Assignment 3

(Team 6)

**Goal**: Predicting whether a person makes over 50K a year.

Sources for Data and its descriptions: UCI Machine Learning Repository https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data

https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.names

**Key findings/Summary/Results**

**Methods and Analysis**

1. Data munging

; Data was imported from the url and was cleaned (removed the observations with “?” mark and checked outliers) before working on it.

2. Exploratory data analysis

; As requested, we used the first half of the data as a training data to explore data further.

[insert Graphs]

3. Classification methods, models and algorithms that we had studied in class.

; We began with conducting logistic model, as it is a simple linear regression model to start when having two classes for a response variable.

Subsetting? **!do choose a justifiable subset according to the requirement!**

4. Use the second half of the data for validation.

;

**Essential R scripts**

[insert R scripts that is important]

1. Data import/cleaning