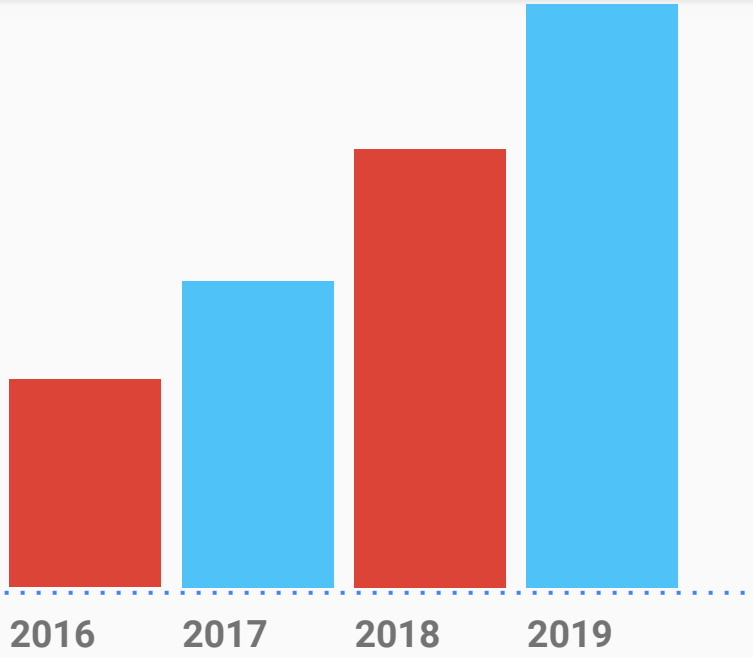
Understanding Customer Defaults

Blackwell Electronics/Credit One

By: Melissa Alemán

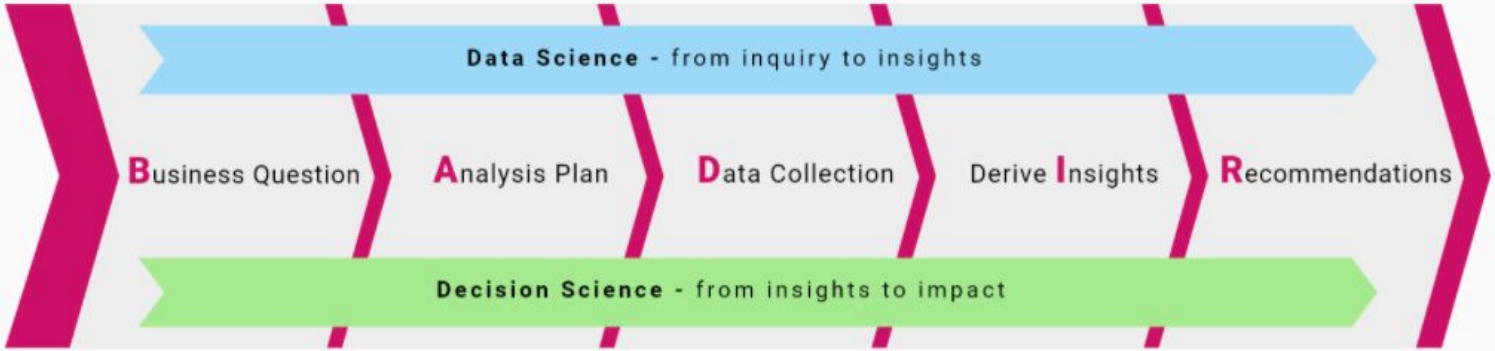
The problem

There has been an increase in customers who have defaulted in the last few years posing major loss of revenue for Blackwell



This process will use BADIR framework

What is BADIR?



Find the real and actionable Business question

Formulate a hypothesis-driven Analysis plan

Collect relevant Data based on the Analysis Plan

Derive Insights using machine learning & statistics

Drive KPI's with actionable Recommendation

GOAL: In an effort to mitigate loss of revenue, Blackwell must understand which customers are likelier to default to be able to better set forth credit limits

Analysis Goals

- Understand how many customers have defaulted or will default based upon data
- Explore how much credit customers should be allowed to use or at minimum if they should be approved for credit
- Utilizing default as the dependent variable, review age, gender, education level and marital status to see if these have bearings on customer default rate

Blackwell Customer Credit Plan

Project Plan

Process the Data

- Gain access to SQL
- Query information: Import data to Python
- Encrypt with password for security protection
- Eliminate errors, null, unneccessary or missing values
- Eliminate unique identifiers since this may pose a problem for machine learning in the future

Explore the Data

- Perform scientific computing by utilizing python, numpy, matplotlib, seaborn, pandas profiling
- Explore correlations via correlation matrix and scatter plots
- Discretize data since credit is on a continuous number line
- Test hypothesis: If default is the dependent variable, might other variables have a bearing on whether a customer will default?
- Test gender, age, education, marital status against default variables

Peform In-Depth Analysis

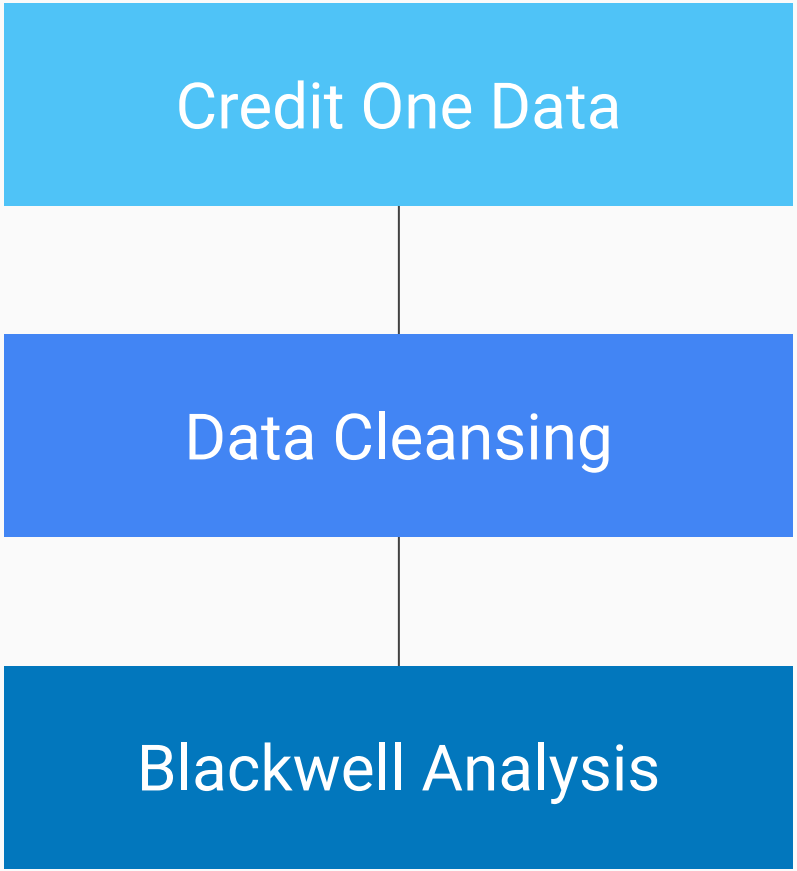
- Utilize machine learning to make predictions based upon hypothesis
- Build models to compare the results to ensure high accuracy
- Potentially train data to use Decision Tree Classifier and Random Forest Classifier

Communicate Results

- Provide recommendations presentation to Guido and management with Blackwell
- Explore options to share the results with Credit One to ensure that customer default rate decreases and changes are made in the future

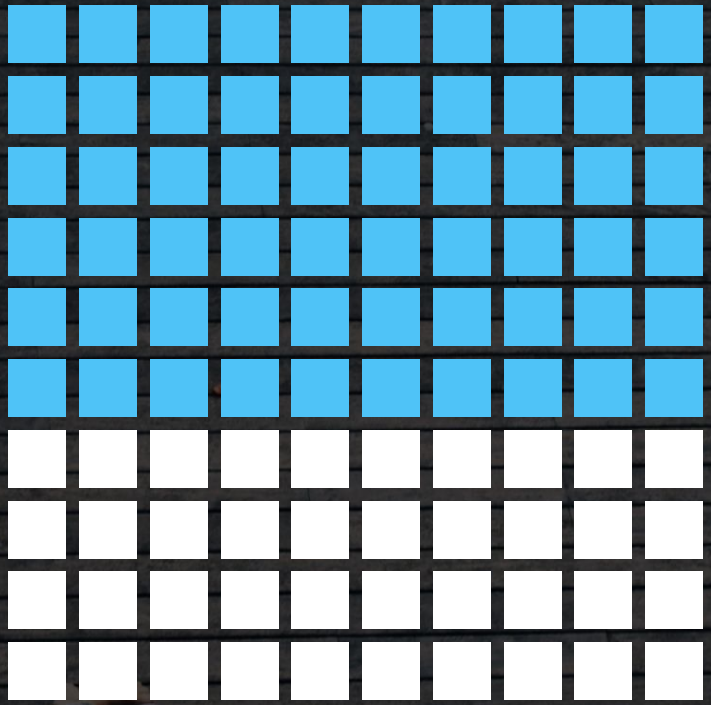
Data Collection

Data will be collected from MySQL and encrypted to assure customer privacy



Insights

- Data cleansing is of supreme priority since many errors can be gleaned at first look
- The dependent variable is likely ‘default’ which will better inform the exploratory data analysis process
- Explore the treatment of non-numeric values
- Opportunities to explore covariance



Action Items for Melissa:

Provide the Blackwell team with report to include the following:

- ***Background***
- ***Recommendations***
- ***Key insights***
- ***Next steps for credit one***