Credit Card Balance Prediction Report

Name of the Student

Institution of Affiliationa

**Project Objective**

There are 400 credit card holders in this dataset, and the purpose is to understand which factors influence a cardholder's Credit Card Balance and forecast the average Balance of a certain individual. A customer analysis at a credit card firm might include such a test. Analyses could reveal whether consumers are at risk of defaulting on their debt or what kind of conduct potential customers can be expected to exhibit. By combining information such as the credit limit and the credit balance, it is possible to calculate the cardholder's credit utilization, which goes into the credit rating.

A multivariable regression analysis will be used to achieve this objective. The exercise will begin with an exploration of the dataset, followed by the selection of features and the analysis of regression, including the linear and logistical models. Finally, the developed regression model will be used to generate a new dataset and estimate the credit balance of cardholders based on demographic data..

Data Availability

The present exercise will study the Credit Card Balance Data. This is a data frame with 400 observations on the following variables:

* ID - Identification
* Income - Income in $10,0000
* Limit - Credit limit
* Rating - Credit rating
* Age - Age in years
* Education - number of years of education
* Gender - Male or Female
* Student - Yes or No
* Married - Yes or No
* Balance - Average credit card balance in $

**Project Assumption**

An individual's average monthly credit card balance is referred to as his or her credit card balance. Due to the Cards variable, which denotes how many credit cards a person has and only has one corresponding Balance value, this assumption was formed.

The Balance is the much you owe on a credit card in a particular month, divided by 12. The average balance recorded will be $500 if a cardholder spends $400 in one month, $500 the next, and then $600 all at once (i.e. any preliminary balances before the maximum are not taken into account, neither is the final balance of zero).

**Describing the features selected**

A notable feature of the sample is the presence of a sizable number of Zero Balance Cards. There are two questions that can be addressed as a result of this:

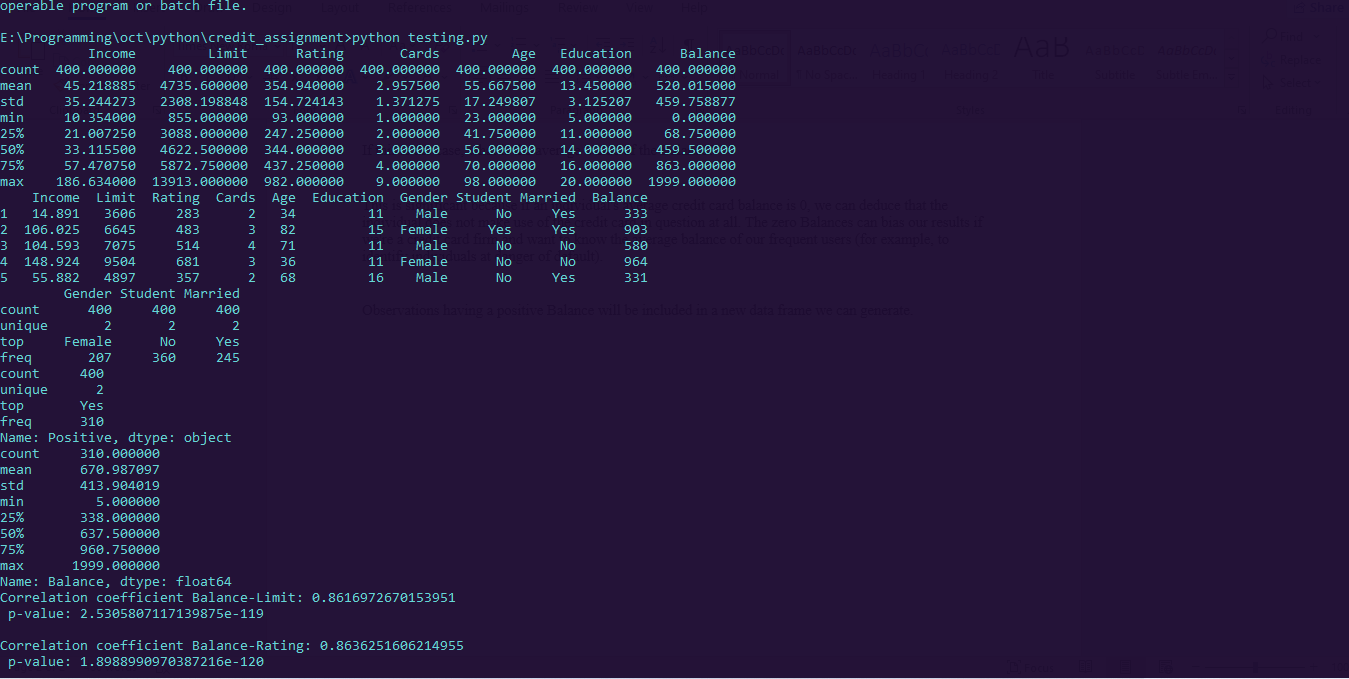
What's the average credit card balance of the person you're looking at?

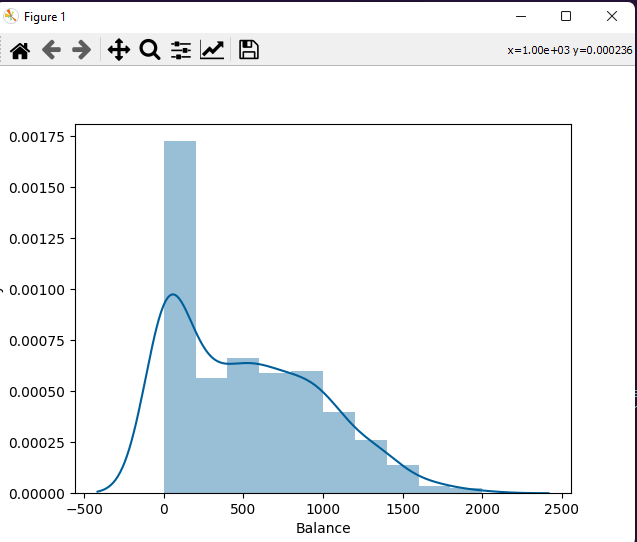
If this is the case, what is the average Value of the Balance?

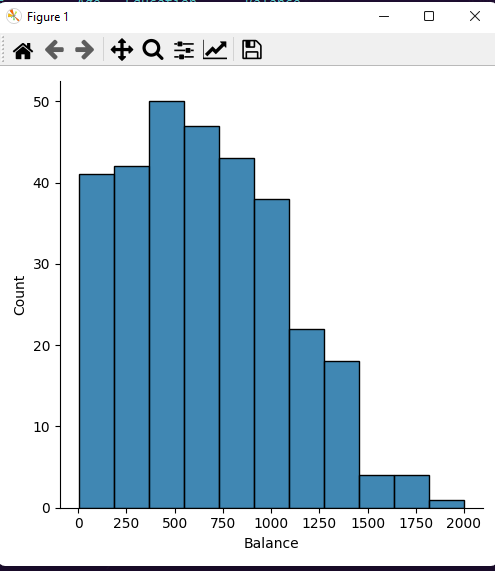
This is significant because if an individual's average credit card balance is 0, we can deduce that the individual does not make use of the credit card in question at all. The zero Balances can bias our results if we're a credit card firm and want to know the average balance of our frequent users (for example, to identify individuals at danger of default).

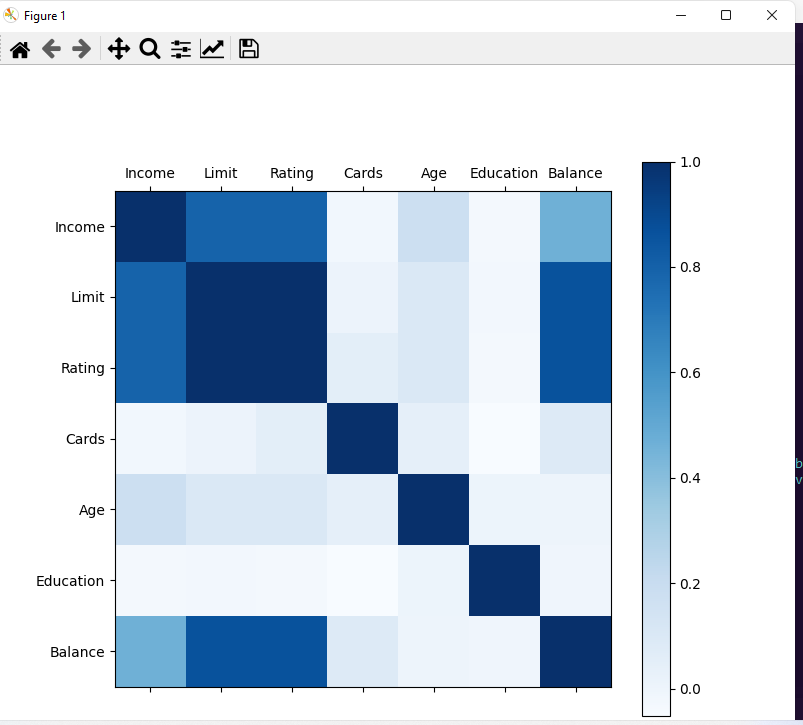
Observations having a positive Balance will be included in a new data frame we can generate.

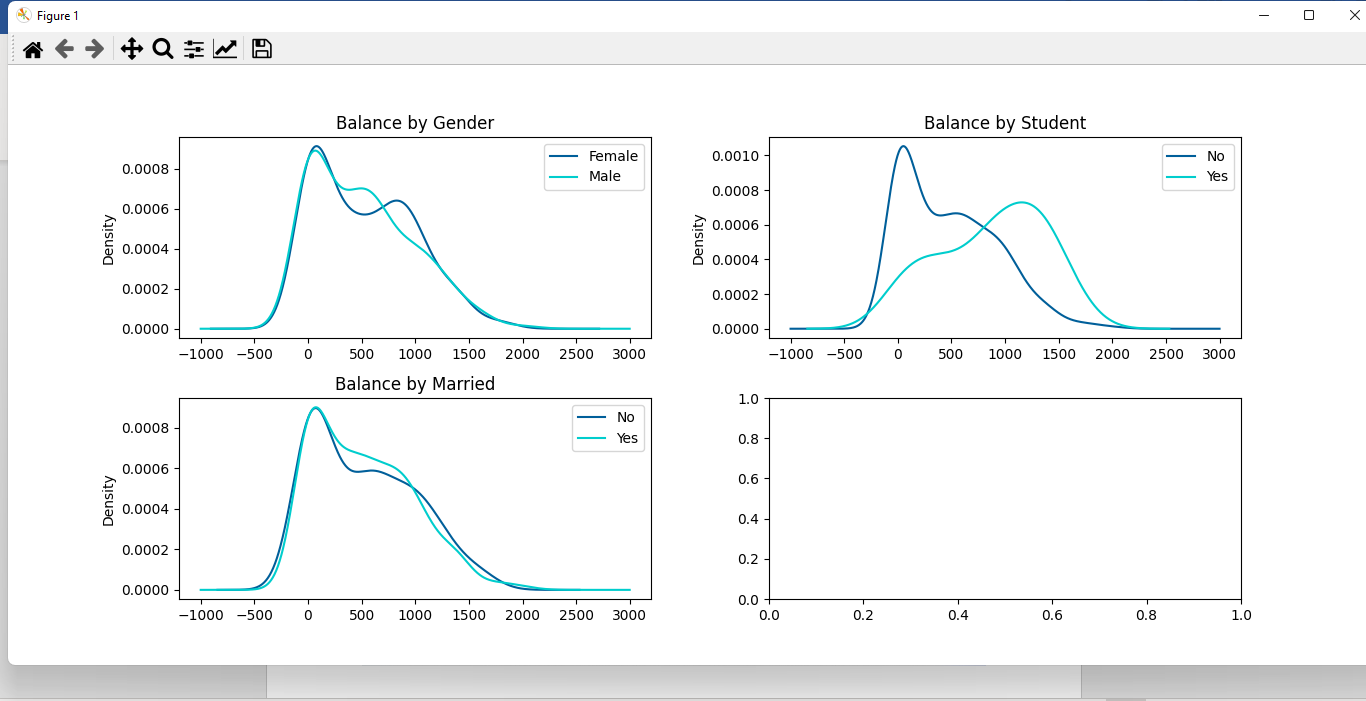
Results











The income-to-rating relationship is, as expected, positive. One possible explanation is that a credit rating is a number provided to people based on their creditworthiness, which includes their personal income.

Students also have lower income values than non-students, as indicated earlier

However, the Income does not increase as one gets older, and students' ages do not differ considerably from non-students'. This finding could make us wonder about the accuracy of the data, or it could prompt us to look for other characteristics of this population to expand our research. Due to the simulated nature of this dataset, more investigations are not possible, but we expect a stronger association between age, student, and income on another population.

We'll look at students and non-students separately now to see how they compare in terms of balance and income. Only non-zero Balances will be considered because the behavior of Balance is of interest to us.

