



Bank marketing Analysis

By:

Harshil maniar
Juhi Kamdar
Kaushil Kundalia
Joy Mehta

IU1641150012
IU1641150020
IU1641150023
IU1641150026

Name of the Application: Bank Marketing Data Analysis

Description:

Taking into consideration the complexity and dynamics of economic and social life, in general, and especially actual environment which the economic entities and, particularly, the financial and economic crisis facing humanity, the role and the importance of marketing in the financial-banking activities is growing. The marketing if done in a way that is most fruitful, can be a useful resource, and is also monetarily

feasible. So, in this project, we are making an application, that is helpful in making good business decisions related to marketing. The bank needs to populate the server with the data of their customers, and the application will provide a predictive analysis which can be further used to filter out non-yielding customers. The analysis is performed individually on each column, so that the bank employees can eliminate some parameters and can work on only specific parameters to market their product.

Objective:

There are two main sources of income for banks : Interest on loans and the monthly or yearly operating fees. Each financial institution wants to maximize their profits but these institutions have individuals have people from all walks of life associated with them. Hence in order to maximize their profits there can be two approaches:

1. Target a small group of high earning individuals and earn a good amount of money from each of them.
2. Target a larger group of low earning individuals and earn a good amount from all of them collectively.

In order to decide on an approach, analysis needs to be performed on the institution's dataset to find out how many people from each segment are associated with them and then tailor the policies of the institution for the segment that brings in more money for the institution.

Scope:

In any financial institutions.

Difficulties with traditional data collection methods/Problem Statement:

Today, people have multiple accounts in different banks and the capital in the accounts is not distributed equally among all bank accounts. As a result, a person having a good job might have a very low capital in one of his bank account or a person with blue collar job might have inherited a good amount of money and might have concentrated the capital into one account. There might be such multiple cases which can produce some unexpected output.

How it is related to 5 V's of Big Data?

In the current scenario, more and more people are switching towards modern ways of saving money that is, in banks. Even in developing countries like India, a major chunk of the weaker sections of the society have opened bank accounts in the last few years. In the world there are about 6 billion people who have bank accounts. Some of them even have multiple accounts which increases the number of bank accounts to over 7-8 billion. The rise and fall of the economy is clearly reflected in the bank accounts of the people. Millions of transactions take place in a day in any given bank. In order to build a marketing strategy, multiple data are collected related to one bank account which are stored as floating values, integer values and strings.

Data Infrastructure Overview:

This is the classic marketing bank dataset uploaded originally in the UCI Machine Learning Repository. The dataset gives you information about a marketing campaign of a financial institution which we have analyzed in order to find ways to look for future strategies in order to improve future marketing campaigns for the bank. It contains data of about 11000 people.

Type of Data

- 1 - **age**: (numeric)
- 2 - **job**: type of job (categorical: 'admin', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
- 3 - **marital**: marital status (categorical: 'divorced','married','single','unknown';)
- 4 - **education**: (categorical: primary, secondary, tertiary and unknown)
- 5 - **default**: has credit in default? (categorical: 'no','yes','unknown')
- 6 - **housing**: has housing loan? (categorical: 'no','yes','unknown')
- 7 - **loan**: has personal loan? (categorical: 'no','yes','unknown')
- 8 - **balance**: Balance of the individual.
- 9 - **contact**: contact communication type (categorical: 'cellular','telephone')
- 10 - **month**: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
- 11 - **day**: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
- 12 - **duration**: last contact duration, in seconds (numeric).
- 13 - **campaign**: number of contacts performed during this campaign and for this client (numeric, includes last contact)
- 14 - **pdays**: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
- 15 - **previous**: number of contacts performed before this campaign and for this client (numeric)
- 16 - **poutcome**: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

Data Processing Architecture

In this project, we have used two parts of Data Analysis, namely: Visualization and Statistical Inference. For Visualization of the data, libraries like seaborn, matplotlib and plotly are used. Whereas, the statistical part is managed using pandas and numpy. The final model which combines visualization and statistical inference is an application. This application is managed by streamline library.

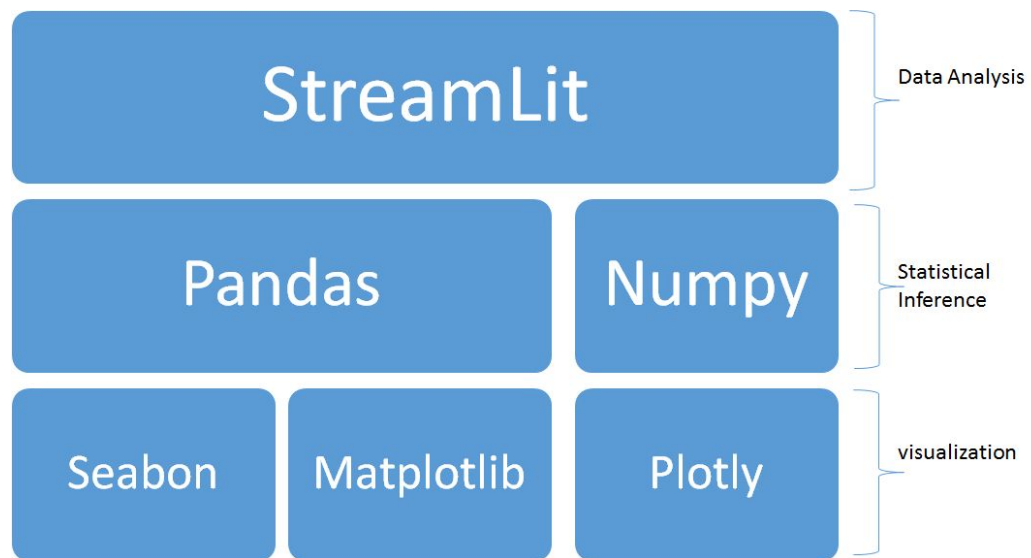


Fig.1 Data processing Architecture

Components (ex. Sqoop,HBase,Oracle)

Pandas, numpy, seaborn

Process

For ease of access and scalability, we have set up a cloud server and uploaded the dataset on the server. This enables us to send daily transactions and updates from anywhere in the world. The analysis is also performed on the cloud server. This enables a person to access the data from any device from anywhere in the world.

This application has a GUI which was made using streamlit python library which is used to bind and display all the graphical and statistical data that is processed on the cloud. Numpy is used to deal with multidimensional data ,Pandas library is used for data manipulation and data analysis. The analyzed data is then displayed using matplotlib and seaborn libraries. Plotly library is used to make interactive charts.

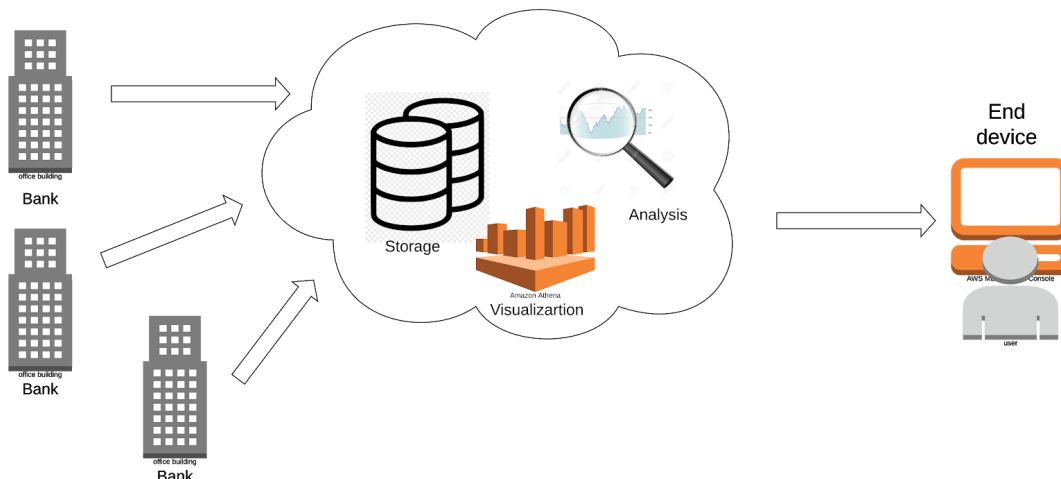


Fig. 2 Process of Big data Analysis for Bank marketing Data

View of Application:

← → ↺ 🏠 167.71.234.104:8501 ... 🌟 📄 📱 ☰

Bank Marketing Analysis!

- The data is related with direct marketing campaigns of a Portuguese banking institution
- The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be (or not) subscribed.

Exploratory data analysis

- As listed above, the below mentioned dataset is regarding the marketing performed by some Portuguese bank
- We need to predict whether the customer will **OPT IN** for a particular offer offered from the bank through telephonic marketing.

← → ↺ 🏠 167.71.234.104:8501 ... 🌟 📄 📱 ☰

Inferences:

- Singles have average minimum balance **HOWEVER** single has the maximum income.
- Married have on average maximum income.

Mean Balance in Account
by Job Occupation

Legend: Negative Balance, Low Balance, Middle Balance, High Balance

← → ↺ 🏠 167.71.234.104:8501 ... 🌟 📄 📱 ☰

Distribution of those who opted in or out

Legend: Opted In, Opted Out

Looking at the distribution of our output variable we can derive the following:

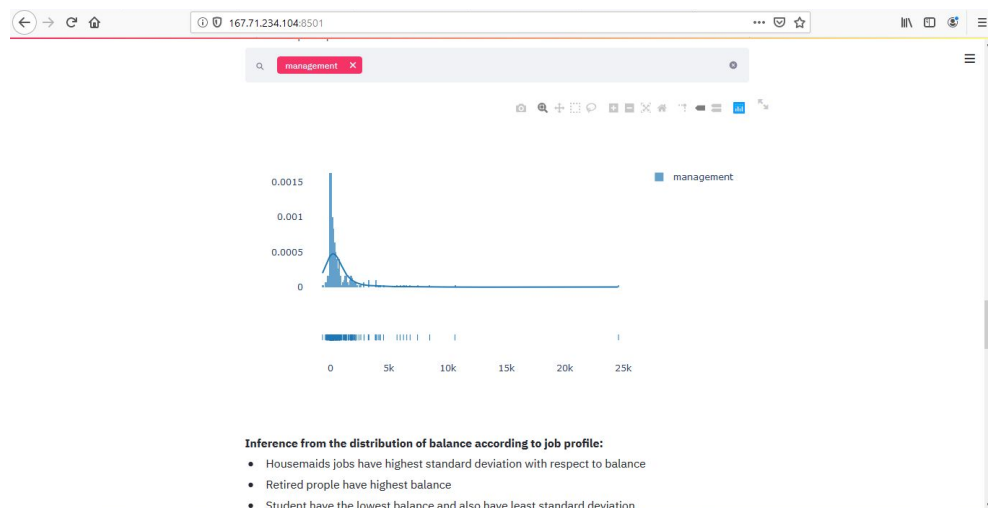
1. The number of people who opted in is considerably greater than those who opted out. (88.3% vs 11.7%)
2. We need to take care so that the model does not become biased towards predicting positive

Number of rows to see
5 28

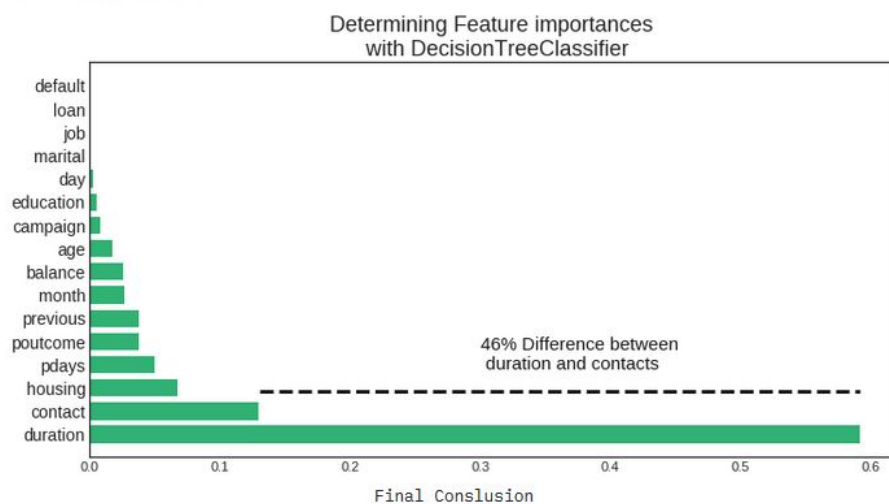
	age	job	marital	education	default	balance	housing	loan	contact	campaign	previous
0	58	management	married	tertiary	no	2143	yes	no	no		
1	44	technician	single	secondary	no	29	yes	no	no		
2	33	entrepreneur	married	secondary	no	2	yes	yes	no		
3	47	blue-collar	married	unknown	no	1586	yes	no	no		
4											

Summary of various columns

	age	balance	day	duration	campaign	pdays	previous
count	45211	45211	45211	45211	45211	45211	45211
mean	40.9362	1,362.2721	15.8864	258.1631	2.7638	40.1978	0.5883
std	10.6188	3,044.7658	8.3225	257.5278	3.6980	100.1287	2.3834
min	18	-8019	1	0	1	-1	0



Final Conclusion



Final Conclusion

167.71.234.104:8501

Correlation of each attribute among each other



Correlation Image

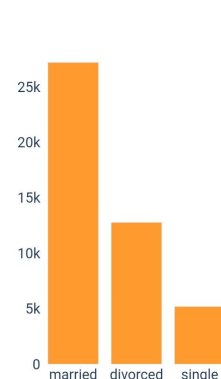
The following inference can be deduced:

- Most of the columns are distinct and not correlates
- This means, we do not need to perform any feature engineering and the raw columns can directly be used to perform analysis.

167.71.234.104:8501

Understanding the marital distribution and their balance

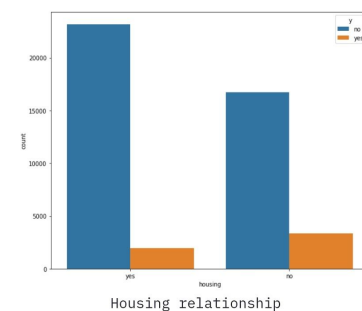
Count by Marital Status



Balance distribution according to marital status:

167.71.234.104:8501

Understanding trend between housing ownership and term deposit:



Inference

- There is not a strong trend or relation between housing and temp deposit.

Relationship between loan and temp deposit:



Business challenges that enable this project

There are two main sources of income for banks : Interest on loans and the monthly or yearly operating fees. Since there are a lot of financial institutions and individuals that are in the profession of lending money, it becomes very important for a bank to analyse and target its customers intelligently. People tend to avoid banks that have a high interest rate or operating fees or those who offer low rate of interest. Hence, to maintain a stable income, financial institutions need to target a group of individuals over others that will benefit them the most and make changes accordingly.

Marketing campaigns are characterized by focusing on the customer needs and their overall satisfaction. Nevertheless, there are different variables that determine whether a marketing campaign will be successful or not. There are certain variables that we need to take into consideration when making a marketing campaign.

1) **Segment of the Population:** To which segment of the population is the marketing campaign going to address and why? This aspect of the marketing campaign is extremely important since it will tell to which part of the population should most likely receive the message of the marketing campaign.

2) **Distribution channel to reach the customer's place:** Implementing the most effective strategy in order to get the most out of this marketing campaign. What segment of the population should we address? Which instrument should we use to get our message out? (Ex: Telephones, Radio, TV, Social Media Etc.)

3) **Price:** What is the best price to offer to potential clients? (In the case of the bank's marketing campaign this is not necessary since the main interest for the bank is for potential clients to open deposit accounts in order to make the operative activities of the bank to keep on running.)

4) **Promotional Strategy:** This is the way the strategy is going to be implemented and how are potential clients going to be addressed. This should be the last part of the marketing campaign analysis since there has to be an in depth analysis of previous campaigns (If possible) in order to learn from previous mistakes and to determine how to make the marketing campaign much more effective.