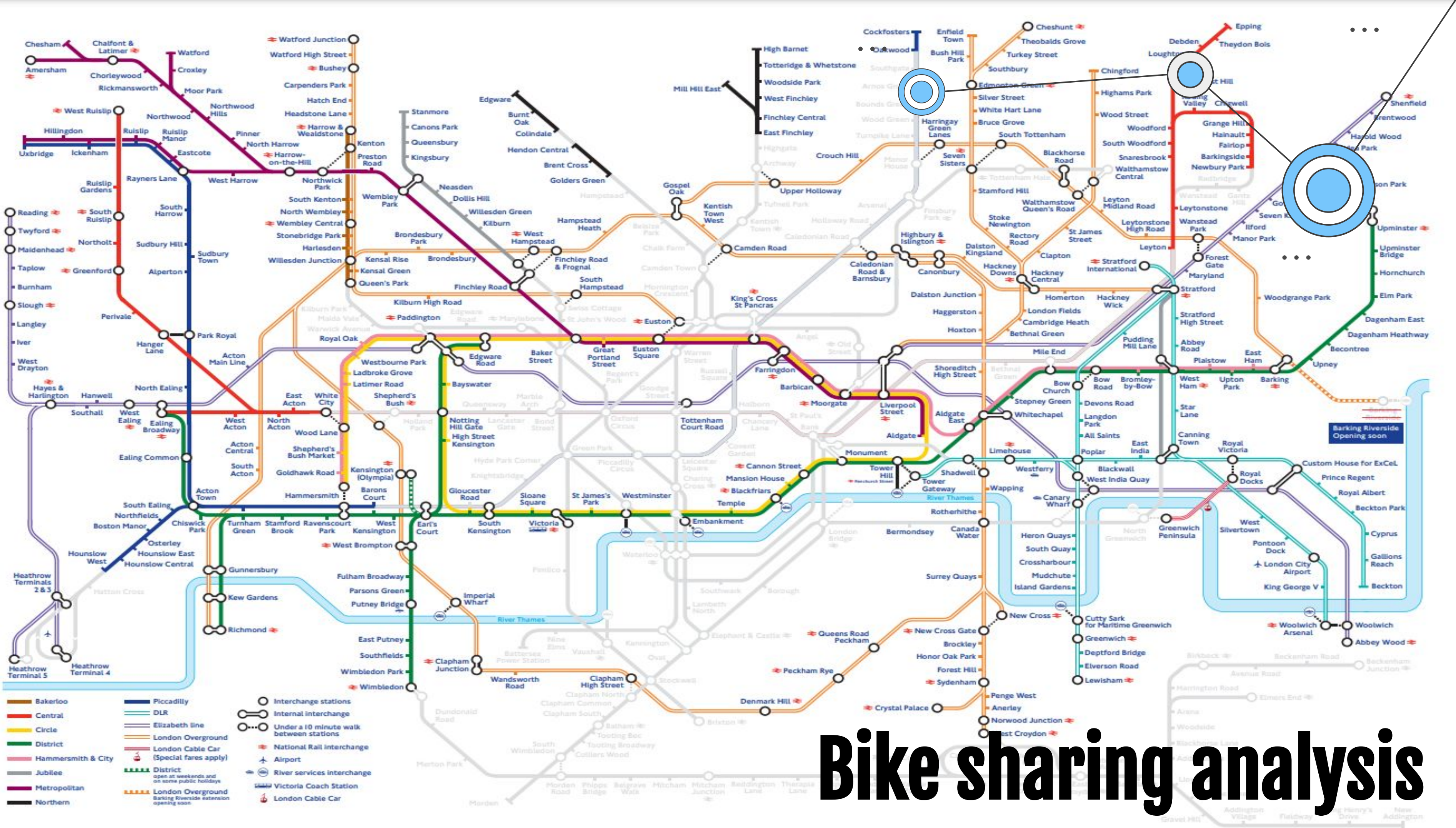
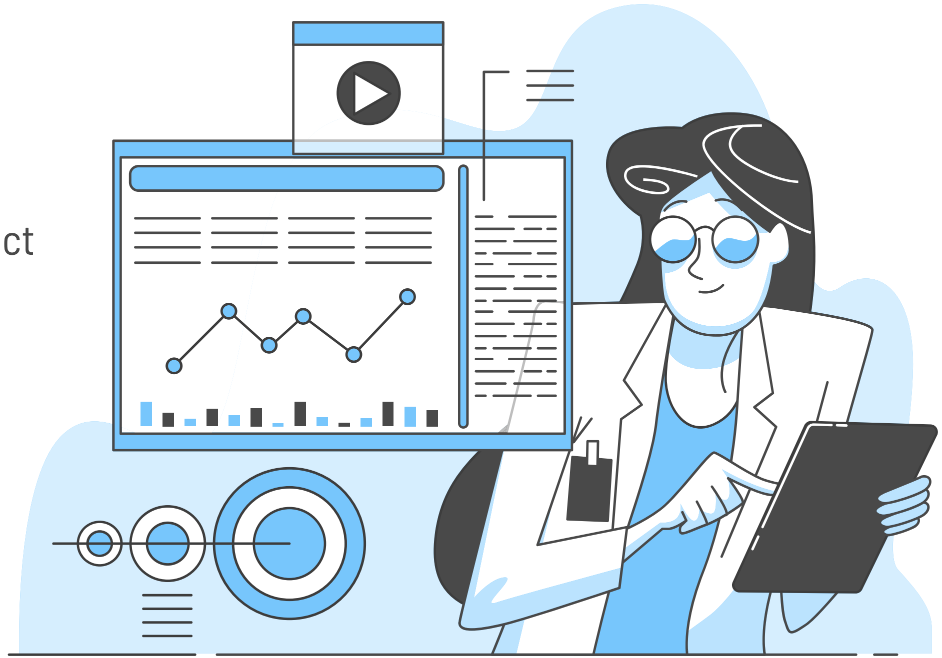
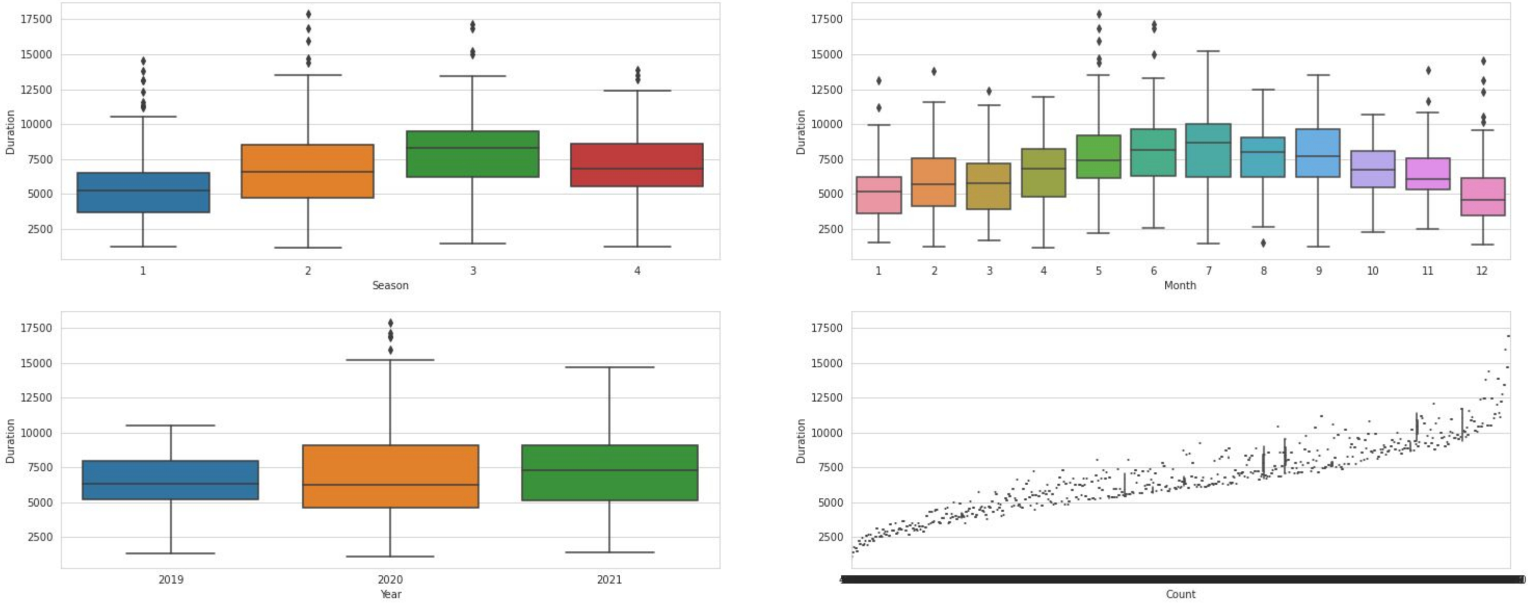


Bike sharing analysis

Table of Contents



Exploratory analysis



We considered short-term bike renting services that last at most 44 minutes [90% of the overall dataset].

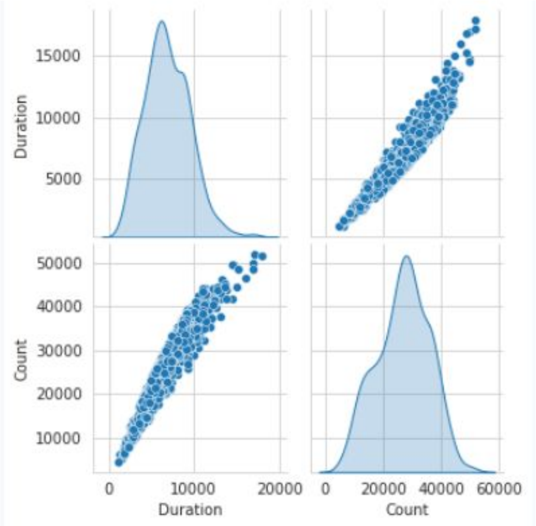
Then, we gathered 5 important variables ‘Month’, ‘Year’, ‘Season’, ‘Count’, and ‘Duration’.

This boxplot shows these features could be helpful in modeling the bike renting problem.

Specifically, we see a linear relationship between ‘Count’ and ‘Duration’, so we will leverage on this.

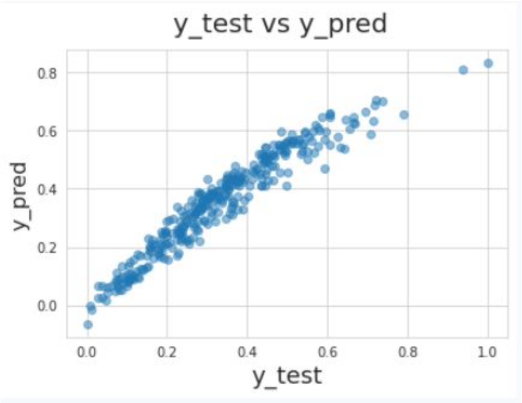
Modelling

02




```
                            OLS Regression Results
==============================================================================
Dep. Variable:               Duration   R-squared:                       0.921
Model:                            OLS   Adj. R-squared:                  0.920
Method:                 Least Squares   F-statistic:                     1086.
Date:                Fri, 01 Jul 2022   Prob (F-statistic):               0.00
Time:                        20:09:55   Log-Likelihood:                 1264.1
No. Observations:                 759   AIC:                            -2510.
Df Residuals:                     750   BIC:                            -2469.
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0627      0.005    -11.766      0.000      -0.073      -0.052
Count          0.8156      0.010     85.593      0.000       0.797       0.834
Season_2       0.0158      0.005      3.272      0.001       0.006       0.025
Season_3       0.0139      0.005      2.768      0.006       0.004       0.024
Year_2020      0.0558      0.004     13.757      0.000       0.048       0.064
Year_2021      0.0278      0.004      6.758      0.000       0.020       0.036
Month_5        0.0126      0.007      1.752      0.080      -0.002       0.027
Month_6        0.0152      0.007      2.135      0.033       0.001       0.029
Month_10      -0.0136      0.006     -2.249      0.025      -0.025      -0.002
==============================================================================
Omnibus:                       87.375   Durbin-Watson:                   1.955
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              136.018
Skew:                           0.785   Prob(JB):                     2.91e-30
Kurtosis:                       4.355   Cond. No.                         8.13
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```



'Count' and 'Duration' seem to be linearly dependent [Left].

Our linear model predicting 'Duration' performs well on training data [Middle].

We can see 1 increase in 'Count' (1 more bike) will lead to a 0.8 minute increase in 'Duration' (bike being used).

R-squared and Adj. R-squared are 0.9 in the test data [code], and also predictions match the true values [Right].

03

Recommendations

The data suggests that an increase in the number of bikes 'Count' [e.g. 100 new bikes] leads to an increase in the bikes being used 'Duration' [80 minutes].

Main suggestion

We see it as an opportunity to be explored, to identify the boundary of when adding more bikes is or isn't beneficial.
Perform a networks-based analysis to identify which stations could present the biggest upside for deploying new bikes.

Secondary suggestion

Investigate the relationship between money acquired through small-term bike renting and long term [10 to 1 M customers].
Is there is a market for longer term bike sharing, and if so what the data suggests [greater than 45 minutes].