Utilizing Quantitative Methods for Anti-money Laundering (AML)

MSFE Practicum: BMO Harris Bank

November 28, 2016

Utsav Adhikari

Shenyuan Chen

Deepa Sathaye

Xiaoyue Sun

Zhan Zhang

Abstract

The goal of the BMO Practicum was to build working knowledge on the current methodologies used by banks to detect money laundering activities in both customer and credit card transaction data as well as use statistical learning to build a predictive model. Because of recent regulations, failure from banks to meet current anti-money laundering detection guidelines results in severe punishments as well as socioeconomic and global security repercussions. Overall, the group developed three risk rules to create a risk model utilizing BigQuery and Python as well as created a predictive models using logistic regression and random forest classifiers to classify high risk customers based upon data provided.

Table of Contents

Introduction

Money laundering dates back to ancient times and has evolved with the development of the monetary system and technology. With new government and banking AML techniques developing, money launderers are also becoming more sophisticated. In addition to traditional money-laundering methods, such as direct small-amount of cash deposits or wiring through checking accounts, money launderers are also taking advantage of modern financial tools such as currency exchanges and third party processers. Because of this, understanding and implementing the current procedures of AML and working to develop new techniques for fraud detection is important. While many procedures in AML focus on transaction monitoring, people argue that this work is not sufficient and thus AML departments need to embrace new quantitative methods[1]. The AML department of BMO Harris Bank is now building up new strategies by pioneering Quantitative Analytics for risk and compliance in the bank.

This practicum project was sponsored by BMO Harris Bank AML department under the guidance of Ms. Ivana Donevska, who is an expert in AML. The goal of the BMO Practicum was to build working knowledge on the current methodologies used by banks to detect money laundering activities in both customer and credit card transaction data and then to apply these techniques with statistical learning. To accomplish this, we worked with synthetic data stored within Google Cloud Platform, used Google Cloud Platform for BigQuery data processing, and Google DataLab for statistical learning and data visualization.

Our main objectives were:

1. Understand the current landscape and procedures in AML
2. Create detection logic based on industry standards (Wolfsberg Principles)
3. Create flexible risk rules and flag risky transactions
4. Build a Risk Model to capture high risk customers behavior
5. Create predictive models to label high risk customers

Understanding Anti-Money Laundering (AML)

This section focuses on the history and regulations associated with money laundering. To develop the risk model and predictive model, we first had to understand some of the hallmarks of money laundering behavior, impact of money laundering to the society, and methods current AML analysts use to fight fraudulent behavior.

Money laundering is the process of making illegally-gained proceeds (i.e. "dirty money") appear legal (i.e. "clean") according to the United States Treasury Department[2]. There are three stages of money laundering: placement, layering, and integration. Placement describes the behavior of moving illicit funds into the financial market. Layering involves trying to distance the money from the criminal activity. The last step is integration where those illicit funds can now be used in the economy legitimately. Not all money laundering utilizes all these steps but these individual steps represents behaviors that can be quantitatively captured and used to detect money laundering. The figure bellows shows a typical money laundering scheme that incorporates the three steps.

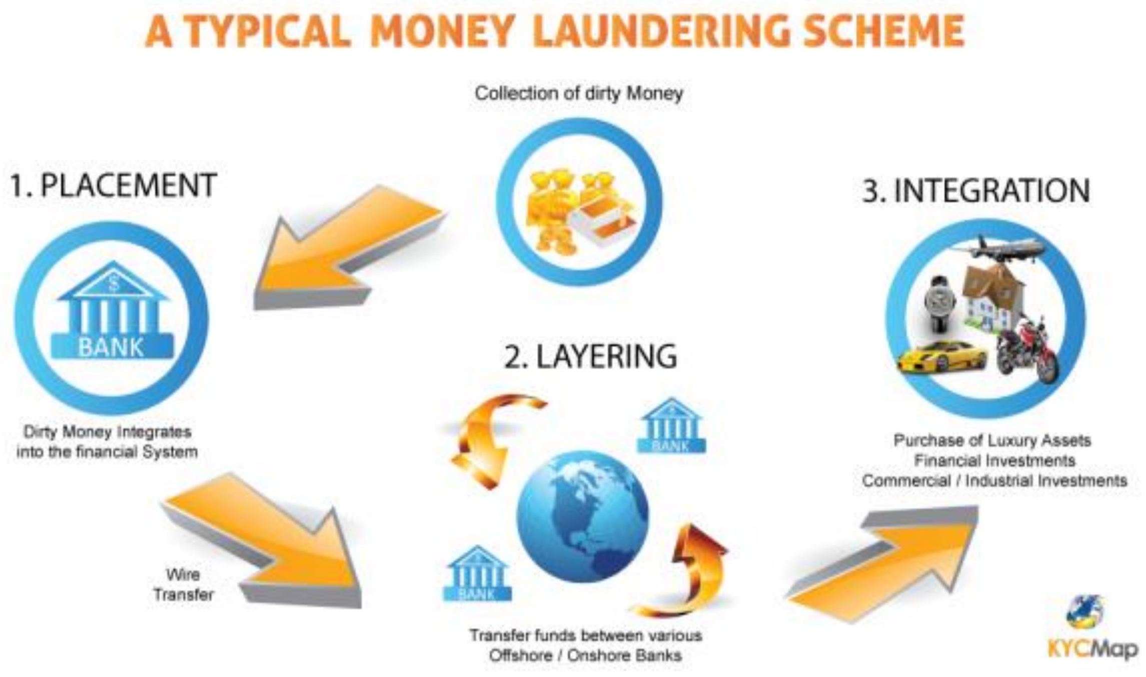
A TYPICAL MONEY LAUNDERING SCHEME
Collection of dirty Money
1. PLACEMENT
BANK
Dirty Money Integrates into the financial System
Wire Transfer
2. LAYERING
BANK
BANK
Transfer funds between various Offshore / Onshore Banks
3. INTEGRATION
Purchase of Luxury Assets
Financial Investments
Commercial / Industrial Investments
KYCMap

Figure 1: Typical money laundering scheme

***source**: http://kycmap.com/what-is-money-laundering/*

Money Laundering has become a more serious issue in recent years. It has economic, political, and social effects[3]. It affects economies because rapid inflow of illegal funds into a country has the potential to harm the purchasing power of a country's currency and cause unintended inflation[3]. Corruption cases involved politically exposed persons (PEP) can create huge political impacts[2]. Also, persistent money laundering allows for expansion of such illegal activities and causes governments to lose money[2].

The statistics are overwhelming regarding the prevalence of money laundering and its effects. According to FATF (Financial Action Task Force) [1], in 2009, the United Nations Office on Drugs and Crime (UNODC) conducted a study to determine the magnitude of illicit funds generated by drug trafficking and organized crimes and to investigate to what extent these funds are laundered. According to UNODC, the estimated amount of money laundered globally in one year is 2 - 5% of global GDP, or $800 billion - $2 trillion in current US dollars.[5] Thus, governments, law enforcement, and banks face greater responsibility than before to detect money laundering.

Prior to 1970, the only charge for money laundering was tax evasion. The **Bank Secrecy Act (BSA)** of 1970 required US financial institutions to assist US government agencies. The **Money Laundering Control Act (1986)** mad money laundering a crime in itself, instead of just an element of another crime. The **Annunzio-Wylie Anti-Money Laundering Act (1992)** strengthened sanctions for BSA violations, required so called "Suspicious Activity Reports" and eliminated previously used "Criminal Referral Forms", required verification and recordkeeping for wire transfers and established the Bank Secrecy Act Advisory Group (BSAAG). The **Money Laundering Suppression Act (1994)** ordered banks to establish their own money-laundering task forces to weed out suspicious activity in their institutions. The **U.S. Patriot Act (2001)** set up mandatory identity checks for U.S. bank patrons and provides resources toward tracking transactions in the underground/alternative banking systems frequented by terrorist money handlers.

Along with these legislations, banks and governments have created international taskforces to combat money laundering. Financial Action Task Force (FATF) is an intergovernmental organization, formed by G7 countries, whose purpose is to develop and promote an international response to combat money laundering. The FATF issued the "40+9 Recommendations" for banks which has become the anti-money-laundering standard[5]. Banks also created their own task force known as the Wolfsberg Group. The Wolfsberg Group, founded by thirteen major banks in 1999, have published several documents that provide AML guidelines to those in the financial industry. Banks, such as BMO Harris, must meet the strict regulations of the government that also closely align with the Wolfsberg Standards. Furthermore, these organizations have had to develop new methods for detection due to the expansion of regulation and the savviness of money launderers.

Data and Tools

For the project, all customer and transaction data was synthetically generated via Python scripts provided by the point of contact on Google Cloud Platform. Open source external libraries, Barnum and Faker, were used to generate the customer and credit card transaction data. Barnum[7] is a library that generates pseudo-random person data. Faker[8] also generates random data used for application testing. Additionally, the scripts included information such as geographic risk and occupational risk based upon reference data. The script created 10,000 customers with around 1.25 million credit card transactions with each customer having about 125 transactions. The customers' information is named 'uber_cust' and the transaction information is named 'cc_trans'. The data was generated for a one-year time frame.

Once the customer data and related credit card transaction data was generated, they were stored in Google Storage. Google Big Query was used for the initial detection logic querying and uses a SQL-like syntax. Python and Google DataLab were used for subsequent risk rule and risk model development as well as visualization of the data. Sklearn which is a machine learning library in python and Matplotlib were used to accomplish these tasks.

There was some data modification done using Google Big Query to add riskier occupation data to the existing customer base. 10% of the customers were updated to be involved with Casino Gambling. Also, customers whose occupation related to money laundering scenarios detailed within the FFIEC BSA/AML Examination Manual[9] were also updated to indicate this behavior. The specifics are detailed below:

1. Mark Gunsmiths as also 'ARMS_MANUFACTURER'
2. Flag 'CASHINTENSTIVE_BUSINESS' such as retail/restaurants/liquor
3. Flag 10% of customer with Casino owners
4. Attach CLIENT_NET_WORTH based upon their transaction data
5. Flag HIGH_NET_WORTH as clients with balance greater than 35,000
6. Flag ATM as 'PRIVATELY_ATM_OPERATOR'
7. Flag Dealership/Car Sales as 'SALES_USED_VEHICLES'

Quantitative Analysis

The section describes the different quantitative analysis done by the group which is broken down into three sections. The end goal was to develop a method for classifying customers as high risk based on their customer attributes and transaction data. Initially, we spent time with pre-classified data and our goal was to determine what decision rules we could create to capture this fraudulent behavior. The rules were based loosely upon the Wolfsberg Principles which are a set of unified rules that banks have come up with classify money laundering behavior. The full scenario descriptions are available in the Appendix.1. The second part of our work was to start with an unclassified dataset and use what we had learned to capture money laundering behavior. With this work, we were mimicking what AML analysts have to do on an everyday basis by looking at the total population and then try to capture risky clients in a reasonable way. The last part of the analysis involved combining risk probabilities from the previous sections to generate a predictive model to classify high risk customers.

Quality Assurance

The purpose of Quality Assurance (QA) in anti-money laundering is to verify the quality of the original labeling in the customer dataset. We have three parts in this section: test the correctness of the customers' original labeling, analyze the result of the detection query on all customers which include vertical analysis and horizontal analysis, and finally building a risk rule which will be applied in customer profiling and model building in the later part of this paper.

Test the correctness of the customers' original labels

We were provided with a table called 'cc_trans' that included all the transaction information about the customers. These transactions included transaction type, merchants the customer traded with, transaction date, and so on. In this table, there is one column called 'USE CASE' which corresponded to the scenarios described in Wolfsberg Principles (Appendix.1). Under this column, customers are labeled as Red, Yellow or Green for one specific scenario. All the transactions of one customer has only one label.

The first part of QA is the process in which we used SQL (Structured Query Language) on Google BigQuery to query the transactions of the customers to test whether the ‘USE CASE’ labels lined up with the scenario descriptions. For example, there was a scenario that looked at the percentage of high risk payment types, such as cash, wire or ACH, for each customer within the transaction data. If a customer had over 50% of these payments, they were labeled "Red" which indicated the riskiest behavior. If a customer had between 30% and 50% of these payments, they were of medium risk and customers who had below 30% were considered lowest risk. Based on this rule, we then used the procedure below to calculate the "Correctly Labeled Rate". The purpose of the "Correctly Labeled Rate" was to allow us to precisely determine whether our detection logic captured the already marked customers.

1. Select the customers that are labeled as ‘USE CASE 2 Red’, ‘USE CASE 2 Yellow’ or ‘USE CASE 2 Green’ from the transaction data. For each customer, count the total number of transactions whose transaction type is cash payment, wire payment or ACH payment and divided this number by the total number of transactions.
2. Apply the threshold percentages to the test ratio to assign a new ‘USE CASE 2 Red’, ‘USE CASE 2 Yellow’ or ‘USE CASE 2 Green’ for each customer. If the new label is the same as the old one, this customer is correctly labeled.
3. For each color group in USE CASE 2, the number of customers correctly labeled divided by the total number of customers in that group yields the “Correctly Labeled Rate”

The following table is the output of the QA query for scenario 2 after using the above procedure. The full output table is available in the Appendix.2. Most of the customers are labeled correctly under this scenario.

Scenario	Actual Labeled based on QA	Originally Labelled	Correctly Labeled Rate
Use Case 2 - Red	39	41	95.10%
Use Case 2 - Yellow	58	70	82.90%
Use Case 2 - Green	621	680	91.30%

Table 1. Correctly Labeled Rate for Scenario 2 - High risk payments by a customer

The code for this section is found in UseCaseQA1.sql.

Vertical analysis and Horizontal analysis

The second part of QA applied the QA procedure to all the customers irrespective of their original labeling to get a label under each scenario. This means that each customer will be marked red, yellow, or green for each scenario. In this part we developed two methods of analysis. Vertical analysis focused on the percentage of red, yellow and green customers in each scenario for the entire customer dataset. As an example, for Use Case 2 there was 20.43% red, 61.14% yellow, and 18.43% green. Horizontal analysis focused on the total number red, yellow, and green labels per customer. As an example, with these labels there was one customer to get six red labels, zero yellow, and six green labels.

The following table is the output of the vertical analysis. From the table, we see the distribution of each group in the different use cases. It helped determine if the decision threshold rules were too strict or not. In this analysis, USE CASE 2, 20.43% of the customers are labeled as red which as mentioned earlier is above the target which was 5% of the overall population. We have to think about if 50% (mentioned in first part of QA) is too strict for this group of customers. On the other hand, USE CASE 1.0 catch 0 red from the customer, we may have to think if the rules applied there might be too loose. This is making an assumption that our sample dataset is uniformly distributed in terms of risky behavior.

Use Case	Actual Labeled based on QA	Percentage in Entire Customer dataset	Use Case	Actual Labeled based on QA	Percentage in Entire Customer dataset
Red 1.0	0	0.00%	Red 6	34	0.34%
Yellow 1.0	85	0.86%	Yellow 6	131	1.33%
Green 1.0	9796	99.14%	Green 6	9716	98.33%
Red 1.1	723	7.32%	Red 7	1639	16.59%
Yellow 1.1	2493	25.23%	Yellow 7	3597	36.40%
Green 1.1	6665	67.45%	Green 7	4645	47.01%
Red 2	2019	20.43%	Red 8	67	0.68%
Yellow 2	6041	61.14%	Yellow 8	26	0.26%
Green 2	1821	18.43%	Green 8	9788	99.06%
Red 3	71	0.72%	Red 9	5176	52.38%

Use Case	Actual Labeled based on QA	Percentage in Entire Customer dataset	Use Case	Actual Labeled based on QA	Percentage in Entire Customer dataset
Yellow 3	390	3.95%	Yellow 9	19	0.19%
Green 3	9420	95.33%	Green 9	4686	47.42%
Red 4	15	0.15%	Red 10	144	1.46%
Yellow 4	1791	18.13%	Yellow 10	44	0.45%
Green 4	8075	81.72%	Green 10	9693	98.10%
Red 5	29	0.29%	Red 11	6849	69.31%
Yellow 5	12	0.12%	Yellow 11	689	6.97%
Green 5	9840	99.59%	Green 11	2343	23.71%

Table 2. Vertical Analysis depicting customer distribution for each Use Case

In horizontal analysis, we looked at the number of red, yellow, and green labels we had marked on each customer. A customer could be labeled ’Red’ for some Use Cases/Scenarios, but also labeled ‘Green’ in other Use Cases. This analysis was relevant because further risk model building used this analysis as a core foundation. The premise was that customers with more risk labels (high risk behavior triggers) had a higher probability of being involved with money laundering. Table 3 aggregates this information by counting the number of customers that have specific number of use case risk group. There is only one customer that have six red labels. Eleven of them have five yellow labels. We ultimately used this table to narrow down the high-risk customer group for further detection.

Number of Use Case Labels	Red	Yellow	Green
0	676	905	0
1	3243	4127	0
2	4484	3531	0
3	1374	1180	0
4	88	127	0
5	15	11	5
6	1	0	116

Number of Use Case Labels	Red	Yellow	Green
7	0	0	745
8	0	0	2915
9	0	0	4049
10	0	0	1783
11	0	0	269
12	0	0	1

Table 3. Horizontal Analysis aggregating Use Case risk groups

The code for this section is found in UseCaseQA2.sql.

Transaction Monitoring and Risk Rule 1:

From the vertical analysis, we observed the distribution of red, yellow and green flags are different from scenario to scenario. To make the data more valuable for setting risk rule and risk model building, we changed the threshold of each scenario to make the distribution of red, yellow and green flags be 5%, 8% and 87% respectively. We analyzed the transactions of each customer under twelve different scenarios of Wolfsberg Principle during the Quality Assurance analysis. Under each scenario a customer will either get a red, yellow or green flag. Therefore, the number of flags received by each customer is twelve. Based on number of red, yellow and green flags, each customer is assigned to a high-risk group, medium risk group or low risk group. This is our first step in getting a reasonable number of high risk customers and will call this as Risk Rule 1. Since, we are only concerned with the high-risk customers, only the criterion for high risk group is listed below.

Risk Rule 1:

High Risk criterion: Customer with at least three red flags and at least one yellow flag.

Based on Risk Rule 1, there are 529 high risk customers. This sample of high risk customers is purely based on the transaction data. In the subsequent analysis steps, we will use these 529 customers for further analysis and try to find fewer number of high risk customers using another layer of data.

Customer Profiling

After narrowing down the number of customers through transaction monitoring, we looked at the customer's information to get feasible number of high risk customers. Through Know Your Customers (KYC), a process associated with customer identification, customers were evaluated based on business risk, geographical locations, occupation, etc. Specifically, for KYC, every customer attributes present in the customer dataset were carefully analyzed. After a careful evaluation, we defined a couple of risk rules to select customer attributes that are associated with high risk customers and finally use them to build a risk model. The code for the customer profiling can be found in CustomerProfileCode.ipynb.

Risk Rule 2:

There are sixty-six customer attributes in the customer dataset. Some of the customer attributes are unique such as name, social security number, email address, credit card number and so on. As our goal is to search for patterns or trends in the data, we started dropping the customer attributes that were unique to the customers.

As we gained a better understanding of AML and how each attribute could be associated with money laundering, we dropped more customer attributes such as location data, consent sharing, dependents count, etc. Eventually, we chose nineteen customer attributes that were be considered for the construction of the risk model. They are listed below:

1. ARMS MANUFACTURER
2. SALES USED VEHICLES
3. PRIVATELY ATM OPERATOR
4. NONREGULATED FINANCIAL INSTITUTION
5. SAR
6. DEMARKET FLAG
7. EXCHANGE CURRENCY
8. MEDICAL MARIJUANA DISPENSARY
9. DIGITAL PM OPERATOR
10. CASINO GAMBLING
11. PEP
12. COMPLEX HI VEHICLE
13. CASHINTENSIVE BUSINESS
14. THIRD PARTY PAYMENT PROCESSOR
15. INTERNET GAMBLING
16. DEALER PRECIOUS METAL
17. AUCTION
18. EMBASSY CONSULATE
19. HIGH NET WORTH

In the next layer of customer profiling, both the customer attributes and transaction data are considered. We look at how the customer attributes selected from the initial screening correlates with the 529 high risk customers purely based on transaction data (i.e. high risk customers generated from Risk Rule 1). To calculate this correlation, we count the number of customers from each customer attribute that are present in the sample from Risk Rule 1. For each customer attribute, the percentage of customers that are present in the sample from Risk Rule 1 is calculated using the following formula:

$$\frac{No. of\ customers\ with\ a\ particular\ customer\ attribute\ in\ the\ sample\ from\ Risk\ Rule\ 1}{Total\ number\ of\ customers\ with\ that\ particular\ customer\ attribute}\ X\ 100\% \qquad (1)$$

Formula 1. Ratio that indicates the relationship of a customer attribute with high risk transaction behavior

For instance, there are 31 customers who are considered to have high net worth in the given dataset. High net worth for this sample is defined as greater than $35,000 in transaction amounts. Out of the 31 customers, six are present in the high risk customer group generated from transaction monitoring. Therefore,

$$Percentage\ of\ high\ net\ worth\ customers\ in\ the\ sample\ from\ Risk\ Rule\ 1 = \frac{6}{31}\ X\ 100\%$$

$$= 19.36\%$$

Note that, 529 is approximately five percent of the total number of customers. Hence, for each customer attribute, we expect to see 5% of the customers from that particular customer attribute to be present in the high risk group from Risk Rule 1. Again, the assumption is that we have a representative population sample that is uniformly distributed. If this percentage is 6% or greater, the customer attribute is kept for further analysis otherwise the customer attribute is dropped and this will be called **Risk Rule 2**. Figure 1 below shows a bar chart of the nineteen customer attributes and their respective percentages:

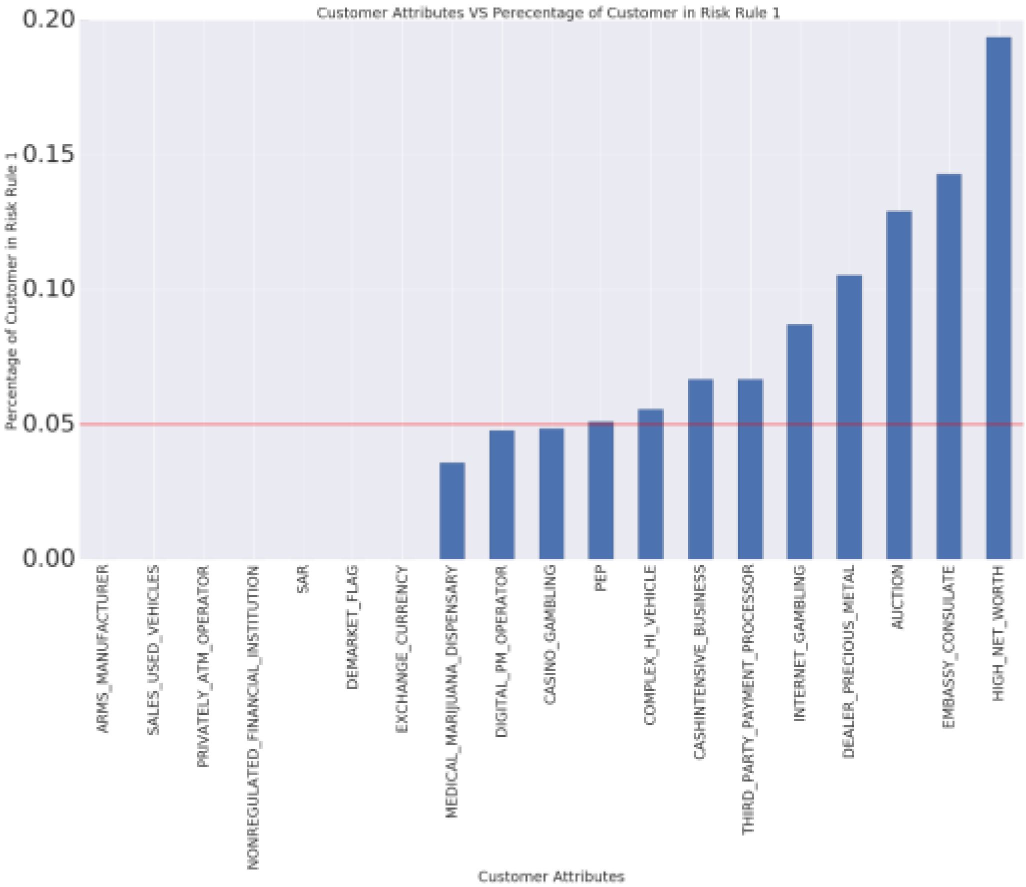

Figure 2. Relationship of Risk Rule 2 customer attributes to Risk Rule 1 customers

Based on the risk rule described above, 7 customer attributes are kept for the further analysis. These 7 customer attributes are listed with their ratio below in table 1:

Customer Attribute	**Percentage**
CASHINTENSIVE BUSINESS	6.77%
THIRD PARTY PAYMENT PROCESSOR	6.77%
INTERNET GAMBLING	8.77%
DEALER PRECIOUS METAL	11.05%
AUCTION	12.90%
EMBASSY CONSULATE	14.29%
HIGH NET WORTH	19.35%

Table 4. High correlated customer attributes with high risk transaction data

Risk Rule 3:

At this point, we have high risk customers based on transaction data and we have identified customer attributes that can be associated with high risk customers. However, there are few other customer attributes that needs special attention. From the BSA/AML Examination Manual, there were other customer attributes that are high risk that our dataset did not represent and thus we manually added these characteristics to build a more robust model, specifically: customer dealer with guns, marijuana dispensaries, and currency exchange facilities. Hence, following three attributes will be added to our model

1. Arms Manufacturer
2. Medical Marijuana Dispensary
3. Exchange Currency

Risk Rule 3 simply incorporates the above customer attributes based on available qualitative information.

Risk Model

Finally, a risk model was designed using the three risk rules. For each risk rule, a range of score is assigned to the customers. The details on how the score is assigned in each risk rule is given below:

Risk Score for Risk Rule 1

From the previous section in Quality Assurance analysis, we have red, yellow, and green flags for each customer. As there are twelve scenarios, there will be twelve flags for each customer. For each red, yellow and green flag a customer receives; two, one and zero points is given respectively to the customer.

Risk Score for Risk Rule 2

For the seven customer attributes selected in Risk Rule 2, we give them points based on the ratio we calculated above using formula 1. If the percentage of customers of a particular customer attribute who are in the sample from Risk Rule 1 is greater than 10%, then we give the customer four points. If this percentage is in between 5% to 10%, we give the customer two points and if it is below 5%, zero points is given to the customer. 10% is almost double the expected value which is why we allocated four points.

Risk Score for Risk Rule 3

All the customer selected based on the attributes in Risk Rule 3 will receive two points.

In summary, points will be assigned in the following manner:

Customer Attributes	**Points**
HIGH NET WORTH	4
EMBASSY CONSULATE	4
AUCTION	4
DEALER PRECIOUS METAL	4
INTERNET GAMBLING	2
THIRD PARTY PAYMENT PROCESSOR	2
CASHINTENSIVE BUSINESS	2
GUNS MANUFACTURER	2
MEDICAL MARIJUANA DISPENSARY	2
EXCHANGE CURRENCY	2
OTHER	0

Table 5. Scoring criteria for customer data

Flag	**Points**
Red	2
Green	1
Yellow	0

Table 6. Scoring criteria for transaction data

Using the scoring system above, we calculate a risk score for each customer using the following formula:

$$Risk\ Score = \sum Transaction\ Scores + \sum Attributes\ Scores \qquad (2)$$

Formula 2: Risk Score

With the formula in 2, a scoring histogram was computed for all the customers:

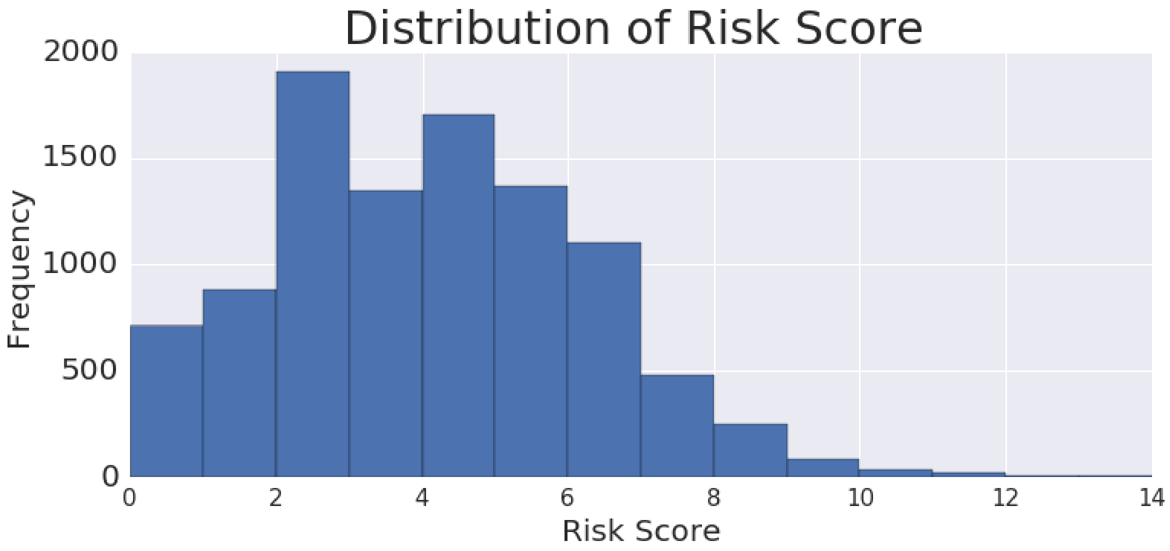

Figure 3. Histogram of Risk Score using Risk Rule 3

The mean value of the distribution is 3.61 with a standard deviation of 2.14. Minimum value in the distribution is 0 and maximum value is 14. The frequency counts for the scores is as follows:

Risk Scores	**Percentage**
[10,14]	0.61%
[7,9]	8.20%
[4-6]	42.25%
[1-3]	41.75%
[0]	7.19%
Total	100%

Table 7. Risk Score range and the percent of customers in the risk score range

The above risk rules and risk model captures a segment of the high risk population. We focused on transaction scenario and attributes that we understood and could substantiate on why they were high risk with the current dataset. This risk model flags the customer for more frequent surveillance. Customers with a risk score greater than or equal to 10 would be the riskiest customers which is 1% of the population. The rest of the 529 customers still be the top 5% of high risk clients that would require frequent surveillance as well. Thus, our group built a risk model using detection logic based on the Wolfsberg Principles and KYC to flag the top 1% and 5% of the high risk customer population. Next, we created a predictive model.

Statistical Learning in Anti Money-Laundering

In this section, we are introducing two machine learning models that could detect money-laundering activities. Before moving forward with the details of our models, we will briefly discuss about the predictor variables and the response variable in our data. The second part is a brief introduction of logistic regression and random forest followed by the results of the random forest model and logistic regression classification model. The third part will be potential models that also fit the context of our task.

Task and Data

Our task is to predict the risk flag of a given customer. The customer could be deemed as high risk or low risk according to the risk score calculation from the previous section. In the predictive modeling part, the task remains unchanged. Therefore, the **response** is our prediction of the risk of a given customer, either high or low.

Our predictors come from the detection logic principles from our point of contact at BMO Harris, which is roughly based on the Wolfsberg Principles. In the QA section, we calculated ratios of these detection logics, such as the percentage of number of overpayments out of total payments in a given month. From the twelve cases available from the Wolfsberg principles, we have calculated eight ratios defined by the detection logic, as our **predictors**. Not all twelve were used because not all scenarios had detection logic that had clear ratios that could be compared across customers.

The usage of these predictors makes sense intuitively because in all of the detection logic, higher percentage implies higher transactional risk. In the risk modeling part, higher transactional risk will lead to higher transaction scores and thus, higher customer risk. Therefore, we are assuming that there is a positive correlation between the **predictors** and **response**.

Our dataset includes the ratios of transactions of 10,000 customers. We will select 10% of the 10,000 for the predictive modeling. 20% of the selected has been used as training data and the remaining 80% of the selected are testing data. UseCaseQA3.sql has the SQL code to build the initial tables used by the Machine Learning in AML.ipynb.

Statistical Learning Models

We used two supervised classification methods to predict the risk of a given customer: *Logistic Regression* and *Random Forest*. We will start with logistic regression.

Logistic Regression

To restate the problem, the response **Risk Flag** falls into one of two categories, **High** or **Low**. Simply, we split the data with a binary classification: 1 for high, 0 for low. Logistic regression fits our task well because it models the probability that the response belongs to a particular category. In this case, logistic regression models the probability that a customer is high risk, given the ratios that we calculated from transaction table.

The value of high risk probability will range between 0 and 1. We claim in our modeling that a calculated probability larger than 0.5 implies high risk for the customer.

The logistic function is defined as such:

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

Formula 3. Logistic Function

Where X is the predictor.

In the case of predicting a binary response with p predictors, we extend the above model to

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}$$

Formula 4. Extension of formula 3

We picked the 500 high risk customers based on the score assigned to them in risk score section. When we pick 500 high risk customers, the remaining 9500 will be low risk model and thus, creating an imbalance between the number of high and low risk customers. To combat the issue with having imbalanced classes we only considered 500 customers from the low risk population. In total, we have 1000 customers for training and testing. 200 of them are used for training and 800 of them are used for testing. The results are as follows:

Training Accuracy: .835

Test Accuracy: .8275

Delta (difference between training and test accuracy): .0075

	Precision	Recall	F1-Score	Support
Low Risk (0)	.94	.70	.80	407
High Risk (1)	.76	.96	.84	393
Average	.85	.83	.82	800

Table 8: Classification Report for the Logistic Regression

Logistic regression model has a training accuracy of 0.84, implying that in training dataset it is able to predict correctly 84% of the training data. In the testing set, it was correct 82.6% of the time. The difference between these two, the delta, is low. It implies that the model hasn’t been over fitted.

In the classification report, the number in the first row, first column implies that 94% of the low risk customer predictions are actually correctly predicted. The number in the second row, first column implies that 76% of the high risk customer predictions are actually correctly predicted. The numbers in the last column are the actual number of low risk and high risk customers.

The confusion matrix tells us how to calculate the ratio in classification report. 286 is the number of low risk customers that has been correctly predicted (true positive). 17 is the number of high risk customers that has been classified as low risk. The classification method predicted 286+17 = 303 customers as low risk. Thus 286/302 = 94% of the predictions are correct. The high risk prediction follows the same logic.

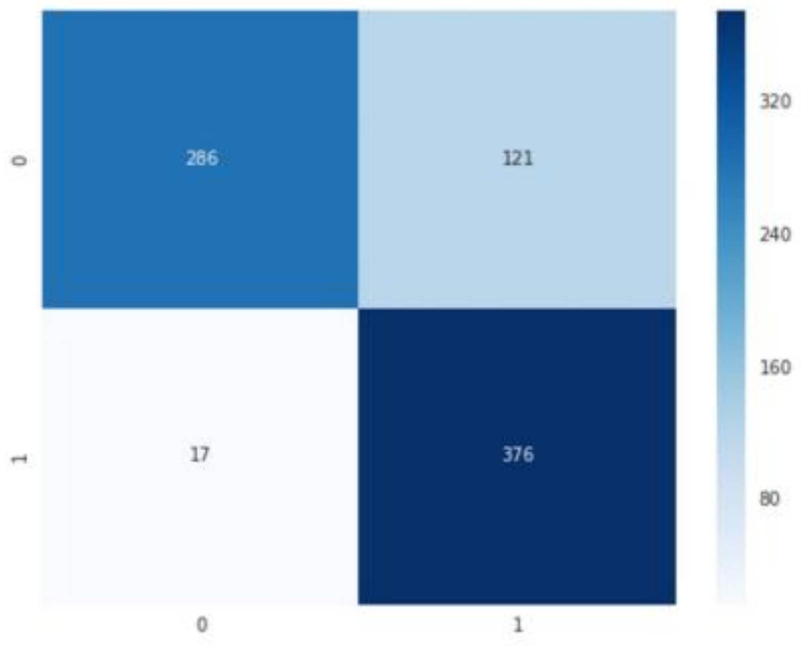

Figure 4. Logistic Regression confusion matrix

The heat map shows us that the false positive (low risk customers that have been classified as high risk) is 121, is a little bit high, implying that there are more alerts than necessary. A high false positive rate will waste banks' resources on checking low risk customers.

Random Forest

The other predictive model we have used is *Random Forest. Random Forest* is a tree based method. The tree based method splits predictor space, usually high dimensional, into different segments. We make predictions of each customer by looking at which segment do its predictors fall into.

We could build one tree out of our data. However, a decision tree usually suffers from overfitting the data and does not give an accurate prediction. People usually construct multiple trees and collect the predictions from all of the trees. The final prediction could be the average of single prediction for a regression problem or the majority vote for a classification problem.

Each time we build a tree, we randomly select 2/3 of the data and train one single tree. We repeat this process as many times as we want (n times) and make a prediction out of the outcomes of each tree. We call this process *bagging***.** A random forest differs from bagging in the construction of a single tree. In a random forest model, each time we split the predictor space, we only consider a random subset of the predictors. This strategy reduces the correlation between one tree and another. In this way, each tree will be very different from another tree and collectively will give a more unbiased prediction.

In our implementation, we used 20 trees to make the prediction to reach optimal test accuracy. Rest of the parameters follow the default of **RandomForestClassifier** function of sklearn. The results for a random forest are as follows:

Training Accuracy: 1

Test Accuracy: 0.8725

Delta: .1275

	Precision	Recall	F1=score	Support
Low Risk (0)	.91	.83	.87	407
High Risk (1)	.84	.92	.88	393
Average	.88	.87	.87	800

Table 9: Results from Random Forest Classifier

Random Forest model has a training accuracy of 1; implying that in the training dataset, it is able to correctly predict all of the training data. In the testing set, it was correct 87% of the time. The delta in this model is 0.1275 which is higher than the delta in the logistic regression. It implies that the model has potentially been over trained.

In the classification report, the number in the first row, first column implies that 91% of the low risk customer predictions are actually correctly predicted. The number in the second row, first column implies that 84% of the high risk customer predictions are actually correctly predicted. The numbers in the last column are the actual number of low risk and high risk customers.

The confusion matrix tells us how to calculate the ratios in classification report. 338 is the number of low risk customers that has been correctly predicted. 33 is the number of high risk customers that has been classified as low risk. The classification method predicted 338+33 = 371 customers as low risk. Thus, 338/371 = 91% of the predictions are correct. The high risk prediction follows the same logic.

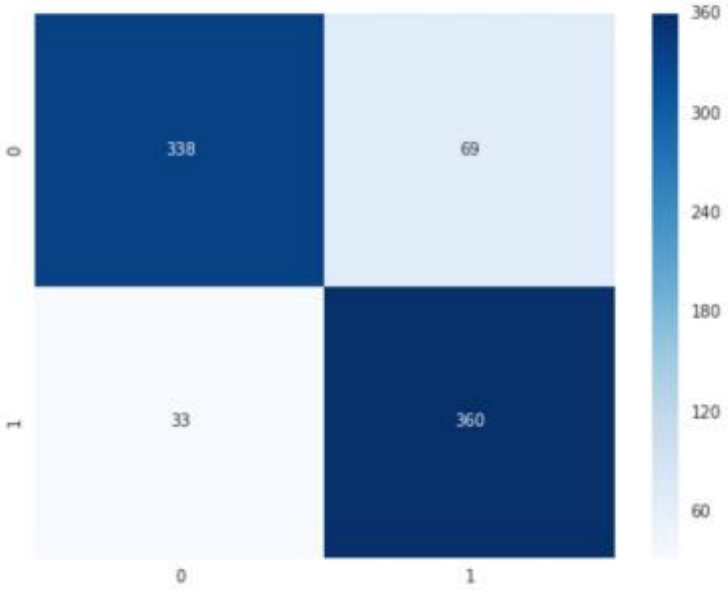

Figure 4. Random Forest confusion matrix

The heat map shows us that the false positive (low risk customers that have been classified as high risk) is 69, is moderate, implying that there are not as many false alerts.

Evaluation and Potential Models

There are other models we can use based on our task. We have many numerical predictors in this section. In previous customer profiling and risk model sections, we have calculated risk score for each customer. Combining these two, we could build a relationship between risk scores and the predictors in this section. In such case, we can use a multiple linear regression model to predict the risk score of a customer by the ratios defined in the detection logic. Finally, we could use the Goodness of Fit test to evaluate and adjust the linear model.

At the same time, taking advantages of the decision tree’s ability to handle qualitative predictors, we could add categorical predictors from customer profiling to our Random Forest model. Some of the attributes of a customer can be built into the Random Forest to predict the risk flag or the risk score.

Conclusion

Overall through this project, the group learned a lot of new quantitative technologies as well as techniques used to detect money laundering in industry. Creating the detection system required thinking not just about data but also the human resource component. Any customers flagged as high risk still require a human AML analyst to do further analysis. Thus, while marking all customers as high risk would ensure no fraudulent behavior was missed, that method is not sustainable for the bank. For the QA, the objective was twofold. We spent learning the technology of BigQuery and SQL. Secondly, we used these practice sessions as inspiration into our own risk model. From there the group used alternative models to detect money laundering activities. Through the different assignments and work, we gained a stronger understanding of the intricacies of money laundering detection. The decision rules from the initial QA were not sufficient to detect a feasible number of high risk clients. Thus we incorporated customer profiling, transaction monitoring, and money laundering guidelines to detect high risk clients more granular and choose just 5% who could be feasibly have in depth monitoring. Finally, we built a predictive model using the probabilities from the transaction monitoring.

References

1. Stabile, Carol. "Machine Learning: Advancing AML Technology to Identify Enterprise Risk." *ACAMS Today* 14.2 (2015): 1-5. http://www.safe-banking.com/DownloadDocument.ashx? documentID=114. 28 Nov. 2016.
2. Goldberg, Warren. "Anti-Money Laundering Laws." *Mortgage Wealth Advisors*. Mortgage Wealth Advisors, 06 Apr. 2014. http://www.mortgagewealthadvisors.com/2014/04/06/anti-money-laundering-laws-affect-mortgage-borrowers/. 28 Nov. 2016.
3. MacDowell, John and Novis, Gary. "Consequences of Money Laundering and Financial Crime." Economic Perspectives, Electronic Journal of the U.S. Department of State 6.2 (May 2001): 1-10. https://www.ait.org.tw/infousa/zhtw/DOCS/ijee0501.pdf. 27 Nov. 2016.
4. Farooqi, Imran ."Anti-Money Laundering." *PwC*. PwC, Jan. 2015. http://www.pwc.com/gx/en/services/advisory/consulting/forensics/economic-crime-survey/anti-money-laundering.html. 28 Nov. 2016.
5. "Money Laundering - Financial Action Task Force (FATF)." *FATF*. Jan. 2015. http://www.fatf-gafi.org/faq/moneylaundering/. 28 Nov. 2016.
6. "United Nations Office on Drugs and Crime." *Money-Laundering and Globalization*. UNODC, Jan. 2016. https://www.unodc.org/unodc/en/money-laundering/globalization.html. 28 Nov. 2016.
7. Moffitt, Chris. Barnum. https://pypi.python.org/pypi/barnum/0.5.1. Web. GPL. https://github.com/chris1610/barnum-proj
8. Joke2k. https://pypi.python.org/pypi/fake-factory. Web. MIT License. http://github.com/joke2k/faker
9. "Online Manual - BSA InfoBase - FFIEC." *Online Manual - BSA InfoBase*. FFIEC, Sept. 2015. https://www.ffiec.gov/bsa_aml_infobase/pages_manual/manual_online.htm>.. 28 Nov. 2016.

Appendix

1. Credit Card Scenarios provided by BMO Harris to build the initial detection logic. These scenario capture some of the Wolfsberg Principles. The below detection logic is to help determine data requirements for a probabilistic model. We'll use the Wolfsberg Principles as a guide.

Detection Logic

1. Overpayments and Refunds to credit card balances

- Scenario Description
  - This scenario purports to detect AML risks in which a money launderer may be using dirty money to overpay credit card balance and get a refund of clean money from the credit card issuer.
  - Overpayment = (Payment made) > (Payment Owed)

  In the example below, there are two overpayments that resulted in negative balance that'd be refunded to the customer (e.g. in 90 days).

Transaction Date	Transaction Type	Transaction Amount ($)	Credit Card Balance
1/1/15	Purchase	100	100
1/15/15	Purchase	100	200
1/23/15	Purchase	100	300
1/31/15	**Payment**	**-600**	**-300 (to be refunded to customer)**
2/10/15	Purchase	100	-200
2/19/15	Purchase	200	0

2/25/15	Purchase	300	300
2/28/15	**Payment**	**-700**	**-400 (to be refunded to customer)**

- Threshold for Overpayment
    - **Red** where # of overpayments is 70% or greater in the # of months in the time period covered by the transaction data (e.g. 7 overpayments in a 10-month period)
    - **Yellow** where # of overpayments is between 30% and 70%.
    - **Green** where # of overpayments is below 30%.

- Threshold for Refunds
    - **Red** where # of Refunds is 50% or greater in the # of months in the time period covered by the transaction data (e.g. 5 refunds in a 10-month period)
    - **Yellow** where # of overpayments is between 25% and 50%.
    - **Green** where # of overpayments is below 25%.

2. Method of payment to card balances

- Scenario Description
    - This scenario purports to detect AML risks in which a money launder may be using money from hard-to-trace source to pay off credit card balance.
    - Categorize the credit payments by the payment methods and look for unusual patterns, e.g. consistent or large usage of ACH by a domestic customer to pay off the credit card
    - The following fields should be part of transaction data schema.

Transaction Type	Credit_Debit_Flag
Payment	C (Credit)
Refund	C
Reversal	C
Award	C

Purchase	D (Debit)
Fee	D
Interest charge	D
Penalty	D
…	…

Payment Method
Cash
Wire
ACH
Paper Check
e-Check
Online Transfer
Payment at ATM
…

- Threshold
  - **Red** where # of payments by Cash, Wire or ACH is 50% or greater in the # of months in the time period covered by the transaction data (e.g. 5 of such payments in a 10-month period)
  - **Yellow** where # of such payments is between 30% and 50%.
  - **Green** where # of such payments is below 30%.

3. Payment source is owned by non-account holder

- Scenario Description
  - This scenario purports to detect AML risks in which a money launder may be using dirty money to pay off credit cards owned by other people

Payment Method	**Payment Source Account ID**	**Payment Source Owner**
Cash	N/A	N/A

Wire	Wire sender's (i.e. the payer) Bank Account Number	Name of the Sender / Payer
ACH	Wire sender's (i.e. the payer) Bank Account Number	Name of the Sender / Payer
Paper Check	Bank Account Number on the paper check	Owner of the account on the paper check
e-Check	Bank Account Number on the e-check	Owner of the account on the e-check
Online Banking Transfer	N/A (Assumption is the online banking transfer is allowed only between accounts owned by the same customer)	N/A
ATM payment	N/A (Assumption is the credit card payment at an ATM is allowed only between accounts owned by the same customer)	N/A
…		

- Threshold
    - **Red** where # of payments by non-account-holder is 40% or greater in the # of months in the time period covered by the transaction data (e.g. 4 of such payments in a 10-month period)
    - **Yellow** where # of such payments is between 25% and 40%.
    - **Green** where # of such payments is below 25%.

4. Payment is frequently made at a location that is materially distant from the account address

- Scenario Description
    - This scenario purports to detect AML risks in which a money launder may be using other people at various locations to pay off credit card

  - Payment method includes cash and ATM payment.
  - This will require the following data elements:
    - Payment location / address
    - Address of record associated with the Credit Card account
    - Distance (in miles) between payment location and account address.
- Threshold
  - "Materially distant" is defined as payment location being more than 100 miles away from the account address
  - **Red** where # of payments at non-account-address is more than 50% or greater in the # of months in the time period covered by the transaction data (e.g. 5 of such payments in a 10-month period)
  - **Yellow** where # of such payments is between 30% and 50%.
  - **Green** where # of such payments is below 30%.

5. Frequent or large transactions at high-risk countries

- Scenario Description
  - This scenario purports to detect AML risks associated with frequent or large credit card transactions at high-risk countries Payment method includes cash and ATM payment.
  - 'High Risk' countries are defined as 'CURRENT_US_COUNTRY_RISK_RATING' of 4 or 5 in the CRR table, for example:

```
select distinct COUNTRY_NAME,current_us_country_risk_rating

from
IDP_PRD_INTERFACE.USPC.V_L3_HRA_US_AML_COUNTRY_RIS
K_RATING

where current_us_country_risk_rating>3

order by country_name
```

- Threshold
    - **Red** where
        - # of transactions at high-risk countries is 45% or greater against the total # of transactions in the same time period
        - $ value of transactions at high-risk countries is 45% or greater against the total $ value of transactions in the same time period
    - **Yellow** where number (or $ value) of transactions at high-risk countries is between 30% and 45% of the total number (or $ value) of transactions in the same time period
    - **Green** where number (or $ value) of transactions at high-risk countries is below 30% of the total number (or $ value) of transactions in the same time period

6. Unusual ATM withdrawals for cash advance

- Scenario Description
    - This scenario purports to detect AML risks associated with unusual pattern of credit card cash advance withdrawals at an ATM
    - This will require the following data elements:
        - ATM location and address
        - Credit card cash advance date/time and amount
        - Account holder address
        - Distance between the ATM location and the Account holder address
- Threshold
    - **Red** where
        - # of cash advances is 20% or greater against the total # of all credit card transactions in the same time period, or
        - $ value of cash advances is 20% or greater against the total $ value of all credit card transactions in the same time period
    - **Yellow** where

        - # of cash advances is between 10% and 20% against the total # of all credit card transactions in the same time period
    - **Green** where
        - # of cash advances is below 10% against the total # of all credit card transactions in the same time period

7. Frequent Credit Card transactions at locations "materially distant" from the account address

- Scenario Description
    - This scenario purports to detect AML risks in which a money launder may be having other people at various locations to use the same credit card
    - Transaction types include ATM cash advances and POS transactions
    - This will require the following data elements:
        - Transaction location / address
        - If ATM or POS transaction, location and address of the ATM / POS
        - Address of record associated with the Credit Card account
        - Distance (in miles) between transaction location and account address.
    - "Materially distant" is defined as payment location being more than 100 miles away from the account address
- Threshold
    - **Red** where
        - 50% or more of the number (or the $ value) of transactions are at a location that is "materially distant" (see #4 scenario) from the account location, or
        - There are multiple transactions at locations 200 miles or greater within an hour.
    - **Yellow** where

        - Between 30% and 50% or more of the number (or the $ value) of transactions are at a location that is “materially distant” (see #4 scenario) from the account location
    - **Green** where
        - Less than 30% of the number (or the $ value) of transactions are at a location that is “materially distant” (see #4 scenario) from the account location

8. Merchant credits without offsetting merchant transactions

- Scenario Description
    - This scenario purports to detect AML risks in which merchant credits are applied to credit card without the matching (or offsetting) merchant transaction
    - This will require the following data elements:
        - Merchant name and address (and unique ID if available)
        - Transaction type is available to identify refund or reversal or merchant credit
- Threshold
    - **Red** where
        - Three or more Merchant credits without offsetting merchant transactions in a given month
    - **Yellow** where
        - One or two Merchant credits without offsetting merchant transactions in a given month
    - **Green** where
        - Zero Merchant credits without offsetting merchant transactions in a given month

9. Hotel room rentals at different hotels over the same time period

- Scenario Description
    - This scenario purports to detect AML risks in which hotel rooms were rented in different hotels in the same time period
    - This will require the following data elements:
        - Credit card transaction information has details about the hotel stay, including hotel name and address, and dates of hotel stays
        - If hotel stays were pre-paid, transaction details should include the actual dates of hotel stays that are pre-paid
        - If available, the merchant credits or refunds for pre-paid hotel stays (e.g. dirty money can potentially be used to prepay for hotel stays, then later cancel the stays for clean-money refunds)
- Threshold
    - **Red** where
        - Two or more simultaneous (or overlapping) stays in different hotels, or
        - Two or more pre-paid hotel stays that were subsequently refunded, or
        - Refund for pre-paid hotel stays without the offsetting pre-paid hotel stays (see #8 scenario)
    - **Yellow** where
        - One simultaneous (or overlapping) stay in different hotels, or
        - One pre-paid hotel stays that were subsequently refunded
- **Green** where
        - Zero simultaneous (or overlapping) stay in different hotels, and
        - Zero pre-paid hotel stays that were subsequently refunded

10. Multiple airline tickets for non-account holders

- Scenario Description
    - This scenario purports to detect AML risks in which multiple airline tickets were purchased for non-account holders
    - This will require the following data elements:

        - Credit card transaction information has details about the airline ticket purchase, including the date of purchase, name of passenger(s), from/to itinerary, and dates of travel

- Threshold
    - **Red** where
        - Four or more purchases of such airline tickets in a given month, or
        - Any refund (or merchant credit) of such purchases without the offsetting transaction
    - **Yellow** where
        - Two or three purchases of such airline tickets in a given month
    - **Green** where
        - One or less purchases of such airline tickets in a given month

11. <u>Unusually large payments for accumulated balance</u>

Exclude the over payments ,

and when you did the check using th e 2 standard deviation above +-

Flagged everything when you did all the

Did not compare the payment with the balance>>

- Scenario Description
    - This scenario purports to detect AML risks in which unusually large payments are made for accumulated balance
    - This will require the following data elements:
        - History of credit card balance by month
        - History of credit card payments by month
- Threshold
    - **Red** where

        - Payment amount is two standard deviation (sigma) or more away from the mean of payments over the preceding six months
    - **Yellow** where
        - Payment amount is between one and two standard deviation (sigma) away from the mean of payments over the preceding six months
    - **Green** where
        - Payment amount is less than one standard deviation (sigma) away from the mean of payments over the preceding six months

12. Out of country transactions

- Use Case Out of Country – 10: approx. 90% of all transactions are out of country
- Use Case Out of Country – 40: approx. 60% of all transactions are out of country
- Use Case Out of Country – 50: approx. 50% of all transactions are out of country
- Use Case Out of Country – 95: approx. 5% of all transactions are out of country
- Use Case Out of Country – only US transactions

Other Scenarios

- Separate transactions by credit and debit
    - See #2 scenario for details on credit / debit transaction types
- Add non-account ids for wires and international ach
- Add locations not on customer balance

2. This chart presents the hit rates by Use Case with the USE_CASE label filter.
   Use Case Scenario refers to the description listed in the Use Case Scenario handout.
   With Filter Percent – This refers to the hit rate of the labelled divided by the actual (clients marked with this particular USE CASE)

Use Case Scenario	Actual Labeled based on QA	Originally Labelled	With Filter Percent
Red 1.0	35	0	0.00%
Yellow 1.0	60	0	0.00%
Green 1.0	615	615	100.00%
Red 1.1	46	0	0.00%
Yellow 1.1	69	0	0.00%
Green 1.1	643	636	98.91%
Red 2	41	39	95.12%
Yellow 2	70	58	82.86%
Green 2	680	621	91.32%
Red 3	43	28	65.12%
Yellow 3	77	29	37.66%
Green 3	659	625	94.84%
Red 4	42	15	35.71%
Yellow 4	80	28	35.00%
Green 4	610	455	74.59%
Red 5	37	1	2.70%
Yellow 5	50	12	24.00%
Green 5	611	610	99.84%
Red 6	45	29	64.44%
Yellow 6	68	42	61.76%
Green 6	630	554	87.94%
Red 7	45	0	0.00%
Yellow 7	61	0	0.00%
Green 7	654	632	96.64%
Red 8	43	43	100.00%
Yellow 8	57	26	45.61%
Green 8	669	669	100.00%
Red 9	44	31	70.45%
Yellow 9	65	11	16.92%
Green 9	646	565	87.46%
Red 10	42	35	83.33%
Yellow 10	70	15	21.43%
Green 10	642	533	83.02%
Red 11	41	39	95.12%
Yellow 11	55	2	3.64%
Green 11	674	68	10.09%
Use Case Out of Country - US	722	722	100.00%

Use Case Out of Country - 40	15	10	66.67%
Use Case Out of Country - 50	38	22	57.89%
Use Case Out of Country 10	22	17	77.27%
Use Case Out of Country 95	65	17	26.15%

3. Table 3: This chart comparing the hit rates by Use Case with and without a USE_CASE label filter.
   Use Case Scenario refers to the description listed in the Use Case Scenario handout.
   With Filter Percent – This refers to the hit rate of the labelled divided by the actual (clients marked with this particular USE CASE)
   Without Filter Percent = This refers to the hit rate of the labelled (all clients that get picked up with the QA SQL) divided by the actual (total number of clients).

Use Case Scenario	**With Filter Percent**	**Without Filter Percent**
Red 1.0	0.00%	0.00%
Yellow 1.0	0.00%	0.86%
Green 1.0	100.00%	99.14%
Red 1.1	0.00%	7.32%
Yellow 1.1	0.00%	25.23%
Green 1.1	98.91%	67.45%
Red 2	95.12%	20.43%
Yellow 2	82.86%	61.14%
Green 2	91.32%	18.43%
Red 3	65.12%	0.72%
Yellow 3	37.66%	3.95%
Green 3	94.84%	95.33%
Red 4	35.71%	0.15%
Yellow 4	35.00%	18.13%
Green 4	74.59%	81.72%
Red 5	2.70%	0.29%
Yellow 5	24.00%	0.12%
Green 5	99.84%	99.59%
Red 6	64.44%	0.34%
Yellow 6	61.76%	1.33%
Green 6	87.94%	98.33%
Red 7	0.00%	16.59%
Yellow 7	0.00%	36.40%
Green 7	96.64%	47.01%
Red 8	100.00%	0.68%

Yellow 8	45.61%	0.26%
Green 8	100.00%	99.06%
Red 9	70.45%	52.38%
Yellow 9	16.92%	0.19%
Green 9	87.46%	47.42%
Red 10	83.33%	1.46%
Yellow 10	21.43%	0.45%
Green 10	83.02%	98.10%
Red 11	95.12%	69.31%
Yellow 11	3.64%	6.97%
Green 11	10.09%	23.71%