Personal Statement:

My career pursuit is to be a data analyst. I believe that data analysis can be used in various fields, whether it is analyzing financial data, business analysis, Internet data product analysis, agricultural analysis, forestry analysis, and so on. No matter which field I enter, I must need more knowledge in that field. Maybe some knowledge can only be understood after entering the industry. Although now I have not yet figured out which field I will focus on immediately after graduation. However, because my undergraduate degree has also been exposed to some financial courses, and I have seen the biography of Buffett and other entrepreneurs and financiers since I was young, which made me admire some experts in the financial field. Therefore, I have a relatively strong interest in analyzing financial data. When it comes to finance and transactions, we inevitably think of banks. After some searching, I found a past competition project that looked pretty good. It is a project proposed by Banco Santander and provides a corresponding data set. The data is anonymous, each row containing 200 numerical values identified just with a number. The purpose of this project is to determine which customers will conduct specific transactions in the future. In my opinion, the analysis of customers is the business of financial institutions, because they need to devote limited manpower and time to places that can bring higher profits. While I think that using these data, I can put into practice the regression model I learned in the first two months, and I can use other algorithms to improve the model after I have a better understanding of machine learning in the future.

Question:

Try to identify which customer will make a transaction.

The data source (s):

https://www.kaggle.com/c/santander-customer-transaction-prediction/data

Proposed Timeline of work: Your tentative plans to finish the following by which date.

EDA:11.16

Data Processing:11.20

Modeling and Validation:11.24

Write up: 12.2