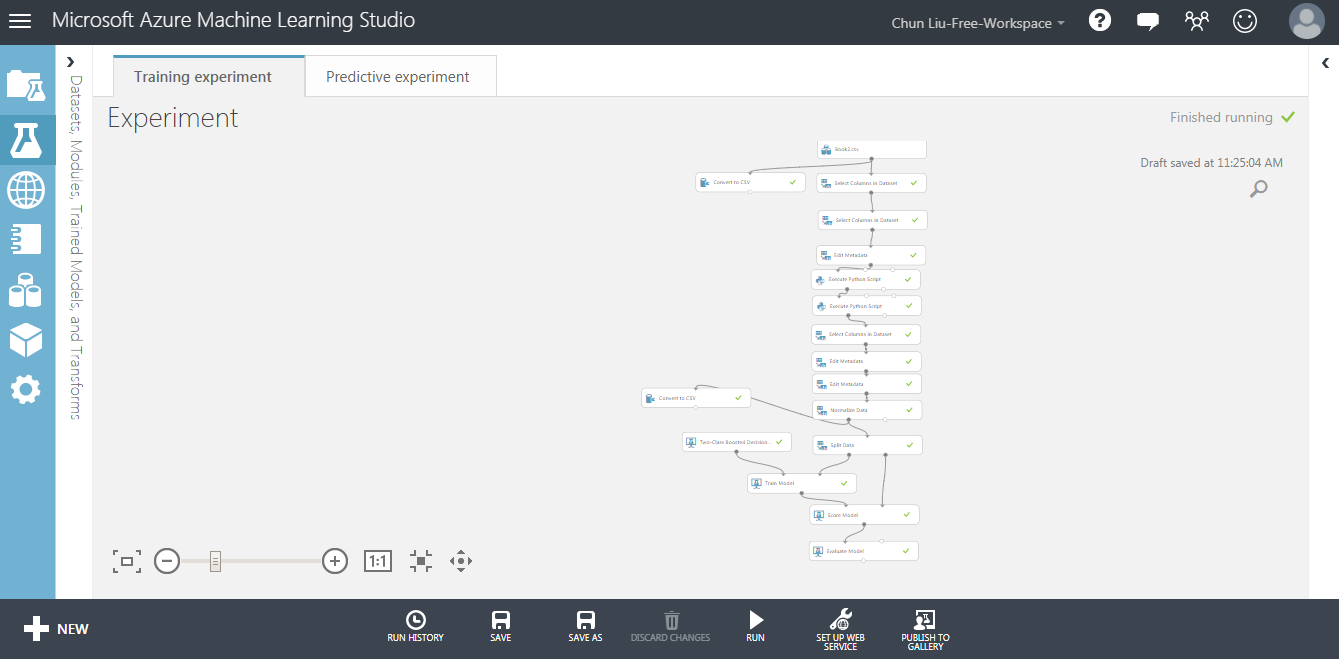
Project Demonstration-Chun Liu

* Machine Learning – Healthcare Claim Error Detection

Predict which healthcare claim had an overpayment.

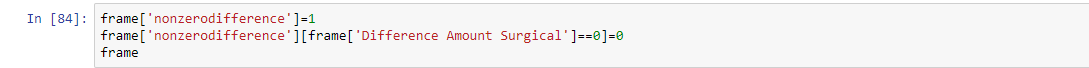
Start with a dataset of healthcare claims with payment error findings. Write code in Python with packages matplotlib, seaborn, pandas, numpy, scikit-learn, etc. Prepare dataset. Explore features in the dataset with visualization. Do feature engineering. Fit data with Logistic Regression Classifier. Evaluate performance of the classification model. Do feature selection, cross validation, regularization, etc. Fit data 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 high it shows symptom of billing and payment problem, which needs further investigation and solution before payment being processed.

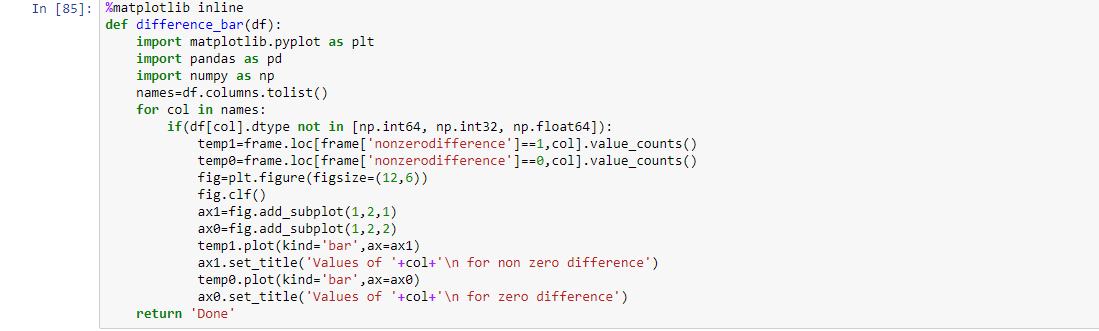
--Azure ML--

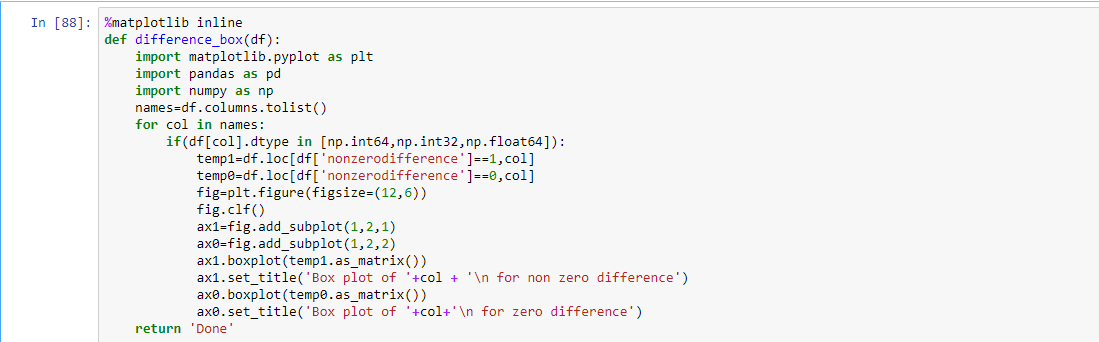


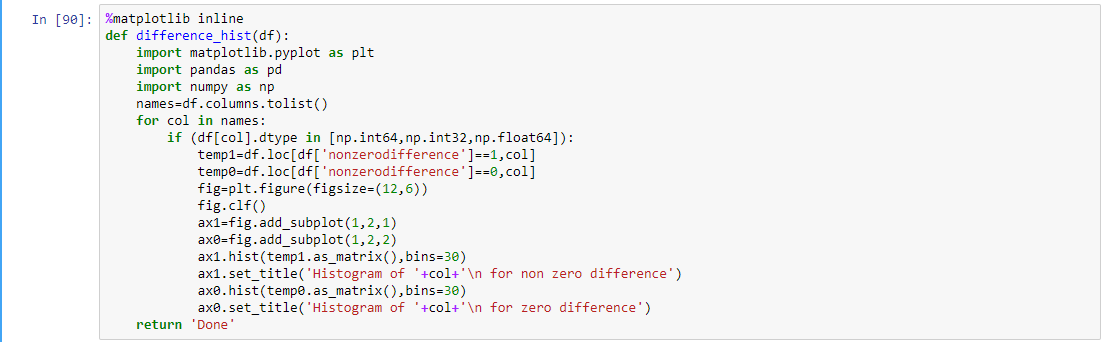
--Jupyter—









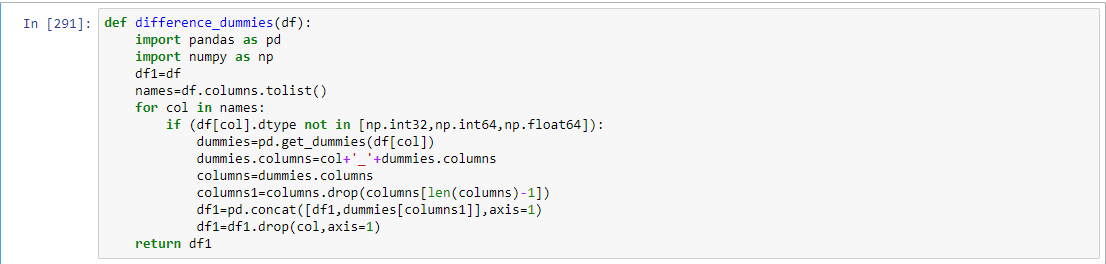






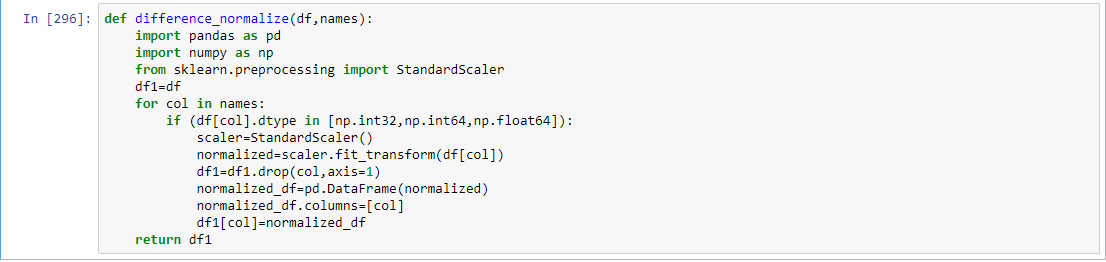






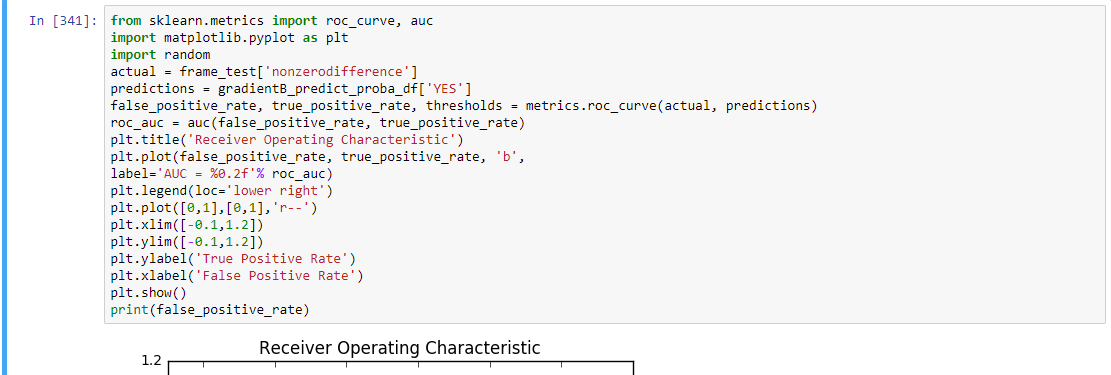






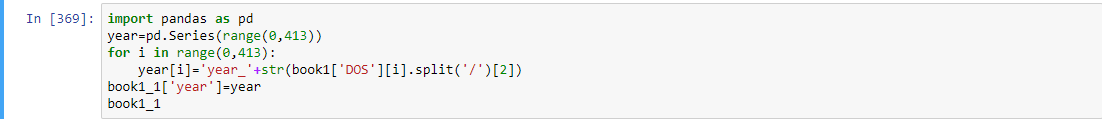




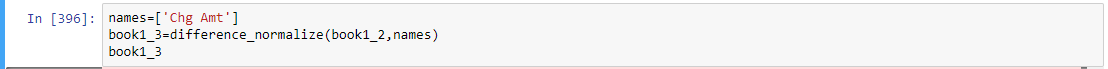


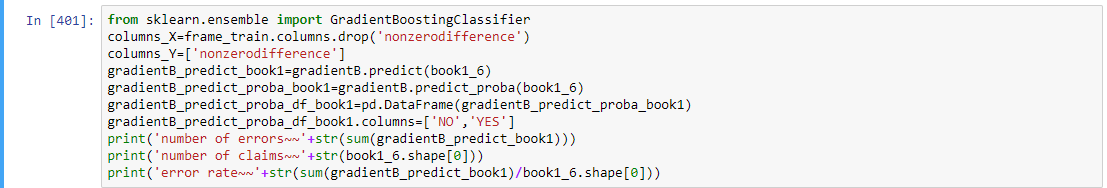












* Apache Spark -- Marketing Analysis

A healthcare financial institution ran a marketing campaign to promote healthcare deposit.

Start with a dataset that contains records of customers contacted in the campaign. Do data cleansing. Write code in Spark-Shell Scala. Load data and create Spark data frame. If a customer invests in the healthcare deposit then it is a success, else it is a failure. Write queries to obtain marketing success rate, explore features with success rate and do feature engineering.

--Spark--

