# Summary of class on 5-April-2017

* First presentation was by Robby Kennedy covering Real Estate Pricing using machine learning. He wanted to investigate value of Real Estate determined by the Hedonic Pricing Theory which states that prices are determined by both external and internal attributes. Real estate professionals use similar techniques when determining price by reviewing heterogeneous data and look for similarities. When pricing real estate the problem is to find recent data to determine values. Upon finding the data the key is to homogenize the data to make it usable. The existing data in a machine learning repository uses data from the city of Boston and was intended for clean air studies. Local data was collected by hand, the Zillow API was not used because certain attributes are not used. Data processing was restricted to the zip code 33432 and involved removing outliers. In this study the M5 linear model worked best.
* Daniel Bruce presented the second topic of the day and involved “Predicting Customer Churn for SaaS Platform. The motivation was that a SaaS platforms spend large amounts of money getting new customers and according to a study on a financial services platform about 1/3 are lost due to churn. The ability to predict if/when customers leave will save the company significant amounts of revenue. An intervention effort to retain customers. The cost of outreach is estimated at $10 per customer versus the potential thousands of dollars lost due to a lost customer. This area needs additional research and more information is needed. It is believed that SaaS companies have the data, since every interaction is logged, but analysis is needed.
* The final presentation was by Richard Bauer whose presentation was titled “The Cost of Being Atypical: Using Anomaly Detection to Discover Healthcare Fraud”. This presentation was he most interesting. In it he described his efforts to catch and predict outliers in Medicare data. In his review he was careful to note that outliers are not an indication of fraud, however he hoped to develop a technique to automatically identify outliers and mark them for additional review. His review was limited to the state of Florida. Of note was that nationally Medicare spending was around 3 trillion dollars and fraud is estimated at 3-10% of the total. Of that number 10-15% is recoverable. The data was split into provider types and medical codes and investigated by procedures. Of note is that one outlier identified, when further researched, was under investigation.