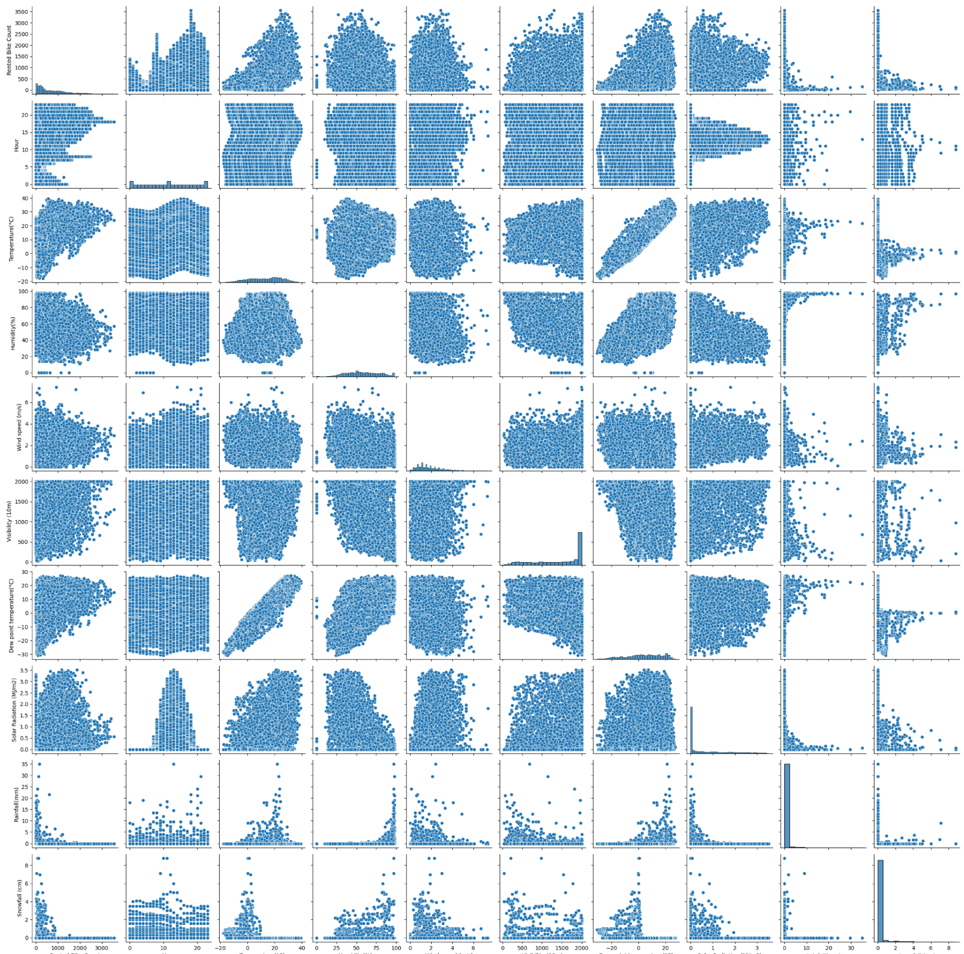
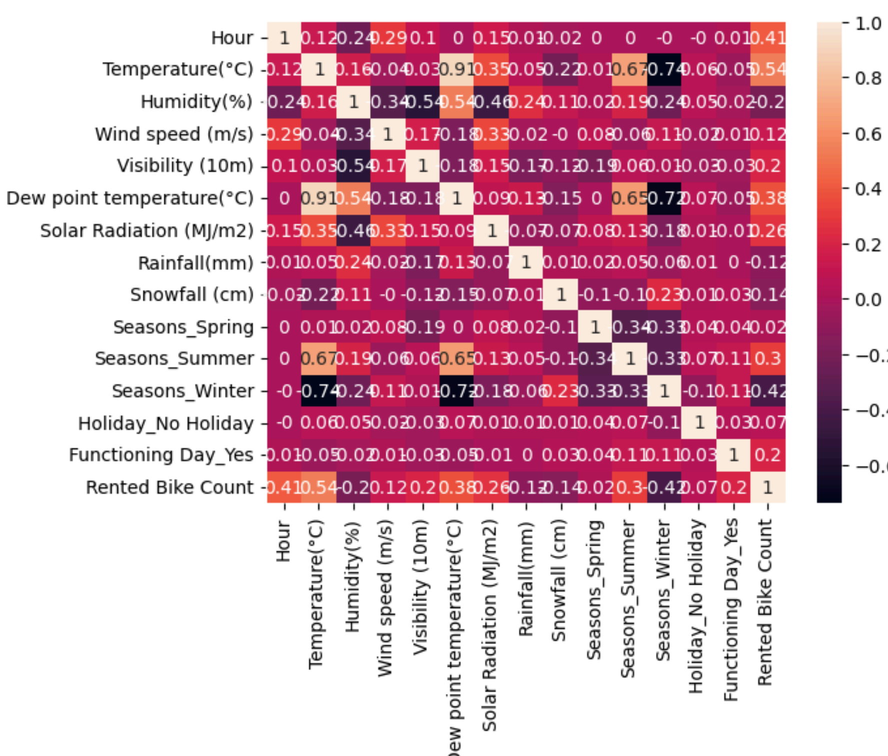
Data Exploration:



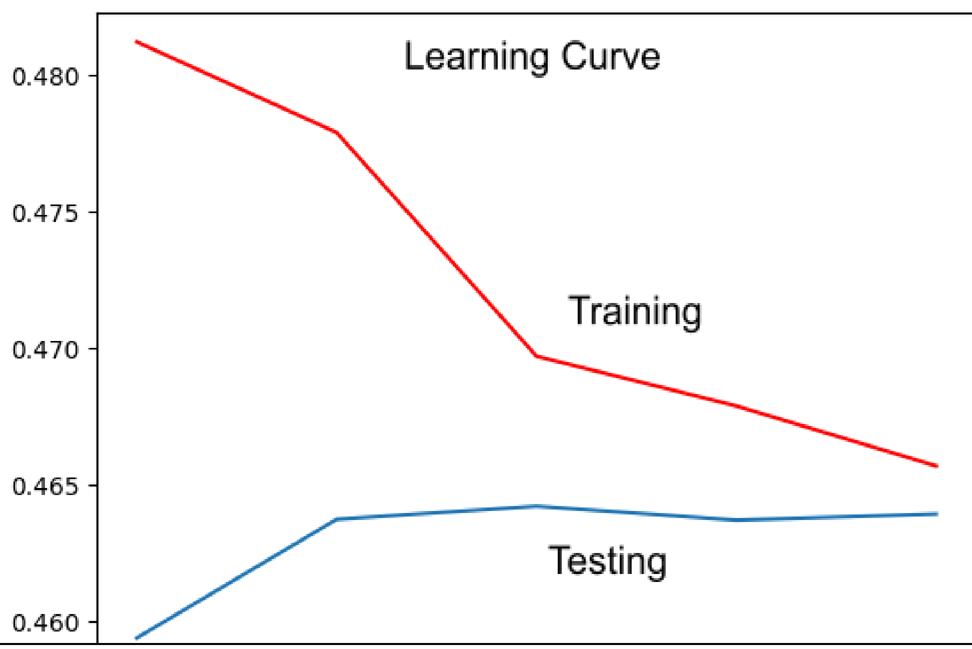


From the first plot, we can see that there are not many linear relationships between the variables and the target variable. Visually speaking it seems that Temperature has the most linear relationship with Rented Bike Counts. This is confirmed when looking at the correlation plot. The correlation plot includes the correlation between each variable with each other. We are only interested in how the variables interact with the target variable. This shows that temperature has the highest correlation with the target variable, with a correlation of around 0.5. Other variables like dew point temperature show moderate correlations, while humidity has a negative correlation, indicating an inverse relationship with bike rentals.

SGD Model Tuning:

Features	**Hyper-parameters**	**Training Error**	**Testing Error**
Temperature(°C), Hour, Dew point temperature(°C), Seasons_Winter, Seasons_Summer, Solar Radiation (MJ/m2)	max_iter = 1000 tol = 1e-3 alpha = 0.0001	MSE: 222293.47 MAE: 351.66 EV: 0.47 R^2: 0.47	MSE: 222873.71 MAE: 352.22 EV: 0.46 R^2: 0.46

Results:
For the features in this model, we selected features that had a correlation greater than 0.2.

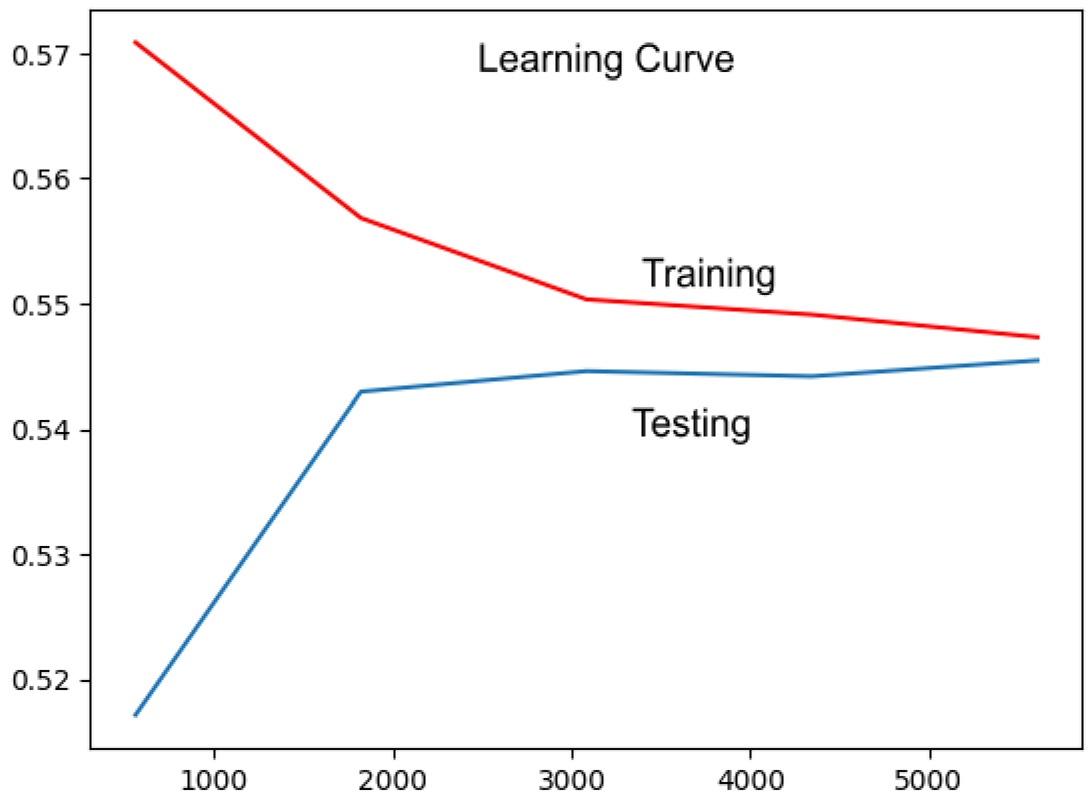


<table>
<tr><td colspan="4">The MSE for training and testing are similar which indicates that this model does not overfit or underfit.
However our R^2 is a little low. This means our model can be improved. Which is why we made a couple more models. The above point in regards to over/under fitting is further highlighted by this plot of the learning curve which shows that the testing set (blue) increases before flattening out around 0.465. The testing and training end at very similar points which once again shows that this model is most likely non overfit or underfit. The next step in this case is to look at adding other features to our model and seeing how the model performs in that case.</td></tr>
<tr><td>Temperature(°C),
Hour,
Dew point temperature(°C),
Seasons_Winter,
Seasons_Summer,
Solar Radiation (MJ/m2,
Visibility(10m),
Functioning Day_Yes</td><td>max_iter = 1000
tol = 1e-3
alpha = 0.0001</td><td>MSE: 195466.12
MAE: 330.15
EV: 0.53
R^2: 0.53</td><td>MSE: 189417.94
MAE: 325.98
EV: 0.54
R^2: 0.54</td></tr>
<tr><td colspan="4">Results:
For the features in this model, we selected features that had a correlation greater than or equal to 0.2.
There are small improvements in the MSE and R^2 in this model but improvements can be made still.</td></tr>
<tr><td>Temperature(°C),
Hour,
Dew point temperature(°C),
Seasons_Winter,
Seasons_Summer,
Solar Radiation (MJ/m2,
Visibility(10m),
Functioning Day_Yes,
Wind speed (m/s),
Rainfall(mm),
Snowfall (cm)</td><td>max_iter = 1000
tol = 1e-3
alpha = 0.0001</td><td>MSE: 191068.21
MAE: 326.46
EV: 0.54
R^2: 0.54</td><td>MSE: 185943.44
MAE: 324.63
EV: 0.55
R^2: 0.55</td></tr>
<tr><td colspan="4">Results:

After adding more features to include those that had a correlation greater</td></tr>
</table>

than 0.1, the model did seem to improve but not by a significant amount. Thus, the next step would be to go back to the features in the previous step and focus on the hyper-parameters as we felt that the decrease in mse and increase in r^2 did not necessarily justify the increase in model complexity.			
Temperature(°C), Hour, Dew point temperature(°C), Seasons_Winter, Seasons_Summer, Solar Radiation (MJ/m2, Visibility(10m), Functioning Day_Yes	max_iter = 1000 tol = 1e-4 alpha = 0.00001	MSE: 195547.87 MAE: 328.96 EV: 0.52 R^2: 0.52	MSE: 189667.69 MAE: 324.96 EV: 0.54 R^2: 0.54
Results: After changing the hyper parameters, it seems to have increased the mse slightly for the testing, and not changed the R^2, which could indicate that hyperparameter tuning might have a limited effect. Nonetheless, it would be prudent to try different combinations of tuning to make sure.			
Temperature(°C), **Hour,** **Dew point temperature(°C),** **Seasons_Winter,** **Seasons_Summer,** **Solar Radiation (MJ/m2,** **Visibility(10m),** **Functioning Day_Yes**	**max_iter = 1000** **tol = 1e-2** **alpha = 0.001**	**MSE: 188772.53** **MAE: 323.70** **EV: 0.54** **R^2: 0.54**	**MSE: 183967.64** **MAE: 322.00** **EV: 0.55** **R^2: 0.55**
Results: After increasing the hyperparameters, it seems that the MSE decreased and the R^2 increased - again not by much - it seems that there is an upper limit to the R^2 value at 0.55			
Temperature(°C), Hour, Dew point temperature(°C), Seasons_Winter, Seasons_Summer,	max_iter = 10000 tol = 1e-4 alpha = 0.00001	MSE: 188491.12 MAE: 323.51 EV: 0.54 R^2: 0.54	MSE: 183544.31 MAE: 321.18 EV: 0.55 R^2: 0.55

<table>
<tr><td>Solar Radiation (MJ/m2,
Visibility(10m),
Functioning Day_Yes</td><td></td><td></td><td></td></tr>
<tr><td colspan="4">Results:

Increasing the total iterations marginally improved the model, however it seems that any form of hyperparameter tuning does not improve the model by a lot. It seems to have reached some kind of plateau at 0.55. The learning curve shows that the testing set (blue) ends close to the training set, which indicates that there is no major overfitting or underfitting. This model performs roughly the same as the model with 1000 iterations, so that model would be the best model of the ones selected. An R^2 value of 0.55 is considerable on the lower side of average, this could be due to non linearity between the features and the target variable. A lot of the variables when plotted against the target showed that the relationships were not exactly linear, which could contribute to the results we are seeing here.

</td></tr>
<tr><td>Temperature(°C),
Hour,
Dew point</td><td>max_iter = 1000
tol = 1e-3
alpha = 0.0001</td><td>MSE: 195198.17
MAE: 329.42
EV: 0.53</td><td>MSE: 189956.27
MAE: 326.14
EV: 0.54</td></tr>
</table>

<table>
<tr><td>temperature(°C),
Seasons_Winter,
Seasons_Summer,
Solar Radiation
(MJ/m2,
Visibility(10m),
Functioning
Day_Yes</td><td></td><td>R^2: 0.53</td><td>R^2: 0.54</td></tr>
<tr><td colspan="4">Results:
This model uses ridge regression in order to regularize the data in order to ensure that we are correct in our analysis of possible overfitting. It uses the features of the second model due to that model having better results.

This model performs the same as model 2. Maybe even slightly worse. The MSE and R^2 are pretty much the same as before.</td></tr>
</table>

We selected the bolded model as our final model using Stochastic Gradient Descent. This model has an R^2 of 0.55 which is the highest amongst all the models that we tested. Furthermore, even though there were models with marginally better mean squared errors, the differences were not significant. Thus, this model represents a good balance between model complexity and model performance.

OLS:
R^2 and adjusted R^2 are both .53. Similar to our SGD models.
Unfortunately this is still just a moderate fit.
The F-statistic is 879.5 and its p-val is 0. This means overall our model is statistically significant and that at least one of our features is relevant.

The RMSE value of 435.8397 means that, on average, the model's predictions deviate from the actual values by around 436 bike rentals.

COEFFICIENTS:
Constant = 705.4033: This is the intercept of the regression line, meaning if all other variables are 0, the predicted rented bike count would be around 705.

SE = 5.282 Very small so the feature is estimated precisely.

P-val = 0. So statistically the feature is significant.

Temperature = 83.1526: For each degree increase in temperature, the model predicts an increase of approximately 83 rented bikes.

SE = 50.461. Quite large so there is uncertainty with this feature.

P-val = 0.099. This is close to 0.05 so marginal significance.

Hour = 189.1477: The hour of the day has a positive impact, meaning that as the hour increases, the rented bike count is expected to increase by 189 bikes, on average.

SE = 5.638. Also small indicating precision.

P-val = 0. Statistically significant.

Dew Point Temperature = 275.0702: This is also positively correlated with bike rentals, though less intuitive. A higher dew point temperature may indicate more humid conditions.

SE = 57.628. This is relatively large so there is some uncertainty.

P-val = 0. Statistically significant.

Seasons_Winter and Seasons_Summer = -122.2094 and -36.5143: They have a negative correlation which means there are fewer bikes rented in these seasons.

P-val = 0 (for both). Statistically significant.

SE = 8.712 and 7.958. Both are small

Solar Radiation = -63.4745: Surprisingly, the model suggests that higher solar radiation decreases bike rentals.

P-val = 0. Statistically significant.

SE = 7.958. This is also small

Humidity = -284.8885: Higher humidity results in fewer bikes rented.

P-val = 0. Statistically significant.

SE = 23.838. Small relative to coefficient

Visibility = 18.1210: As visibility increases, bike rentals increase slightly.

P-val = 0.006. Close to 0 so statistically significant.

SE = 2.661. Also small

Functioning Day_Yes = 156.2189: If it is a functioning day (i.e., bikes are available), rentals are expected to increase by 156 on average.

P-val = 0. Statistically significant.

SE = 5.464. Also really small relative to coefficient.

Notable t-values: (t-val > 2)

Hour: 33.551

Dew Point Temperature(°C): 4.773

Seasons_Winter: -14.068

Seasons_Summer: -4.589

Solar Radiation (MJ/m2): -8.718

Humidity(%): -11.941

Visibility (10m): 6.727

Functioning Day_Yes: 28.588

These t-values indicate their coefficients are statistically significant.

temperature, hour, dew point temperature, seasons, humidity, visibility, and functioning day all are statistically significant so they have an impact on bike rentals.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      Rented Bike Count   R-squared:                       0.531
Model:                            OLS   Adj. R-squared:                  0.530
Method:                 Least Squares   F-statistic:                     879.5
Date:                Sat, 07 Sep 2024   Prob (F-statistic):               0.00
Time:                        23:41:44   Log-Likelihood:                -52629.
No. Observations:                7008   AIC:                         1.053e+05
Df Residuals:                    6998   BIC:                         1.053e+05
Df Model:                           9
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
const                       705.4033      5.282    133.545      0.000     695.049     715.758
Temperature(°C)              83.1526     50.461      1.648      0.099     -15.766     182.071
Hour                        189.1477      5.638     33.551      0.000     178.096     200.199
Dew point temperature(°C)   275.0720     57.628      4.773      0.000     162.103     388.041
Seasons_Winter             -122.2094      8.687    -14.068      0.000    -139.239    -105.180
Seasons_Summer              -36.5143      7.958     -4.589      0.000     -52.114     -20.915
Solar Radiation (MJ/m2)     -63.4745      7.281     -8.718      0.000     -77.747     -49.202
Humidity(%)                -284.8858     23.858    -11.941      0.000    -331.654    -238.118
Visibility (10m)             18.1821      6.667      2.727      0.006       5.113      31.252
Functioning Day_Yes         156.2189      5.464     28.588      0.000     145.507     166.931
==============================================================================
Omnibus:                     1089.035   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2172.183
Skew:                           0.951   Prob(JB):                         0.00
Kurtosis:                       4.955   Cond. No.                         27.1
==============================================================================
```

Conclusion: Although our models account for overfitting, our R^2 values are still around 0.5. The linear regression models are not very good at capturing everything. Perhaps a more complex model would be better at predicting this data than Linear Regression.