TFL Bike Hire Data Analysis

Raye HS Yoo

21 July 2023

Outline

Summary

Business Problem

Data & Methods

Result

Conclusion

Summary

This data analysis project provides the most impactful features to contribute to the number of bike hires in London through multiple regression analysis. The data shows that the hours (trip starting), weather and temperature affect the bike hire demand significantly. Transport for London (“TFL”) can use this data to plan the bike supply to improve its business profit model.

Business Problem

TFL is currently facing issues to predict bike-hire demand post-pandemic in London. Londoners are going back to the offices and tourists are coming back to London significantly. To solve this problem, I will find out the most impactful features of hiring bikes in London through multiple regression analysis to predict bike-hire demands for the future years.

Data & Methods

- **Data source**: London bike sharing dataset, kaggle (LINK)
  - The number of bikes hired every hour in London (00:00 4th Jan 2015 – 23:00 3rd Jan 2017)
  - Historical data for bike sharing in London 'Powered by TfL Open Data'
  - A merged dataset from https://cycling.data.tfl.gov.uk/, freemeteo.com, and https://www.gov.uk/bank-holidays

- **Understanding data**
  - timestamp: Timestamp field for grouping the data by hour
  - cnt : The **count** of a new bike hires
  - t1 : Official temperature in Celsius
  - t2 : "Feels like" temperature in Celsius
  - hum : Humidity in percentage
  - wind_speed : Wind speed in km/h
  - weather_code : Category of the weather
  - is_holiday: Boolean field - 1 holiday / 0 non-holiday
  - is_weekend : Boolean field - 1 weekend / 0 weekdays
  - season - Category field meteorological seasons: 0-spring; 1-summer; 2-fall; 3-winter

weather_code category description:
1 = Clear ; mostly clear but have some values with haze/fog/patches of fog/ fog in vicinity
2 = scattered clouds / few clouds
3 = Broken clouds
4 = Cloudy
7 = Rain/ light Rain shower/ Light rain
10 = rain with thunderstorm
26 = snowfall
94 = Freezing Fog

Data & Methods

- **Data Analysis Approach :** Multiple Regression Analysis

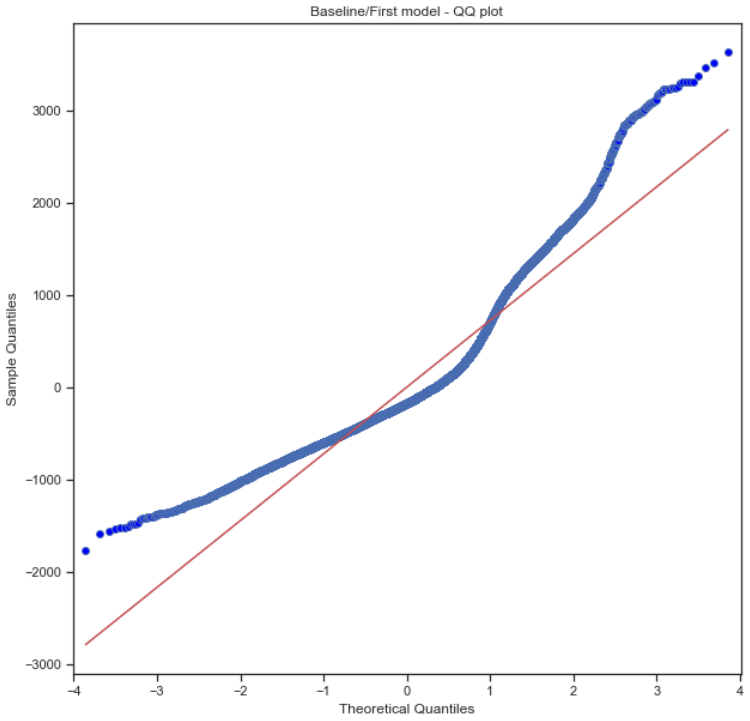
R-squared: 0.312
Skewness: 1.277
Kurtosis: 4.859

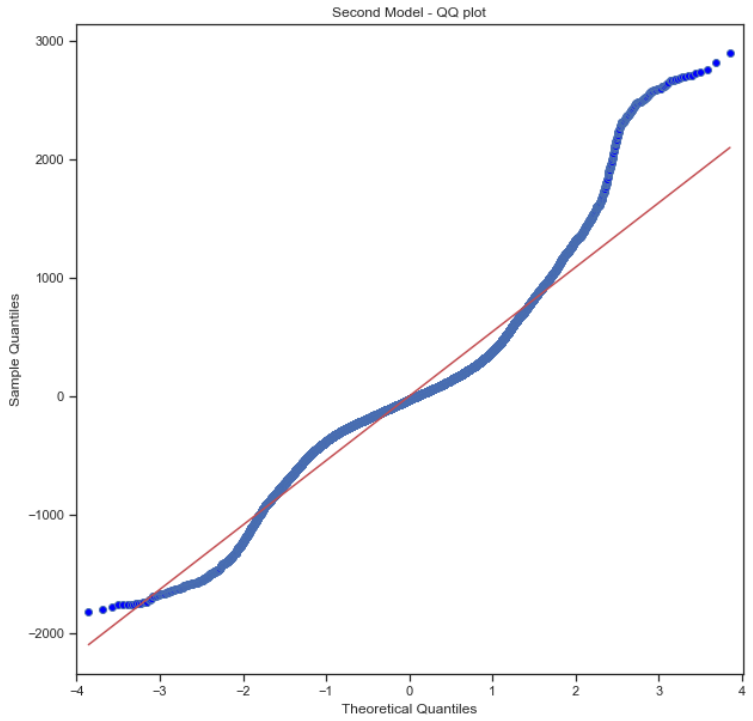
Checking Multicollinearity
& Removing variables
Creating dummy variances

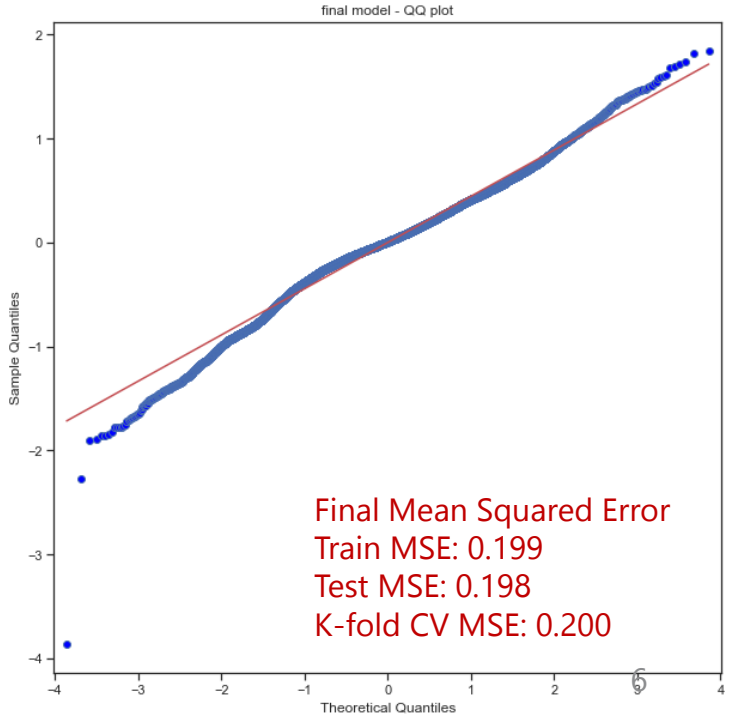
R-squared: 0.611
Skewness: 0.657
Kurtosis: 6.514

Log transformation &
Scaling
Removing variables with
high p-value (> 0.05)

R-squared: 0.802
Skewness: -0.336
Kurtosis: 4.368

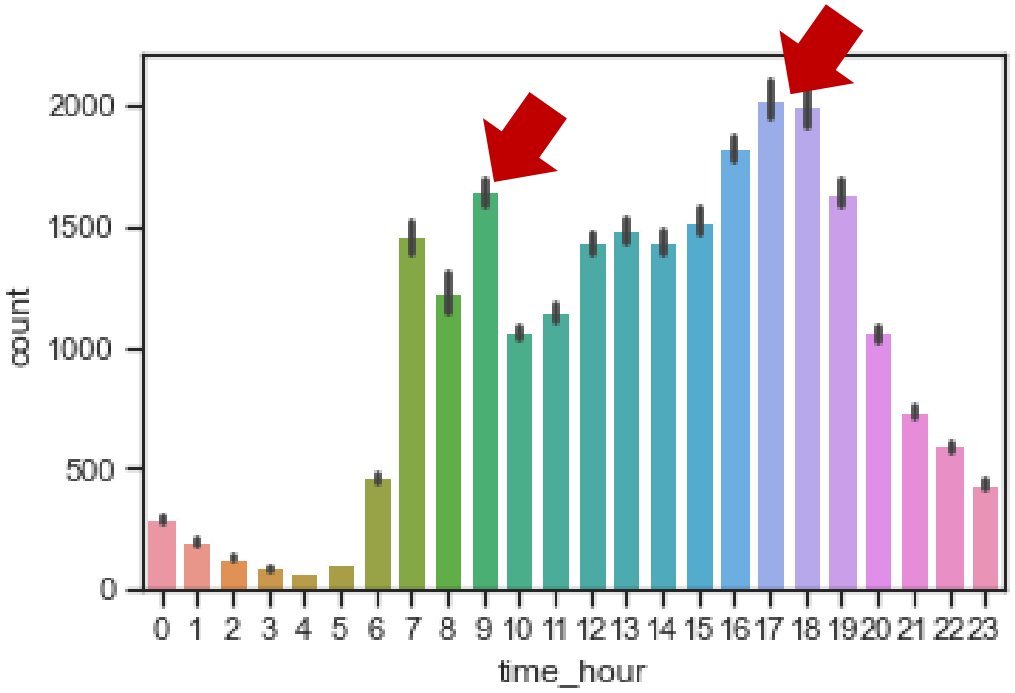






Result

The most impactful feature is **the trip starting time** (hour)
The commuting hours are directly related to the number of bike hire



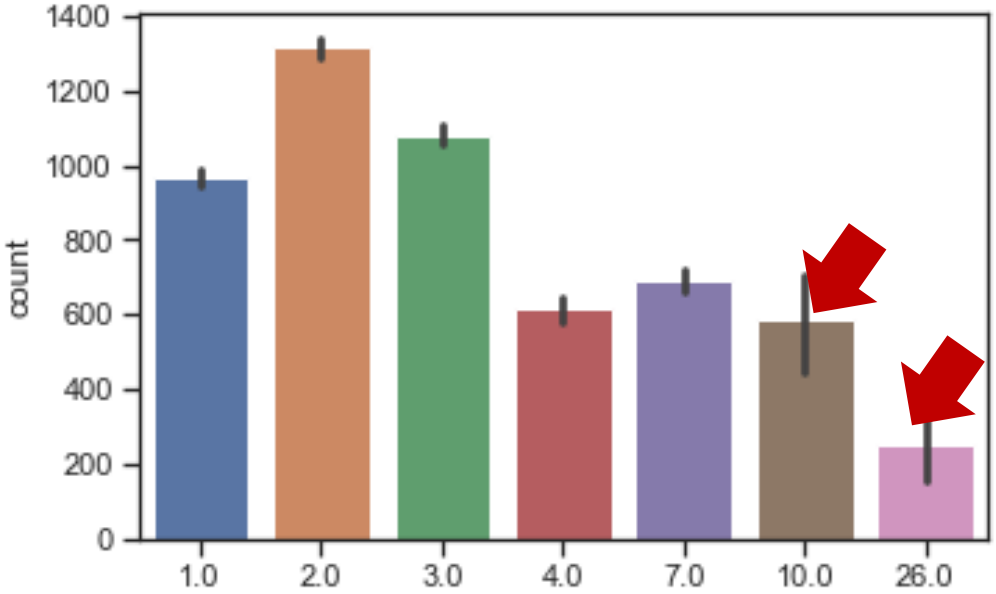
Top Coef values:

time_17	1.459423
time_18	1.458170
time_16	1.434914
time_9	1.397967
time_19	1.367520

Result

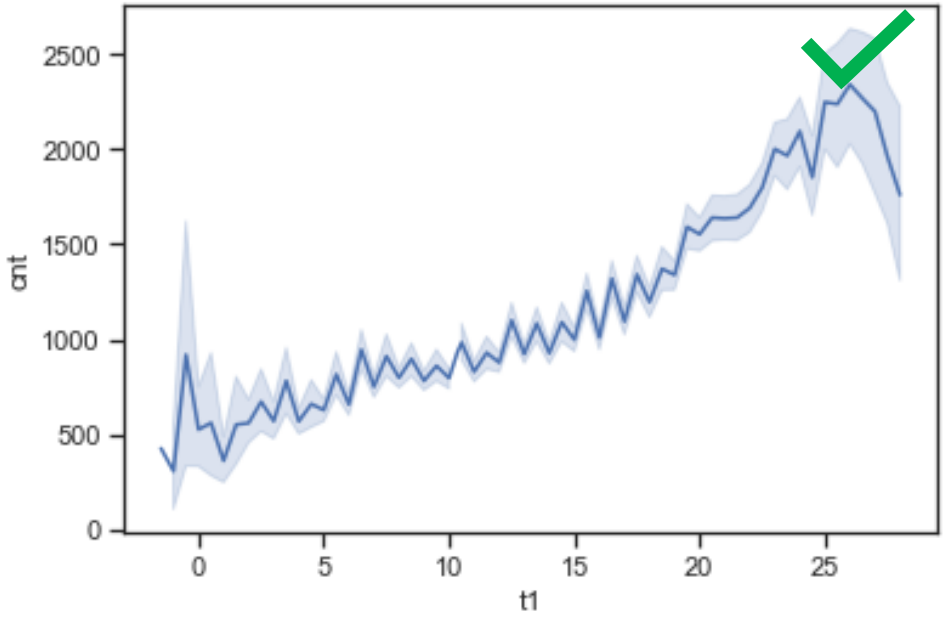
Bad Weather

affects negatively :
10: Rain with thunderstorm
26: Snowfall



Higher temperature

impacts positively

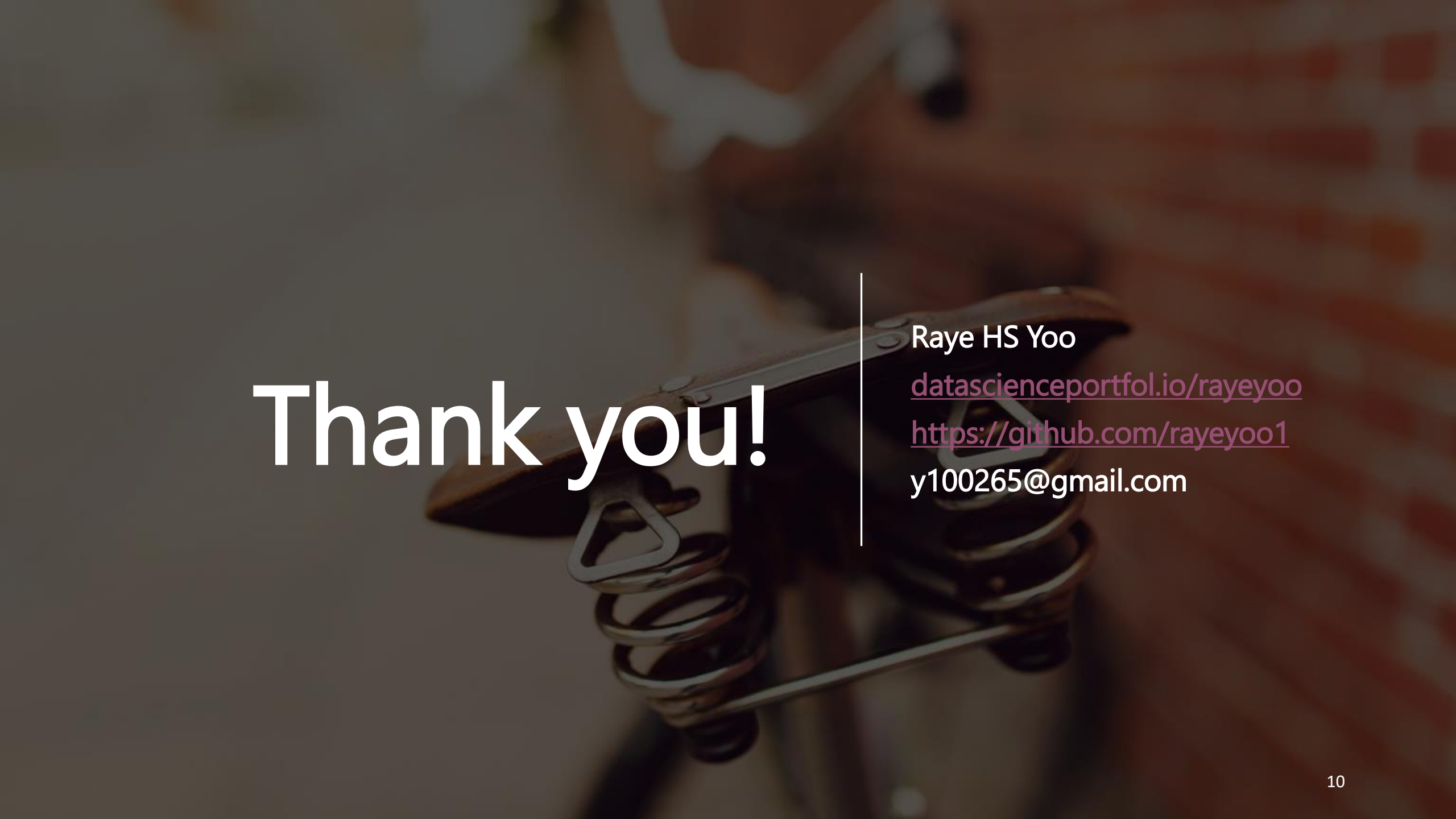


Conclusion

Supply the hire bikes mainly around the business district areas in London based on the commuting hours. also, recommend considering the weather and temperatures to adjust the number of bikes based on the analysis result

Further action

- Obtain the location data analysis: To be able to predict not only the demand but also effective bike relocations to increase bike hire. This will also help to supply the right number of bikes to be hired from the right places
- Adopt new types of vehicles and demand: E-bikes & scooters were released in late 2020 in London, we should look at the recent data to see the trends to predict more accurate customer demands



Thank you!

Raye HS Yoo
datascienceportfol.io/rayeyoo
https://github.com/rayeyoo1
y100265@gmail.com