

MAKERERE



UNIVERSITY

SEMESTER ONE 2024/2025 ACADEMIC YEAR

SCHOOL COMPUTING AND INFORMATICS TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

MASTER OF SCIENCE IN COMPUTER SCIENCE

MCS 7103

MACHINE LEARNING

EXPLORATORY DATA ANALYSIS

BRENDA NUWASIIMA

2024/HD05/21941U

2400721041

EXPLORATORY DATA ANALYSIS FOR BIKE HIRING(SHARING) IN SEOUL

[Seoul Bike Sharing Demand](#)



[UCI Machine Learning Repository](#)

<https://archive.ics.uci.edu/datasets/Seoul+Bike+Shari...>

29 Feb 2020 — The **dataset** contains a count of public **bicycles** rented per hour in the **Seoul Bike** Sharing System, with corresponding weather data and holiday information.

Background

There is a rise in urbanization in many cities across the world which has resulted in extreme congestion, pollution, high cost of living and environmental concerns. This has prompted many to seek sustainable transportation solutions to reduce traffic, lower carbon emissions, and improve public health.

Bike hiring has been successfully implemented in cities like Amsterdam, Paris, Copenhagen, Seoul, New York City, San Francisco, London, Barcelona, Montreal, and Melbourne. In Kampala, Makerere university in partnership with Star Digital Bikes introduced bicycles to ease movement on the hill and promote exercise while sustaining the environment. [1]

A deeper understanding of the bike-sharing market will offer action points and recommendations for government, policymakers and entrepreneurs and to improve city life.

Introduction

This report will show the data analysis carried out on the above data set, conclusions from exploration and steps taken.

STEP 1: DATA WRANGLING.

```
data = pd.read_csv('/content/drive/MyDrive/colab Notebooks/SeoulBikeData.csv', encoding='utf-8')
```

data.head(5)

	Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0	Winter	No Holiday	Yes
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0	Winter	No Holiday	Yes

Figure 1: Snapshot of the data , these are the first 5 columns.

QUESTIONS

1. What is this data set about?
2. What are the attributes?
3. Is the data clean?

Data.info()

This data set has 8760 entries, 0 to 8759 and the following Columns.

1. Date 8760 non-null object
2. Rented Bike Count 8760 non-null int64
3. Hour 8760 non-null int64
4. Temperature(°C) 8760 non-null float64
5. Humidity(%) 8760 non-null int64
6. Wind speed (m/s) 8760 non-null float64
7. Visibility (10m) 8760 non-null int64
8. Dew point temperature(°C) 8760 non-null float64
9. Solar Radiation (MJ/m2) 8760 non-null float64
10. Rainfall(mm) 8760 non-null float64
11. Snowfall (cm) 8760 non-null float64
12. Seasons 8760 non-null object
13. Holiday 8760 non-null object
14. Functioning Day 8760 non-null object

Datatypes: float64(6), int64(4), object(4)

conclusion

- This dataset has 14 features (columns) and 8760 rows.
- 'Rented Bike Count' is the key attribute for this set.
- The data does not have any null values or duplicated records.
- The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

STEP 2: DATA EXPLORATION

QUESTIONS

1. Which season has the most bike rentals?
2. In which month or day are most bike rentals?
3. How do weather conditions affect bike rental count?
4. Which attribute has the biggest effect on our variable?
5. How do the types of days or seasons affect bike rental count?

SEASONS

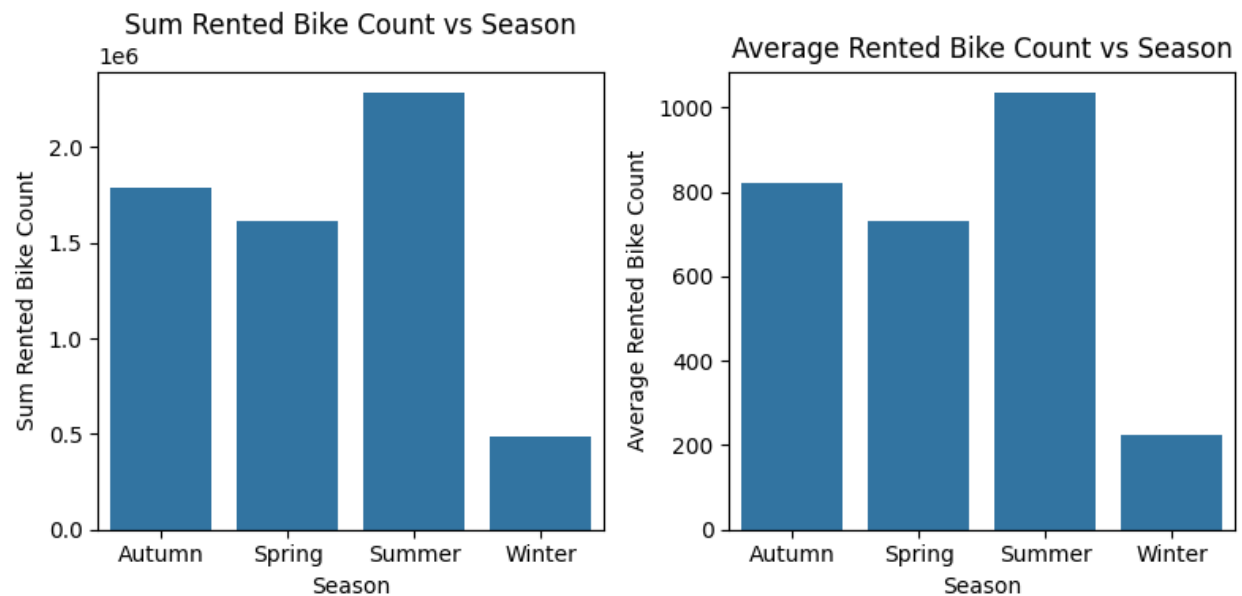


Figure 2: Sum and Average bike count in each season

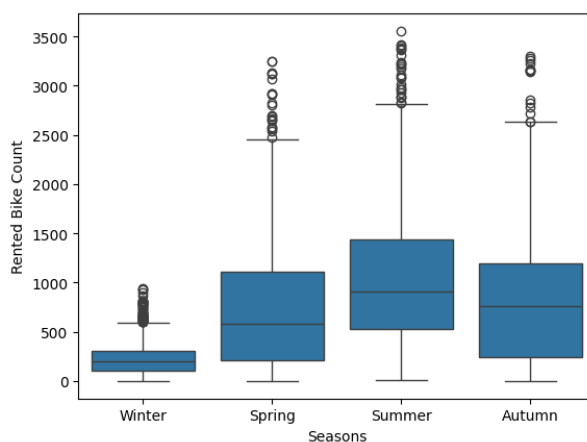


Figure 3: Distribution of bike count

According to the graphs shown above, most bikes are rented in summer and most days range between 500 – 1500 per day. The lowest number is in winter with ranges less than 400 a day.

CALENDER

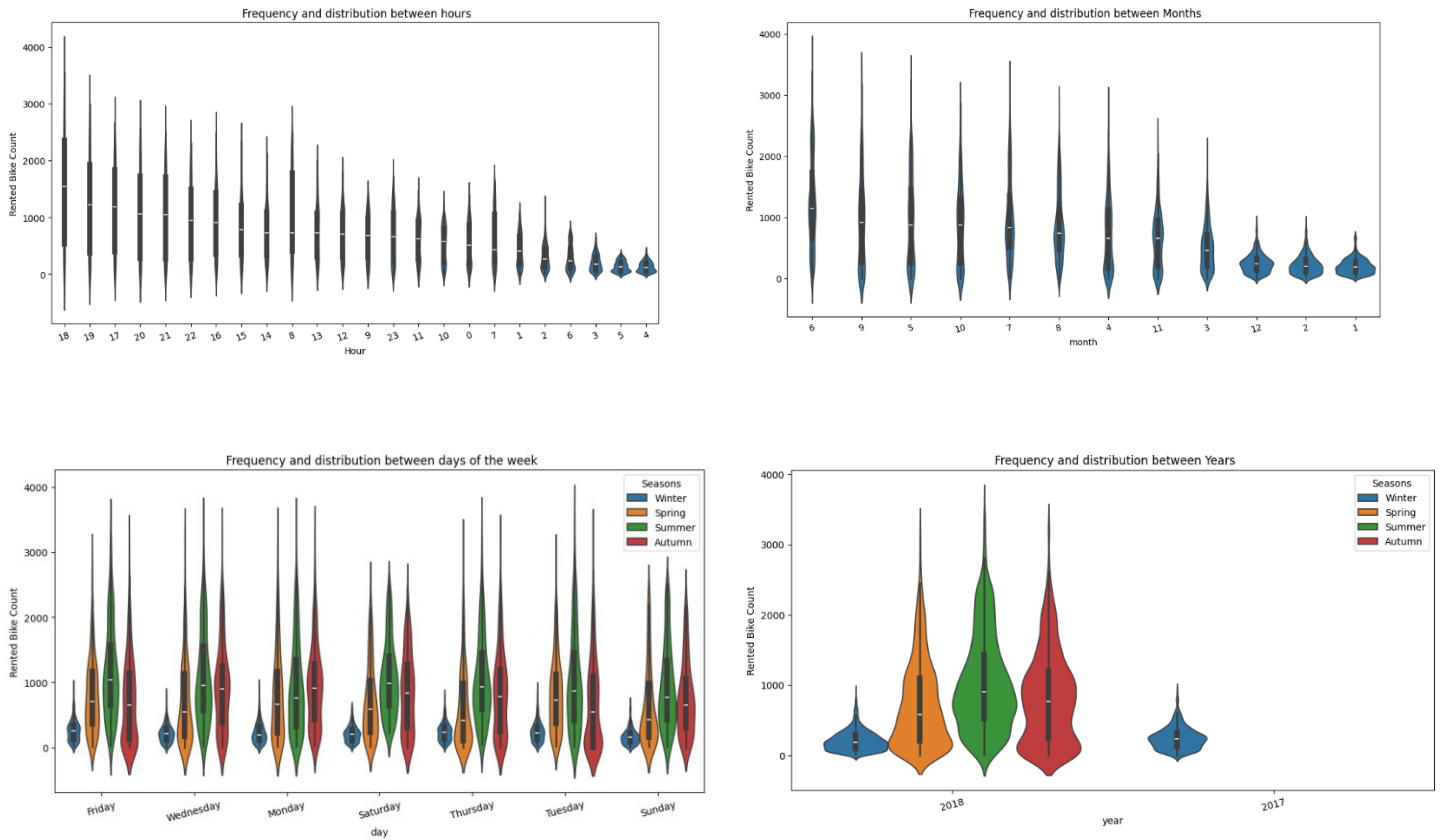


Figure 4: Distribution and frequency of rentals by the calender.

- According to these analysis visuals, bikes were most rented in 2018, The month of June and September have peak. The afternoon hours of 17:00 – 19:00 have the most rentals and Friday.
- No bikes are rented on the Non-functioning Day hence that column is redundant in prediction.
- More bikes are rented on non-holiday days i.e. Holiday – 215895 bikes, No Holiday 5956419 bikes

WEATHER.

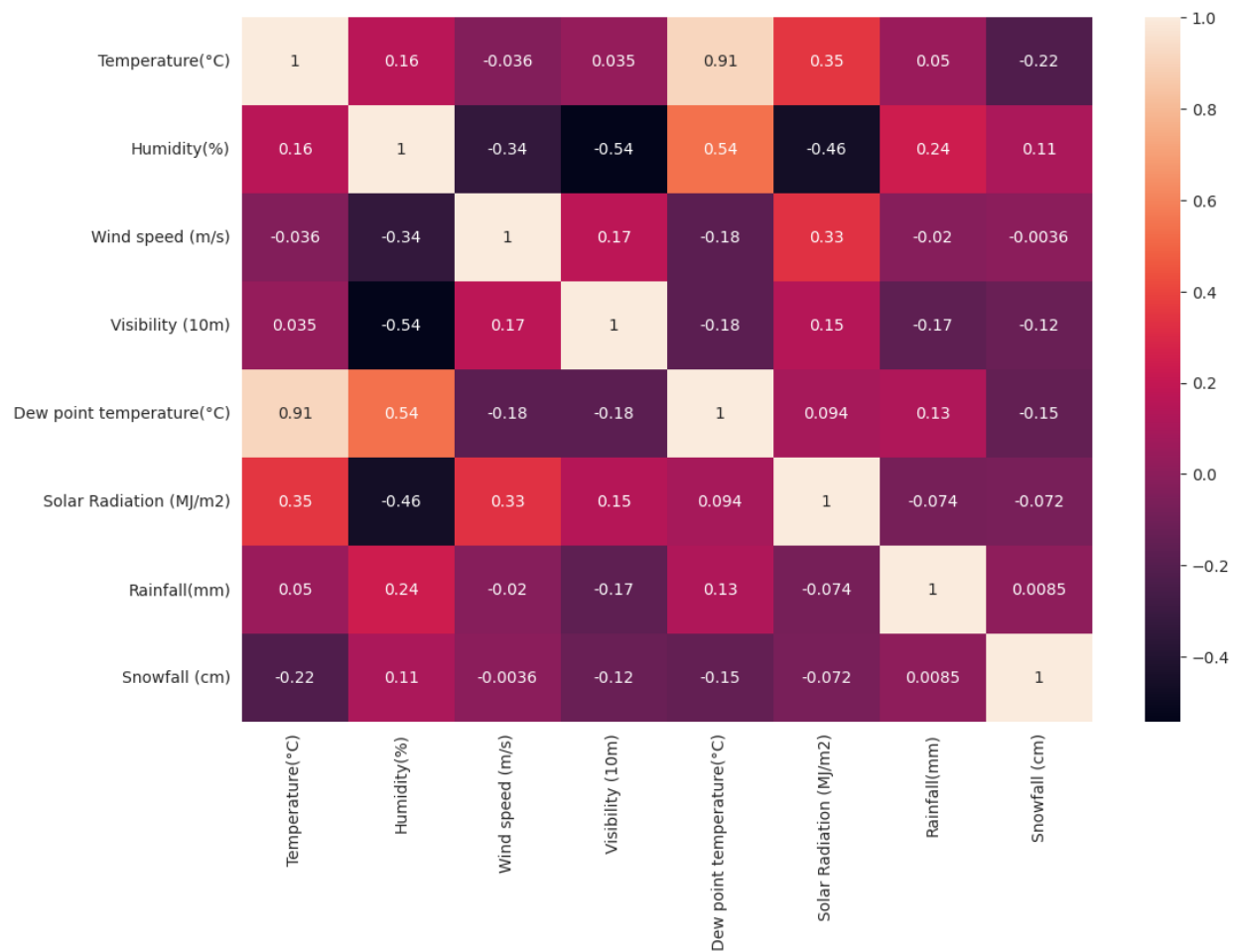


Figure 5:Figure 5: Correlation between weather conditions

Temperature and Dew point temperature have a high correlation while Rainfall and snowfall have high correlation.

Rented Bike Count Vs Weather

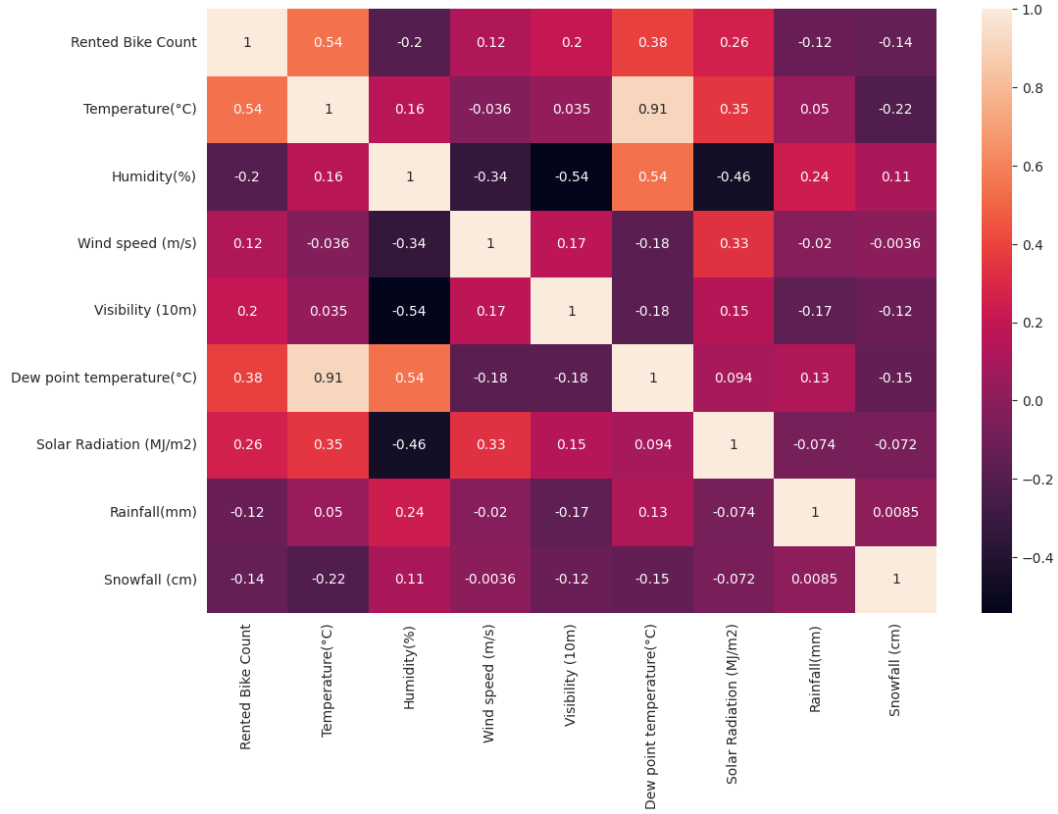


Figure 6: Correlation between Rented Bike Count and weather conditions

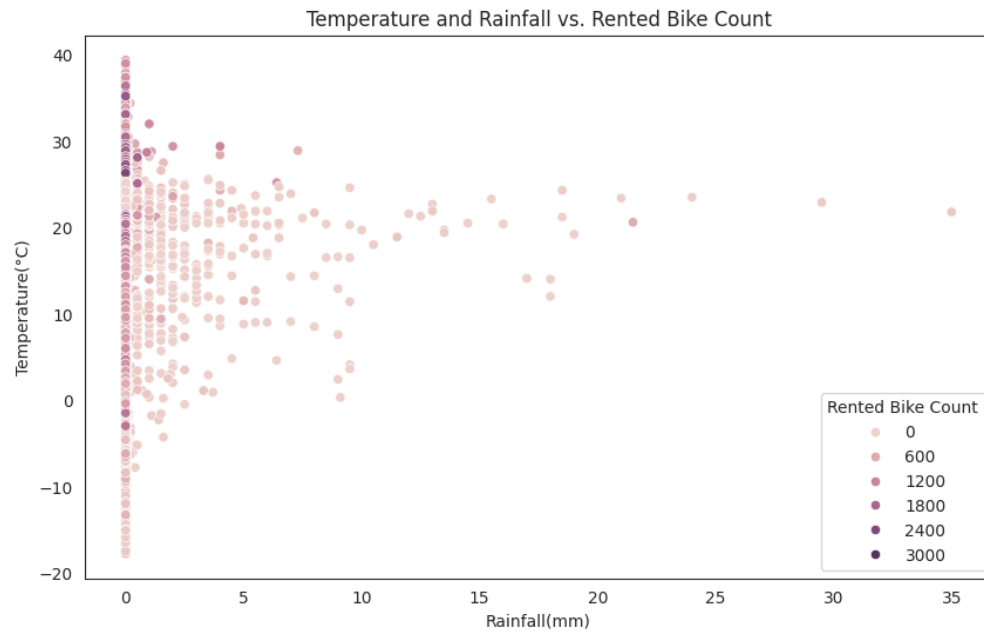


Figure 7: How Temperature and rainfall influence bikes rented.

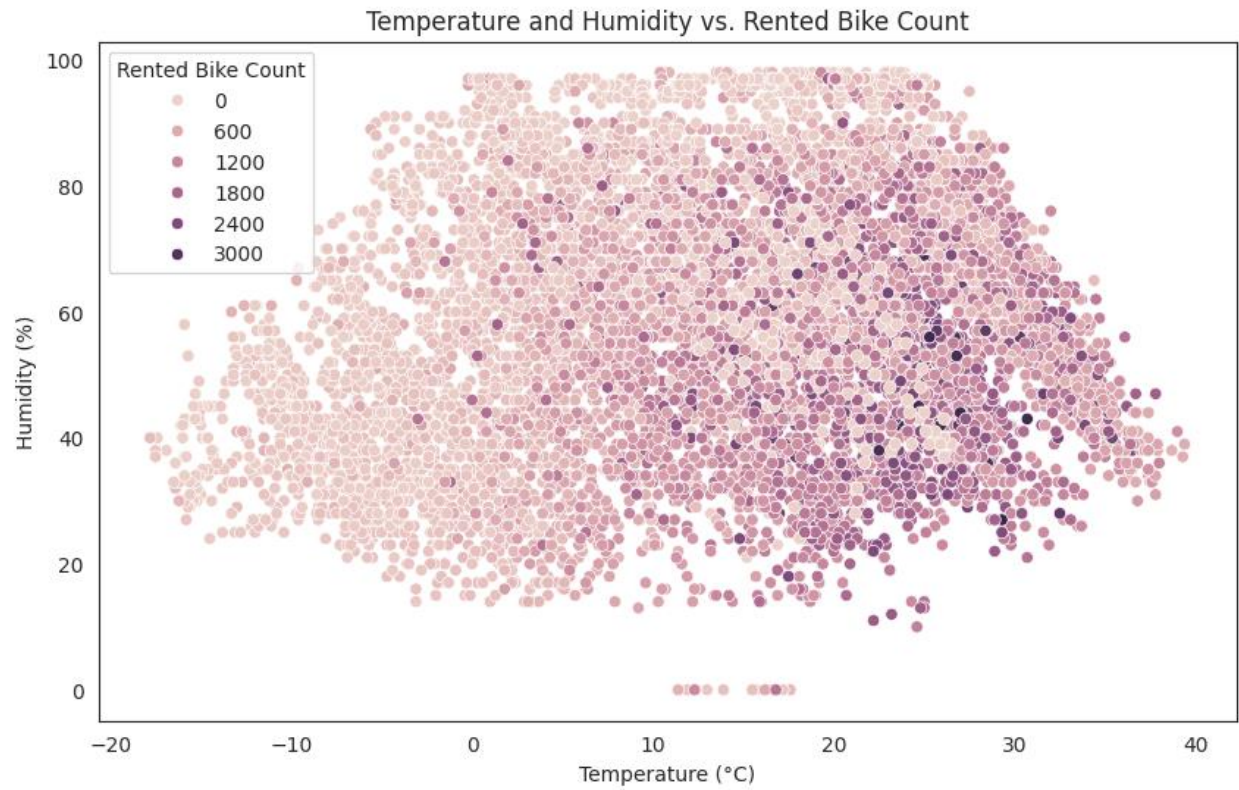


Figure 8: Humidity, Temperature and Rented Bikes

Here concentrating on temperature in the two graphs above shows that the higher the temperature the higher the number of bikes rented.

Rented Bike Count Vs All attributes

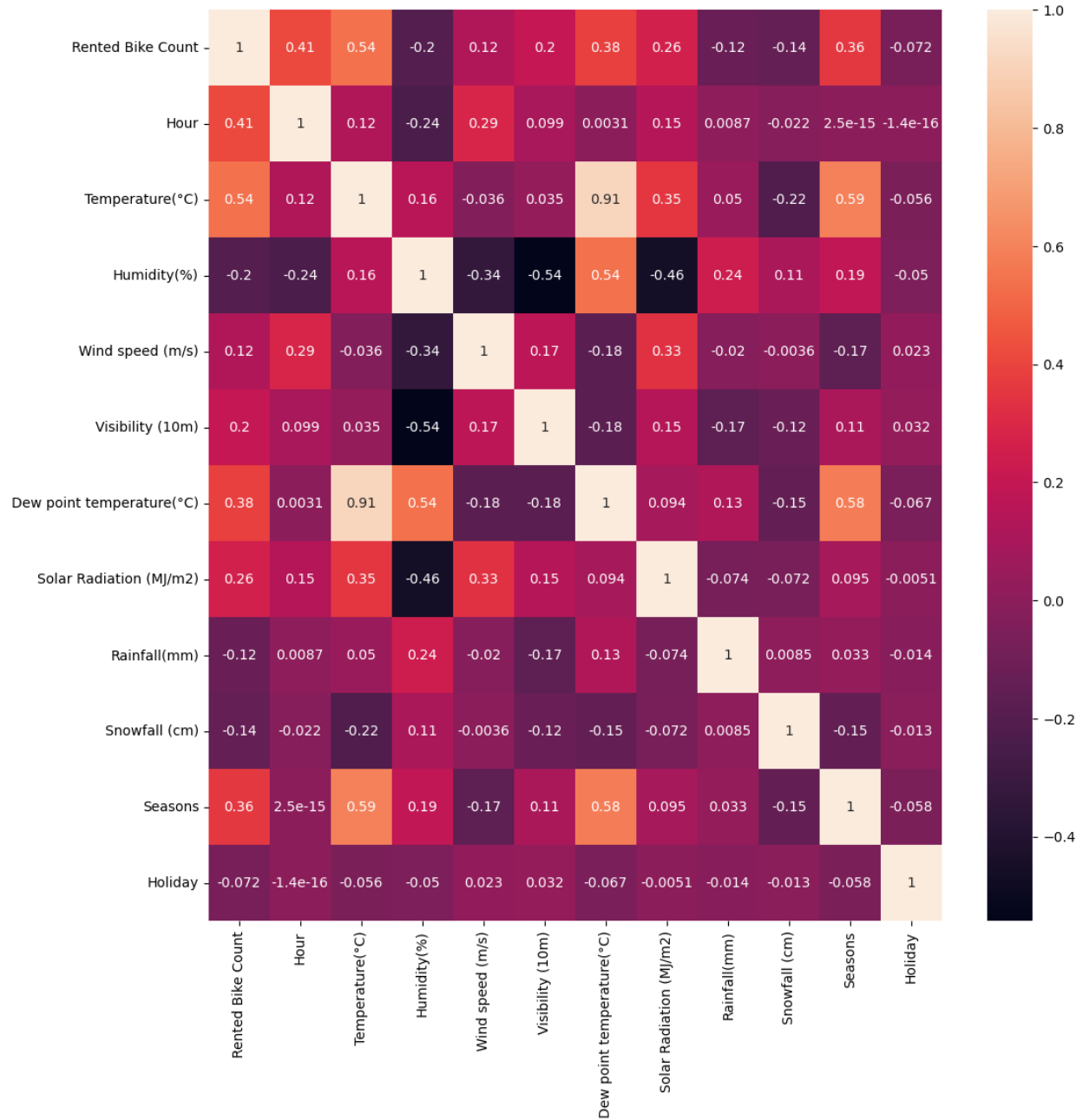


Figure 9: Correlation between all Variables

Strong Positive Correlation:

- Temperature has the highest correlation to depict the number of bikes rented. Strong positive correlation with Rented Bike Count is also noted with Humidity, Wind speed, and Visibility.

- This suggests that as these weather conditions increase, the number of rented bikes also tends to increase.

Weak Correlation:

Dew point temperature' has a weak correlation with Rented Bike Count which means that suggests that the dew point temperature doesn't significantly influence bike rentals.

Conclusion

The heatmap indicates that weather conditions, particularly temperature, humidity, and wind speed, have a significant impact on the number of rented bikes.

The correlation between 'Rented Bike Count' and these weather parameters indicate that favorable weather conditions are associated with higher bike rental demand.

Temperature, hour of the day, dew point and season in the order are great determinant factors for bikes rented on a particular day