The .RMD files show the code I used to do analysis.
The .PDF files show the result of analysis.
The LCDataDictionary.PDF files show the profiles of all lendingClub loans.

Reading sequence would be

1. LoanRiskAnalysis_EDA.pdf or LoanRiskAnalysis_EDA.Rmd
2. LoanRiskAnalysis_InterestRate.pdf or LoanRiskAnalysis_InterestRate.Rmd
3. LoanRiskAnalysis_LoanStatus.pdf or LoanRiskAnalysis_LoanStatus.Rmd

Summary of data analysis
Part one: I explored three questions in LoanRiskAnalysis_EDA part.
■ What factor will influence the interest rate?
■ What are the distribution of loan status?
■ What are the purpose of applying a loan with LendingClub?

Part two: I engineered feature and built linear regression in LoanRiskAnalysis_InterestRate part.
■ Performed correlation analysis for numeric features and hypothesis test (unpaired t-test, ANOVA) for category features.
■ Built multi-variable linear regression model for interest rate prediction.
■ Selected high impact features through Lasso regularization.
■ Evaluated the performance of interest rate model through adjusted coefficient of determination.
■ Fitted linear regression model with regularization to control for multicollinearity and also built decision tree, random forest, boosting decision tree to predict interest rate for each loan.
■ Achieved 0.3 RMSE by boosting decision tree model on test data set

Part three: I engineered feature and built logistic regression in LoanRiskAnalysis_LoanStatus part.
■ Performed correlation analysis for numeric features and hypothesis test (unpaired t-test, ANOVA) for category features.
■ Built logistic regression model for interest rate prediction.
■ Selected high impact features through Lasso regularization.
■ Evaluated the performance of loan status model by ROC/AUC.