**Data Analytics Assignment**

We will use lending data from 2007-2010 and be trying to classify and predict whether the borrower paid back their loan in full.

Here are what the columns represent:

* credit.policy: 1 if the customer meets the credit underwriting criteria of lendco.in, and 0 otherwise.
* purpose: The purpose of the loan (takes values "credit\_card", "debt\_consolidation", "educational", "major purchase", "small\_business", and "all\_other").
* int.rate: The interest rate of the loan, as a proportion (a rate of 11% would be stored as 0.11). Borrowers judged by lendco.in to be more risky are assigned higher interest rates.
* installment: The monthly installments owed by the borrower if the loan is funded.
* log.annual.inc: The natural log of the self-reported annual income of the borrower.
* dti: The debt-to-income ratio of the borrower (amount of debt divided by annual income).  
  fico: The FICO credit score of the borrower.
* days.with.cr.line: The number of days the borrower has had a credit line.
* revol.bal: The borrower's revolving balance (amount unpaid at the end of the credit card billing cycle).
* revol.util: The borrower's revolving line utilization rate (the amount of the credit line used relative to total credit available).
* inq.last.6mths: The borrower's number of inquiries by creditors in the last 6 months.
* delinq.2yrs: The number of times the borrower had been 30+ days past due on a payment in the past 2 years.
* pub.rec: The borrower's number of derogatory public records (bankruptcy filings, tax liens, or judgments)

**Questions:**

1. What factors effect the interest rate of loan taken?
2. What factors are used to evaluate risk level of customers?
3. What trends are observed w.r.t delinquencies - what kind of customer should the company avoid as inferred form given data
4. Which loan combination is best for company e.g. interms of higher loan interest rate, higher repayments, low default
5. what is correlations between FICO and credit history/ behavior. Are there any common trends you can identify between FICO and credit and repayment history?
6. Are there any trends that can be identified about customer from 'revolving utilization rate ' & 'revolving balance'
7. Which 'purpose' ( loan category) is most beneficial to business
8. How does the number of days with a credit line ('days.with.cr.line') relate to the loan repayment status?
9. Are borrowers with a longer credit history more likely to meet the credit policy criteria?