**Task 2- Intermediate**

**PUBLIC BIKE SHARING DATA ANALYSIS REPORT**

**By Meghna Ghosh**

**Data Analytics Intern at CodEvo Solutions**

**Id: 592zr1wq**

**Date: 25/08/2024**

**CONTENTS**

|  |  |  |
| --- | --- | --- |
| S. No | Content | Page no. |
| 1.  2.  3.  4.  5.  6. | About this Report  Methodology  Analytical Tools and Techniques  Visualization and Findings  Recommendations  Conclusion | 3  4  5-6  7-8  9-10  11 |

**ABOUT THIS REPORT**

This report, "Public Bike Sharing Data Analysis," examines bike-sharing systems to understand user behaviour, usage trends, and operational efficiency. As a key part of sustainable urban transportation, bike-sharing offers flexible mobility. By analysing extensive data from these programs, this report aims to uncover insights that can improve service delivery, enhance user satisfaction, and inform policy decisions.

**Objective**

The primary objective of this report is to analyze the public bike-sharing data to extract meaningful insights that can drive data-informed decision-making. The analysis seeks to answer key questions such as:

* When and where are bike-sharing services most utilized?
* What are the peak usage hours?
* Which stations experience the highest demand?
* How do factors like weather, day of the week, and time of day influence bike usage?
* What are the demographic characteristics of the users?

**Scope**

This report covers a wide range of topics related to bike-sharing data, including data collection, cleaning, and preparation, as well as the visualization and interpretation of key findings. The scope extends to an in-depth analysis of usage patterns, demographic insights, and potential areas for operational improvements.

**Significance**

The significance of this report lies in its potential to impact urban mobility positively. As cities around the world strive to reduce traffic congestion and promote environmentally friendly transportation options, bike-sharing systems play a vital role. By analyzing the data generated by these systems, this report contributes to a better understanding of how they are used and how they can be optimized to meet the needs of urban residents more effectively.

**METHODOLOGY**

Data Collection

The dataset was sourced from Kaggle and included extensive records of bike-sharing usage, covering trip details and user demographics.

Data Cleaning

Data cleaning was performed in Microsoft Excel to ensure accuracy. Key steps included:

* Handling Missing Values: Missing entries were filled or removed based on their significance.
* Filtering Irrelevant Data: Records with zero duration or unrealistic distances were excluded.
* Correcting Data Entries: Inconsistent or erroneous entries were corrected.
* Ensuring Data Consistency: Standardization of dates, times, and station names ensured uniformity.

Data Organization

The cleaned data was organized into categories like temporal, geographical, and demographic data, facilitating targeted analysis.

Data Importation and Visualization

The organized data was imported into Power BI for visualization. Data modeling established relationships between tables, and visualizations like heat maps and bar charts were created to identify trends and patterns.

Challenges and Solutions

* Missing Data: Addressed through imputation or exclusion.
* Data Inconsistency: Resolved by standardizing data entries.
* Large Dataset Management: Data was processed in chunks, and only relevant subsets were imported into Power BI to optimize performance.

**Analytical Tools and Techniques**

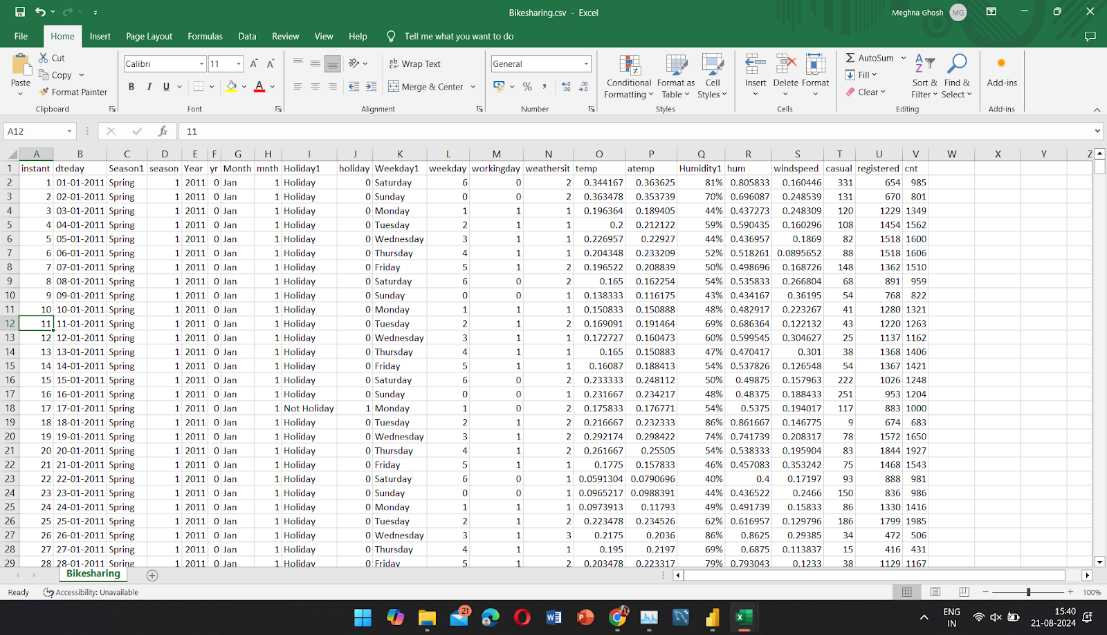
In this report, a combination of Microsoft Excel and Power BI was used to conduct the analysis of the public bike-sharing data. Each tool played a crucial role in ensuring that the data was meticulously cleaned, organized, and analyzed, ultimately leading to insightful visualizations that informed the findings and recommendations.

Microsoft Excel for Data Cleaning and Organization

Microsoft Excel was the primary tool used for the initial stages of data analysis, particularly for cleaning and organizing the raw data sourced from Kaggle. The extensive capabilities of Excel in handling large datasets and performing data manipulations were critical to preparing the data for further analysis.

Key Functions and Features Used:

The analysis utilized Microsoft Excel for data cleaning and Power BI for visualization and in-depth analysis, combining the strengths of both tools to derive meaningful insights.

Data Cleaning with Excel:

* Text Functions: TRIM, CLEAN, and SUBSTITUTE functions were used to standardize text data like station names and user demographics.
* Data Validation: Ensured consistency in categorical variables by restricting entries to predefined lists.
* Conditional Formatting: Highlighted inconsistencies, errors, and outliers for quick identification.
* Handling Missing Data: Applied imputation techniques using IF and AVERAGEIF functions to fill missing values, while irrelevant entries were filtered out.
* Data Summarization: Pivot tables and descriptive statistics (AVERAGE, MEDIAN, STDEV) were used to summarize trends and patterns within the dataset.

Visualization and Analysis with Power BI:

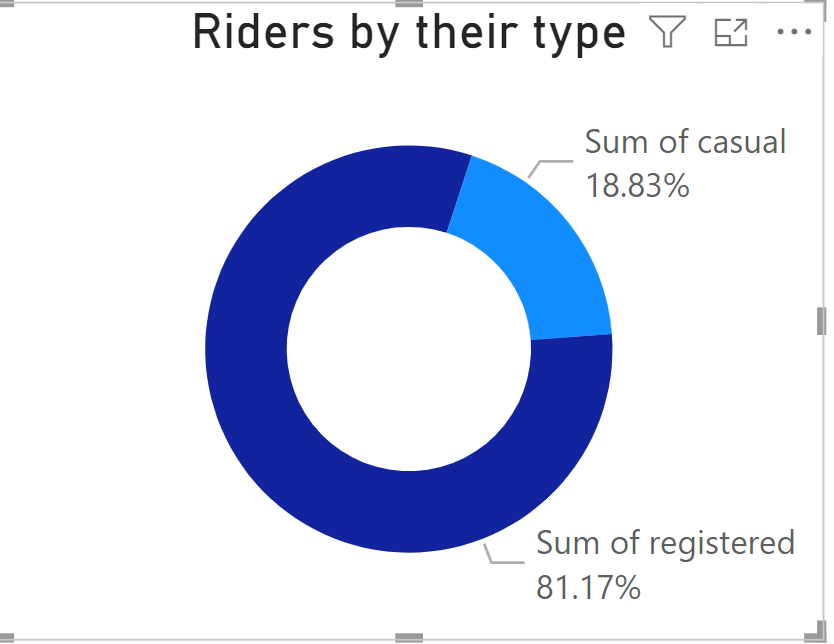
* Data Modeling: Established relationships between tables (e.g., trip data, demographics) to enable multi-dimensional analysis.
* Calculated Columns and Measures: Used DAX to create custom calculations, such as average trip duration by time of day.
* Interactive Visualizations: Created dashboards with bar charts, line graphs, heat maps, and scatter plots, allowing users to drill down into specific data segments.
* Time Series Analysis: Analyzed trends over time to identify peak usage periods and seasonal variations.
* Geographical Mapping: Visualized start and end stations on maps to reveal spatial patterns in bike usage.

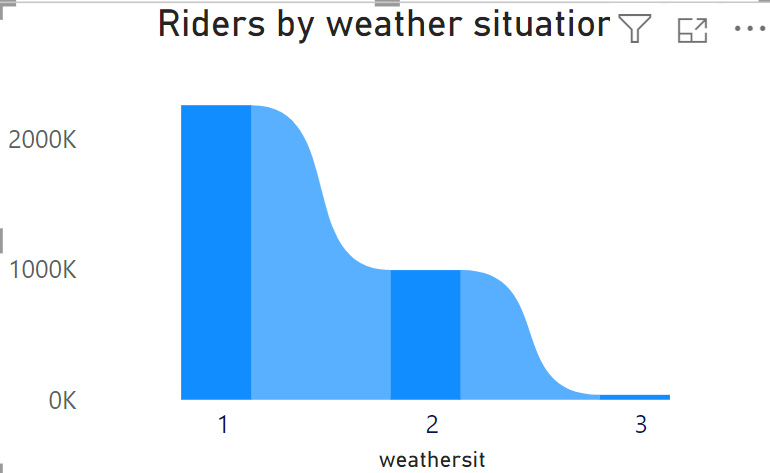
Advanced Techniques:

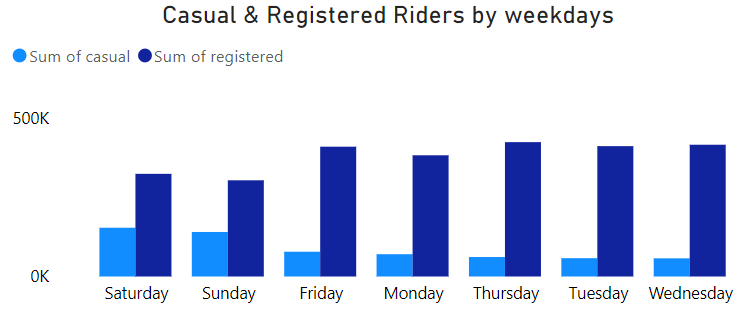
* Clustering: Segmented users based on usage patterns to gain deeper insights into customer behavior.
* Trend Analysis: Utilized Power BI’s forecasting features to predict future usage trends, informing strategic decisions.

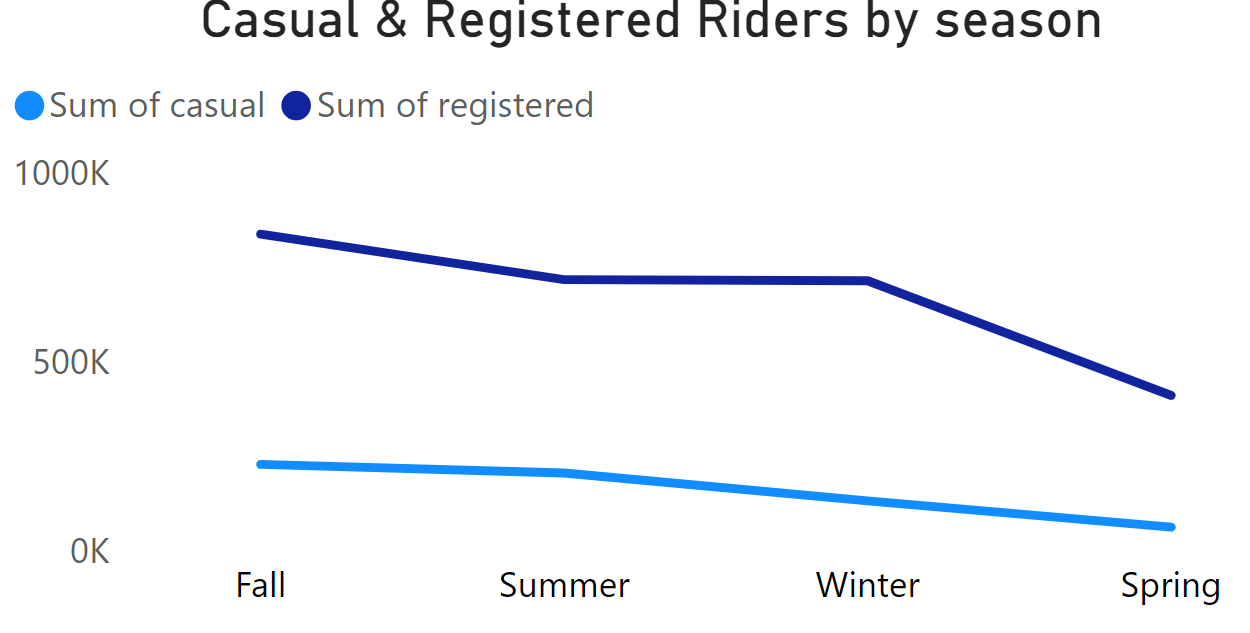
The seamless integration of Excel and Power BI allowed for a smooth workflow, from data preparation to visualization, ensuring accurate and actionable insights from the public bike-sharing data.

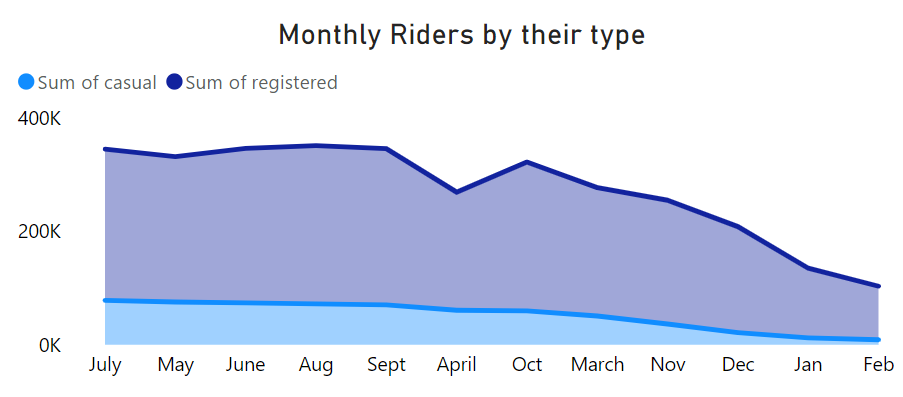
**Visualization and Findings**



1. The pie chart illustrates the distribution of rider types within the dataset, revealing that a significant majority, 81.1%, are registered riders, while the remaining 18.8% are casual riders. This highlights the predominant use of the bike-sharing service by registered users, indicating a strong base of regular, committed riders compared to occasional or one-time users.
2. The analysis of bike rental counts across different weather conditions shows that demand remains relatively consistent during clear, cloudy, or misty periods, with a slightly higher rental count during better weather conditions. However, on rainy and snowy days, there is a noticeable decline in the average number of rentals. It's important to note that the dataset appears to lack observations during extreme weather conditions.
3. The clustered chart below illustrates the distribution of casual and registered riders across different days of the week. It reveals that registered riders predominantly use bikes on weekdays, while casual riders are more active on weekends, particularly on Saturday and Sunday. This suggests that registered users likely use the service for routine, possibly work-related commutes, whereas casual riders tend to rent bikes for leisure or recreational activities during the weekend.



1. The line chart displaying bike rental counts across different seasons reveals a clear seasonal trend. Rental counts are generally lower in spring, with a noticeable peak in fall. This suggests that seasonality plays a significant role in influencing bike rental demand, making it an important factor to consider when analyzing rental patterns.



1. The stacked area chart shows a clear pattern in monthly rider activity, with the highest number of riders observed in September and a significant dip in April. This trend indicates that rider demand varies throughout the year, possibly due to seasonal changes or other external factors. The peak in September suggests it may be a popular time for biking, while the dip in April could reflect less favorable conditions or other influences that reduce bike usage during that month.

**Recommendations**

Based on the analysis of the public bike-sharing data, several key insights have been identified that can inform strategic decisions to enhance the service. Here are the recommendations derived from the findings:

Focus on Registered Users:

With 81.1% of riders being registered users, efforts should be made to maintain and further develop this strong base. Consider implementing loyalty programs, exclusive discounts, or additional services to encourage continued usage and increase rider retention.

Weather-Responsive Strategies:

The analysis indicates that bike rentals decrease during rainy and snowy days. To mitigate this, the service could introduce weather-adaptive pricing, offering discounts during less favorable weather conditions to encourage usage. Additionally, enhancing bike infrastructure, such as providing rain covers or heated bike stations, could make the service more appealing in inclement weather.

Targeted Marketing for Casual Riders:

Casual riders show higher activity on weekends, suggesting they use the service primarily for leisure. Marketing campaigns should focus on promoting weekend-specific offers or events that align with recreational activities. Additionally, partnerships with local attractions or tourist spots could attract more casual users, especially on weekends.

Seasonal Campaigns:

The data shows a peak in bike rentals during the fall and a dip in spring. To capitalize on seasonal trends, marketing campaigns and promotional events should be concentrated in the fall to maximize ridership. Conversely, efforts should be made to boost spring usage, perhaps by introducing early-season discounts or spring-themed biking events to attract more users.

Monthly Rider Activity Adjustments:

Given the peak in rider activity in September and the dip in April, resource allocation should be adjusted accordingly. More bikes, maintenance services, and staff could be deployed during peak months to accommodate higher demand, while April could see promotional efforts or special events aimed at increasing ridership.

Improved Data Collection During Extreme Weather:

The absence of data during extreme weather conditions suggests a gap in understanding rider behavior in these scenarios. Improved data collection and analysis during such periods could provide deeper insights and help in developing strategies to maintain or even boost ridership during extreme weather events.

By implementing these recommendations, the bike-sharing service can enhance user experience, optimize operations, and ultimately drive higher engagement and satisfaction among both registered and casual riders.

**Conclusion**

The analysis of public bike-sharing data has revealed key insights into user behavior, seasonal trends, and the impact of weather on rentals. The distinction between registered (81.1%) and casual riders highlights the importance of focusing on loyal, weekday users who rely on the service for regular commutes. Enhancing their experience through loyalty programs could boost retention.

Weather and seasonality significantly influence rental demand, with declines during adverse conditions and peaks in the fall, especially in September. These trends suggest opportunities for weather-adaptive pricing and strategic marketing to optimize usage throughout the year.

Casual riders' preference for weekends indicates potential for targeted marketing, such as leisure-oriented packages. Operational adjustments, like increasing bike availability during peak months, can enhance service reliability.

The analysis also identified gaps in data collection during extreme weather, suggesting the need for more comprehensive data to inform strategies for maintaining ridership in challenging conditions.

Overall, this report emphasizes the importance of understanding user behavior and external factors to optimize service offerings, improve customer satisfaction, and achieve sustainable growth. Leveraging tools like Excel and Power BI has been crucial in uncovering actionable insights, guiding strategic decisions, and ensuring long-term success in the bike-sharing service.