PROMPT

Next, you will play the role of the client and the data scientist.

Using the topic that you selected, complete the Business Understanding stage by coming up with a problem that you would like to solve and phrasing it in the form of a question that you will use data to answer. **(3 marks)**

You are required to:

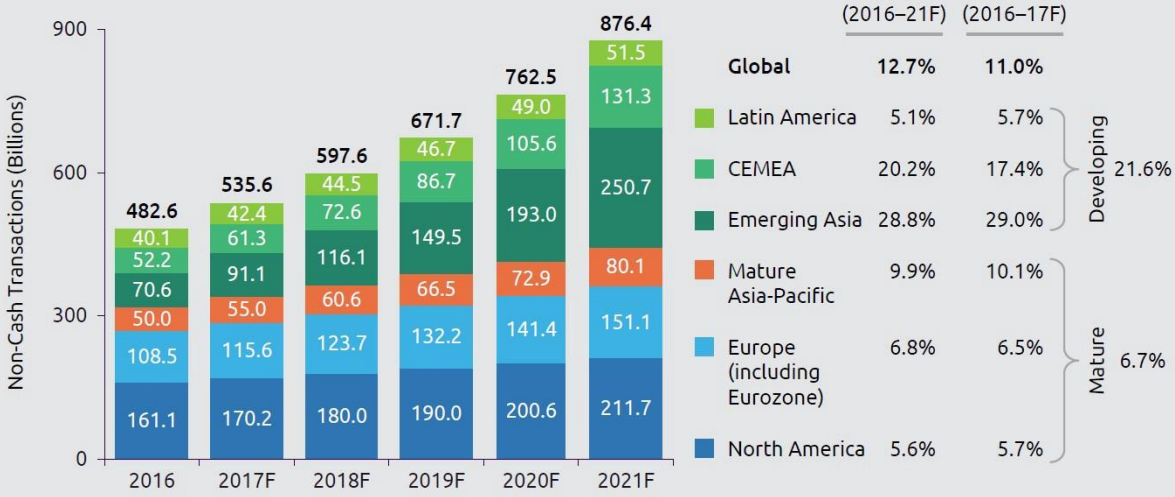
1. Describe the problem, related to the topic you selected.
2. Phrase the problem as a question to be answered using data.

For example, using the food recipes use case discussed in the labs, the question that we defined was, "Can we automatically determine the cuisine of a given dish based on its ingredients?".

Answer:

In today's world, we are on the express train to a cashless society. According to the World Payments Report, in 2016 total non-cash transactions increased by 10.1% from 2015 for a total of 482.6 billion transactions! That's huge! Also, it's expected that in future years there will be a steady growth of non-cash transactions as shown below:

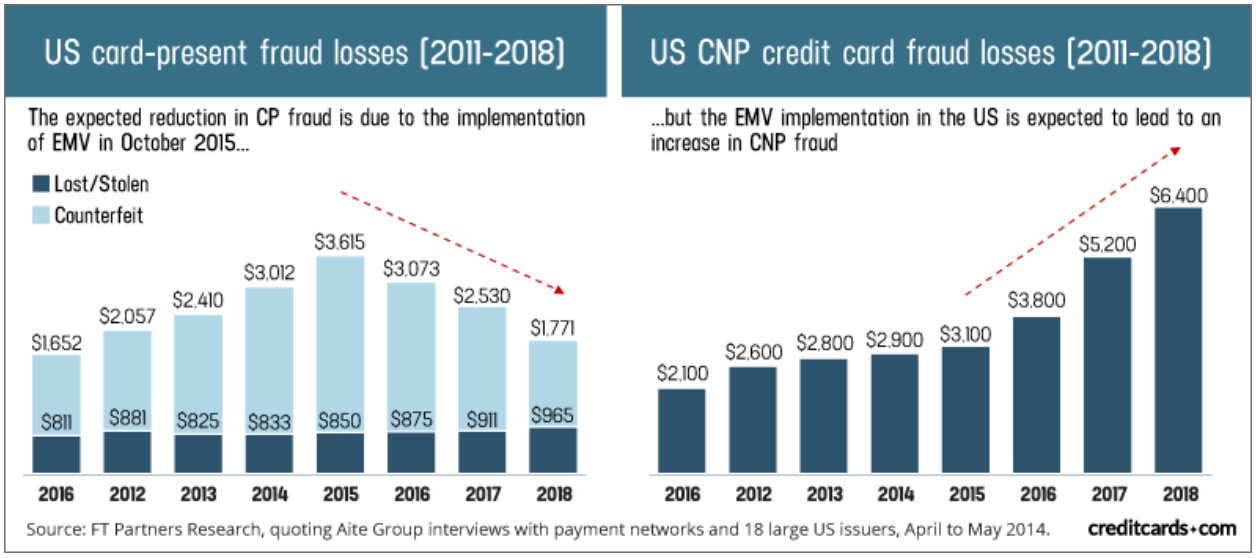
Figure 1.5 Number of Worldwide Non-Cash Transactions (Billions), by Region, 2016–2021F



Credit Card fraud is the cause of considerable losses every day.. Hackers and crooks around the world constantly find new ways of defrauding people of their money.. Consequently, relying exclusively on rule-based, conventionally programmed systems for detecting financial fraud would not provide the appropriate time-to-market. But that doesn't mean that the crooks have won. Machine learning shines as a unique solution for just this type of problem.

The main challenge in finding the appropriate solution is working through a suitable model. The challenge of modeling fraud detection as a classification problem is due to the fact that in real world data, the majority of transactions are not fraudulent. Investment in technology for fraud detection has increased over the years so this shouldn't be a surprise, but this brings us a problem of imbalanced data.

We still have a very high amount of money lost from credit card fraud:



The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be fraud. This model is then used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

PROMPT

Briefly explain how you would complete each of the following stages for the problem that you described in the Business Understanding stage, so that you are ultimately able to answer the question that you came up with. (5 marks):

Answer:

Analytic Approach

Data Requirements

Data Collection

Data Understanding and Preparation

Modeling and Evaluation

You can always refer to the labs as a reference with describing how you would complete each stage for your problem.

Analytic Approach: After analysis the business requirement we will go to binary classification model for find out new transaction is fraudulent or not.

Data Requirements: We need historical data regarding credit card transaction with labeled data.

Data Collection: Client will provide historical data regarding credit card transaction.

Data Understanding and Preparation: The data understanding section of the data science methodology answers the question: Is the data that you collected representative of the problem to be solved? Before continuing with our analysis, it is important to data preparation with required feature & label column.

Modeling and Evaluation: For classify the Credit Card Transaction we will use following Classification model.

1.Logistic Regression

2.Naive Bayes Classifier

3.Decision Trees

4.Random Forest

5.GBT

After applying those model we have to evaluate for model is appropriate. In this connection we may use Classification Accuracy, ROC term.