Assignment 9

In this module, you will write one Python notebooks to create required outputs. You will also participate in module 9 forum discussion.

Notebook 1: Sklearn

Watch the lecture videos. The assignment is in the same format as presented in the lecture but on a different data set.

Requirements:

1. Watch the lecture videos and fully understand the lecture notebook
2. Analyze the "app_usage" data set by following the example in the lecture
3. Create the heat map as in the attached "heatmap.png".
4. Display the R and adjusted R of each model. Make sure that the features are added in the descending order of their individual R squared values. Create an output as in the attached "models.png".
5. Read this (http://abbottanalytics.blogspot.com/2004/12/find-correlated-variables-prior-to.html) and improve your model by removing some features. The answer may vary. Write in your notebook to explain what features you removed and why you remove those features.
6. Find a model with 3 features and reasonable R squared value
7. Summarize the feature selection process in your notebook

*. An assignment template is provided to give you detailed guidance of completing the assignment

Submissions:

You will export your notebook to both .html and .py formats. You will submit the following file to Blackboard. In your html file, you should **include only required outputs** of your python script without error messages.

1. Firstname_Lastname_sklearn.zip (zip the .html and .ipynb files)

Attachments:

heatmap.png	:	Sample correlation map
models.png	:	Sample output of R and adjusted R of all models.
App_usage.csv	:	VPN access data set
Scikit_learn_assignment_template.ipynb:		The assignment template