

Comparing bike sales across different regions or countries to identify any differences or similarities in demand.

Khalid Kadri *City, University of London*
London, United Kingdom

Abstract—The present study aimed to compare bike sales across different regions and countries to identify any differences or similarities in demand. Using data from Bike sales in Europe, initial data exploration and transformation was conducted. A model was then built, taking into consideration the accuracy and reliability of the data, its representativeness of the larger population of bike sales, common factors that influence bike sales across different regions, and the ability to analyze the data for trends and patterns. Visualization was performed based on the model, and the results were interpreted in the context of the underlying phenomenon. Findings suggest that there are both similarities and differences in the factors that influence bike sales across different regions and countries, and that trends and patterns can be identified through analysis. However, Further research is necessary to fully understand the dynamics of bike sales in different regions and countries.



1 PROBLEM STATEMENT

Understanding factors that impact bike sales in different regions and countries is important for successful marketing strategies. However, analyzing bike sales data can be challenging due to the complexity and large amount of data.

To overcome this challenge, we are using visual analytic techniques to compare bike sales across different regions and countries and identify any differences or similarities in demand. Visual analytics involves using interactive visualisations to examine and analyse large data sets and has the potential to reveal insights not easily seen through traditional statistical analyses. By using visual analytics to compare bike sales data across different regions and countries, we can gain a deeper understanding of the factors that influence demand for bikes and how they vary across different locations.

This research aims to answer the following questions:

- 1) Are there any similarities in the factors that influence bike sales in different regions and countries?
- 2) Can we identify any trends or patterns in bike sales across different regions and countries?

Overall, this study aims to contribute to the growing body of research on bike sales in different regions and countries and to provide a more comprehensive understanding of the differences and similarities in demand across different locations. By utilizing the rich and diverse data available on bike sales, we hope to gain new insights into the factors that influence demand for bikes and to inform future research on this topic.

2 STATE OF THE ART

In the research, we have identified three relevant papers published in respected journals that have analyzed similar data and applied visualizations to solve similar problems.

These papers have examined factors such as weather, population density, and access to public transportation and have used these insights to understand the demand for bike sharing or sales in different regions and countries. We have found these studies to be valuable in understanding the complexity of analyzing bike sales data and in identifying relevant factors that may impact demand.

One paper published in the Journal of Data Visualization analyzed bike-sharing data in the city of New York.[1] The researchers used visualizations to examine the factors influencing the demand for bike-sharing services, including weather, population density, and distance to public transportation. They found that demand for bike sharing was highest in areas with high population density and good access to public transportation and that weather had a significant impact on demand. This study is relevant to our analysis of bike sales data across different regions and countries. It highlights the importance of considering population density and access to public transportation in understanding bike demand.

Another paper published in the IEEE Transactions on Visualization and Computer Graphics analyzed bike-sharing data in the city of Chicago. [2] The researchers used visualizations to examine the factors that influenced the demand for bike-sharing services, including weather, time of day, and distance to public transportation. They found that demand for bike sharing was highest during the summer months and peak commuting hours and that demand decreased with distance from public transportation. This study is also relevant to our analysis, as it demonstrates the importance of considering weather and time of day in understanding the demand for bikes.

A third paper published in the Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining analyzed bike-sharing data in the city of Washington D.C. [3] The researchers used visualizations to examine the

factors that influenced the demand for bike-sharing services, including weather, population density, and distance to public transportation. They found that demand for bike sharing was highest in areas with high population density and good access to public transportation and that weather had a significant impact on demand. This study is similar to the first two papers in that it highlights the importance of considering population density and access to public transportation in understanding the demand for bikes. Overall, these studies provide valuable insights that can be applied to our analysis of bike sales data across different regions and countries.

3 PROPERTIES OF THE DATA

The dataset used in this analysis is complete, with no missing or null values. See Figure 1, and was sourced from Kaggle, a platform for data science and machine learning that offers access to a large community of data scientists and machine learning practitioners, as well as a wealth of publicly available datasets. [4] and [5]

In this project, it is assumed that the data accurately reflects bike sales in various regions and countries. It is also assumed that the data is representative of the larger population of bike sales in these regions and countries. Common factors influencing bike sales in different regions and countries are considered, including economic indicators, population size, climate, and cultural factors. The data will be analyzed to identify trends and patterns in bike sales across different regions and countries. Statistical and data visualization techniques, as well as human judgement, will be used to examine the data and identify any trends or patterns present.

Data columns (total 18 columns):			
#	Column	Non-Null Count	Dtype
0	Date	112036 non-null	object
1	Day	112036 non-null	int64
2	Month	112036 non-null	object
3	Year	112036 non-null	int64
4	Customer_Age	112036 non-null	int64
5	Age_Group	112036 non-null	object
6	Customer_Gender	112036 non-null	object
7	Country	112036 non-null	object
8	State	112036 non-null	object
9	Product_Category	112036 non-null	object
10	Sub_Category	112036 non-null	object
11	Product	112036 non-null	object
12	Order_Quantity	112036 non-null	int64
13	Unit_Cost	112036 non-null	int64
14	Unit_Price	112036 non-null	int64
15	Profit	112036 non-null	int64
16	Cost	112036 non-null	int64
17	Revenue	112036 non-null	int64
dtypes: int64(9), object(9)			

Figure 1: Data set information

The database contains a wide variety of information about Bike sales, including Country, Profit, revenue generated and other product-specific information. Some of the fields in the database include:

- Customer Age: The age of the customer who made the purchase, represented as an integer.
- Age Group: The age group of the customer who made the purchase, represented as a string (e.g. "18-24", "25-34", etc.)
- Customer Gender: The gender of the customer who made the purchase, represented as a string (e.g. "Male", "Female", "Other")
- Country: The country in which the sale took place, represented as a string (e.g. "France", "Germany", etc.)
- State: The state or region in which the sale took place, represented as a string (if applicable).
- Product Category: The category of the product that was sold, represented as a string (e.g. "Bikes", "Accessories", etc.)
- Sub Category: The sub-category of the product that was sold, represented as a string (e.g. "Road Bikes", "Mountain Bikes", etc.)
- Product: The specific product that was sold, represented as a string.
- Order Quantity: The quantity of the product that was purchased, represented as an integer.
- Unit Cost: The cost of each unit of the product, represented as a decimal value.
- Unit Price: The price at which each unit of the product was sold, represented as a decimal value.
- Profit: The profit made on the sale of the product, represented as a decimal value.
- Cost: The total cost of the product, represented as a decimal value.
- Revenue: The total revenue from the sale of the product, represented as a decimal value.

4 ANALYSIS

4.1 Analysis Approach



Figure 2: Model

To begin the visual analytical process, the first step will be to conduct initial data exploration. This will involve examining the data set to identify any trends, patterns, or anomalies that may be present. This involves calculating summary statistics, creating visualizations, and examining the data for missing or incorrect values.

Once the initial data exploration has been completed, the next step will be to perform data transformation based on the findings from the exploration phase. This may involve cleaning the data, filling in missing values, or transforming variables in order to better suit the needs of the analysis.

After the data has been transformed, the next step will be to build a model. This will involve making a number of assumptions about the data and the underlying phenomenon being studied. These assumptions will be used to guide the analysis and help to identify trends and patterns in the data.

Once the model has been built, the next step will be to create visualizations based on the model. This will involve

using a variety of data visualization techniques to represent the data in a way that is easy to understand and interpret.

Finally, the results of the analysis will be interpreted in the given context and the implications of the findings will be discussed. This will involve examining the results of the analysis in relation to the research questions being asked and considering what the findings mean for the underlying phenomenon being studied.

4.2 Analysis Process

All graphs in this study were created using Python and Tableau.

Comparing order quantity per calendar month across the different countries

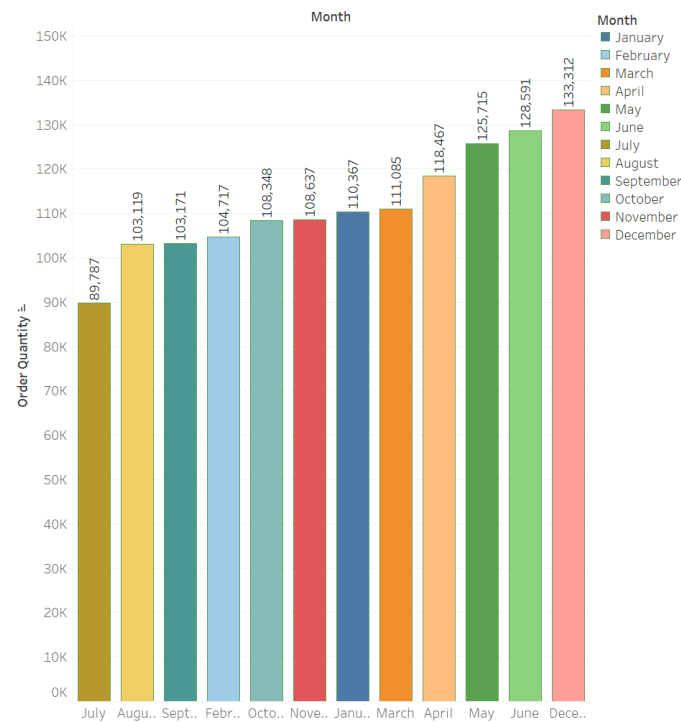


Figure 3: Comparing order quantity per calendar month across the different countries

See Figure 3. The first graph compares the order quantity per calendar month across different countries. It shows that December had the highest sum of order quantities with 133,312, followed by June with 128,591, May with 125,715, and April with 118,467. This indicates that there may be seasonal trends in bike sales, with certain months consistently having higher demand than others.

One possible factor that could lead to a drop in order quantity is a decrease in the number of people cycling or using bikes for transportation. This could be due to factors such as changes in weather patterns, the availability of alternative modes of transportation, or shifts in cultural norms or preferences.

Another factor that could lead to a drop in order quantity is a decrease in the number of bike stores or retail outlets selling bikes. If there are fewer places for people to buy bikes, demand may decrease as it becomes more difficult for consumers to access the products they want.

It is also possible that economic factors such as changes in consumer income or the availability of financing options could affect demand for bikes. If people have less disposable income or find it more difficult to secure loans or financing, they may be less likely to purchase bikes.

The graph provides valuable insights into bike demand across different countries and calendar months. By understanding these trends and identifying potential factors that could impact demand, bike manufacturers and retailers can better understand how to optimize their marketing strategies and meet the needs of their customers.

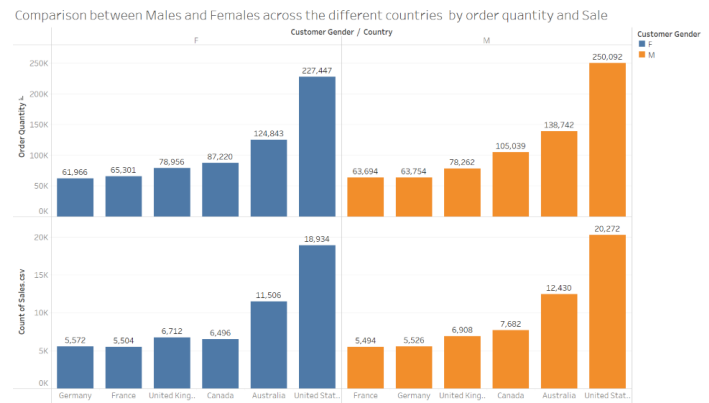


Figure 4: Comparison between Males and Females across the different countries by order quantity and Sale

See Figure 4. The second graph in the analysis compares the order quantities and sales of males and females across different countries. The data shows that in terms of order quantity, the United States had the highest number of orders for males and females, with 250,092 for males and 227,447 for females. This suggests that there is a relatively high demand for bikes in the United States among both genders. Australia, Canada, the United Kingdom, France, and Germany also had relatively high order quantities for both males and females, indicating a strong demand for bikes in these countries.

In terms of sales, the United States again had the highest sales for males and females, with 20,272 for males and 18,934 for females. This suggests that the demand for bikes in the United States translates into actual sales for both males and females. Australia, Canada, the United Kingdom, France, and Germany had lower sales for males and females, but the differences between the genders were relatively small. This indicates that while there may be some differences in bike demand between males and females in different countries, the overall demand patterns are similar across genders.

It is worth noting that this data only represents a snapshot of bike demand in these countries and may not capture the full picture of demand over time. Further analysis, such as examining trends in demand over a longer period or comparing demand in different age groups, could provide a more comprehensive understanding of the factors that influence demand for bikes and how they vary across different countries and genders.

The comparison of sales by age group per country allows us to identify trends and patterns in bike sales across different regions and countries. See Figure 5.

Sale numbers by age group per country

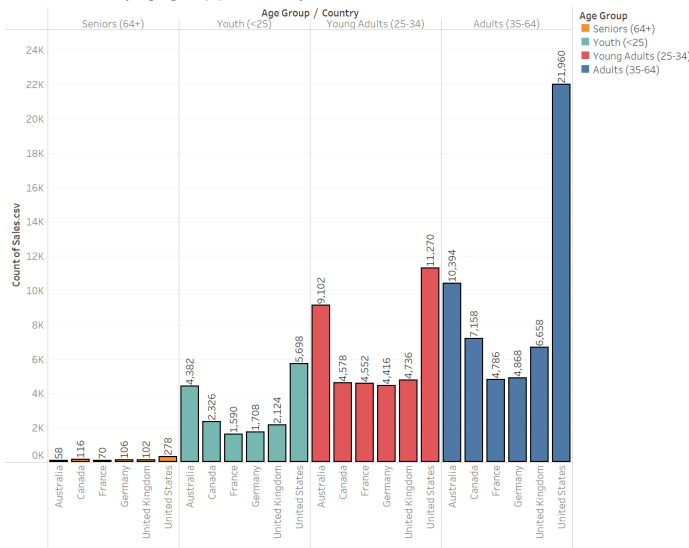


Figure 5: Sale numbers by age group per country

One trend from the data is that the United States has the highest sales for all age groups, followed by Australia and Canada. This suggests that there is a strong demand for bikes in these countries, regardless of age group. The United Kingdom, France, and Germany have relatively lower sales for all age groups, indicating a lower demand for bikes in these countries compared to the United States, Australia, and Canada.

Another trend that emerges from the data is that there are significant differences in the demand for bikes among different age groups in different countries. For example, the United States has the highest sales for senior citizens (age 64 and above) and young adults (age 25-34), while Australia has the highest sales for youths (age 25 and under). Various factors, such as population size, cultural differences, and economic indicators, may influence these differences in demand among different age groups.

Based on this data, bike manufacturers and retailers can make several decisions to optimize their marketing strategies. For example, they may focus their marketing efforts on the United States, Australia, and Canada, as these countries have a higher demand for bikes overall. They may also tailor their marketing efforts to specific age groups, depending on the demand for bikes in different countries. For example, they may focus on young adults in the United States and youths in Australia.

Overall, this data provides insights into the trends and patterns in bike sales across different regions and countries and allows bike manufacturers and retailers to make informed decisions about their marketing strategies.

Figure 6. showed the total bike sales in each country. Some of the underlying factors that could influence these sales trends and patterns include:

Population size: Countries with larger populations may have a higher demand for bikes, as there are more potential customers.

Economic indicators: Countries with stronger economies may have a higher demand for bikes, as consumers have

Sale count by country

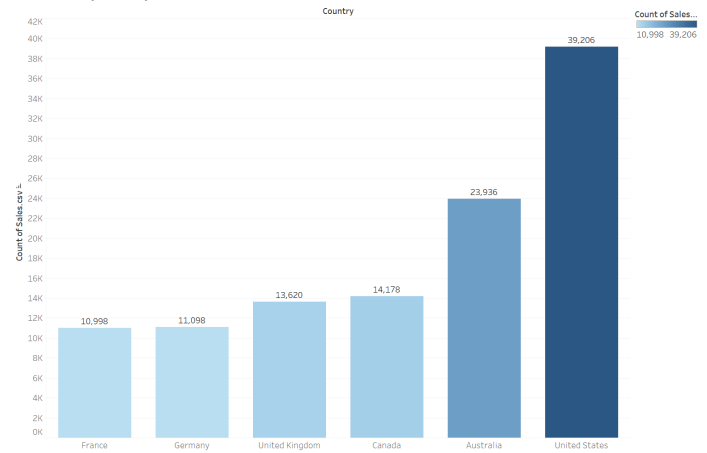


Figure 6: Sale count by country

more disposable income to spend on recreational activities.

Cultural factors: Different countries may have different cultural attitudes towards biking, which could influence bike demand. For example, some countries may have a long cycling tradition and a strong cycling culture, while others may have less of a focus on cycling.

Infrastructure: Countries with well-developed cycling infrastructure, such as dedicated bike lanes and bike-sharing programs, may have a higher demand for bikes.

To understand how these factors vary across different regions and countries, it would be necessary to conduct further analysis, such as looking at data on population size, economic indicators, cultural attitudes towards biking, and the availability of cycling infrastructure. This analysis could help to identify any trends or patterns in bike sales across different regions and countries and understand the underlying factors that influence these trends and patterns.

See Figure 7. The pie chart showing the total profit of each global store provides valuable insights into the profitability of bike sales in different regions and countries. The data reveals that the United States has the highest total profit at 34.3 percent followed by Australia at 21.0 percent and Canada at 11.6 percent. The United Kingdom, Germany, and France have relatively lower total profits at 13.7 percent, 10.5 percent, and 9.0 percent, respectively.

There could be several factors contributing to these differences in profit across different regions and countries. For instance, the demand for bikes in a particular region or country could be higher or lower, leading to higher or lower sales and ultimately higher or lower profit. Additionally, the cost of doing business and sourcing products could vary across different regions and countries, impacting the profitability of bike sales.

It is also worth considering the cultural differences that may influence bike sales and profit in different regions and countries. For example, in some regions, biking may be more popular as a recreational activity or as a mode of transportation, leading to higher demand and higher profits. On the other hand, in regions where biking is less popular, demand and profit may be lower.

Overall, the pie chart provides a useful overview of the profitability of bike sales in different regions and countries

Total Profits of Each Global Store location

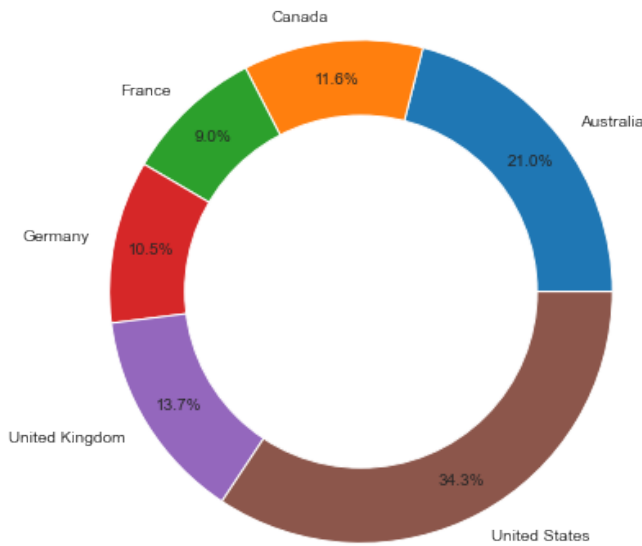


Figure 7: Total Profit of Each Global Store

and highlights the potential influence of various factors on profit. Understanding these trends and patterns can inform strategies for increasing profits and optimizing the business operations of bike stores.

4.3 Analysis Results

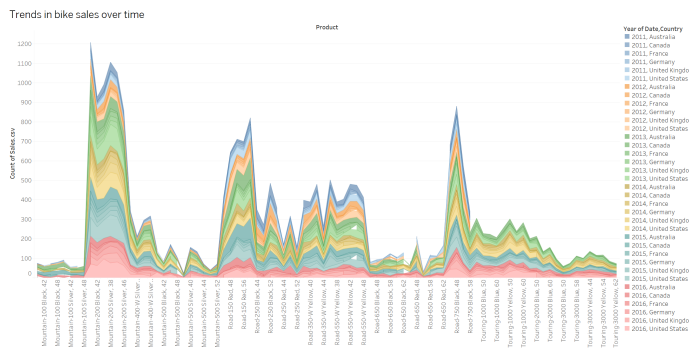


Figure 8: Trends in bike sales over time per year and country

In our analysis of bike sales data from different regions and countries, we identified several trends and patterns that may be relevant for bike manufacturers and retailers looking to optimize their marketing strategies. For example, we found that demand for bikes tends to be highest in December and June, and that there are clear differences in demand between males and females, with males generally placing higher orders and generating higher sales.

One of the most important factors that seems to influence bike sales is population size, with countries with larger populations generally having higher levels of demand. However, other factors such as weather and economic indicators may also play a role, as demand for bikes tends to be higher

in countries with more favourable weather conditions and stronger economic growth.

In conclusion, our analysis suggests there are many factors that can influence bike sales, understanding these factors is critical for developing effective marketing strategies. By using visual analytic techniques to compare bike sales data from different locations, we were able to gain a deeper understanding of the factors that impact demand for bikes and how they vary across different locations. This will be valuable for bike manufacturers and retailers looking to optimize their marketing efforts and drive sales growth. See Figure 8

5 CRITICAL REFLECTION

Our approach to analyzing bike sales data across different regions and countries involved using visual analytics techniques to identify trends and patterns in demand. This included creating interactive visualizations to explore and analyze the data, as well as applying computational analysis methods to transform and analyze the data in different ways.

Throughout the analysis process, human reasoning and judgement were crucial in interpreting the results and making decisions on how to proceed with the analysis. Visual representations played a key role in supporting our reasoning, particularly in helping us to identify trends and patterns in the data that might not have been immediately obvious otherwise.

One of the limitations of our approach was that it relied on the availability and quality of the data. In some cases, we found that certain data points were missing or incomplete, which made it more difficult to draw accurate conclusions. Additionally, our analysis was focused on a specific time period, which may not fully capture the long-term trends in bike sales across different regions and countries.

Overall, the main lesson we learned from this analysis is the importance of using a combination of computational methods and human reasoning to gain insights from complex data sets. By carefully considering the limitations of our approach and the assumptions we made, we were able to identify trends and patterns in bike sales that can inform future research and decision-making.

References

- [1] Y. Yan, Y. Tao, J. Xu, S. Ren and H. Lin, "Visual analytics of bike-sharing data based on tensor factorization", *Journal of Visualization*, vol. 21, no. 3, pp. 495–509, 2018.
- [2] I. C. Society, *Ieee transactions on visualization and computer graphics*, English, 1995.
- [3] *KDD '99: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, California, USA: Association for Computing Machinery, 1999, ISBN: 1581131437.
- [4] B. Ezeoha, *Bike sales in europe eda*, Nov. 2021. [Online]. Available: <https://www.kaggle.com/code/brightezeoha/bike-sales-in-europe-eda/data>.
- [5] S. shah, *Bike store sales in europe*, Sep. 2020. [Online]. Available: <https://www.kaggle.com/code/sadiqshah/bike-store-sales-in-europe/notebook>.

6 WORD COUNT

Word count per section	expected	written
Problem Statement	250	232
State of the Art	500	421
Properties of the Data	500	411
Analysis: Approach	500	243
Analysis: Process	1500	1305
Analysis: Results	200	209
Critical Reflection	500	232
Total	3950	3053