Human Centered Visual Analytics

Coursework 1

Tahsin Saira 2152766

"Analyze and Visualize data in a programming language (Python and Altair) with a chosen Dataset"

Dataset Information:

Seoul Bike Sharing Demand Data Set:
https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand

Recently many big cities have adopted the use of rental bikes to improve mobility comfort and also for sustainable purposes. It is crucial to make the rental bikes accessible and available to the general public at the appropriate time since it reduces wait-time, overcrowdings etc. Eventually, maintaining a steady supply of rental bikes for the city emerges as a top priority. Predicting the number of bikes needed to maintain a steady supply of rental bikes at each hour's interval is essential.

The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

Objective:

From the dataset, we want to establish some trends. Which season is the most popular for renting out bikes? What is the average number of bikes being rented out for each season? Do people use them for work-purpose, or do they use them for recreational purposes only? What type of effect does the temperature have on the rental bikes?

Analyzing these trends can help the bike rental company prepare for rush hours and cut down on maintenance cost. Keep up with bike demands, provide better services. Track trends for next year etc.

Design:

The graphical depiction of information and data in a pictorial or graphical manner is known as data visualisation (Example: charts, graphs, and maps). Tools for data visualisation offer a simple approach to spot and comprehend trends, data patterns, and outliers. Tools and methods for data visualisation are crucial for processing vast volumes of data and making data-driven decisions.

1. Data Visualization Discovers the Trends in Data
2. Data Visualization Provides a Perspective on the Data
3. Data Visualization Saves Time

When forming visual designs, it is important to follow some rule of thumb interface design. For my designs, I will be considering these rules such as:

1. What type of visual format would be preferable for a dataset depends on what type of data it contains. Is it continuous? Ordered? Categorical?[3]
2. Understanding how humans process information: [3]

- Human Perception: Context is important as it helps make sense of ambiguity from prior knowledge [3].
- Pragnatz: Meaning, simplest possibility wins [3].
  For this dataset, we can take a quick look at the data and draw some basic assumptions that the number of rented bikes would most probably vary depending on seasons or months.

3. Shneiderman's mantra in practice [3]:
   i. **Overview:** This visualization should display an **average** overview of the monthly data as well as distinct seasonal difference in the number of rented bikes. We will use a scatter plot for this. The filters are off at this stage.
   ii. **2 - Zoom, filter, and details-on-demand:** When we hover over a plot on the scatterplot, we can see additional zoomed in information such as: Which Season, Date, Rented Bike Count.
   iii. **3 - More Filtering and Relate:** After clicking on a certain plot, the visualization enables to identify for each season with a boxplot. The boxplot tells us the min and max number of bikes needed depending on the temperature. The box plot will also show any outliers as well. We put more filtering on it. We can plot Number of Rented Bike Count on the X axis, and Temperature(°C) on the Y axis. Moreover, we can find the ratio between people using the bikes on non-holidays for work purpose and people using them on holidays for recreational purposes. So, we color them according to Holiday and Non-Holiday.
4. After effective visualization, it is also important to measure its effectiveness [1]. The overall data trends should make sense.
5. Asking the right questions to find the underlying pattern in important as well.[2]
6. Prioritizing colors over text, because human brain responds faster with color. [3]

The prototype 1 is a scatter plot and Prototype 2 is a Box Plot:

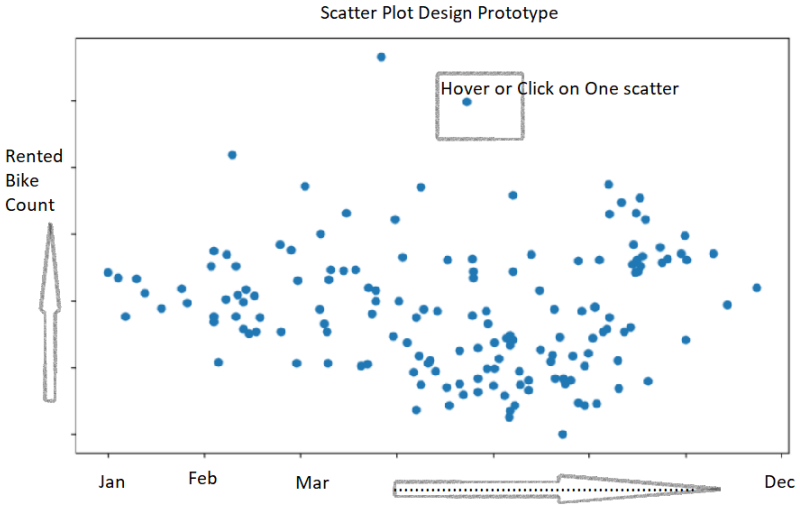


Fig: Prototype 1 of a scatter plot

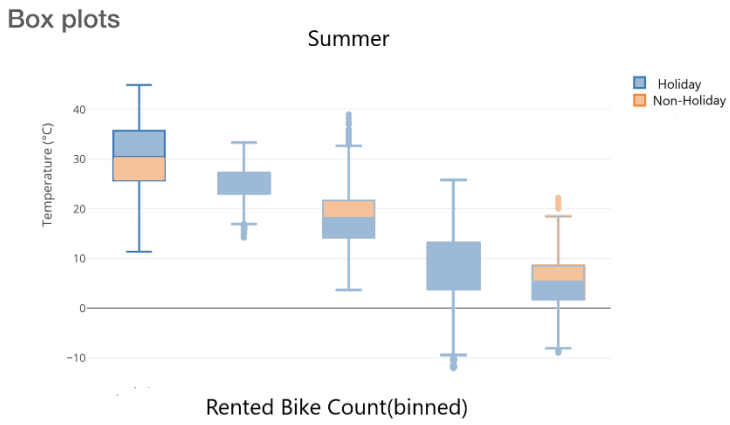
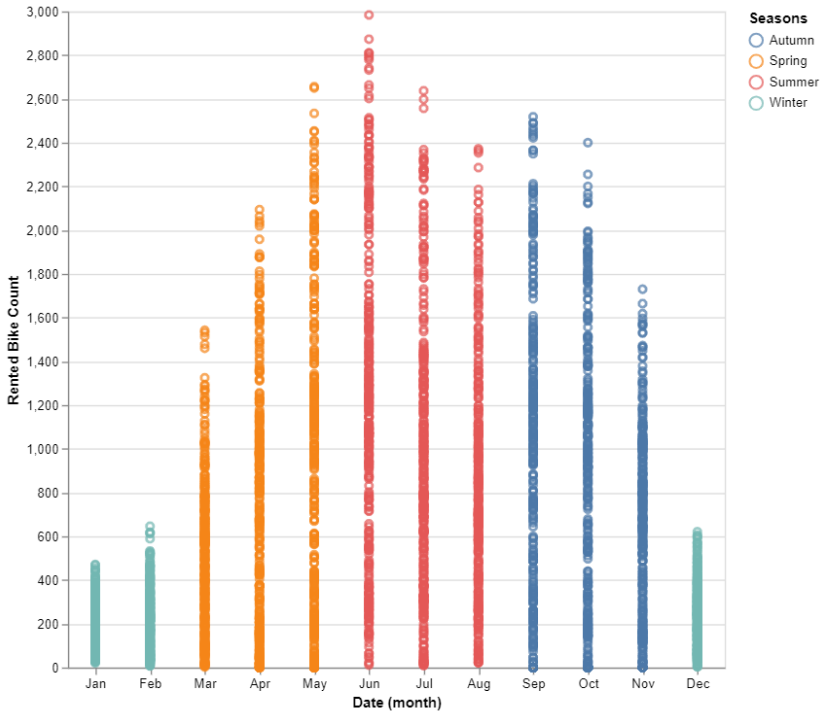
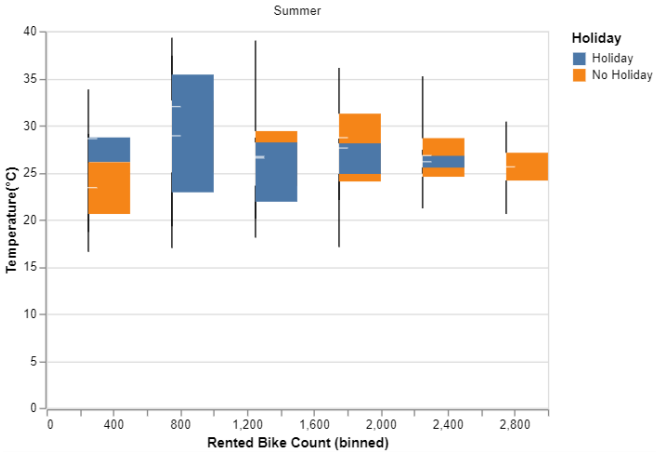


Fig: Protype 2 is a box plot of each season (Depending on the Scatter plot clicked)

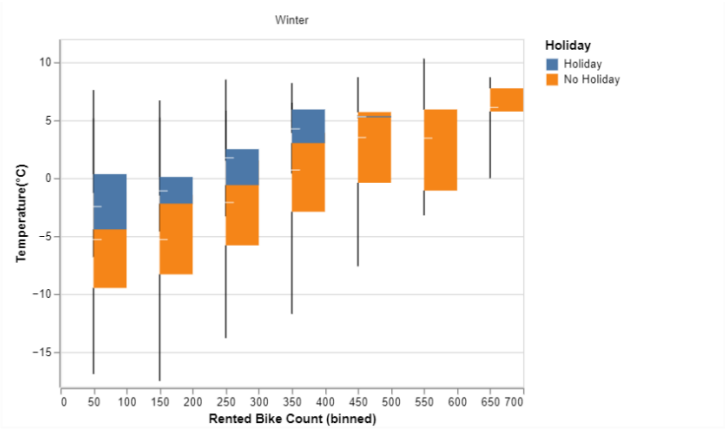
Findings: After running the code, we get the following finding:



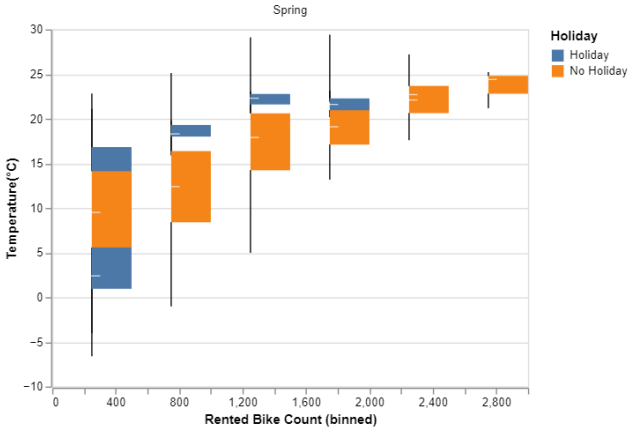
1. From the scatter plot, we can definitely see a clear trend that during Summer and Spring time, the bikes are being rented out the most. On the other hand, during Winter, hardly any bikes are being rented out.



2. During Summer, when the temperature is around 25 degrees Celsius, we need around 2800 bikes but when the temperature is over 33 degree Celsius, we need around 800 bikes only. Insight: People prefer taking the public transport system such as bus or subway which has air conditioning in it during summertime to commute to work.



3. During Wintertime, the maximum number of bikes needed is around 700 only. No more than that is needed. Also, very few people use the bikes for work purpose. Most bikes are being rented for Holiday or recreational purposes only.



4. During Springtime, depending on the temperature, the number of bikes rented vary a lot. Around 2,800 bikes on average needed daily when the temperature is over 22 degrees. And most people are using it for Holiday purposes as well.

In conclusion, I have used the Date, Temperature, Holiday, Rented Bike Count, seasons these 5 columns or attributes to find the underlying trends for rented bikes in Seoul. We can see a clear trends between months and seasons and conclude how many bikes on average are needed for each month or season. This finding can help companies prepare for rush hour, keep up with high demand, cut down on maintenance cost when bike demand is low and overall provide better service.

Reference

1. Zhu, Y. (2007). Measuring Effective Data Visualization. In: , *et al.* Advances in Visual Computing. ISVC 2007. Lecture Notes in Computer Science, vol 4842. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-76856-2_64
2. Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilović. 2019. How Data ScientistsWork Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? Proc. ACM Hum.-Comput. Interact. 3, GROUP, Article 237 (December 2019), 23 pages. https://doi.org/10.1145/3361118
3. Human Centered Visual Analytics Module Materials.