Visualization

1. Target Audience

There are two target audiences for the app, the general public and bike-sharing managers.

General Public – They are individuals who use the Seoul Public Bike as a method of transportation and hence would be using the app to understand if the bikes will be available when they need them. These individuals include such as workers who ride the bike to work and students who ride the bike to school. The problem these individuals face is whether or not the bikes would be available for them to use when they need them. There are a few important things to note about them:

- Devices used: phone
- Motivation: to know when there are fewer bikes rented so they can rent one
- Others: they lack the knowledge to reading different graphs

Bike-sharing managers – Bike-sharing managers are Seoul City's public officers that are managing bike-sharing. Their purpose is to have enough bikes out so the public can use them, but not have too many out on the street so they don't block the road and sidewalks making it inconvenient. They would typically use both a laptop and a phone.

- Devices used: phone, laptop
- Motivation: they are interested in determining which variables are related to the number of bikes rented
- Others: they have a good understanding of graphs and would be interested in exploring the data

2. Questions addressed

- What time of the day are the bikes used the most?
- What month of the year should more bikes be available to be used?
- What are other factors are relevant to the changes in bike-sharing demand?

3. Reasoning behind overall visualization (format of the app)
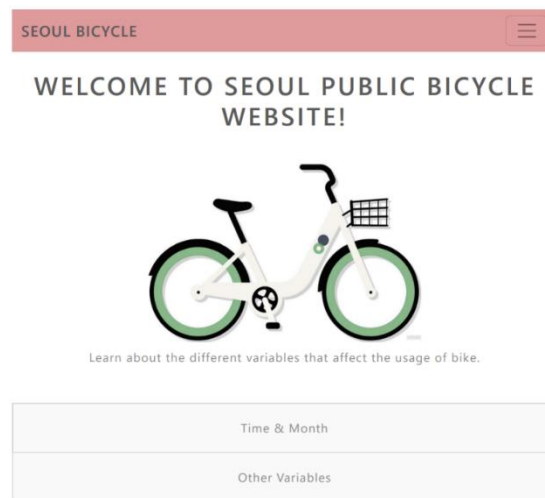

Figure 1

The dashboard is divided into two different tabs: the 'Time & Month' tab and the 'Other Variables' tab. The 'Time & Month' have graphs made up of the average bike rented for the month or the hour of the day. The 'Other Variables' have graphs made up of the number of bikes rented for the other variables aside from the ones used in the first tab. The tabs are there to organize the graph as there are two target audiences for the app, and the first target audience, the public, would rarely likely be interested in other variables but more in time-related variables. So, the graphs they would most likely use were put in the first tab and the graphs they would less likely use but would help provide insights for the bike-sharing managers were put in the second tab. Some of the points mentioned in Stephen Few's "Information Dashboard Design" is that it needs to fit in the desired screen so information can be monitored at a glance. Since the device that is expected to use by the users is a smartphone, a small screen, I have made the dashboard to only display one graph in one row of the screen.

4. Reasoning behind visualization of Graphs

Before describing the different elements of each visualization, it is important to note that for all of them the "Data-ink ratio" was considered. The "Data-ink ratio" introduced by Edward R. Tufte, an artist and statistician professor at Yale University, describes the non-erasable ink that makes up data information on the view. The key principle is to "limit the amount of noise and unnecessary formatting so the reader can focus on what the data has to say".[1] Taking it into consideration for all visualizations the blue background and the gridlines were removed. Gridlines are lines that cross the chart to show axis division. Although gridlines can be helpful to understand unlabelled data points faster, an interactive graph allows one to hover over a data point to see the value making it unnecessary.[2]

Moreover, all visualization has a title to help the viewers understand what kind of information they can expect to get from the visualization and the axes are labeled accordingly so the viewers know what data are displayed. Finally, no data was taken out or modified. All relevant data are shown to not mislead the viewers. In "Graphics Lies, Misleading Visuals" Alberto Cairo mentions the truncated Y-axis that gives a more dramatic effect on the data eventually misleading the viewer. To avoid any distortion in the data the scale of the Y-axis unit is only modified so the visualization represents data accurately.

4.1 1st Visualization: Weekday Line Chart

The data used for the first visualization is 'Hour' which represents the hour of the day and 'Count' the number of bikes rented per hour. The purpose of the visualization is to show how many bikes are rented on average for each hour of the day.
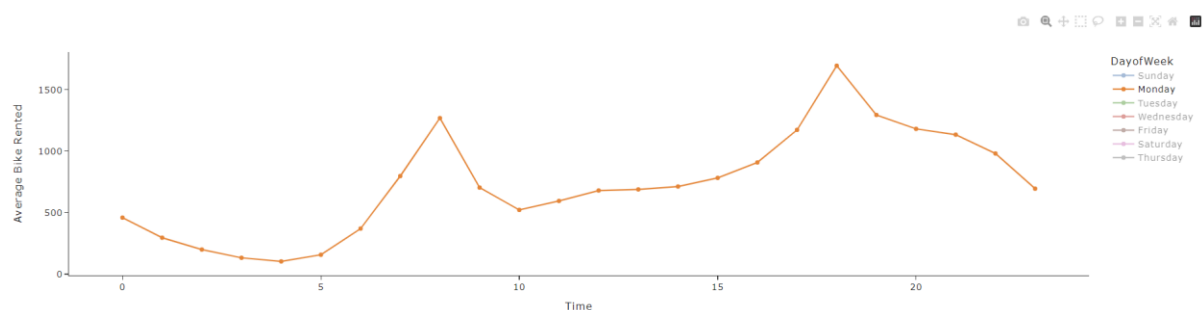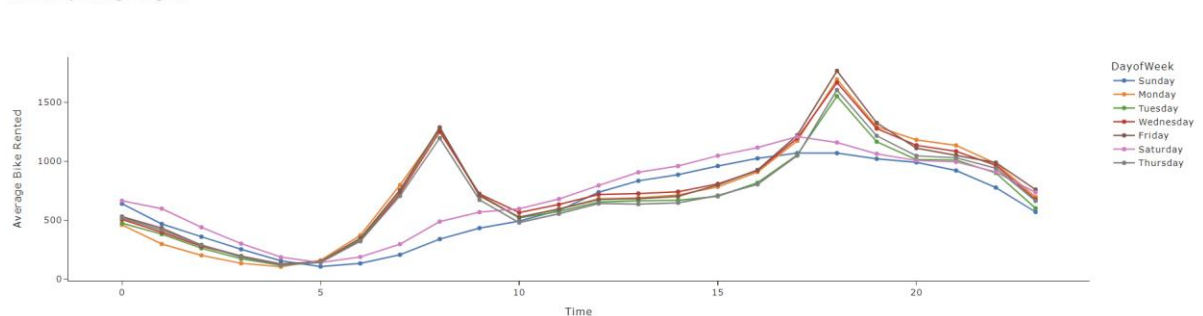

Figure 2

Figure 3

---

[1] (Tracy, 2021)
[2] (Castillo, 2020)

4.1.1 Chart Type – Line Chart

Line charts are used to show the quantitative values over a continuous interval or time period. It is considered an effective chart to compare lots of data all at once.
Hence is a great chart type to show the change in public bike rent trend for a different point in time.

4.1.2 Visualization

Initially, the Y-axis, which is the number of bikes rented, was not an average but the total rented for the different hours of the day. The data was going to be unhelpful to the viewers as it would not be a good predictor for the future, and it would have also been misleading since for each day of the week the number of bikes rented can be very different. For example, the number of bikes rented at 7 o'clock on Sunday is much fewer compared to the number of bikes rented on a Monday. To avoid misleading the viewer by hiding the relevant details, the 'DayofWeek' is added to show the average bike rented for each hour of the day for each day of the week. The markers are also added to make it easier for the viewers to identify the hour of the day on the graph without the grid lines. After the altercation, when plotted there are multiple lines on top of each other. Presenting too much data in a graph can "make it hard to identify salient data or stories"[3], specifically for this visualization the sheer number of lines makes it hard to focus on any one data point or trend. While this can be problematic, since it is an interactive chart, one can simply isolate a day of the week by double-clicking on the legend to solve the problem. Lastly to make it more convenient the randomly ranked legend was put in order from Sunday to Saturday.

4.1.3 Evaluation

The chart helps answer the question of when the bikes are rented the most in a day. The users can select the day of the week and hover over the time of the day to see the average bike that is rented. For the general public, this can help them decide if it would be a good idea to use the public bike as there might not be one available, and for the managers, it can help them decide if more bikes should be available at the time. It is also great to compare the trends of the different days of the week. One of the most important aspects of the visualization is that this graph would be used regularly by the general public. Hence. It being a line graph that doesn't require too much knowledge to read and understand the chart, makes it suitable for the target audience. Moreover, to not scare off the viewers by presenting too much data at once, initially only Monday's data is plotted and to help viewers interact with the visualization a short instruction is given on how to interact with the graph.
However, the visualization isn't perfect. Something that can be improved is having a dropdown list of days or weeks. This will make it much easier to interact with the graph.

---

[3] (Hogle, 2018)

4.2 2nd Visualization: Month Bar Graph

The data used for the second visualization is 'Month' which represents each month in a year and 'Count' the number of bikes that are rented per hour. The purpose is to show how many bikes are rented on average for each month.
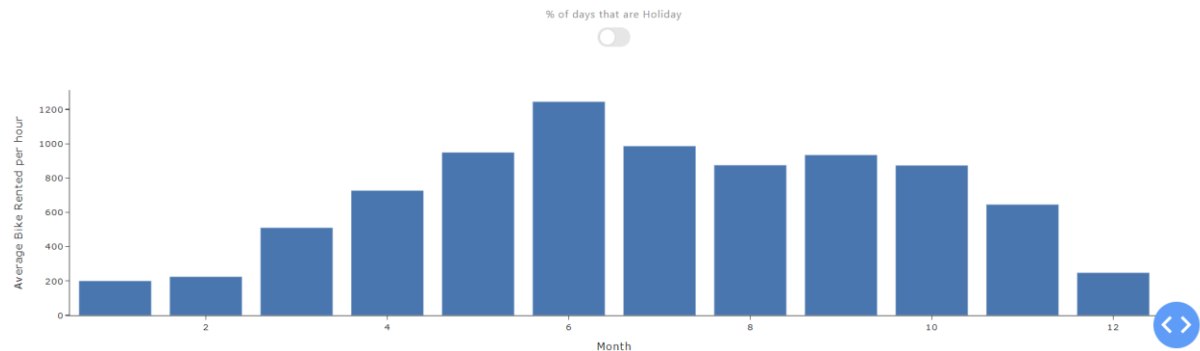
AVERAGE BIKE RENTED EACH MONTH

Figure 4
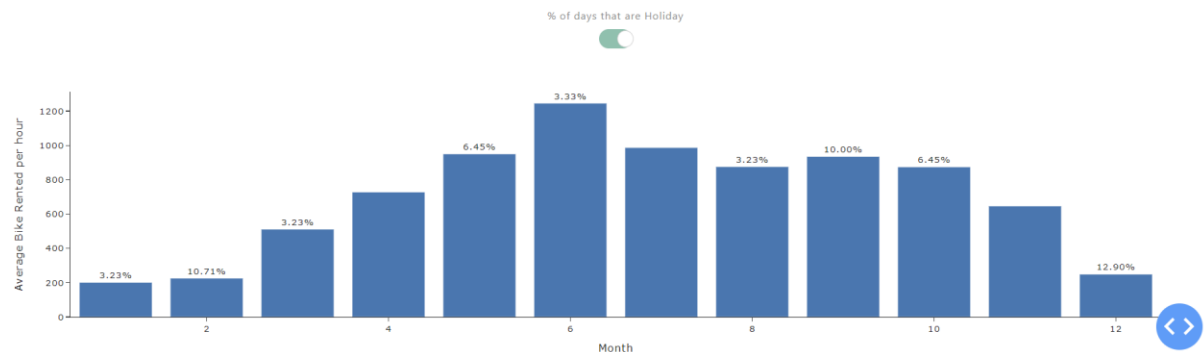
AVERAGE BIKE RENTED EACH MONTH

Figure 5

4.2.1 Chart Type – Bar Graph

Bar graph presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. It is used to show the discrete, numerical comparisons across categories making it a suitable graph to show the average number of bikes rented for each month.

4.2.2 Visualization

Similar to the first visualization of the original graph, the total number of bikes rented per hour each month was considered unhelpful for the users, as the average is better in understanding how much bike is likely to be rented per hour in a month. Other altercations made were removing the legend and the colors of each bar. This goes back to the "Data-ink ratio" mentioned above. The colors are kept uniform to remove unnecessary inks and the label is removed as the X-axis name is enough to show their difference. There is also a Boolean Switch that will show the percentage of holidays for each month on the bars.

4.2.3 Evaluation

The visualization help answer the question about in what month there is the highest use of bike per hour. The bar graph is one of the simplest graphs to read making it suitable for

my target audiences. The labels of my variables are also appropriately labeled, and the overall graph is very simple, making it an easy graph to read.

The X-axis doesn't have labels for all months because it was considered unnecessary as the months are lined in order and one can know the month by clicking on the bar. Moreover, it is important for dashboards to not only answer questions but also provide additional insight[4] and this visualization fulfills it through having a Boolean Switch. When the switch is on it will show the percentage of holidays in each month so one can seek to understand if holidays contribute to more rental of bikes. The Boolean Switch allows the data to be hidden and be seen only when one chooses to see it, so it doesn't interfere with the audience's understanding of the data by having too much data but can also provide additional insight after. While the overall data is good, one small improvement that can be made is to have the months written down in words, so "Jan", "Feb" etc in the X-axis.

4.3 3rd Weather Variables Scatter Plot

The third visualization is different from the first two, it has a tab that allows the Y-variable of the graph to change. The X-variables of the graph are the number of bikes rented per hour and the list of Y-Variables are Humidity, Windspeed, Temperature, Snowfall, and Rainfall. The purpose is to help determine what other non-categorical variables are important in predicting the demand for the bike.

HOW DOES OTHER VARIABLES EFFECT THE NUMBER OF BIKE RENTED
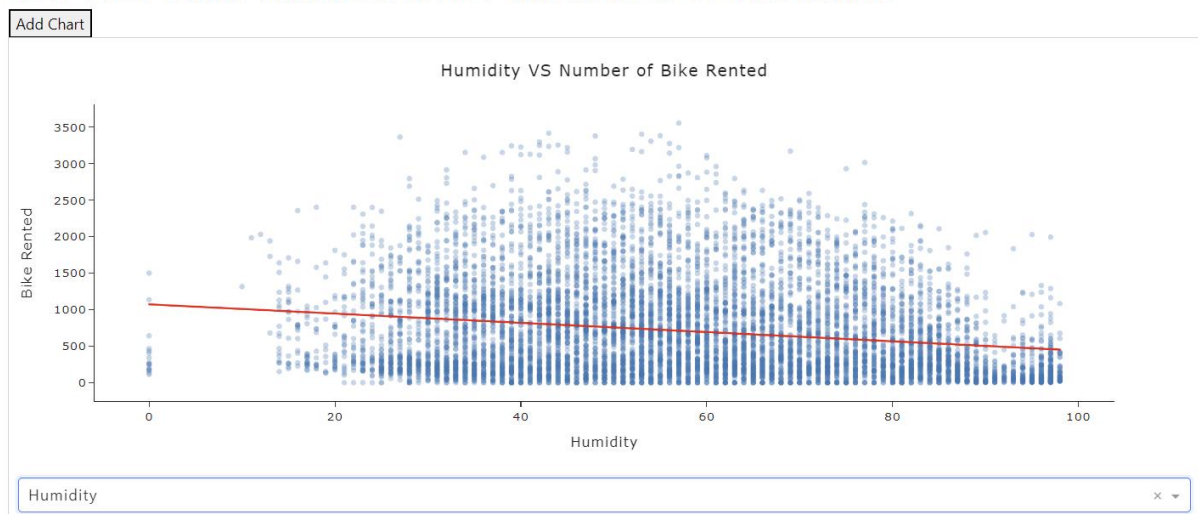
4.3.1 Chart Type – Scatter Plot
A scatter plot is a great chart to show the relationship between two data by simultaneously plotting the two numerical data along the vertical and horizontal axis. It shows if the data has a positive, negative, or neutral correlation.

4.3.2 Visualization
There is a total of 8760 data points plotted on the scatter plot making it overcrowded and overwhelming. Hence, to make them more differentiable and help the viewers understand the density of the data, transparency for each dot of the data was

---
[4] (Stephen few on Data Visualization: 8 core principles)

increased. However, with only the scatter plots plotted it can be hard to understand what the chart wants to say, so a trend line, also known as the line of best fit, is drawn to explicitly show the correlation. Since the trend line is more important here to understand the chart than the individual plots, the trend line is changed to red to emphasize it and to have it stand out from the points plotted. Furthermore, the title of the graph and the Y and X axis label change accordingly to the data plotted.

4.3.3 Evaluation

Although the scatter plot isn't one of the easiest graphs to understand for non-statisticians, the visualization is made for bike managers who have a good grasp of them. The main message that is supposed to be delivered to the viewers is the correlation between the two variables and by clicking on the trend line one can easily know what the correlation is.

The drop-down list allows the viewers to choose the variable they are interested in knowing concerning the number of bikes rented, and with the title changing with the variable they can easily know what variables are used. Furthermore, the "Add Chart" function allows viewers to create another chart below it to compare it. Compare is one of Stephen Few's 8 Core Principles, and while it is suggested to have graphs side by side, as mentioned previously because it is intended for users who would use it in a smartphone having two crowded graphs side by side would have made it unable for the viewers to gain the insights they need. So the "Add Chart" function adds a chart below, which still allows the user to compare but also so they can see all graphs clearly. Although the overall graph is good and easy to understand, a slight improvement that could have been made is the Y-axis. The Y-axis had different variables plotted and it would be necessary to have the units written down as well to help readers better understand the measurement of the Y-axis data.

Reference

Alberto, C. (2015) *Chapter 5 graphics lies, misleading visuals - university of california ...* Available at: https://faculty.ucmerced.edu/jvevea/classes/Spark/readings/Cairo2015_Chapter_GraphicsLiesMisleadingVisuals.pdf (Accessed: February 2, 2023).

Castillo, D. (2022) *4 ways to improve your graphs using Plotly, Dylan Castillo*. Dylan Castillo. Available at: https://dylancastillo.co/4-ways-to-improve-your-plotly-graphs/ (Accessed: February 2, 2023).

Charts and Graphs - a complete guide (no date) *storytelling with data*. Available at: https://www.storytellingwithdata.com/chart-guide (Accessed: February 2, 2023).

Chavan, S. (2021) *Plot beautiful and interactive graphs using Plotly, Medium*. Analytics Vidhya. Available at: https://medium.com/analytics-vidhya/plot-beautiful-and-interactive-graphs-using-plotly-64ee1951a038 (Accessed: February 2, 2023).

Deep, M. (2020) *Little known ways to make your data visualization awesome, Medium*. Towards Data Science. Available at: https://towardsdatascience.com/little-known-ways-to-make-your-data-visualization-awesome-890d71b6e365 (Accessed: February 2, 2023).

Dowling, E. (2016) *The Data School - Tableau Tip Tuesday - using transparency in scatterplots, The Data School RSS*. Available at: https://www.thedataschool.co.uk/emily-dowling/tableau-tip-tuesday-using-transparency-scatterplots (Accessed: February 2, 2023).

Few, S. (2006) *Information dashboard design*. Sebastopol, California: O'Reilly.

Fung, K. (2022) *Scatter plot, Junk Charts*. Available at: https://junkcharts.typepad.com/junk_charts/scatter_plot/ (Accessed: February 2, 2023).

Hogle, P. (2018) *Misleading data visualizations can confuse, deceive learners : Learning solutions, The Learning Guild*. Available at: https://www.learningguild.com/articles/misleading-data-visualizations-can-confuse-deceive-learners/?rd=1 (Accessed: February 2, 2023).

Ribecca , S. (no date) *The Data Visualisation Catalogue*. Available at: https://datavizcatalogue.com/index.html (Accessed: February 2, 2023).

Stephen few on Data Visualization: 8 core principles (no date) *Tableau*. Available at: https://www.tableau.com/blog/stephen-few-data-visualization (Accessed: February 2, 2023).

Tracy, S. (2020) *Covid-19 in charts: Examples of good & bad data visualisation, Analythical by Stephen Tracy - An Analytics & Research Blog*. Analythical by Stephen Tracy - An Analytics & Research Blog. Available at: https://analythical.com/blog/covid19-in-charts (Accessed: February 2, 2023).