Report: Predict Bike Sharing Demand with AutoGluon Solution
Kevin Barry

Initial Training
What did you realize when you tried to submit your predictions? What changes were needed to the output of the predictor to submit your results?
Based on the project instructions it indicated that I could not submit any predictions that were less than zero. Therefore once I extracted the predictions into the pandas Series I could iterate over the predictions to count how many or in my case I ran the describe function first to determine if there was at least one negative prediction. If so I went ahead and counted how many and then set those to zero and recheckec with the describe function.

What was the top ranked model that performed?
I obtained the same top ranked model for the first two model executions which was WeightedEnsemble_L3, due to my hyperparameter tuning and a smaller set of models, LightGBM_BAG_L1/T4 came out as the top model on the hpo execution.

Exploratory data analysis and feature creation
What did the exploratory analysis find and how did you add additional features?
Based on the histograms, the temp, atemp and humidity were fairly normally distributed, windspeed and count both appeared to be right skewed while the remaining variables datetime, holiday and workingday were as expected in all three cases - holiday being a one or zero only had two bars and the same for workingday.
In terms of the additional feature, I extracted the weekday number from the datetime feature. My thought here was that by doing this I could see if there were days that had more rentals such as workdays, Mon-Fri when workers are in the city or if Sat & Sun might be useful in terms of people renting for relaxation and touring around the city.

How much better did your model preform after adding additional features and why do you think that is?
The model did perform better with the additional feature but only very slightly. This was probably due to the weekday not being that good as a predictor for this dataset.

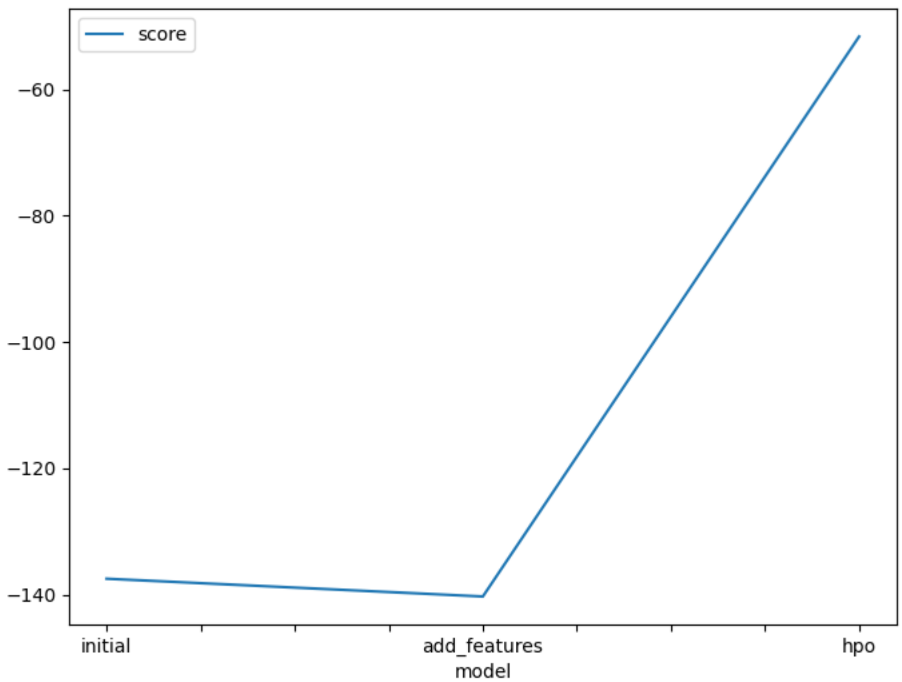
Hyper parameter tuning
How much better did your model preform after trying different hyper parameters?
The model did not perform as well post the hyperparamter tuning, part of my tuning removed some of the available models that appearted to have been contributing greatly to the model's accuracy.

If you were given more time with this dataset, where do you think you would spend more time?
With more time I would spend it in tuning the hyperparamters and focusing only on the more accurate model types.

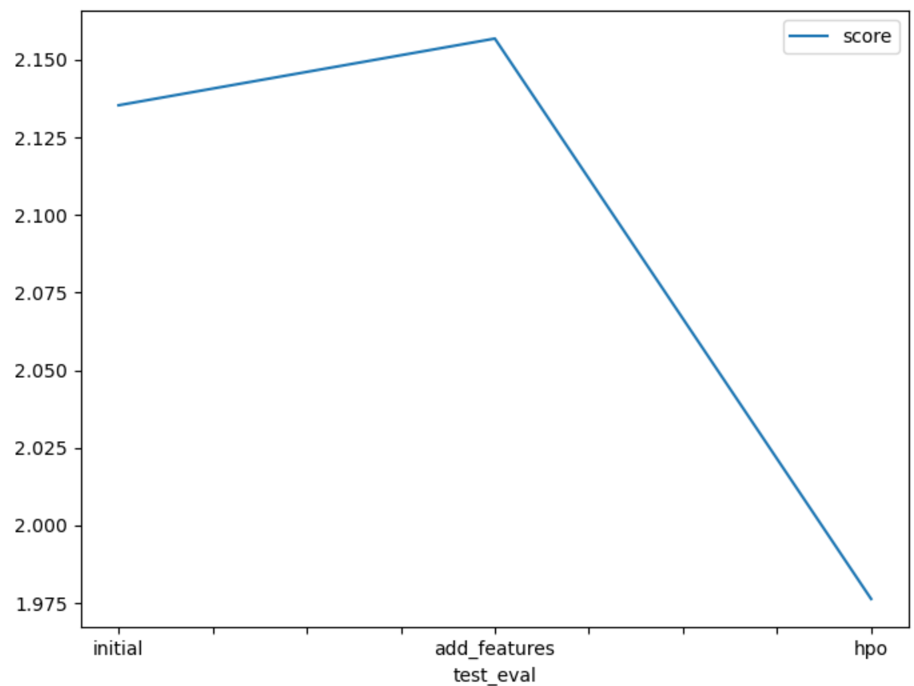
Create a table with the models you ran, the hyperparameters modified, and the kaggle score.

	model	No. of Models	Types of Models	time	score
0	initial	15	7	600	-137.475170
1	add_features	14	7	600	-140.290140
2	hpo	6	2	600	-51.638148

Create a line plot showing the top model score for the three (or more) training runs during the project.



Create a line plot showing the top kaggle score for the three (or more) prediction submissions during the project.



Summary
In summary the project consisted on acquiring the data from Kaggle via the Kaggle API, performing exploratory data analysis followed by a model fitting of the data as-is. This first model scored a 2.13 on Kaggle and an RMSE of -137.475.
The second model consisted of adding the weekday number as a new feature which was extracted from the datetime feature, this model scored an RMSE of -140.29 and a Kaggle score of 2.15.
For the final model I tuned the hyperparameters for GBM and NeuralNetworks. For GMB I tuned the learning rate/num of leaves & boosting type. For neural networks I tuned the options number of epochs/learning rate/layers/activation & dropout prob. This model did not perform as well as the previous models due to limiting to the two models chosen.
If I spent more time on this project I would invest it in the hyperparmeter portion of the project.