Project Assignment

Overall Requirements

The project code and report are due at the same time and shall be submitted together.

Report Requirements

Overall

- 10 pages or more (excluding the appendix)
- Shall be a Microsoft Word Document
- High quality writing is expected with good grammar and punctuation. The document should be thoroughly proof read and reviewed by the team prior to submission.
- Provide visualizations throughout the report which can be any form of plots, tables, graphics, etc.

Sections

Only the following sections in the following table shall be included in the report:

Report Sections	
Item	**Description**
Title Page	Including the name of your project, the team members who contributed to the project, and your project group number assigned on blackboard.
Abstract	See abstract assignment for details. The abstract can be an update from the abstract which was turned in previously.
Introduction	Describe the project in more detail expanding upon the abstract. • Expand upon the project objective described in the abstract. • Describe the data set in more detail. For example, where did the dataset come from and what is it used for. Describe the predictors. Summarize your data with plots and statistics. • Describe the business or scientific problems you are trying to solve in detail. The abstract should have provided a summary of the problems you are trying to solve, here is your chance to fill in the details.

Report Sections	
Item	**Description**
Methodology	Transition from the specific business or scientific questions you are trying to answer into methods to answer those questions. What data wrangling was needed and how did you approach it? What data preprocessing did you use (PCA, standardization, normalization, etc)? Did you explore the data using visualizations to gain insight? What machine learning models did you use (decision trees, linear regression, logistic regression)? How did you do feature selection? How were model hyper parameters chosen? Did students analyze validation data misclassification gain insight about model misclassification. Did students use models for inference. What insights were gained through inference. Support results with visualizations.
Results	Describe the results of the project. What accuracy did the models achieve. What were your test, train, and validation errors? Did you have problems with over or under fitting. How did you identify over or under fitting. What specific scoring criteria did you use. What insights did you learn about the dataset using inference. Compare and contrast different approaches tried. Present both good and bad results in your comparisons. Use visualizations to support your results.
Conclusion	Summarize the project results. What conclusions can you draw from your research. Did you successfully answer the business or scientific questions you wanted to answer. Support your claims with data and visualizations.

Report Grading (100 pts)

Report Grading Rubric		
Item	**Description**	**Weight %**
Writing Quality / Following Instructions		10
Abstract		10
Introduction		20
Methodology		20
Results		20
Conclusion		20

Code Grading (50 pts)

Overall

The code is expected to be well organized and well commented so that it's easy to follow. Try to break the code up into logical sections that perform major tasks. For example, perform all of the data wrangling in one section, develop machine learning models in their own sub sections, inference in a different section. Code shall follow the following guidelines:

- All code shall run on databricks without error.
- All machine learning, inference, analysis, data wrangling, and exploration shall use spark. In short, the data science work flow shall be performed using spark.
- Visualizations can be performed using any python package which will run on spark. Graphics, tables, plots, visualizations of ANY kind can be generated using any python package which will run on databricks.

You can install packages on databrocls but must provide the code which installs the packages in your notebook. For example, if there is some package you want to use for visualizations which is not available by default on databricks, you can install and use that package on databricks as long as you supply working code which installs the package on databricks in your notebook.

Code Grading

An overall grade will be performed by reading through the code and subjectively evaluating according to the following criteria:

- Code effectively split up into modules:
- Effective use of comments. Is the code easy to follow.
- Does the code correlate with the information in the report
- Did students follow submission instructions
- Does the code run without error on databricks

Submission:

- All submissions shall be submitted as a zip file.