**Task 2- Intermediate**

**A blue and black bicycle

Description automatically generated**

**PUBLIC BIKE SHARING DATA ANALYSIS REPORT**

**By Aishwarya Kande**

**Data Analytics Intern at CodEvo Solutions**

**Id: xl7uklin**

**Date: 02/09/2024**

**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,” looks at bike-sharing systems to understand how people use them, the patterns of usage, and how well they work. Bike-sharing is an important part of eco-friendly city transport because it provides flexible ways to get around. By studying a lot of data from these programs, this report aims to find insights that can make the service better, increase user happiness, and help make policy decisions.

**Objective**

The main goal of this report is to study public bike-sharing data to find useful insights that can help make better decisions based on data. The analysis aims to answer important questions like:

* When and where are bike-sharing services used the most?
* What are the busiest hours?
* Which stations have the highest demand?
* How do factors like weather, day of the week, and time of day affect bike usage?
* What are the characteristics of the users?

**Scope**

This report looks at many aspects of bike-sharing data, such as how the data is collected, cleaned, and prepared. It also includes how the key findings are shown and understood. The report goes into detail about usage patterns, user demographics, and possible ways to improve operations.

**Significance**

This report is important because it can help improve city transportation. Cities everywhere are trying to reduce traffic and support green travel options, and bike-sharing systems are a key part of this. By studying the data from these systems, this report helps us understand how they are used and how they can be made better to meet the needs of people living in cities.

**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 Line charts 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**

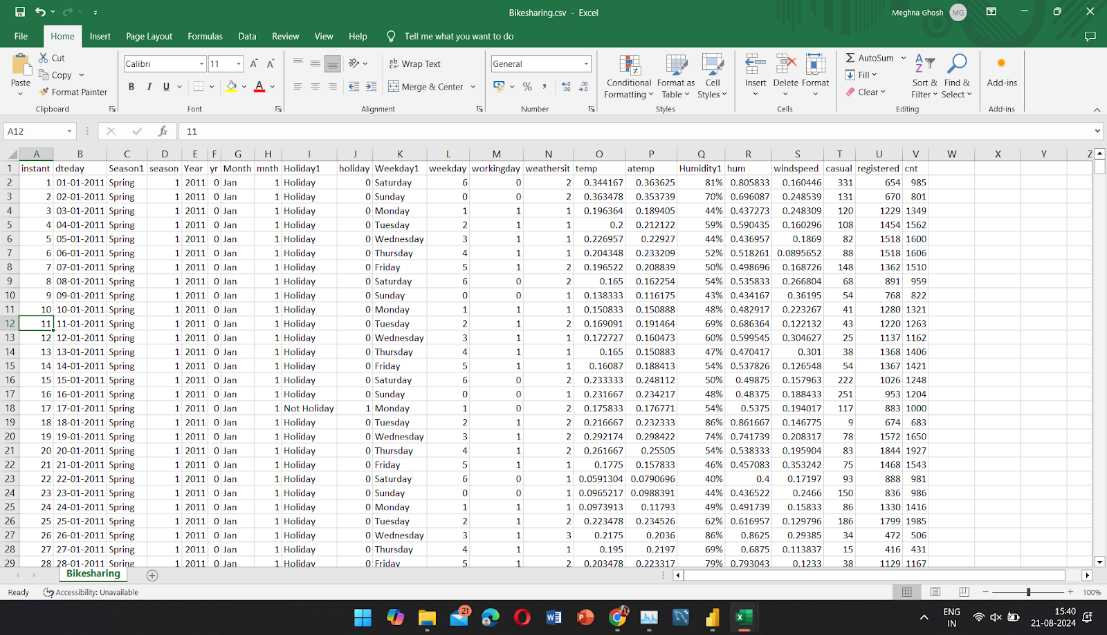
In this report, we used Microsoft Excel and Power BI to analyse public bike-sharing data. Both tools were essential for carefully cleaning, organizing, and analysing the data. This process led to clear visualizations that helped shape the findings and recommendations.

Microsoft Excel for Data Cleaning and Organization

Microsoft Excel was the main tool used at the beginning of the data analysis. It was especially useful for cleaning and organizing the raw data from Kaggle. Excel’s powerful features for handling large datasets and making data changes were essential for getting the data ready for further analysis.

Key Functions and Features Used:

I used Microsoft Excel to clean the data and Power BI to create visualizations and do detailed analysis. By using both tools, we were able to get valuable 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:

* 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.

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.This pie chart showing the percentage of total ridership for each month.A pie chart with numbers and a number of different colored circles

Description automatically generated

Peak months: **August (10.67%)**, followed by **June (10.77%)**, **July (10.52%)**, and **May (10.48%)**.

Lowest ridership occurs in **January (4.19%)** and **February (6.41%)**.

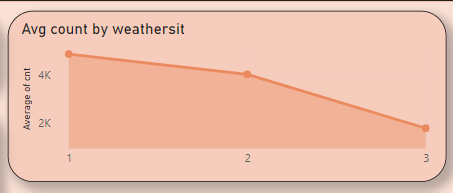
A graph on a screen

Description automatically generated

2.Bar chart comparing casual and registered user ridership across months in 2011 and 2012.

**Registered users consistently dominate** the ridership, especially in summer months.

Casual ridership is notably lower, but shows spikes during peak months (e.g., **July and August**).

3.This Line chart displaying the average bike count by weather condition.

Clear weather (Weathersit = 1) has the highest average ridership (~4.5K).

Ridership drops significantly as weather conditions worsen (e.g., cloudy or rainy).

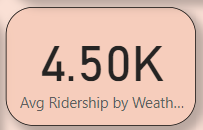
A screenshot of a computer

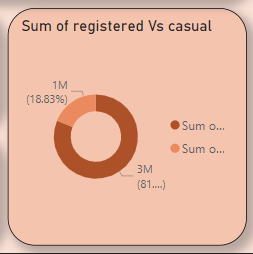
Description automatically generated

4.A table summarizing ridership across seasons with total counts for casual, registered, and working days.

**Summer** shows the highest total ridership, followed by Fall and Spring.

**Fall has the highest number of working days**, with the highest count of casual users.

5.A KPI displaying the average ridership (~4.5K) by weather condition. Good weather plays a crucial role in maintaining higher ridership.

6. A donut chart comparing the total ridership between registered users and casual users.

**Registered users make up the bulk of ridership (81%)**, while casual users account for a smaller share (18%).

**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:

* Adjust operational resources like bike availability and staff according to the seasonal demand, focusing on summer months.
* Promote casual membership or subscription programs during high-demand months to convert casual riders into registered users.
* Adjust bike-sharing promotions or pricing during unfavourable weather conditions (e.g., rain) to maintain ridership.
* Special promotions for casual users in the fall can further capitalize on ridership peaks in this season.
* Weather forecasts could be integrated into bike-sharing apps to help users plan their trips.
* Consider expanding operations during peak months or in regions with similar weather and ridership conditions.
* Introduce special offers for casual users during weekends or holidays to boost casual ridership.
* Optimize bike availability and maintenance schedules around peak months (May to September).
* Implement membership drives and promotions for casual users during peak months and holidays.
* Offer discounts or incentives during poor weather conditions to mitigate drops in ridership.
* Launch targeted campaigns during the fall and summer to attract more casual riders.

**Conclusion**

* Summer months see the highest bike usage, indicating a strong seasonality in ridership.
* A substantial gap between casual and registered users shows opportunities to convert casual riders.
* Weather is a major factor in determining ridership, with clear weather driving the most usage.
* The ridership pattern is strongly seasonal, with summer being the most active.
* Ridership is sensitive to weather conditions, necessitating contingency plans for adverse weather.
* There is consistent demand, with opportunities to optimize the system for peak performance.
* Registered users form the backbone of the service, but there’s potential to grow casual ridership.
* The data clearly shows ridership is strongly seasonal, with clear weather driving the most activity.
* Registered users are the primary source of ridership, indicating opportunities for growth in the casual user base.
* Aligning resources with peak times can lead to better performance and user satisfaction.