IE 594 Data Science 1

Final Exam

Fall, 2016

Please read the instructions for submitting the results for the final exam very carefully. <u>**Any violation of the instructions will be subject to withdrawal of some points from your grade.**</u>

Instructions:

1. The final exam MUST be done individually.
2. The data set for the exam is also uploaded on BB. (Auto.csv)
3. You MUST upload only <u>**one**</u> Jupyter Notebook file on BB. Avoid sending zipped files, pdfs, images, etc.
4. Feel free to include texts as <u>**Markdown**</u> or links to your Plotly images in your file.
5. The files MUST be named as your UIN. Do not include anything else in the name of the file.
6. For questions 2 and 3, you are free to use different built-in functions/methods in different Python packages for modeling.
7. The due date for submitting the final file is Friday 12/9 at 9 PM.
8. Needless to say that the students with outstanding results (Specially in question 1) will earn <u>**Happy Points**</u>.

The Final Exam's description is presented in the next page.

Final Exam:

A market research company has created a data set on different popular cars in the world. You are approached to deliver the following data analytics products:

1. **Data Exploration:** Create one (or more if necessary) graph(s) which enables the user to pick different variables in the data set and explore their individual or group behavior.
(40 points)
   - *All the visualizations for this section **MUST** be done in Plotly.*
   - *Feel free to use different control widgets (Dropdown menus, Button, Slider, etc.) to make an interactive data visualization product.*
   - *You will be graded based on your **creativity** and the **quality of the graph**(s) for deriving insights form the data set.*

2. **Regression:** Develop a regression model to predict cars' **mpg** based on other features.
(30 points)
   - *Make sure that you include all necessary tests to validate your results.*
   - *For the visualizations in this section, you are free to use other packages than Plotly.*
   - *Make sure that you investigate regression assumptions.*
   - *Make sure that you report your conclusion about the model's performance based on different criteria.*

3. **Classification:** Create a new variable as a flag for those cars with acceleration above 15 or less than/equal to 15. Name the new variable as "**Target**". Develop a classifier to classify cars based on their different features and the Target as the response variable.
(30 points)
   - *Make sure that you take a reliable procedure to avoid overfitting.*
   - *For the visualizations in this section, you are free to use other packages than Plotly.*
   - *Make sure that you report your conclusion about the model's performance based on different criteria.*
   - *Feel free to compare different models (with different parameters) to gain acceptable results.*