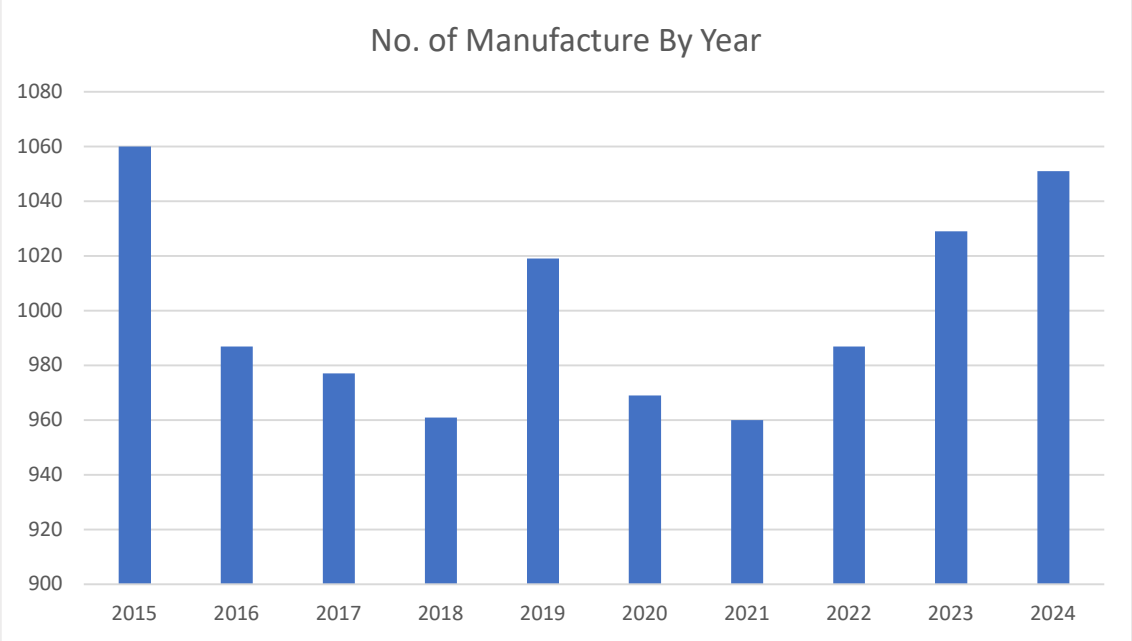
5. Purchasing trends among all types of cities are the same. In our dataset, there are four types of city like: Metro, Tier 1, Tier 2, and Tier 3. In all types of city customers are buying the same number of bikes, and the percentage is 25% in all types of cities.



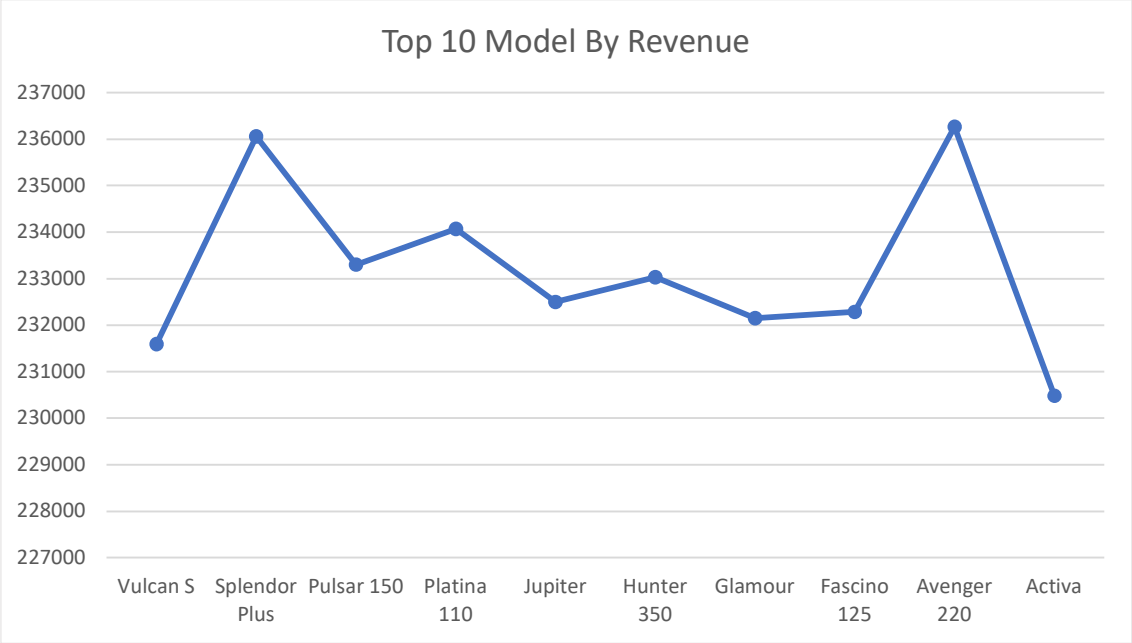
6. Punjab is the top state in terms of the number of bike sales; the number of units sold in Punjab is 1051. Second position is held by Maharashtra; the number of units sold in Maharashtra is 1030. West Bengal is the lowest-performing state in terms of bike sales; the number of units sold in West Bengal is only 971. The second-lowest position is held by Madhya Pradesh; the number of units sold is 973.



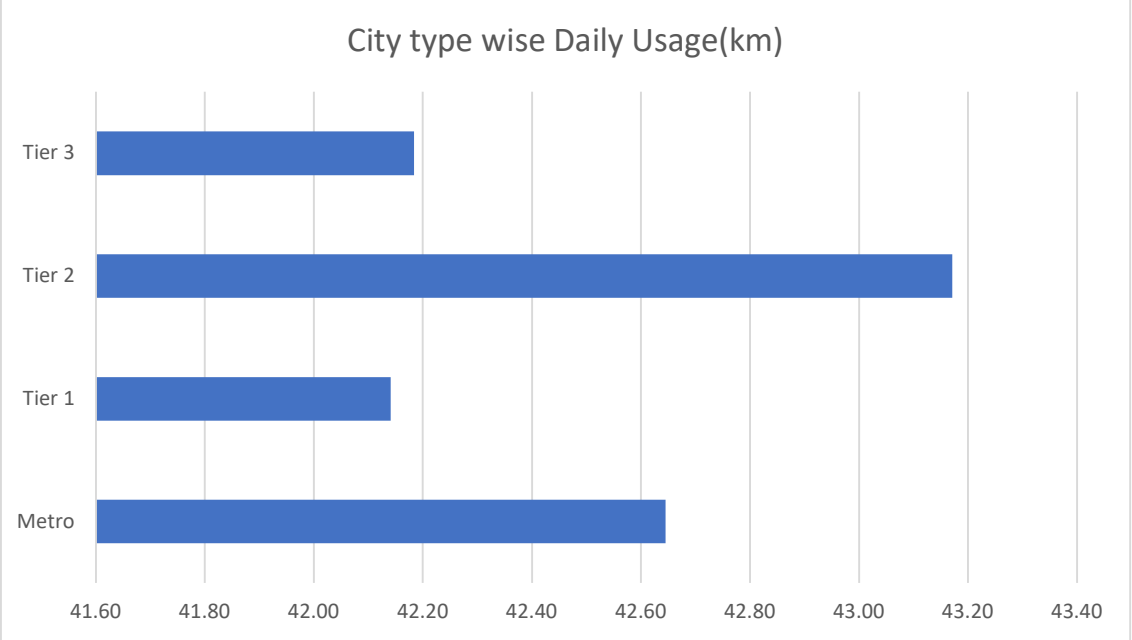
7. The above graph shows the top 10 mileage bikes in the Indian market. The NTorq 125 from TVS has the highest mileage, which is 72.25 km/l. Followed by the Unicorn from Honda has the second-highest mileage, which is 70.21 km/l.



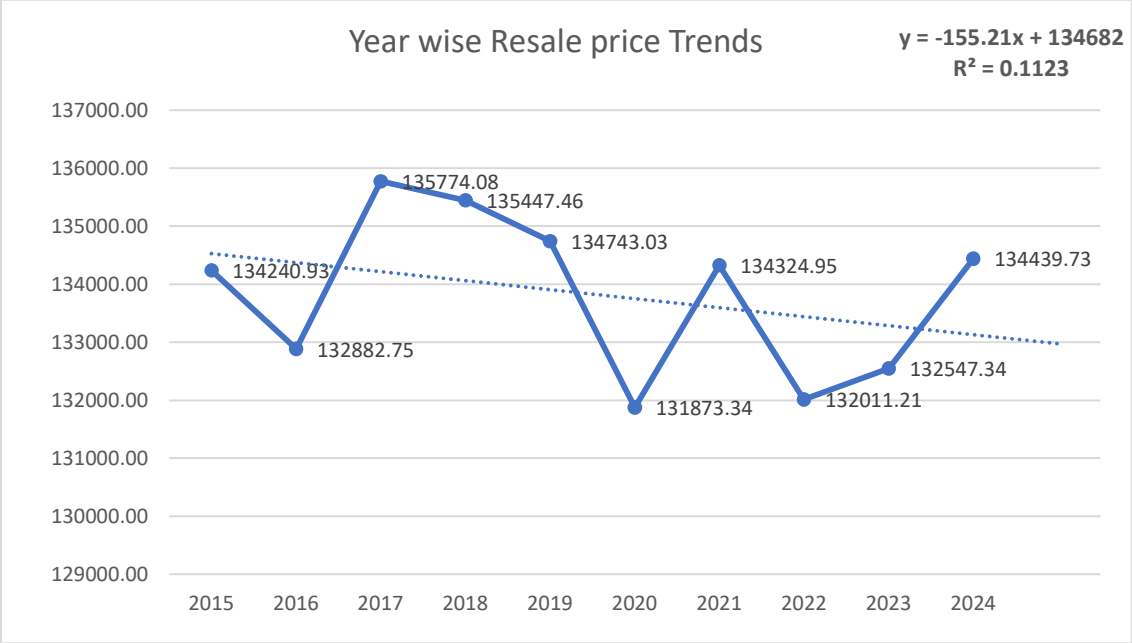
8. From the above graph, it is clear that from 2021, the manufacture of bikes has been increasing. In 2015, the number of bikes manufactured was 1060, and then the manufacturing of bikes started decreasing. In 2019, we can see a sharp increase in manufacturing, which was 1019. In 2020, it fell to 969. In 2024, the number of units manufactured was 1051.



9. The Avenger 220 from Bajaj in average, generated the highest revenue. Which is ₹2,36,263. In the second position, the splendor plus from Hero is there, which generated revenue of ₹2,36,058. The lowest revenue-generating bike is the 390 Adventure from KTM, which generated revenue of ₹2,18,125.



10. When it comes to city-wise usage of bikes, we can see that Tier 2 citizens are using bikes more than the Metro city and the Tier 1 city. Data shows that the Tier 2 citizens used a bike on an average of 43.17 km, whereas the citizens from the Metro city used 42.65 km. In the Tier 1 city, citizens used a bike on an average of 42.14 km.



11. In this graph, I tried to find the trends of resale value over the year. I used the linear regression to find the trend line. I found that the average resale value of bikes in 2025 may be ₹1,33,000. The regression equation is also shown on the graph. R^2 value is 0.1123.

Regression Statistics	
Multiple R	0.010418164
R Square	0.000108538
Adjusted R Square	-0.000291619
Standard Error	66622.9325
Observations	10000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	4815703843	1203925961	0.271239097	0.89664381
Residual	9995	4.4364E+13	4438615135		
Total	9999	4.43688E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	130612.669	3338.651476	39.12138477	0	124068.2399	137157.0982	124068.2399	137157.0982
Vehicle Age	153.3832289	229.2108153	0.669179719	0.503396298	-295.9161227	602.6825805	-295.9161227	602.6825805
Mileage	23.13656887	31.89498461	0.725398339	0.468224674	-39.38402329	85.65716104	-39.38402329	85.65716104
Engine Capacity	0.830084996	2.551297598	0.325357966	0.744916975	-4.170972021	5.831142013	-4.170972021	5.831142013
Avg distance	4.881352859	30.74370145	0.158775705	0.873848799	-55.38249249	65.1451982	-55.38249249	65.1451982

12. I tried to find whether the resale value depends on Vehicle age, Mileage, Engine Capacity and Avg Distance with the help of multiple linear regression. Interestingly, I found that the regression model is statistically insignificant with a near-zero R^2, indicating that vehicle age, mileage, engine capacity, and usage distance alone cannot explain resale price.

Regression Statistics	
Multiple R	0.905412622
R Square	0.819772015
Adjusted R Square	0.819753989
Standard Error	28280.90229
Observations	10000

ANOVA

	df	SS	MS	F	Significance F
Regression	1	3.63723E+13	3.63723E+13	45476.18183	0
Residual	9998	7.99649E+12	799809434.1		
Total	9999	4.43688E+13			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-457.114508	690.2990458	-0.6621978	0.507859728	-1810.239586	896.0105701	-1810.239586	896.0105701
Price	0.598612996	0.002807076	213.2514521	0	0.593110562	0.604115429	0.593110562	0.604115429

13. Here, I tried to find whether the actual price has any influence on the resale price or not. In this case, I applied simple linear regression to find out the relationships. And I found the R-Squared is 0.82. This means approximately 82% of the variance in Resale Price can be explained by the original Price alone.

Conclusions

Based on the analysis of the dashboard and statistical outputs, the following conclusions can be drawn:

1. **Price over Performance:** For valuation purposes, the original "sticker price" is the only reliable metric. Factors like mileage and engine capacity, while technically important, are not currently being "priced in" by the secondary market as independent variables.
2. **Technological Neutrality:** The market has successfully integrated Electric and Hybrid vehicles. Their resale volumes are on par with traditional Petrol bikes, suggesting high consumer confidence and a stabilised secondary market for EVs.
3. **Brand Strength:** Bajaj and Yamaha maintain the highest average resale values. Sellers of these brands can expect better value retention compared to the broader market average.
4. **Insurance Gap:** Approximately **34% of the bikes in the market have expired insurance**, and another 32% have "Not Available" status. This represents a significant risk for buyers and a potential value-add opportunity for dealers who provide certified, insured inventory.
5. **Depreciation Forecast:** The trendline analysis for "Year of Manufacture" suggests a stable but slightly downward-tilting average resale price for newer models, likely due to increased market competition and a higher volume of entry-level EV/Hybrid models entering the data.