A New Bike-Sharing Venture

Our Research Question

How can data from bike-sharing operations from two regions inform operating decisions when opening a new branch?

Our Data

- Two primary data sets:
  - Bikeshare data from Porto, Portugal (1/2011 -1/2013)
  - Bikeshare data from Seoul, South Korea (12/2017 - 12/2019)
    - Includes: date , time, temperature, humidity, wind speed, precipitation, number of customers
- Ancillary data: climate data from Porto, Seoul, New York City and Los Angeles

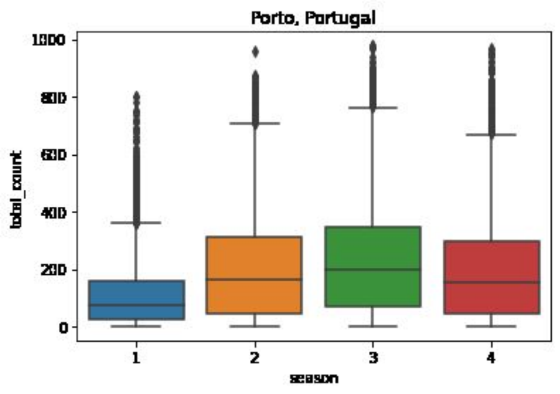
Data Cleaning and Standardizing:

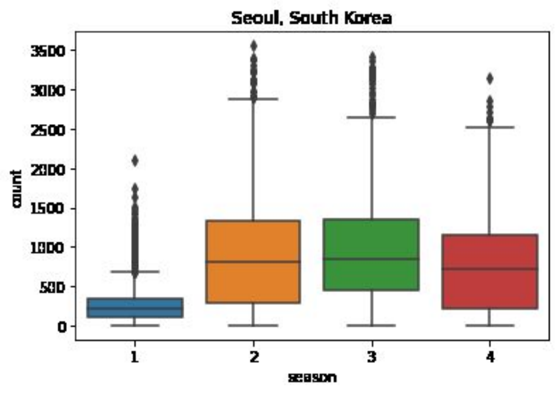
- Standardized Porto and Seoul data to the same date format and then separated dates into 3 columns
- Added a column that corresponds to the day of the week (0 = Sunday)
- Converted Seoul holidays and season data into numeric categories
- Note: Porto weather data is normalized while Seoul weather data is not normalized
    - We can not directly compare values but we can compare weather trends

A Seasonal Look

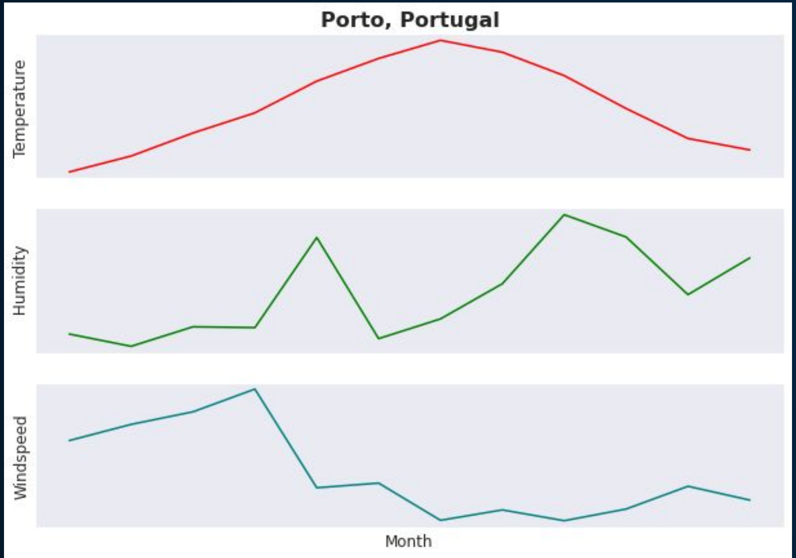
- Both markets demonstrate similar trends throughout the seasons. Ridership is lowest in the winter, with higher, and more widely-spread distributions in the spring, summer and fall.

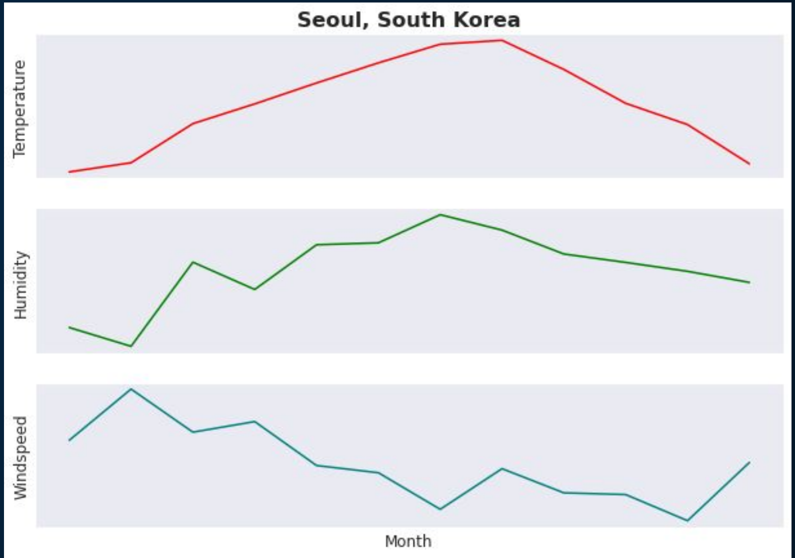
- We used this plot type to highlight the differences in the distribution of riders for each season.



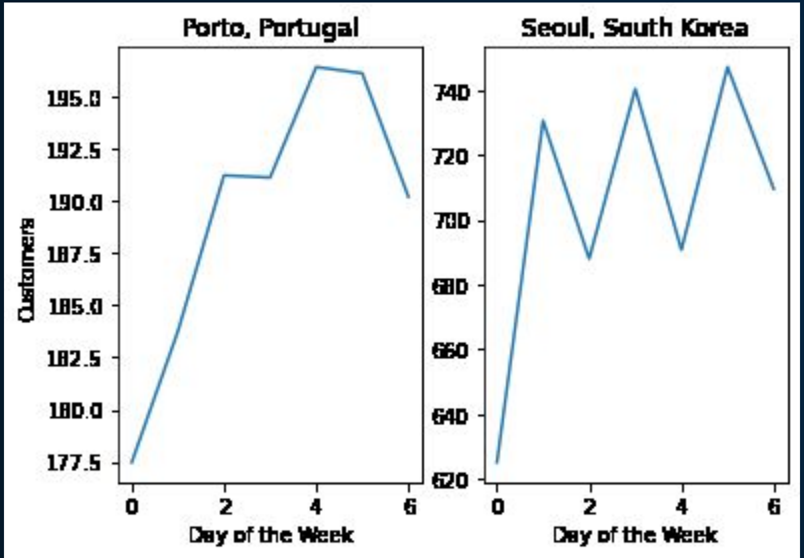


Weather Trends



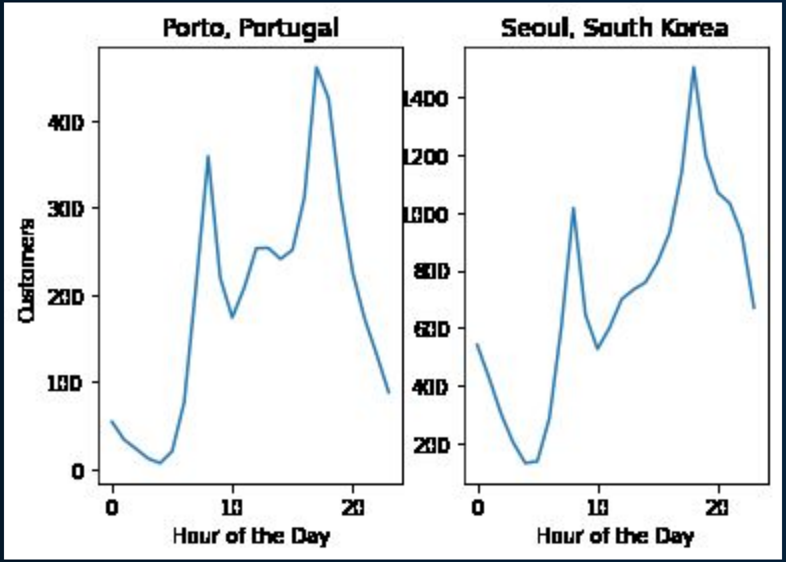


A Weekly Look



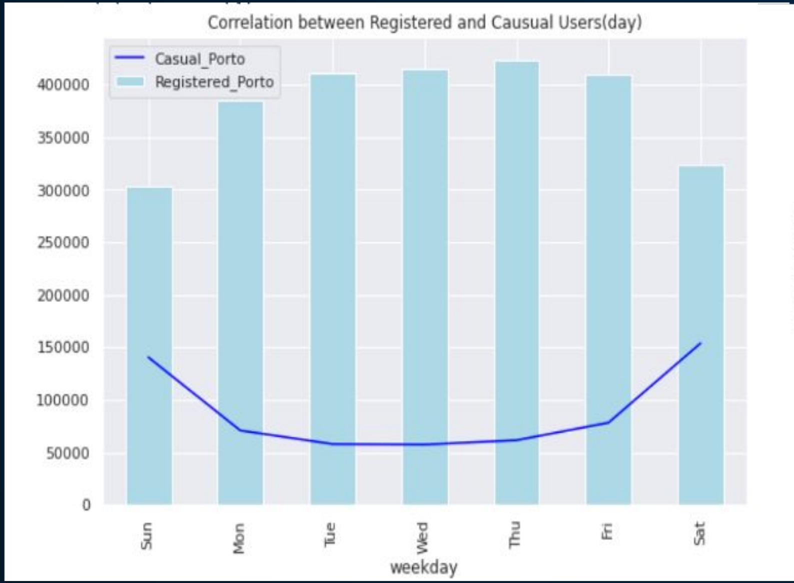
Both markets demonstrate a similar trend of customers throughout the week. Sunday is the least busy day with ridership increasing throughout the week.

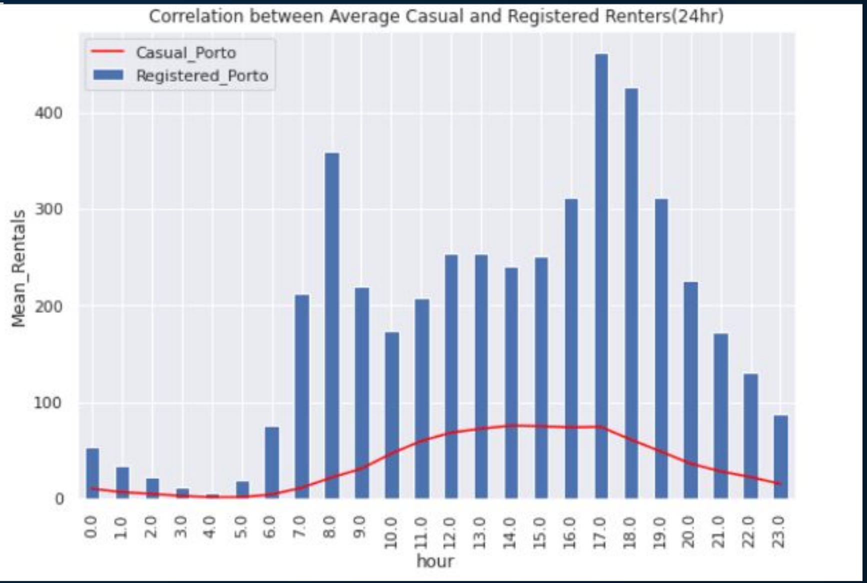
A Daily Look



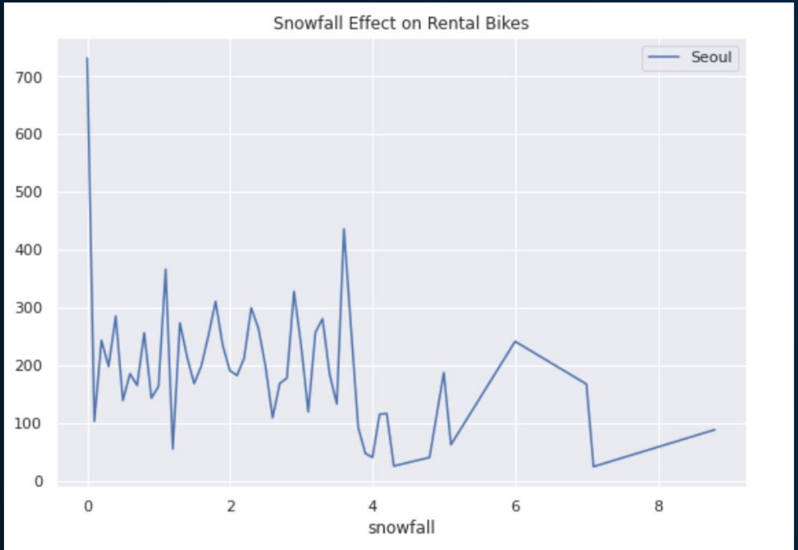
- Both markets demonstrate a nearly identical, bimodal trend of customers throughout the day (in the morning and in the evening).
- Comparable trends across markets validate using this data to inform operations decisions.

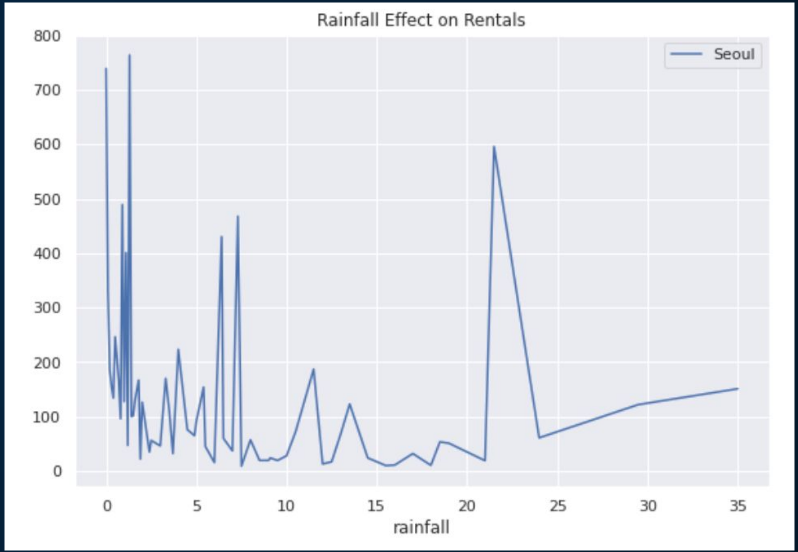
Registered & Non-Registered Users



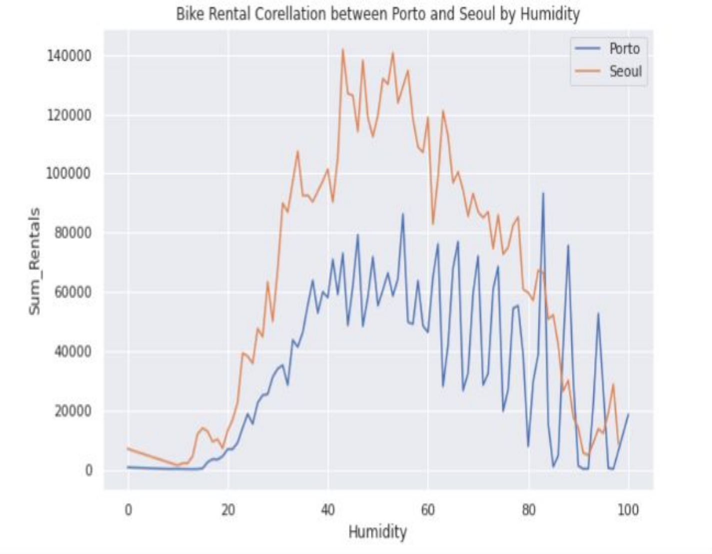


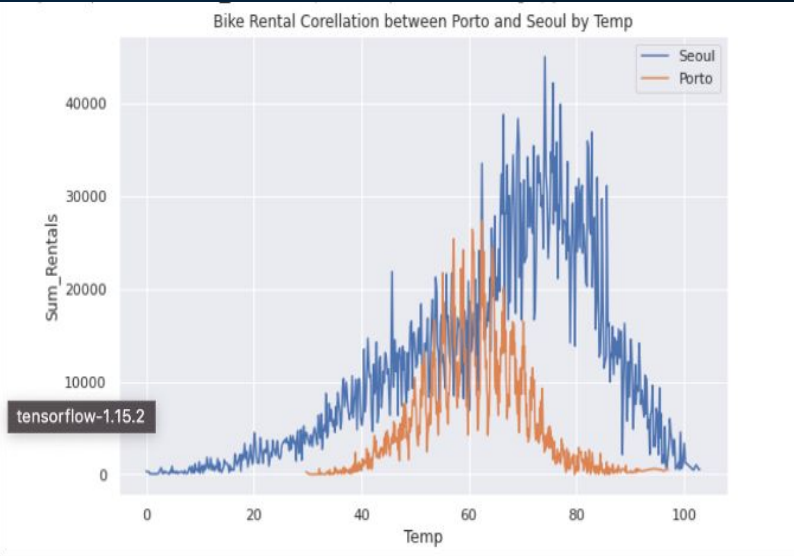
Effect of Precipitation





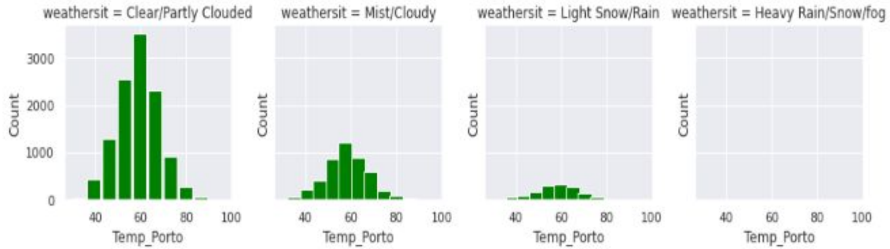
Effects of Temperature and Humidity



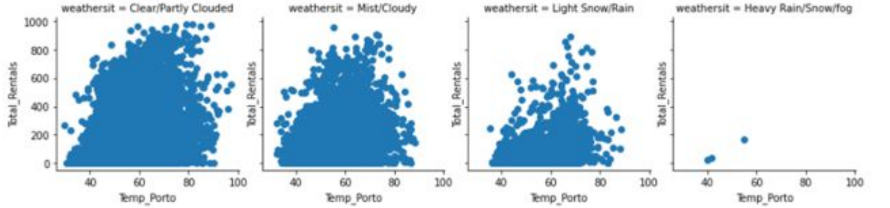


Variables Effects on Bike Rentals Comparison

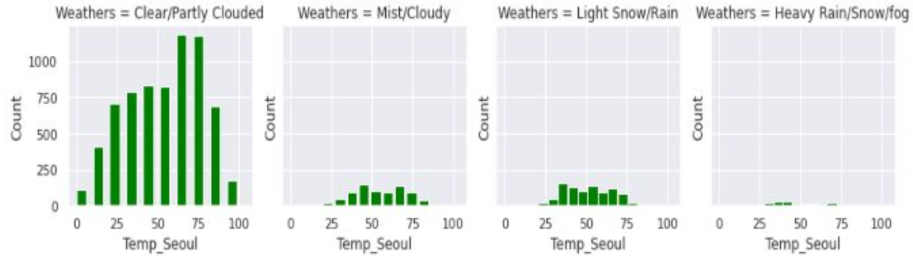
```
weathersit
Clear/Partly Clouded    11413
Heavy Rain/Snow/fog         3
Light Snow/Rain          1419
Mist/Cloudy              4544
Name: total_count, dtype: int64
<seaborn.axisgrid.FacetGrid at 0x7fd3ee55ef10>
```



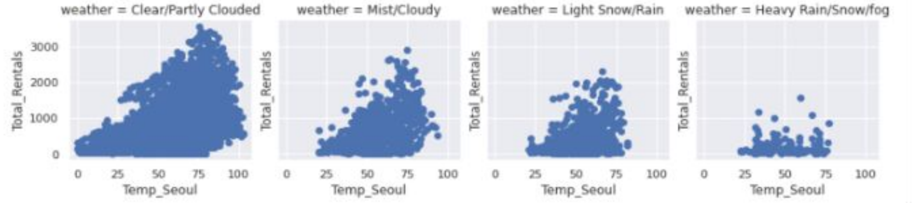

```
weathersit
Clear/Partly Clouded    2338173
Heavy Rain/Snow/fog         223
Light Snow/Rain          158331
Mist/Cloudy              795952
Name: total_count, dtype: int64
<seaborn.axisgrid.FacetGrid at 0x7f40c9343c10>
```




```
Weathers
Clear/Partly Clouded    6914
Heavy Rain/Snow/fog      156
Light Snow/Rain          904
Mist/Cloudy              783
Name: count, dtype: int64
<seaborn.axisgrid.FacetGrid at 0x7fd3e908a550>
```




```
weather
Clear/Partly Clouded    5317385
Heavy Rain/Snow/fog       29350
Light Snow/Rain          347616
Mist/Cloudy              474688
Name: count, dtype: int64
<seaborn.axisgrid.FacetGrid at 0x7fd3f0561510>
```



Regression Analysis: Porto

```
smf.ols('total_count ~ windspeed + temp + atemp + hum + C(season) + C(weathersit) + C(weekday) + C(month) + C(hour)', data = data_porto).fit()
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            total_count   R-squared:                       0.632
Model:                            OLS   Adj. R-squared:                  0.631
Method:                 Least Squares   F-statistic:                     594.5
Date:                Fri, 14 May 2021   Prob (F-statistic):               0.00
Time:                        21:48:48   Log-Likelihood:            -1.0636e+05
No. Observations:               17379   AIC:                         2.128e+05
Df Residuals:                   17328   BIC:                         2.132e+05
Df Model:                          50
Covariance Type:            nonrobust
```

Regression Analysis: Seoul

```
smf.ols('count ~ rainfall + solar_rad + dew_point + visibility + windspeed + hum + temp + C(weekday) + C(hour) + C(season)', data = data_seoul).fit()
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  count   R-squared:                       0.635
Model:                            OLS   Adj. R-squared:                  0.633
Method:                 Least Squares   F-statistic:                     303.1
Date:                Fri, 14 May 2021   Prob (F-statistic):               0.00
Time:                        21:57:20   Log-Likelihood:                -64685.
No. Observations:                8760   AIC:                         1.295e+05
Df Residuals:                    8709   BIC:                         1.298e+05
Df Model:                          50
Covariance Type:            nonrobust
------------------------------------------------------------------------------
```

Regression: Porto Weather Variables

```
model_temp_porto = smf.ols('total_count~ temp + atemp + hum + windspeed+ C(season)+ C(weathersit)', data =data_porto).fit()
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            total_count   R-squared:                       0.281
Model:                            OLS   Adj. R-squared:                  0.281
Method:                 Least Squares   F-statistic:                     971.6
Date:                Fri, 14 May 2021   Prob (F-statistic):               0.00
Time:                        22:39:02   Log-Likelihood:            -1.1217e+05
No. Observations:               17379   AIC:                         2.244e+05
Df Residuals:                   17371   BIC:                         2.244e+05
Df Model:                           7
Covariance Type:            nonrobust
```

Regression: Seoul Weather Variables

```
smf.ols('count ~ rainfall + solar_rad + dew_point + visibility + windspeed + hum + temp + C(season)', data = data_seoul).fit()
```

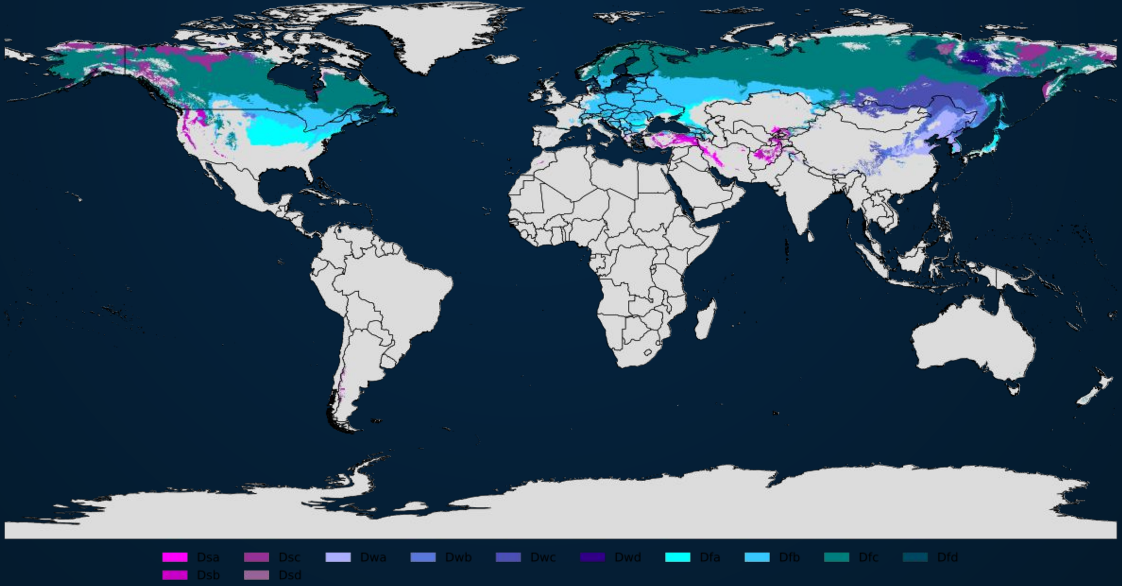
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  count   R-squared:                       0.422
Model:                            OLS   Adj. R-squared:                  0.421
Method:                 Least Squares   F-statistic:                     638.9
Date:                Fri, 14 May 2021   Prob (F-statistic):               0.00
Time:                        22:42:28   Log-Likelihood:                -66699.
No. Observations:                8760   AIC:                         1.334e+05
Df Residuals:                    8749   BIC:                         1.335e+05
Df Model:                          10
Covariance Type:            nonrobust
```

What do the regression models tell us?

- There is a significant relationship between weather variables and ridership.
- The model performance suffers when we just include the weather variables.
  - Temporal data is important to consider in future models (Hour and weekday).
- We should look for other variables that we could include to improve model performance.

Climate Comparison: Seoul

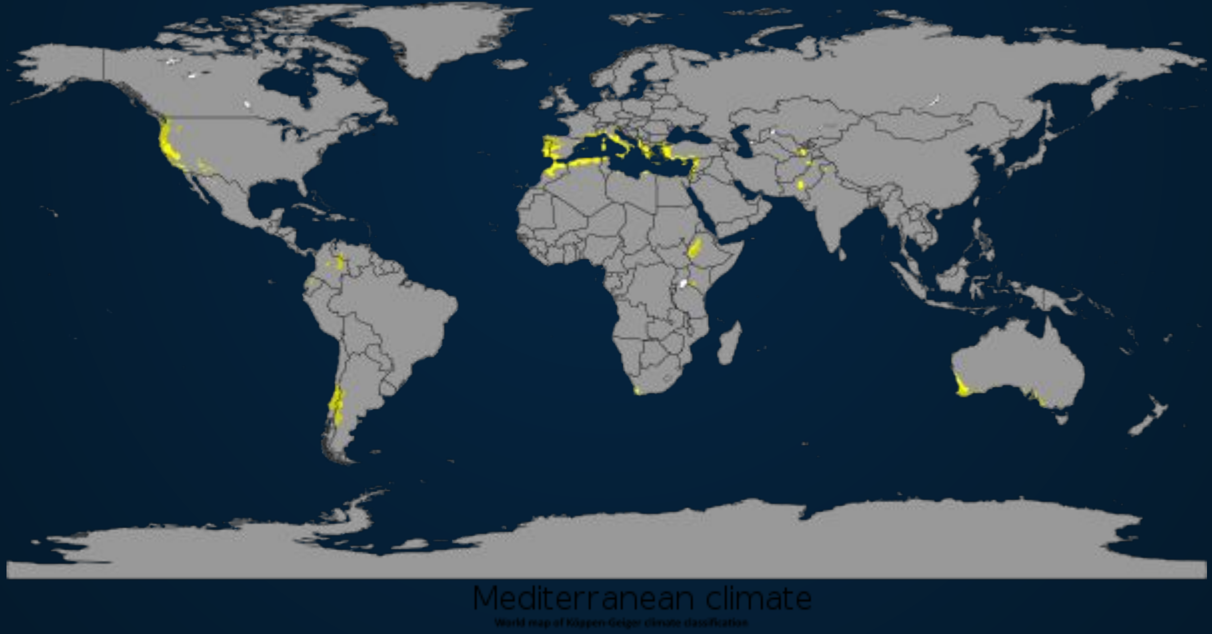
- The climate is a humid, continental climate characterized with dry winters.
- The most pleasant seasons are spring and autumn. Most of Seoul's precipitation falls in the summer monsoon period between June and September.



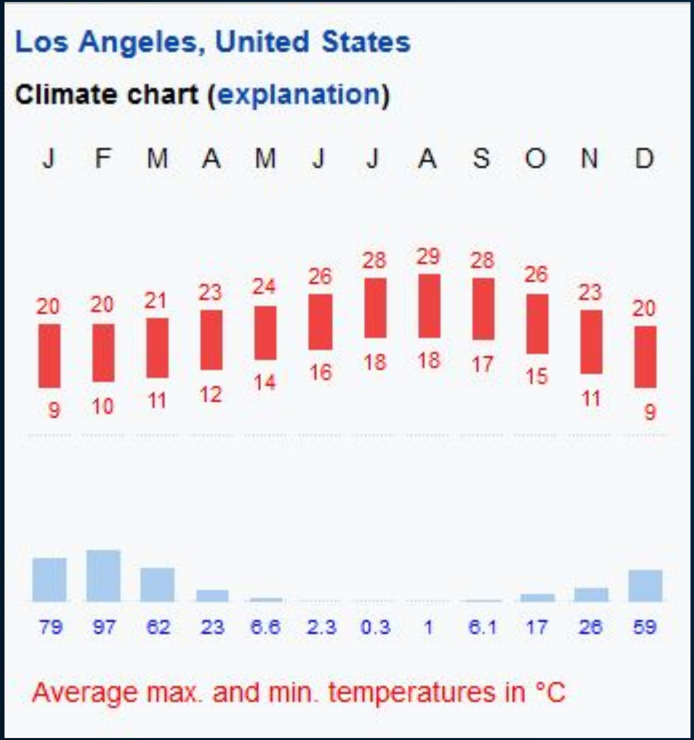
Source: Beck et al.: Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data 5:180214. doi:10.1038/sdata.2018.214 (2018)

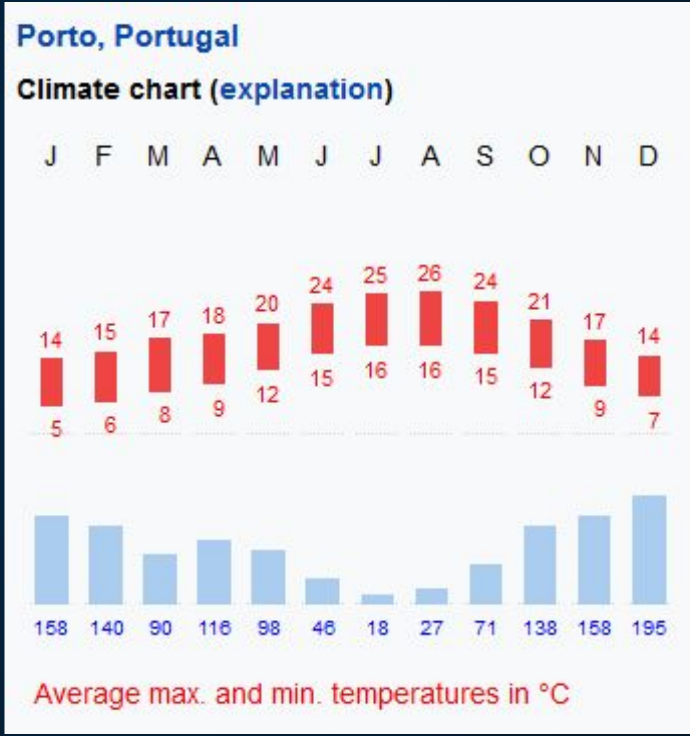
Climate Comparison: Porto

- The climate is mediterranean climate characterized by dry summers and mild, wet winters.
- Usually occur on the western sides of continents and is sparsely found.



Climate Comparison





Conclusions: Customer Behavior

- Customers tend to use the bike-share programs most in the morning and evening hours.
- Customers use the programs more frequently and with higher uncertainty in the spring, summer and fall.
- The more extreme the weather the less predictable customer behaviour is.
- Day, time and weather hold a significant explanatory power in determining ridership.
- Removing day and time variables decrease the explanatory power of the model.
- Registered users are more consistent at the weekly and daily level.

Conclusion: Business Decision Making

The analysis takes a deep dive into the weather and climate factors. This combined with the known climate facts of each different regions, the company can lead decision making in not only the launch of operations but also other areas like marketing.

Thank you!