International Journal of Advanced Research and Interdisciplinary Scientific Endeavours, Vol. 2(4), 2025
DOI: 10.61359/11.2206-2524

Data Analysis For Bikes Dataset Using Tableau

Deekonda Santhosh*, Kolanu Raviteja†, Majjiga Varshith Yadav‡,

Dr. Diana Moses§

Email Correspondence*: ravitejakolanu00@gmail.com

[1,2 &3] UG Scholar Department of AI&DS, Methodist College of Engineering and Technology, Hyderabad, India.

[4] Professor, Department of CSE, Methodist College of Engineering and Technology, Hyderabad, India.

Abstract:

This dataset contains 150 records of motorcycle sales across different locations. It includes 11 attributes, such as Bike_ID, Date, Location, Brand, Model, Bike_Name, CC (engine capacity), Dealer, Price, Units_Sold, and Total_Revenue. The dataset captures sales transactions from various motorcycle brands like KTM, Kawasaki, Harley-Davidson, Yamaha, and BMW, recorded in multiple cities, including London, Toronto, and New York. Each record represents a unique bike sale with details on pricing, the number of units sold, and the total revenue generated. Additionally, the dataset provides insights into customer preferences, popular bike models, and the impact of different pricing strategies on sales performance. It enables businesses to analyze dealership effectiveness, assess regional demand variations, and identify high-performing models. This dataset can be used for sales analysis, market trends, brand performance evaluation, and revenue forecasting, making it a valuable resource for motorcycle manufacturers, dealers, and market analysts looking to optimize their strategies and improve profitability.

Keywords: Digital Signal Processing, Noise Suppression, Noise Estimation, Wiener Filtering, Speech Recognition.

1. Introduction

This paper presents an analysis of a motorcycle sales dataset comprising 150 records across various locations. The dataset includes 11 attributes, such as Bike_ID, Date, Location, Brand, Model, Bike_Name, CC (engine capacity), Dealer, Price, Units_Sold, and Total_Revenue. The data captures sales transactions from various motorcycle brands, including KTM, Kawasaki, Harley-Davidson, Yamaha, and BMW, across multiple cities like London, Toronto, and New York. Each record details a unique bike sale, providing insight into pricing, units sold, and total revenue. The analysis of this dataset is valuable for understanding customer preferences, identifying popular bike models, and evaluating the impact of pricing strategies on sales performance. It also enables businesses to assess dealership effectiveness, analyze regional demand variations, and identify high-performing models. This information can be used for sales analysis, market trend identification, brand performance evaluation, and revenue forecasting, making it a valuable resource for motorcycle manufacturers, dealers, and market analysts.

*UG Scholar Department of AI&DS, Methodist College of Engineering and Technology, Hyderabad, India.
†UG Scholar Department of AI&DS, Methodist College of Engineering and Technology, Hyderabad, India.
‡UG Scholar Department of AI&DS, Methodist College of Engineering and Technology, Hyderabad, India.
§Professor, Department of CSE, Methodist College of Engineering and Technology, Hyderabad, India.

Article Timeline: Received: Apr 15, 2025; Revised: Apr 28, 2025; Published: Apr 30, 2025

2. Literature Survery

An, Jingrui, et al. The study explores challenges in Free-Floating Bicycle Sharing (FFBS), such as congestion, parking issues, and mismanagement. It proposes a real-time information management platform using data visualization to enhance coordination among governments, companies, and users. The solution introduces collaborative innovation, responsibility quantification, and data-driven decision-making to improve governance efficiency and sustainability. These insights help optimize FFBS operations and set a foundation for future sharing economy models [1].

Rennie, Nicola, et al. The study analyzes bike-sharing demand patterns using the Capital Bikeshare dataset, focusing on outlier detection caused by extreme weather or public transport disruptions. By clustering stations and applying functional depth analysis, it enhances demand forecasts and planning efficiency. Findings show that demand outliers are more frequent near city centers and during summer, impacting bike rebalancing strategies. The study suggests using automated outlier alerts for better inventory management and explores potential applications in dockless bike-sharing systems. [2]

Warmayana, I. Gede Agus Krisna, Yuichiro Yamashita, and Nobuta Oka. This study applies machine learning models (Decision Tree, Random Forest, and Neural Networks) to predict bike rental demand using daily and hourly datasets. The Random Forest model performed best for daily forecasts, while Neural Networks showed potential for hourly predictions. Key insights include higher demand in summer and peak usage during weekday mornings and evenings. These findings help optimize bike allocation, reduce idle capacity, and improve marketing strategies by targeting high-demand periods. Future research should integrate real-time traffic, social events, and advanced models to enhance prediction accuracy.[3]

Buning, Richard J., and Vijay Lulla.This study analyzes visitor usage patterns in urban bikeshare programs using GPS-based data from 353,733 trips over five years. Findings show that visitors primarily use bikeshare for leisure and exploration, while residents use it for daily commuting. Differences in trip behaviors include route selection, rental duration, and physical activity intensity. The study highlights bikeshare's role in sustainable tourism by connecting public transport with destinations and enhancing urban exploration, offering insights for tourism management, bikeshare programs, and urban planning.[4]

Nguyen, Hai Hoang. This thesis explores the impact of historical data on marketing strategies for Divvy, a bike-sharing service. By analyzing publicly available data through quantitative methods, it provides insights into optimizing Divvy's marketing mix. Key findings highlight customer journey trends, emphasizing seasonal usage patterns and the need for operational adjustments in winter. Recommendations include enhancing membership flexibility, modifying bike designs for cold weather, and strategic advertising at high-traffic stations. Future research should expand beyond the 7P's framework and integrate financial data for a more comprehensive competitive analysis.[5]

George Krull/Grant Thornton. This case study introduces students to data analytics and visualization using Tableau through real-world data from Divvy, Chicago's bike-share program. It caters to both novice and advanced users, guiding them through data exploration, variable creation, and visualization techniques. Advanced users integrate external data, conduct research, and design infographics for meaningful storytelling. With over three million rows of data, this case provides hands-on experience in handling large datasets, making it ideal for accounting information systems and data analytics courses.[6]

Svartzman, Gabriela Gongora, Jose E. Ramirez-Marquez, and Kash Barker. QoE by integrating performance metrics with user perceptions, using Citibike as a case study. By analyzing both operational data and social media feedback, the framework provides a comprehensive evaluation of service effectiveness. Insights gained helped improve city services beyond traditional reliability measures, ensuring they align with user expectations.[7]

Veldscholten, Sander. This study explores the transformation of public transportation from a top-down approach to a customer-centered model driven by data. By leveraging available datasets, including internal OVCK data, regiotaxi service records, and open data sources, Keolis can gain deeper insights into travel patterns in the Twente region. Using the Knowledge Discovery in Databases (KDD) method, data was cleaned, analyzed, and used to develop a decision support system in its prototype stage. Future enhancements could integrate machine learning for travel demand estimation and customer satisfaction data, making the system a comprehensive tool for transportation planning and KPI monitoring.[8]

Kwigizile, Valerian, Jun-Seok Oh, and Keneth Kwayu. This study examines the integration of crowdsourced cycling data into bicycle exposure estimation methods to enhance spatial and temporal accuracy. Traditional methods for measuring bicycle volume are often costly and lack comprehensive coverage. By incorporating data from Strava Metro, this research evaluates probabilistic and machine learning models, finding that the Random Forest (RF) model provides the most accurate predictions. The addition of Strava data significantly improved RF model performance, increasing its ability to explain variations in hourly bicycle volume from 65% to 71%. The findings highlight the potential of crowdsourcing as a cost-effective and scalable tool for urban planners to improve non-motorized transportation infrastructure.[9]

Zeid, Abe, Trisha Bhatt, and Hayley A. Morris. This study presents a machine learning (ML) model for predicting daily bike rental demand in bike-ride sharing systems, addressing the challenge of bike unavailability and revenue loss. Using historical data from Boston and New York City (NYC), the model incorporates environmental and seasonal factors to enhance accuracy. The results demonstrate that the model provides reliable forecasts, enabling bike-sharing companies to make informed business decisions, such as optimizing workforce allocation for bike restocking. This research contributes to improving operational efficiency, traffic management, and the environmental benefits of urban bike-sharing programs.[10].

3. Materials and Methods

Dataset: Bikes Dataset

This dataset provides records of bike sales across various locations. It includes information on sales trends, brand popularity, pricing, and dealership performance. The dataset consists of 150 records with multiple attributes, including Bike ID, Date, Location, Brand, Model, Bike Name, CC, Dealer, Price, Units Sold, and Total Revenue. These attributes help in analyzing sales performance across different locations and market conditions.

However, the dataset contains some inconsistencies, such as missing values in certain columns and data types that may require transformation before analysis. Some numerical columns are stored as text, which requires preprocessing to ensure accurate analysis.

The Bikes Dataset includes sales data from 10 major cities around the world. New York has the highest number of entries, followed by London, Toronto, and Paris. Other cities like Mumbai, Tokyo, Dubai, Los Angeles, Berlin, and Sydney are also included. This mix of locations helps compare bike sales, prices, and brand popularity in different parts of the world.

This dataset is valuable for various stakeholders. Bike manufacturers and dealerships can use it to forecast demand, optimize inventory, and adjust pricing strategies. Business analysts and market researchers can

E-ISSN: 3048-7021 (Online)

leverage it to study the impact of location-based trends, brand popularity, and consumer preferences. Data analysts and researchers can explore sales trends, uncover hidden patterns, and develop predictive models. Additionally, policymakers and transportation departments can utilize this data to study market demands and improve urban mobility strategies.

This project will analyze bike sales trends using Python, Tableau, and machine learning models to uncover key factors affecting purchasing behavior. By identifying significant correlations, the analysis aims to help businesses and city planners enhance bike sales strategies and optimize dealership operations.

Software: Tableau

Tableau's key features include a drag-and-drop interface, real-time data analysis, forecasting, and advanced analytics such as trend identification and clustering. It also supports data blending, allowing users to combine data from various sources for more comprehensive analysis. With its ability to map geographic data, Tableau is highly effective for geospatial analysis, providing location-based insights.

The application of Tableau spans across a wide range of industries, aiding businesses in decision-making by providing insights into past and present performance, as well as customer behavior. It enables automated reporting, ensuring that stakeholders have timely access to critical data. Tableau also facilitates collaboration through Tableau Server and Tableau Online, allowing teams to share and discuss insights efficiently. Additionally, Tableau Prep helps clean and structure data before analysis, and calculated fields allow users to create customized metrics to suit specific analytical needs, enhancing the overall data analysis process.

Data Visualization

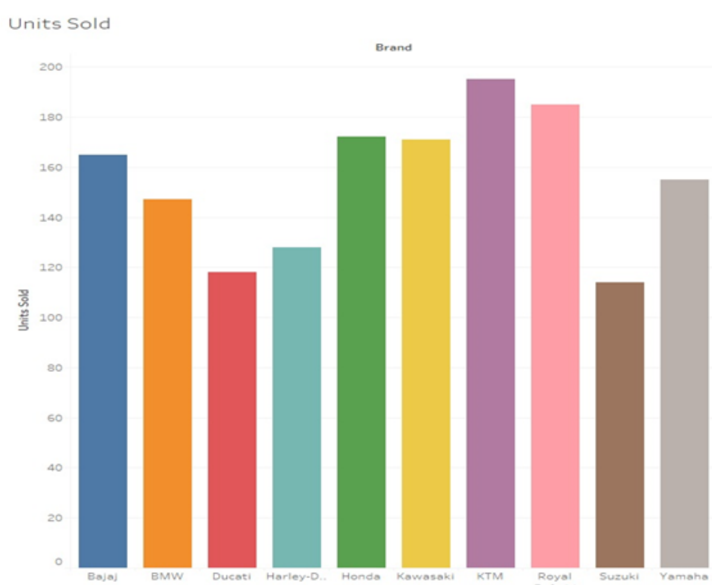

Figure-1 Units Sold by Motorcycle Brand

This bar chart represents the number of units sold for different motorcycle brands. The X-axis lists the various brands, including Bajaj, BMW, Ducati, Harley-Davidson, Honda, Kawasaki, KTM, Royal Enfield, Suzuki, and Yamaha, while the Y-axis indicates the total number of units sold.

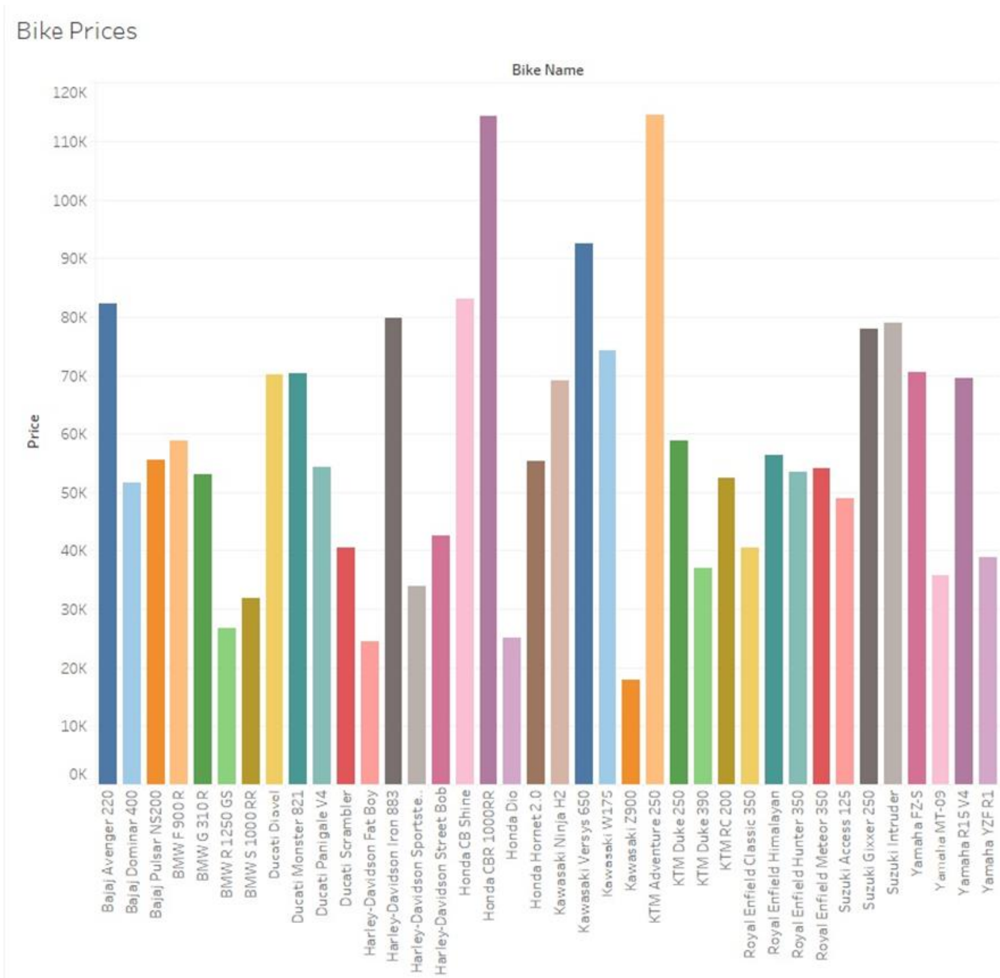

Figure-2 Price Distribution of Motorcycle Models

This bar chart represents the prices of various motorcycle models, with the X-axis displaying the bike names and the Y-axis indicating their prices. The most expensive bikes, exceeding 100K, likely belong to premium brands such as BMW, Ducati, and Harley-Davidson, catering to high-performance and luxury segments. Many models fall within the 50K–80K range, indicating a popular price point for mid-range motorcycles from brands like Honda, Kawasaki, and Royal Enfield. Additionally, some bikes are priced below 20K, representing budget-friendly commuter models, likely from Bajaj and Suzuki. This chart provides insights into price distribution, helping manufacturers, dealers, and consumers understand market trends and affordability.

E-ISSN: 3048-7021 (Online)

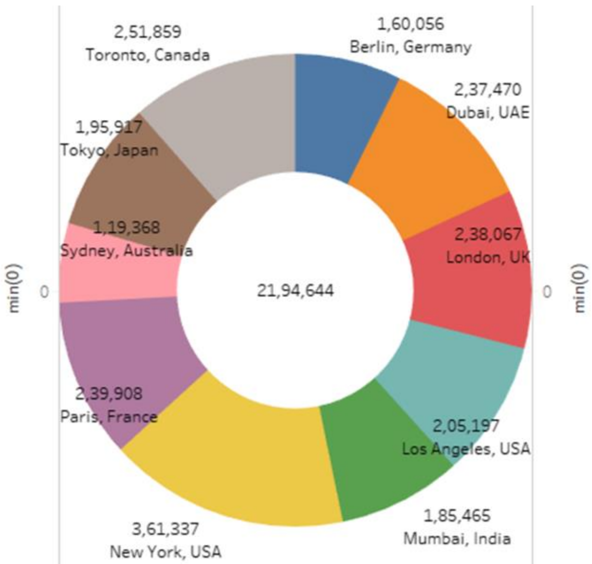

Figure-3 Brand-wise Total Revenue

This donut chart represents the total sales revenue generated from different global cities based on the dataset. Each segment of the chart corresponds to a specific location, with the size of the segment indicating that city's contribution to the overall revenue. The center of the donut displays the total revenue, amounting to 21,94,644. Among the cities, New York, USA stands out with the highest revenue contribution of 3,61,337, followed by cities like Paris, Toronto, London, and Los Angeles. Smaller segments, such as those for Berlin, Mumbai, and Sydney, indicate comparatively lower sales figures. The use of distinct colors for each location makes it easy to differentiate and analyze the revenue distribution across various international markets. This visualization provides a clear overview of which cities are driving the highest sales and helps in identifying key areas of performance.

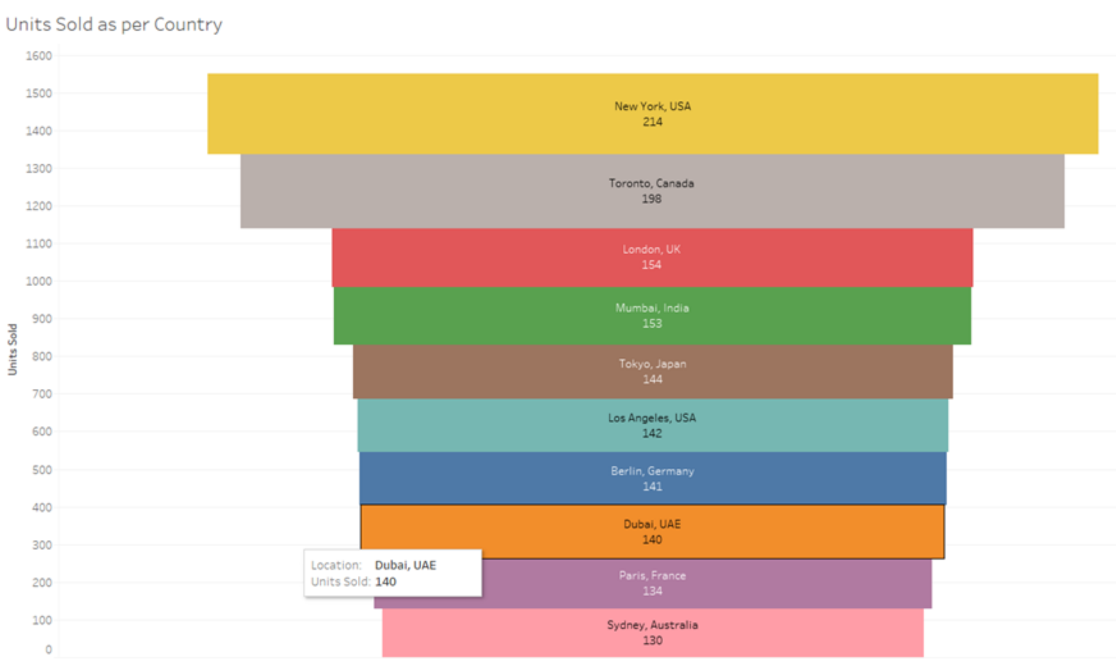

Figure-4 Units Sold as per Country

This horizontal bar chart illustrates the number of motorcycle units sold across various cities. The chart is titled "Units Sold as per Country," and it clearly highlights the differences in sales volume between locations. At the top of the chart, New York, USA leads with the highest number of units sold, totaling 214. It is followed by Toronto, Canada with 198 units, and London, UK with 154 units. Other cities like Mumbai, Tokyo, Los Angeles, Berlin, Dubai, Paris, and Sydney show relatively lower sales figures, ranging from 153 down to 130 units. The bars are color-coded for each city, and the values are labeled directly on the bars for easy comparison. This visualization provides a quick and effective overview of market performance by location, helping identify top-selling regions and potential areas for growth.

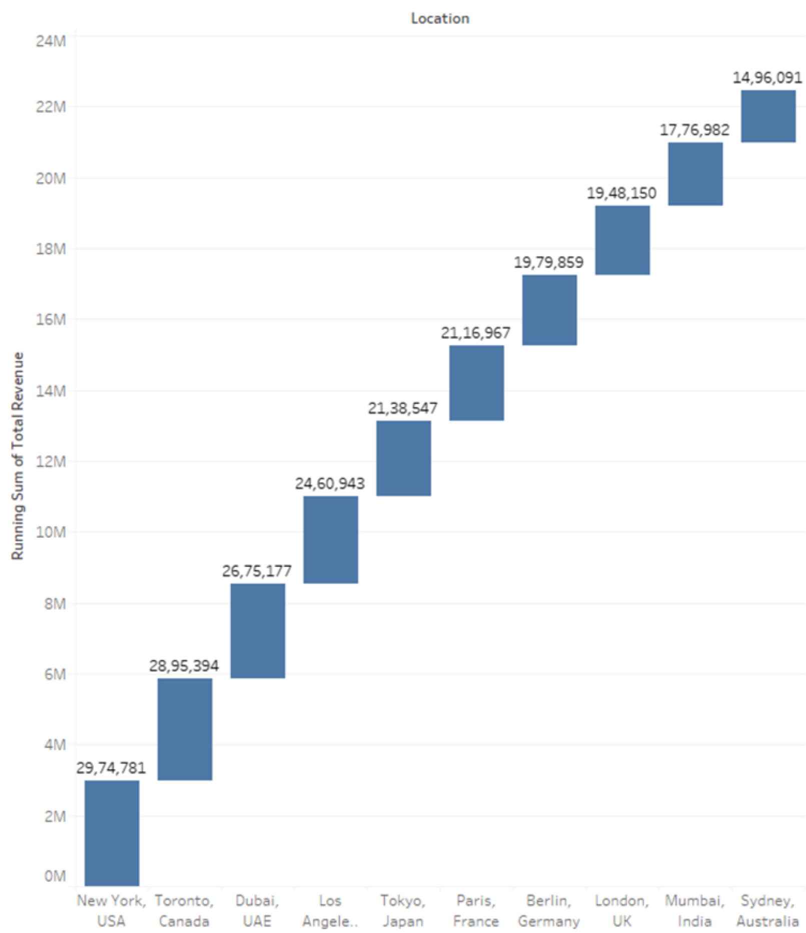

Figure-5 Running Total of Total Revenue Across Global Locations

This Gantt-style chart visualizes the running total of total revenue across various global locations. Each bar represents a city's contribution to the cumulative revenue, helping us understand how each location adds up to the overall total. The cities are arranged sequentially along the X-axis, starting with New York, USA, which has the highest individual contribution of ₹29,74,781. The cumulative revenue continues to increase with subsequent contributions from Toronto, Dubai, Los Angeles, Tokyo, and others. By the time all the locations are accounted for, the total running sum reaches ₹13,66,64,431. This chart is particularly effective for visualizing incremental growth and comparing the revenue performance of different locations in a progressive manner. It also helps identify which locations are the strongest contributors and which are lagging behind.

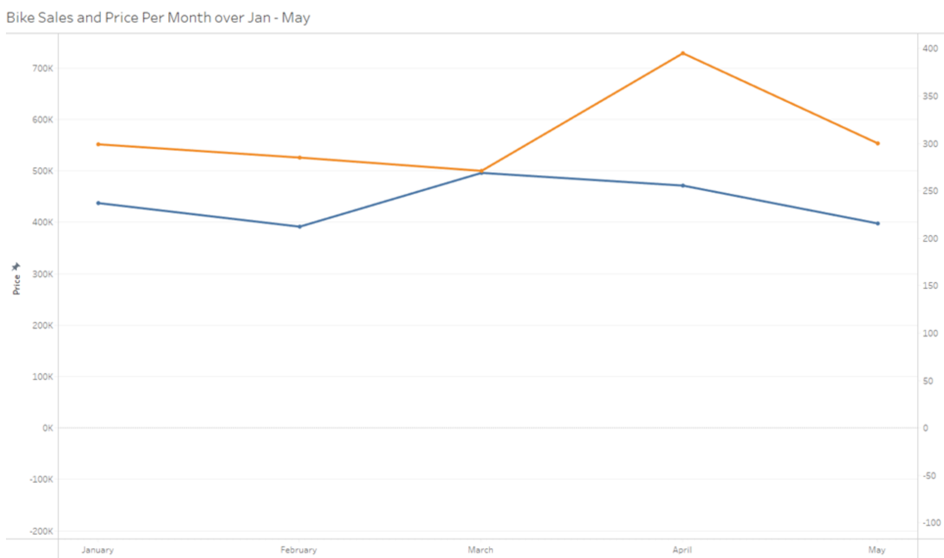

Figure-6 Trend of Bike Sales and Average Price per Unit (Jan–May)

The dual-axis line chart illustrates the trend of bike sales and average price per unit over the five-month period from January to May. The orange line represents bike sales, while the blue line denotes the average price. In the initial months—January to March—sales remained relatively steady with a slight decline, while prices showed a minor dip in February before rising again in March. In April, bike sales peaked sharply, indicating a significant surge in customer demand or successful promotional efforts. However, sales dropped again in May, returning to levels similar to those at the beginning of the year. Meanwhile, the price trend remained more stable throughout the period, with slight fluctuations but no extreme spikes or drops. This chart highlights that April marked a clear divergence between sales and pricing trends, making it a focal point for further analysis.

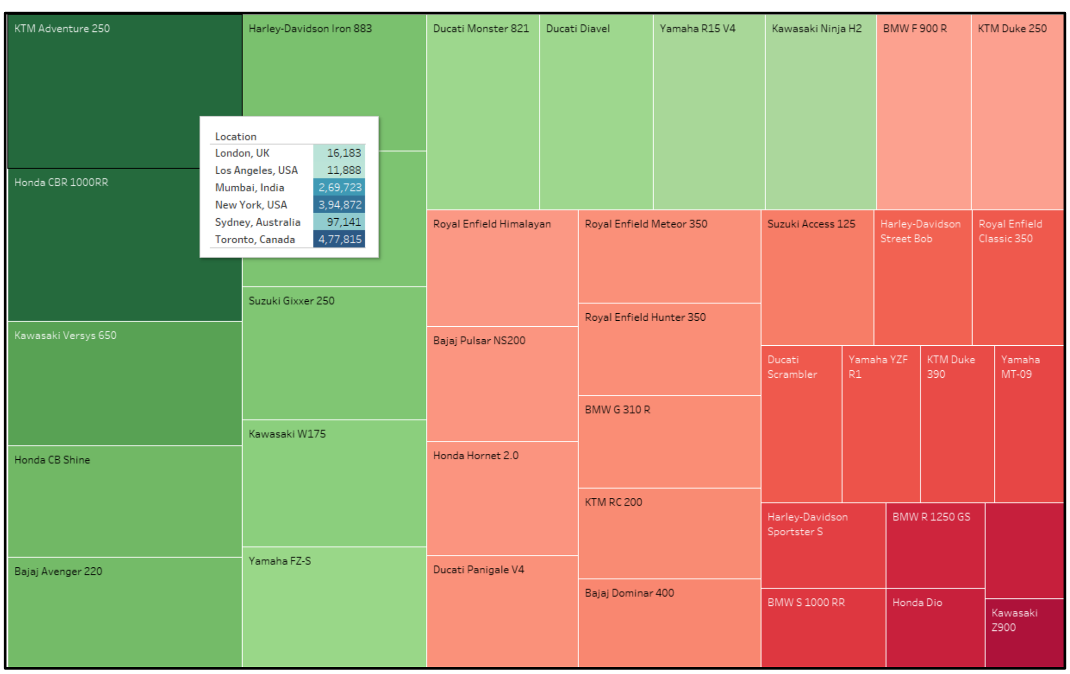

Figure-7 Revenue Distribution Across Bike Models

This Treemap chart visually represents the revenue distribution across various bike models, with each rectangle sized according to the model's sales and colored to indicate performance—green shades for high performers and red for lower ones. Top-selling models like KTM Adventure 250 and Honda CBR 1000RR stand out with larger, darker green blocks, indicating strong market success, while smaller, red-toned blocks like the Kawasaki Z900 and Honda Dio reflect weaker sales. The chart also includes tooltips that show revenue contributions from different locations, highlighting Toronto, New York, and Mumbai as major contributors. Overall, this visualization effectively shows which bike models and regions are driving revenue and which may need attention.

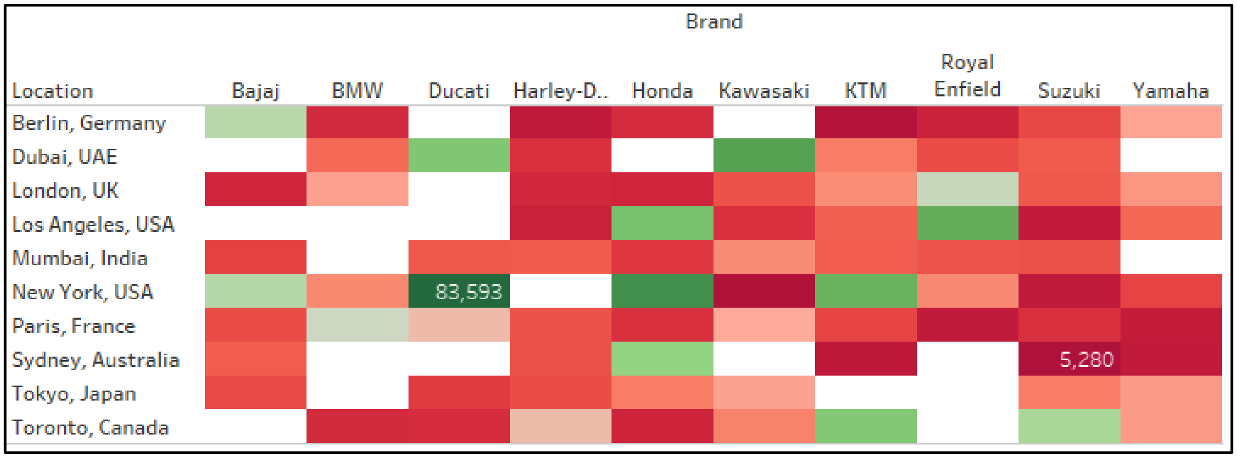

Figure-8 Comparative Heatmap of Bike Brand Performance by City

This heatmap provides a comparative overview of bike brand performance across different global locations, with each cell representing the total revenue generated by a brand in a specific city. The color gradient—from red (low revenue) to green (high revenue)—enables quick visual insights into where each brand is performing well or underperforming. Notably, BMW in New York shows the highest individual brand-city revenue at ₹83,593, while Suzuki in Tokyo has the lowest at ₹5,280. The chart is a powerful tool for identifying market strengths and weaknesses, regional brand preferences, and strategic opportunities for expanding or optimizing sales efforts.

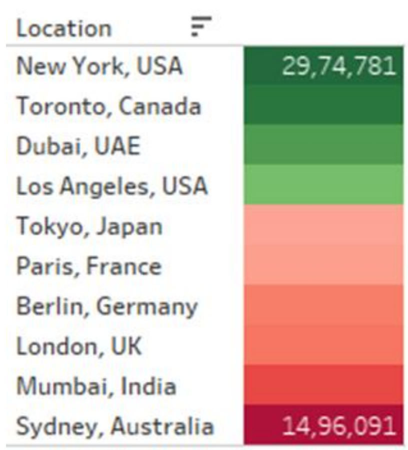

Figure-9 Heatmap of Total Revenue by Location

This heatmap provides a clear comparison of total revenue generated by each location, using a color gradient from red to green to reflect performance levels. New York, USA leads with the highest revenue of ₹29,74,781, followed by Toronto, Dubai, and Los Angeles, indicating strong sales in North American and Middle Eastern markets. In contrast, Sydney, Australia ranks lowest with ₹14,96,091, signaling potential for market development. The visualization effectively highlights which locations are the top contributors and which ones may require strategic attention for revenue improvement.

4. Results

The analysis of the dataset reveals key insights into motorcycle sales trends, pricing, and brand performance across multiple locations. KTM and Royal Enfield emerged as the top-selling brands, indicating strong customer preference and brand loyalty. Mid-range motorcycles, priced between 50K–80K, dominated the market, suggesting that consumers prioritize affordability and performance. Premium brands like BMW, Ducati, and Harley-Davidson had relatively lower sales volumes, likely due to their higher price points catering to niche buyers. Additionally, sales distribution varied across different locations, highlighting regional preferences and demand fluctuations

5. Discussion

The findings suggest that price plays a significant role in influencing motorcycle sales. Brands offering a balance between cost and performance, such as Honda, Kawasaki, and Yamaha, performed well, reinforcing the importance of affordability in industry. The lower sales of high-end motorcycles indicate that luxury bikes cater to a limited but dedicated market segment. Regional trends also impact sales, as consumer preferences may vary based on economic factors, transportation needs, and brand availability. The dataset's insights can help manufacturers and dealers optimize inventory, refine pricing strategies, and enhance marketing efforts to target specific consumer groups. Additionally, predictive analytics using this data could aid in forecasting future demand, identifying sales patterns, and improving business decision-making. Overall, this dataset serves as a valuable resource for understanding the motorcycle industry's market dynamics, providing actionable insights for manufacturers, dealers, and analysts aiming to improve sales performance and market competitiveness.

6. Conclusion

This dataset offers a comprehensive overview of the motorcycle industry, capturing key insights into sales performance, pricing strategies, and brand popularity across multiple locations. The data reveals that brands like KTM and Royal Enfield lead in sales volume, indicating strong consumer demand and market presence. Meanwhile, premium brands such as BMW, Ducati, and Harley-Davidson cater to a niche segment with high-priced models, reflecting a focus on luxury and performance enthusiasts. Pricing plays a crucial role in sales trends, as mid-range motorcycles (priced between 50K–80K) appear to be the most popular, while budget-friendly bikes from Bajaj and Suzuki offer affordability for daily commuters. Additionally, the dataset enables businesses to analyze dealership effectiveness, assess regional variations in demand, and identify emerging market trends. From a forecasting perspective, this dataset can be utilized for predictive modeling, revenue forecasting, and customer preference analysis, helping manufacturers and dealerships optimize their strategies. By leveraging data-driven insights, businesses can enhance decision-making, improve market competitiveness, and maximize profitability in the motorcycle industry.

7. References

[1] An, Jingrui, et al. "Collaborative governance in the sharing economy. A case of free-floating bicycle sharing with visualized analyzation." The Design Journal 22.sup1 (2019): 777-788.

[2] Rennie, Nicola, et al. "Analysing and visualising bike-sharing demand with outliers." Discover Data 1.1 (2023).

[3] Warmayana, I. Gede Agus Krisna, Yuichiro Yamashita, and Nobuta Oka. "Predictive Analysis for Optimizing Targeted Marketing Campaigns in Bike-Sharing Systems Using Decision Trees, Random Forests, and Neural Networks." Journal of Digital Market and Digital Currency 2.1 (2025): 67-92.

[4] Buning, Richard J., and Vijay Lulla. "Visitor bikeshare usage: tracking visitor spatiotemporal behavior using big data." Journal of Sustainable Tourism 29.4 (2020): 711-731.

[5] Nguyen, Hai Hoang. "Providing the marketing mix through the analysis of historical data: case: Divvy." (2021).

[6] Data, Divvying Up, et al. "George Krull/Grant Thornton."

[7] Svartzman, Gabriela Gongora, Jose E. Ramirez-Marquez, and Kash Barker. "Social media analytics to connect system performability and quality of experience, with an application to Citibike." Computers & Industrial Engineering 139 (2020): 106146.

[8] Veldscholten, Sander. "Spatial and temporal public transport data visualization: data analysis using a decision support system for alternative public transport services." (2018).

[9] Kwigizile, Valerian, Jun-Seok Oh, and Keneth Kwayu. Integrating crowdsourced data with traditionally collected data to enhance estimation of bicycle exposure measure. No. TRCLC 2017-03. Western Michigan University. Transportation Research Center for Livable Communities, 2019.

[10] Zeid, Abe, Trisha Bhatt, and Hayley A. Morris. "Machine learning model to forecast demand of Boston Bike-ride sharing." European Journal of Artificial Intelligence and Machine Learning 1.3 (2022): 1-10.

8.Conflict of Interest

The authors declare that there are no conflicts of interest to report in this article.

9.Funding

No external funding was received to support or conduct this study.