Visualization of Seoul Bike dataset in Streamlit

Ashutosh B. Rajan , Sudhansu Koirala
Student, Department of Computer Science & Engineering
Kathmandu University, Dhulikhel, Nepal
ashutoshbr@pm.me , sudhansukoirala@gmail.com
https://github.com/ashutoshbr/streamlit-dashboard-seoulbike

***Abstract*— This paper presents a detailed exploration of the Seoul Bike dataset, employing a diverse array of visualization techniques to unravel valuable insights embedded within the data. The dataset, rich in temporal and spatial dimensions, encompasses a wealth of information related to bike-sharing patterns in the vibrant city of Seoul. Leveraging the power of visual charts, our study aims to unearth hidden patterns, trends, and correlations that can contribute to a deeper understanding of Seoul Bike dataset. Through the application of various visual charts, including line graphs, scatter plots, pie chart, bar chart, bubble chart and interactive dashboards, this study aims to unravel hidden correlations and trends within the data.**

We aim to discern factors influencing bike usage, peak demand periods, and the impact of external variables such as weather conditions.

Keywords: Data Visualization, Dashboard

I. Introduction

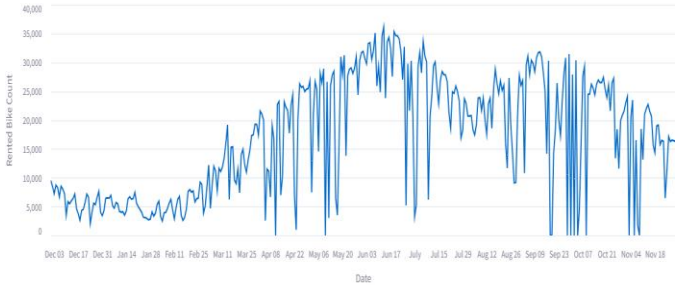

Fig. 1.1. Line chart representing rented bike count per day

The dataset encompasses 8,760 records, each characterized by 14 distinct attributes. These data points span from December 2017 to November 2018, capturing a comprehensive snapshot of bike rental activity. In the course of this timeframe, an impressive 6.17 million bikes were successfully rented.

Upon analyzing the line graph derived from this dataset, it becomes evident that the initial stages of data collection saw a modest rental rate, with figures ranging from 3,000 to 4,000 bikes. However, a noteworthy shift occurred after February 2018, leading to a significant surge in bike rentals. The peak of this upward trajectory was reached at an impressive 36,149 rentals per day. Subsequently, the trend stabilized, settling within a range of 20,000 to 30,000 bike rentals per day.

II. Distribution According to Season

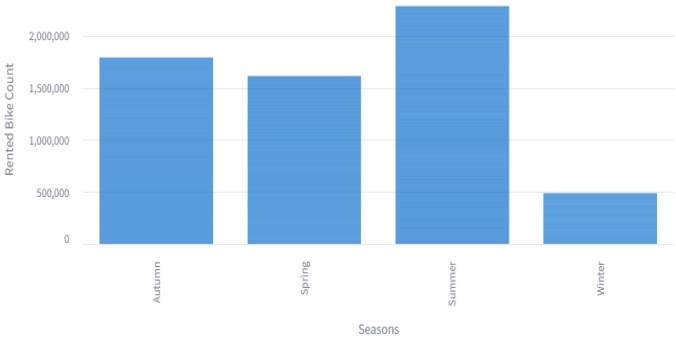

Fig. 2.1 Bar chart showing the distribution of rented bike according to seasons

Upon generating a bar chart to illustrate the distribution of rented bikes across seasons, a distinct pattern emerges. The summer season stands out as the peak period for bike rentals, with the highest count recorded. In contrast, the winter season exhibits the lowest bike rental activity.

Delving into the specific figures, the bike rental counts for each season are as follows: Autumn records 1,790,002 rentals, Spring observes 1,611,909 rentals, summer leads with 2,283,234 rentals, and Winter trails with 487,169 rentals.

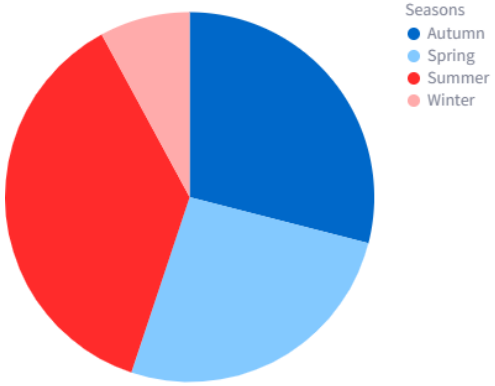

Fig. 2.2 Pie chart showing the distribution of rented bike according to seasons

The data can also be visualized using the Pie chart. This shows that people are less likely to go for cycling during winter.

III. Impact of Time

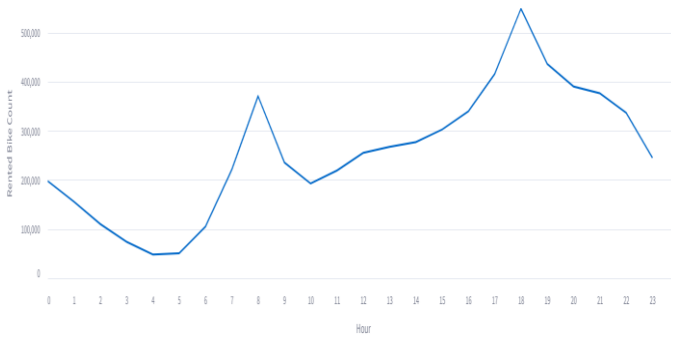

Fig 3: Line plot between time(hr) on rented bike count

In the dataset, each day comprises 24 data points representing hourly rental counts. By grouping these hourly counts together and plotting a Line Plot, it can be observed that over the 24-hour period, the highest number of bikes is rented around 6 P.M., with the second-highest peak occurring around 8 A.M. This information holds significant importance from a business perspective, as shopkeepers can capitalize on these peak hours to generate substantial profits.

IV. Impact of Rainfall

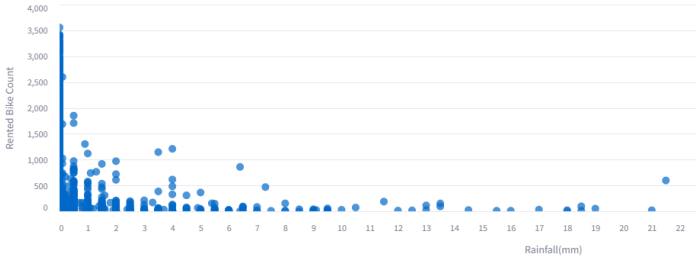

Fig. 4. Scatter chart representing influence of rainfall on the rented bike count

From the scatter plot above, it can be observed that when there is no rainfall, the rented bike count is unaffected, but even with small amount of rainfall, it can be observed that the bikes rented is drastically decreased. The scatter plot above gives the exponential drop in the rented bike count when there is increase in rainfall. Similar plot can also be observed when rented bike count is plotted against snowfall.

V. Impact of Holiday

From the below bar chart, it can be observed that the majority of bikes are rented, during workdays and very small number of bikes are rented during Holidays. This pattern suggests that individuals in Seoul are not only using rented bikes for exercise and health purposes but may also be relying on them as a means of transportation to commute to work.

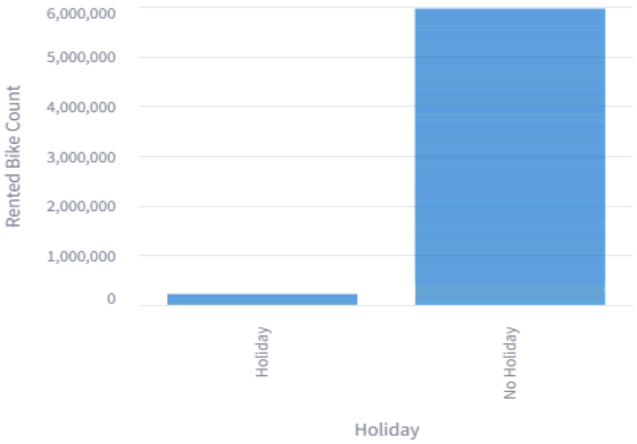

Fig. 5. Bar chart representing the effect of holidays on rented bike count

VI. Impact of Wind

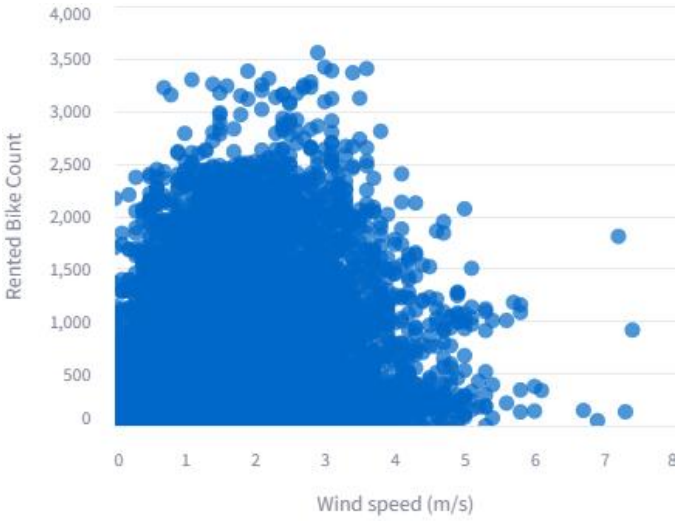

Fig. 6. Scatter chart representing impact of wind speed on rented bike count

In the scatter plot above, showing the relationship between rented bike count and wind speed, it is noticeable that the majority of data points cluster within the 0 to 5 m/s wind speed range. This observation indicates that bike rentals are more prevalent when wind conditions are lower, and wind speeds up to 5 m/s are generally tolerated. Beyond this threshold, it is less likely that the bikes will be rented.

VII. Impact of Solar Radiation

In a similar fashion, when the rented bikes were plotted against solar radiation, it was observed that as solar radiation increased, there was a corresponding decrease in the rented bike count. The radiation was tolerable up to 3.5 MJ/m2, and beyond this threshold, not a single bike was rented. This observation suggests that individuals in Seoul prioritize their health and skin, avoiding outdoor activities during periods of high radiation.

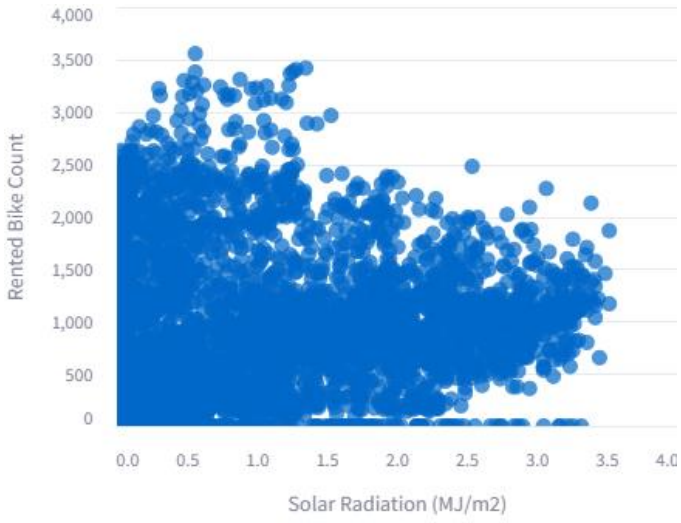

Fig. 7. Scatter chart representing impact of solar radiation on rented bike count

VIII. Impact of Visibility and Humidity

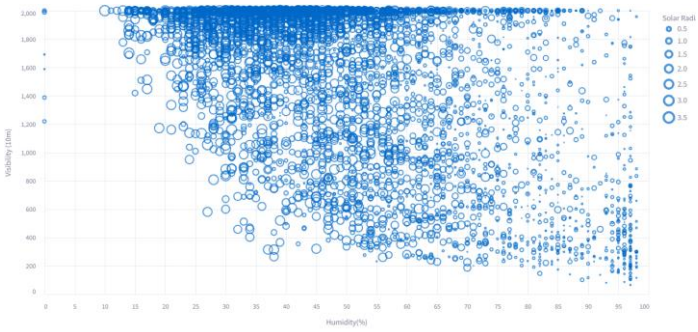

Fig. 8. Bubble chart representing impact of humidity and visibility

The bubble chart above illustrates the relationship between visibility and humidity, with the size of the bubble indicating solar radiation. Through this representation, it becomes clear that visibility and humidity are negatively correlated. Utilizing Pearson's coefficient method, a correlation of -0.543 was determined, signifying an inverse relationship—when humidity is high, visibility is low, and vice versa.

As humidity increases, the size of the solar radiation bubbles decreases. Additionally, a clustering of data points is observed around high visibility, suggesting that during periods of average humidity, when radiation is moderate and visibility is high, there is a significant increase in bike rentals.

IX. Methodology

A dynamic dashboard was developed using Streamlit [1], featuring multiple interactive charts that provide diverse insights. Users can seamlessly export these charts in various formats, such as PNG. Most of the charts were available in Streamlit itself, except the pie chart for which Altair chart [2] was used. It gave us a wide array of charts with more customizability. charts Moreover, the Orange Data Mining [3] tool proved invaluable for calculating several key statistical parameters, offering a rough estimate of the parameters with the most significant influence.

X. Conclusion

The dataset analysis revealed correlations between various environmental factors and the number of rented bikes. By graphing multiple parameters, we identified an optimal range that maximizes bike rentals. This information is crucial for shop owners aiming to enhance profits and optimize their bike rental strategies.

XI. References

[1] API reference - Streamlit docs. API Reference - Streamlit Docs. (n.d.). https://docs.streamlit.io/library/api-reference

[2] Vega-Altair: Declarative visualization in python. Altair. (n.d.). https://altair-viz.github.io/index.html

[3] Data Visualization | Orange Data mining (n.d.). https://orangedatamining.com/