# CAPITAL BIKESHARE

Name: Ze Weng Ng

Imperial Business School

# Introduction – Capital Bikeshare Bike Rental Data

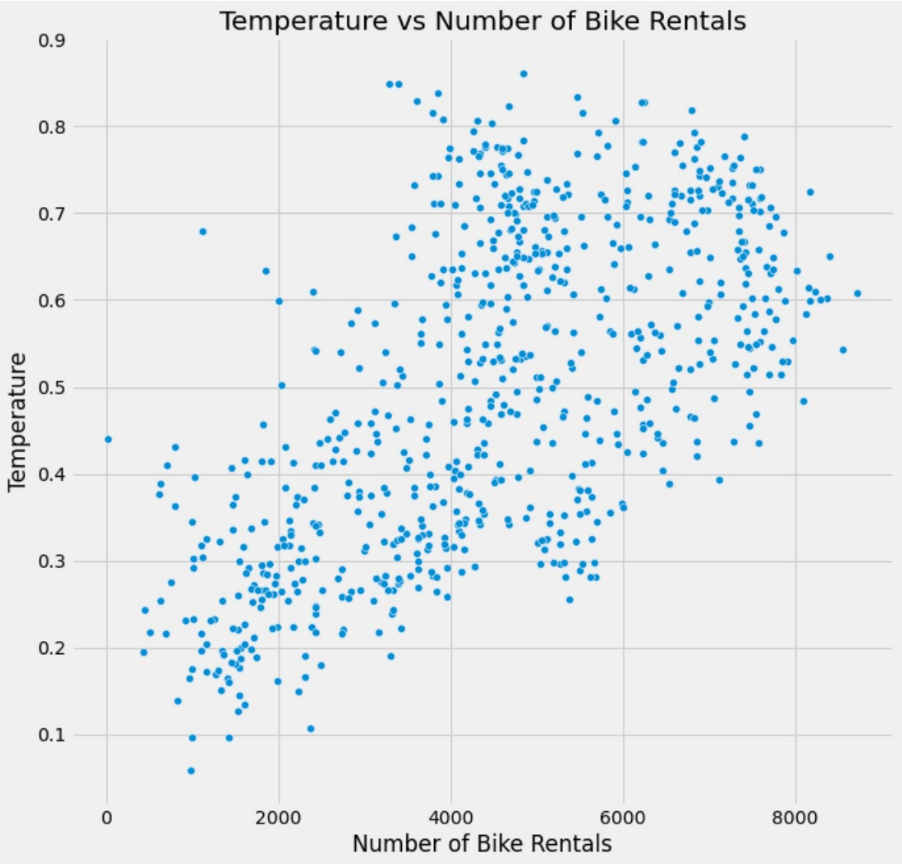
- The dataset I used was obtained from Data Science Dojo. I will be exploring contains the daily count of rental bikes between years 2011 and 2012 collected by Washing based bicycle sharing company Capital Bikeshare as shown in the table preview below. See appendix 1 for column definitions.
- The bike-sharing rental process is understood to be highly correlated to the environmental and seasonal settings so this dataset contains primarily data about the season, weather conditions, humidity etc.
- I will be using this dataset to identify opportunities for growth and suggestions on how Capital Bikeshare can use this data to make better business decisions.

| | instant | dteday | season | yr | mnth | holiday | weekday | workingday | weathersit | temp | atemp | hum | windspeed | casual | registered | cnt |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2011-01-01 | 1 | 0 | 1 | 0 | 6 | 0 | 2 | 0.344167 | 0.363625 | 0.805833 | 0.160446 | 331 | 654 | 985 |
| **1** | 2 | 2011-01-02 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0.363478 | 0.353739 | 0.696087 | 0.248539 | 131 | 670 | 801 |
| **2** | 3 | 2011-01-03 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0.196364 | 0.189405 | 0.437273 | 0.248309 | 120 | 1229 | 1349 |
| **3** | 4 | 2011-01-04 | 1 | 0 | 1 | 0 | 2 | 1 | 1 | 0.200000 | 0.212122 | 0.590435 | 0.160296 | 108 | 1454 | 1562 |
| **4** | 5 | 2011-01-05 | 1 | 0 | 1 | 0 | 3 | 1 | 1 | 0.226957 | 0.229270 | 0.436957 | 0.186900 | 82 | 1518 | 1600 |

# Correlation between Bike Rentals and Temperature

- It was observed in the dataset that temperature and ‘feeling temperature’(how we experience the temperature) have the highest correlation to the number of bike rentals on a particular day with each variable having a positive correlation of 0.63. Refer to Appendix 2 for further information on other correlation values.
- The relationship between temperature and number of bike rentals can be illustrated using a scatterplot as shown in figure 1 where we observe a positive relationship between temperature and number of bike rentals.

Figure 1. Temperature Vs Number of Bike Rentals Scatterplot



# When do we observe the highest bike rentals?

- I performed an analysis of the average bike rentals in the different seasons throughout the year and found that spring tends to has the lowest average number of bike rentals as shown in figure 2. This can be explained by figure 3 which shows that spring has the lowest average temperature (surprisingly not winter) and this is inline with my observation that temperature has a high positive correlation with number of bike rentals.

<u>Business Suggestion:</u>

- Based on the observations made on temperature, I think there is potential to grow the number of bike renters in Spring. Since we know that temperature plays a big influence on whether someone decides to cycle, I would suggest that during spring we could tailor our social media post to teach people how to keep warm while cycling this winter season.
- We may also explore the possibility of partnering with outdoor companies such as North Face to get exclusive discounts for Capital Bikeshare users on winter cycling equipment like cycling jackets and socks to encourage people to cycle during colder months.

Figure 2. Bar Chart showing the average bike rentals in different seasons

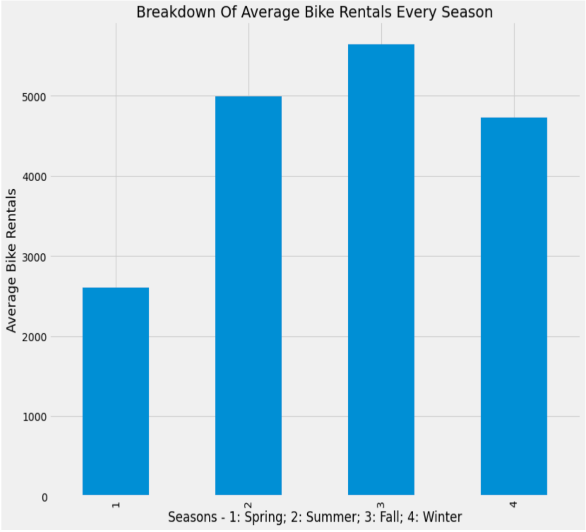
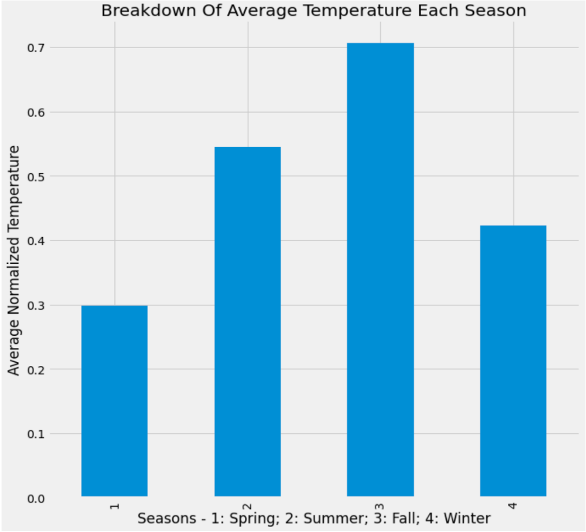



Figure 3. Bar Chart showing the average temperature in different seasons



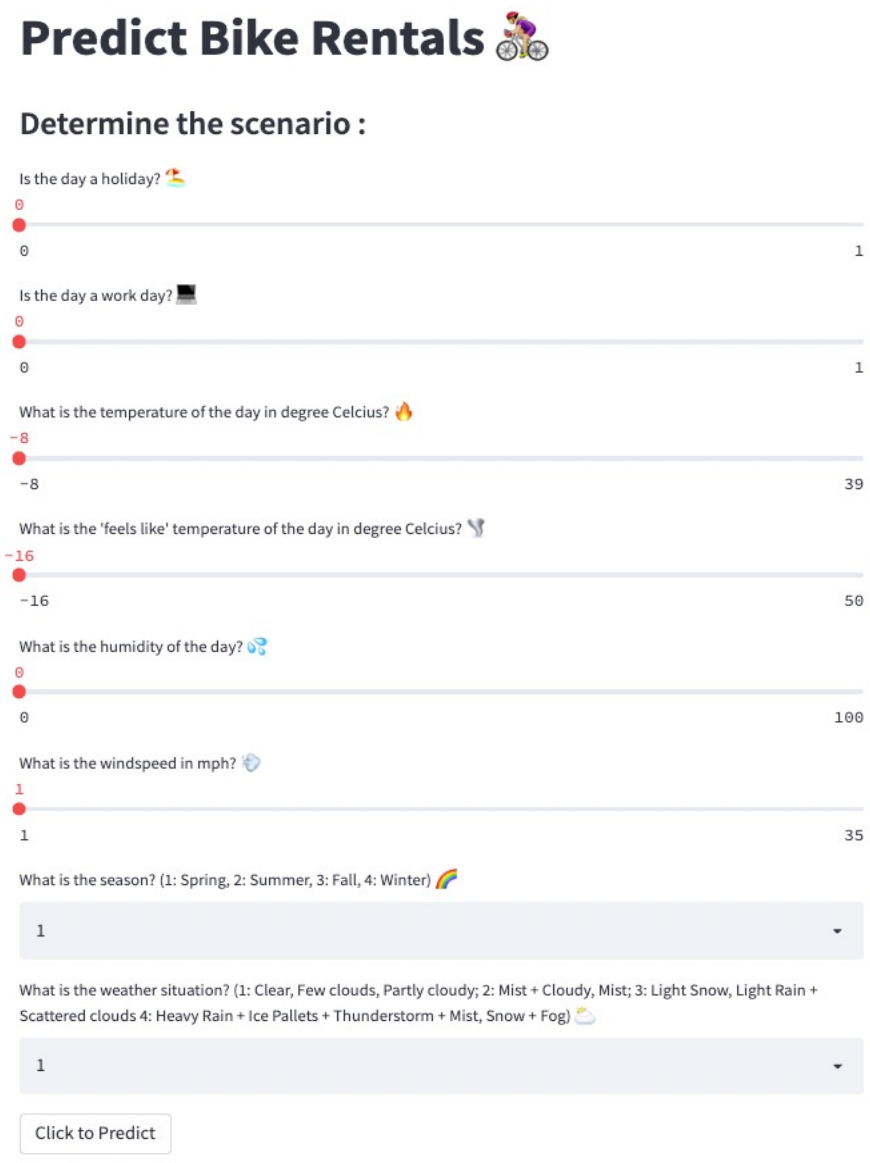
# Machine Learning Prediction

- I've also utilized Machine Learning techniques to create an application on Streamlit that can be used to predict the number of bike rentals based on the environmental and seasonal setting as shown in figure 4. (Linked here)
- Through my investigation, I determined that the Random Forest Regressor Model with a max depth of 10 yielded the best R^2 value of 0.59 and root mean squared error of ±1373.29 which was then used to develop the Streamlit app.

Business Suggestion:

- Using this app, Capital Bikeshare will be able to predict the number of bike rental based on weather forecast data and plan the number of bikes a certain region will need. For example, if region A is forecasted to be raining heavily and have lower temperatures compared to region B which is forecasted to have sunny warm weather, we can use this app to predict the demand in both regions and then allocate resources accordingly to meet the higher predicted demand in region B.
- This model can also be useful when Capital Bikeshare needs to perform repairs/updates to the bikes and may choose the days with lower demand to perform maintenance on bikes without compromising the ability to meet the demand for bicycles.

Figure 4. Preview of my Streamlit App



# Conclusion

- Based on my analysis, I determined that temperature plays the most significant factor in determining the number of bike rentals for a particular day. I believe that there is potential to capture more cyclist during colder months and Capital Bikeshare can use this insight to adapt their social media marketing strategy during colder months like winter and spring while exploring potential partnerships with outdoor winter clothing companies to encourage users to rent bicycles during colder months and increase demand.
- Using the Random Forest Regressor Model, I developed an application that the business can use to predict the number of bike rentals for a particular day thus allowing Capital Bikeshare to deploy their resources productively. (see link)
- For further information on the analysis that was performed please feel free to inspect my Jupyter Notebook file name ’Predicting Bike Rental (Capstone)’ linked here.

# Appendix 1 – Data Dictionary

**Data Dictionary**

| Column Position | Atrribute Name | Definition | Data Type | Example | % Null Ratios |
|---|---|---|---|---|---|
| 1 | instant | Record Index | Quantitative | 190, 7, 17180 | 0 |
| 2 | dteday | Date (Format: YYYY-MM-DD) | Quantitative | 2012-12-23, 2012-01-01, 2012-06-24 | 0 |
| 3 | season | Season (1: springer, 2: summer, 3: fall, 4: winter) | Quantitative | 1, 2, 4 | 0 |
| 4 | yr | Year (0: 2011, 1:2012) | Quantitative | 0, 1 | 0 |
| 5 | mnth | Month (1 to 12) | Quantitative | 1, 6, 12 | 0 |
| 6 | hr | Hour (0 to 23) - Not in day.csv dataset | Quantitative | 4, 6, 14 | 0 |
| 7 | holiday | Weather day is holiday or not | Quantitative | 0, 1 | 0 |
| 8 | weekday | Day of the week | Quantitative | 0, 6, 3 | 0 |
| 9 | workingday | Working Day: If day is neither weekend nor holiday is 1, otherwise is 0 | Quantitative | 0, 1 | 0 |
| 10 | weathersit | Weather Situation (1: Clear, Few clouds, Partly cloudy, Partly cloudy; 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist; 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds, 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog) | Quantitative | 1, 2, 3 | 0 |
| 11 | temp | Normalized temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-8, t_max=+39 (only in hourly scale) | Quantitative | 0.08, 0.22, 0.34 | 0 |
| 12 | atemp | Normalized feeling temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-16, t_max=+50 (only in hourly scale) | Quantitative | 0.0909, 0.2727, 0.303 | 0 |
| 13 | hum | Normalized humidity. The values are divided to 100 (max) | Quantitative | 0.53, 0.8, 0.31 | 0 |
| 14 | windspeed | Normalized wind speed. The values are divided to 67 (max) | Quantitative | 0.194, 0, 0.2985 | 0 |
| 15 | casual | Count of casual users | Quantitative | 0, 2, 57 | 0 |
| 16 | registered | Count of registered users | Quantitative | 1, 0, 118 | 0 |
| 17 | cnt | Count of total rental bikes including both casual and registered | Quantitative | 1, 2, 175 | 0 |

# Appendix 2- Correlation Heatmap