**Team Name:** The Bejamins

**Team Members:** Ziyaad, Cora, Brian, Ken, and Lisa

**Project Description/Outline:**

LendingClub claimed that they have a Machine Learning model that could predict default rates better than using FICO score system.

What we are trying to analyze is, does LendingClub use of Machine Learning model gives them a higher predictability?

If we build a similar model, will it hold up over time? In conclusion, is it really a superior Machine Learning model? Is the R-Squared consistent?

**Visualization**

Create 4 regression models

Timeframes:

* + 2007 – 2011 (LC started in 2006)
  + 2014 (this is the year that the LC went public) (to test)
  + Q3 2017 (to test)
  + Q4 2018 (to test)

Y-axis: interest rate

X-axis: grade (A, B, C, D, E, F)

Create bar charts using Tableau or LendingClub’s site

**Datasets to be Used:**

Lending Club CSV: <https://www.lendingclub.com/info/download-data.action>

**Workflow:**

Step 1: Parse the data by grade and interest rate just the 2 columns.

Step 2: Train 2007 – 2011 data and test it on 2014, Q3 2017, and Q42017.

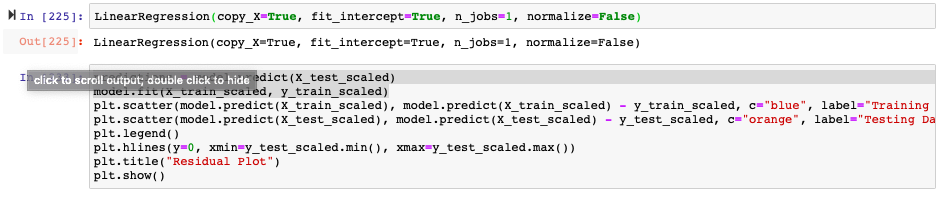
* Take the 2007- 2011 and train the regression = you want 3 different regressions.
* Use this website <https://www.lendingclub.com/info/download-data.action> to get the datasets.

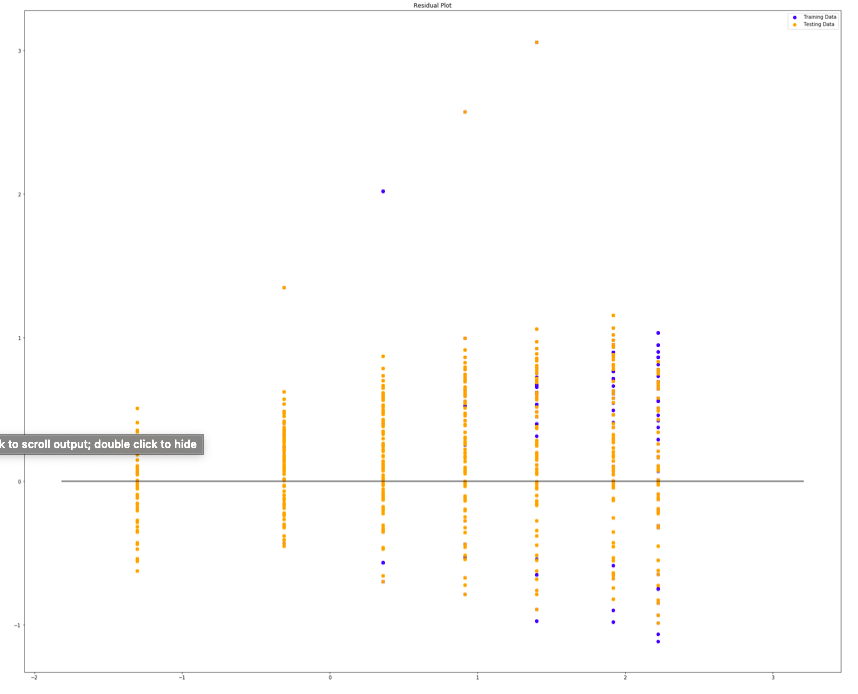
Step 3: Put the results in HTML

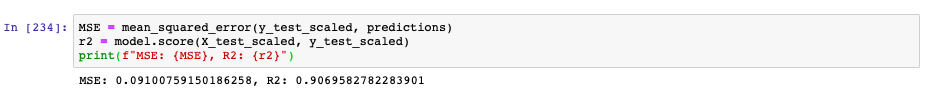
**My comments:**

* So I tested by doing the 2007 – 2011 data, but I don’t know how we can use the “grade” to predict using regression model only because the A, B, C, D, E, F, G will pick up fixed interest rates for each grade.
* I tried the regression model on 2007 – 2011 and the outcome was this:









MSE: 0.09100759150186258, R2: 0.9069582782283901 (which is good). Also 2007 – 2011 show 85% fully paid and 15% charge off. I am not sure what the industry standard but assume that this is good?!

