# Data Summary

Approximately 30,000 Credit One customers’ balance history was reviewed through an Exploratory Data Analysis (EDA) using the BASIR Data Science Framework. The following parameters were studied for correlation: Credit amount given, gender, education, marital status, age, past 6 month payment status, past 6 month bill statement, past 6 months payment history, and Client’s behavior (defaulting on loan or not). The goal of this EDA is to determine if any biographical or balance history impacted a Client’s likelihood to default on their loan.

# Did you learn anything of potential business value from this analysis?

When evaluating a data science EDA, it is difficult to know where to start and where to stop. From this analysis I learned to create a **working process** to perform the EDA. I started with the correlation and the covariance tables to determine which features appeared to be most important. The features with the highest correlation helped guide which visuals I should focus on.

# What are the main lessons you've learned from this experience?

1. You can spend an endless amount of time restructuring the data without a plan. It’s best to focus on the business question as detailed in the BADIR Data Science Framework.
2. I learned how to extract data from an SQL database and to export the data into a CSV file.
3. I learned new data frame operations that allowed me to perform mathematical operations with features.
4. I learned how to create new data frames that would allow me to transform the data into numerical or Boolean features to form additional analyses and insights into the customer’s behavior.
5. I learned to make a table with Pandas Crosstab and to group various features together.

# What recommendations would you give to Guido regarding your findings?

It was difficult to see any strong correlations between any of the features and the Client’s Default Behavior. However, the following insights were seen regarding customer’s biographical data and behavior:

1. Most loans are provided to females (60%), and to persons that are either married (45%) or single (53%). Males had slightly more percentage of defaults (24%) on their loans compared to females (21%).
2. Chart, histogram

   Description automatically generatedIn terms of education, the greatest percentage to default on loans are high school (25%) > university (24%) > graduate school (19%) > other (7%).
3. The greatest number of loans appear to be for customers that were given lower credit amounts (as shown in the figure on the right). This appeared to relate to the greatest amount of people that defaulted on their loans. **My recommendation is to review creating a minimum loan amount to help reduce the number of defaulted loans.**

Chart, scatter chart

Description automatically generated

1. There appeared to be a consistent positive linear correlation around one month’s bill amount and the subsequent month’s bill amount (as shown in figure on the left). That is, if a customer had a high bill in one month they were likely to have a high bill in the months following. **My recommendation is to review if there could be an early pattern forming to determine the client’s likelihood to default on their loan.**