Analytical plan: Final lab

Components of analytic Plan	
Phase 1: Discovery	The data for this lab is the housing loan database assembled by federal agencies pursuant to the Home Mortgage Disclosure Act (HMDA). This database identifies the census tract location of almost every housing loan and housing loan application made in the United States each year. The data provided for analysis in this lab is an extract for the year 2010.
Data	
	The data set contains 230000
	subscribers (23000 training sample,207000 test sample).
Business Problem Framed	How to predict the chances of a borrower for a loan for a small amount of data filled in the online form.

Initial Hypotheses	Using advanced regression
	techniques can help to assess the predict success client when applying for a mortgage loan.
Phase 2: Data Preparation	
Variables description	Main table - lardb1, it contains information on different States.The sample was taken across the state 27,Minnesota . The other tables are used to replace codes in lardb1 words.
Phase 3: Model Planning	
Analytic Techniques	**Random Forest**
	Classifier and SGDCClassifier,
	Cross-Validation and ROC-AUC curve for validation and comparisson both models.
Tools	Python: sklearn, pandas.
Phase 4: Communicate Results	

Key Finding	1)The model developed has predictive power at least 0.79-0.81 at score. ROC-AUC curve squar is 0.67-0.68	
	2)The model developed has predictive power at least as good as the bank's current churn model	
Comparison with initial Hypothesis	The result is consistent with the	
	hypothesis.	
Business impact	This model, on the one hand, enables the client to assess their chances of getting a mortgage of a certain size, without spending a lot of time and without a large amount of data about yourself.The Bank in turn receives a database of potential customers.	