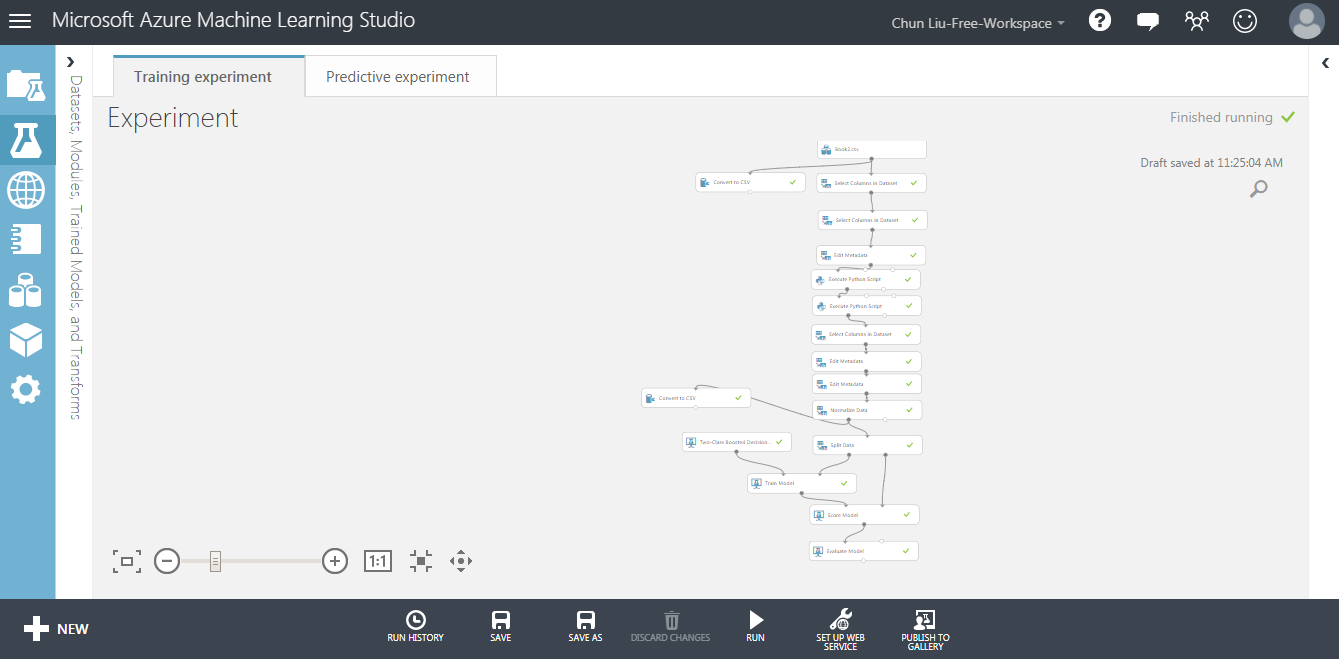
Project Demonstration-Chun Liu

* Machine Learning -- Healthcare Insurance Claim Error Detection

Predict which healthcare claim had an overpayment.

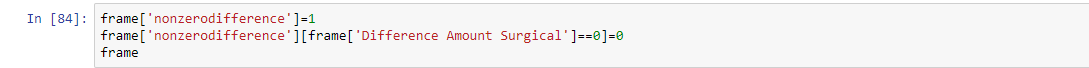
Start with a dataset of healthcare insurance claims. Create statistical universe and sample with criteria. Obtain payment error findings for each healthcare insurance claim sampled. Explore features in the sample dataset with visualization. Do feature engineering. Fit training dataset with Logistic Regression Classifier. Evaluate performance of the classification model. Do feature selection, cross validation, regularization, etc. Fit training dataset with Gradient Boosting Classifier. Evaluate performance of the classification model. Compare models and find the optimized one. Predict new datasets with the optimized classification model and obtain claim error rates. When claim error rate is beyond the threshold it shows symptom of billing and payment problem, which needs further investigation and solution before payment being processed. Programming: Spark Scala, SQL, Azure ML, Python with packages matplotlib, seaborn, numpy, pandas, scikit-learn, etc.

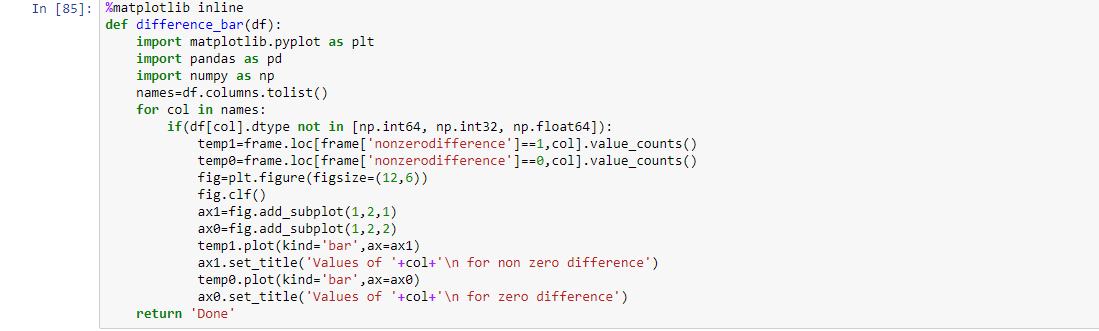
--Azure ML--

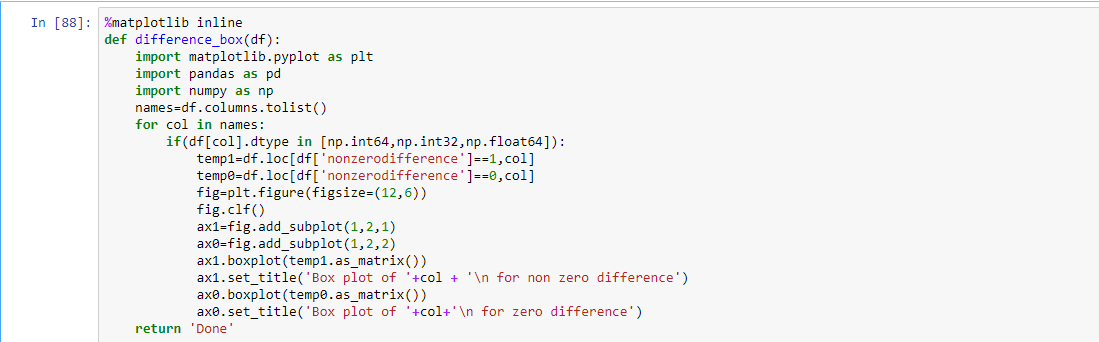


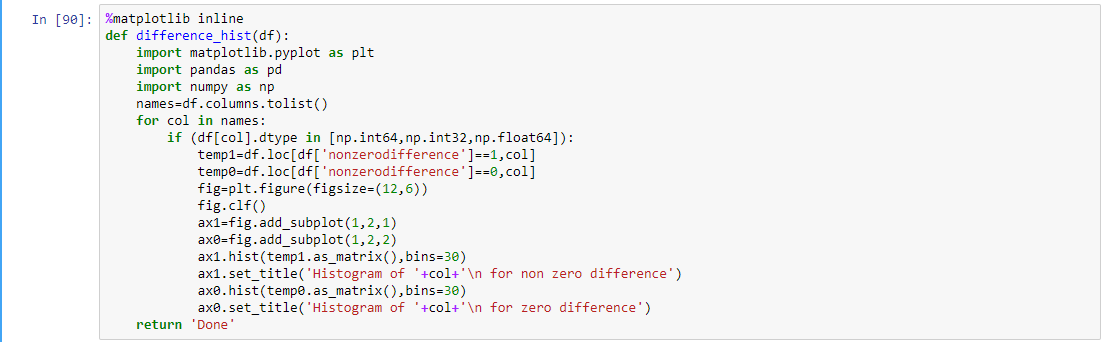
--Jupyter--









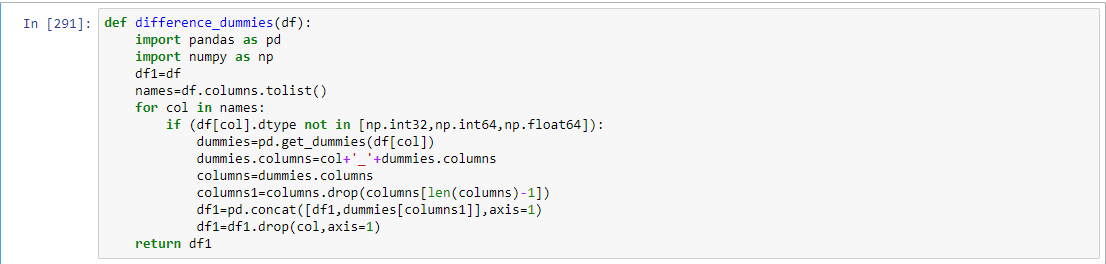






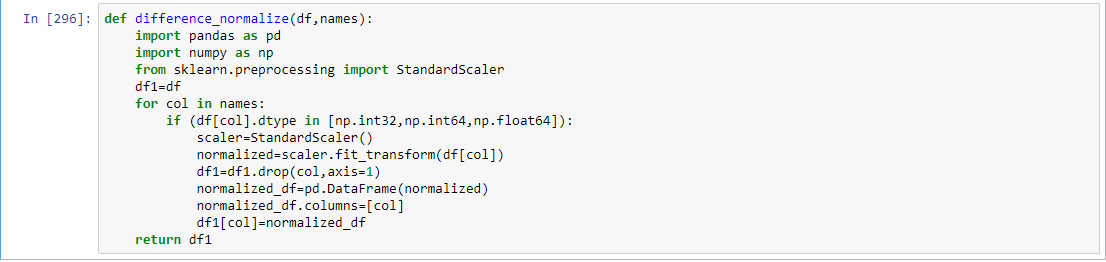






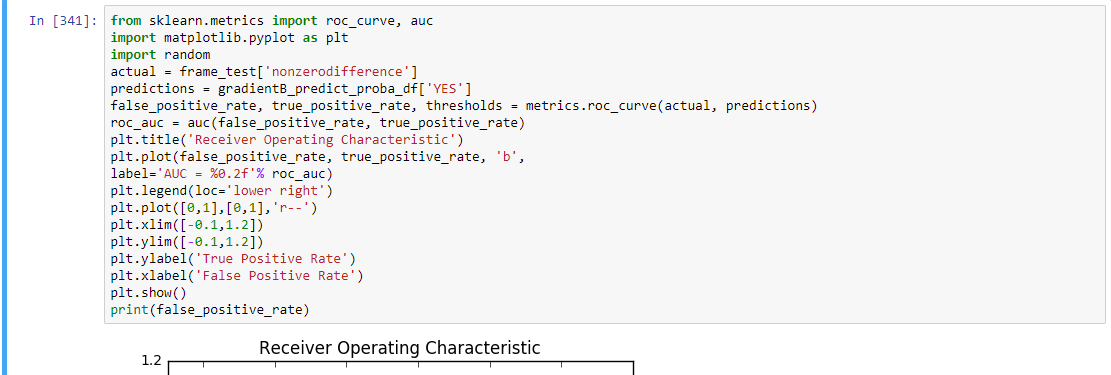






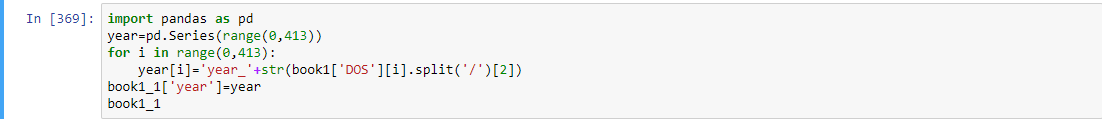




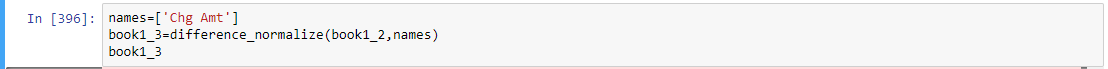


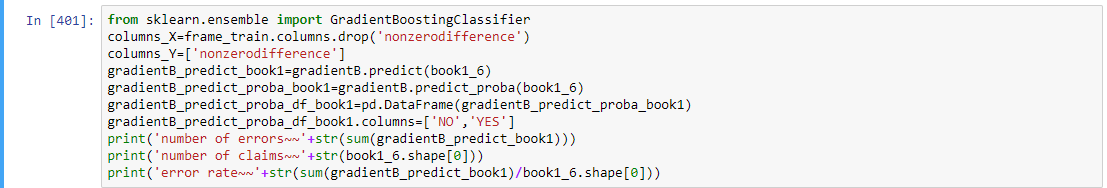












* Apache Spark -- Marketing Analysis

Do data mining on customer records of a marketing campaign.

Start with a dataset that contains records of customers contacted in the campaign. Do data cleansing. Load data and create Spark data frame. Write queries to obtain marketing success rate, explore features and do feature engineering, etc. Programming: Spark-Shell Scala.

--Spark--

