EPSILON 3.0

THE ANALYTICS CONCLAVE

By – Team Thinkers

CASE STUDY

Dunder Mifflin Bank is a financial institution principally engaged in the business of providing finance to the public, whether by making loans or advances or otherwise. It provides loans to the borrower for various purposes ranging from home, auto, student, and small business loans to various others.

It conducts its business operations in the following countries:

- United States of America
- United Kingdom
- Brazil
- India

You are Jim Halpert (their recently appointed Chief Analytics Officer) has received an urgent request from Michael Scott (CFO) to come up with a detailed analysis of the bank's lending data addressing few concerns.

Bank Credit Analysis

In bank credit analysis, banks consider and evaluate every loan application based on merits. They check the creditworthiness of every individual or entity to determine the level of risk that they subject themselves by lending to an entity or individual.

Clients with a high level of risk are less desirable since they present with a high likelihood of defaulting on their loan obligations. Low-risk clients are more likely to get their loan applications approved since the lender considers them creditworthy.

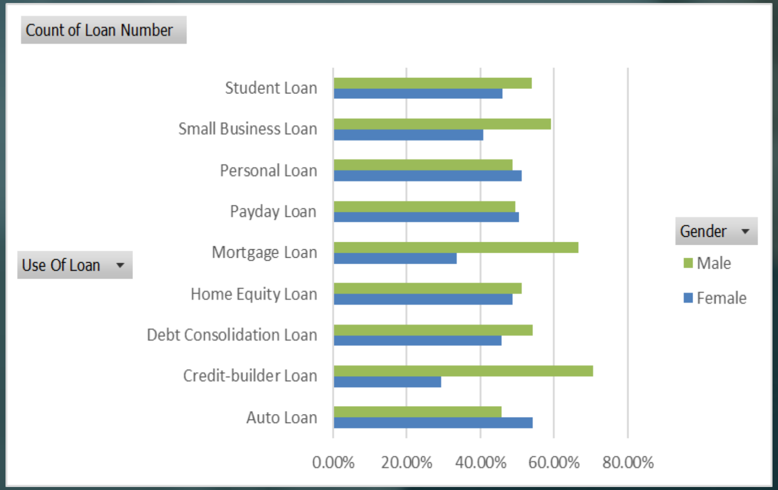
DATA CLEANING AND PRE PROCESSING

- There were no duplicate values present in our dataset and all the values for the loan number were unique and hence it served as the primary key for our dataset.
- We dropped the columns that we found redundant and were not required for analysis like **Loan ID, Application signed hour, Application signed weekday.**
- **Transformation of data -**
- We transformed the column **'Loan Application Start Date'** and extracted the year from it for better analysis. **New column formed - 'Year of Loan sanctioned'.**
- We performed discretization on the **'Age'** column and divided it into intervals for better analysis. **New column formed - 'Age group'.**
- ❑ 17 - 40 = Adult
- ❑ 40 - 60 = Middle age adult
- ❑ 60 - 75 = Senior citizen
- For better analysis, we converted the **'Probability of default'** column into a percentage so that we can compare the values better. **New column formed - 'Percentage_defaulting'.**
- We performed discretization on the **'Percentage_defaulting'** column and divided it into intervals for better analysis**. New column formed - 'Defaulting_on_loan'.**
- ❑ 0-20 = Low probability
- ❑ 20 - 60 = Moderate probability
- ❑ 60 - 100 = High Probability
- We formed another column containing the total income of the customer for better analysis. **New column formed - 'Total Income' = Income from employer + Income from pension + Income from other sources.**
- We formed another column which states whether the customer has adhered to the maturity date given by the bank or not. **New column formed - 'Adherence to maturity date' which takes binary value of 0 or 1**,
- ❑ 0 : 'Maturity date last' >= 'Maturity date original'
- ❑ 1: 'Maturity date last' <= 'Maturity date original'
- We formed another column which states whether full loans (the whole amount applied by the customer) were given by the banks to the customers or not. **New column formed - 'Full loans sanctioned' which takes binary value of 0 or 1,**
- ❑ 0: 'Applied Amount' != 'Amount'
- ❑ 1: 'Applied Amount' = 'Amount'

ASSUMPTIONS

- The **'Defaulting_on_loan'** considers all the details of the customers and tells us the probability of customer defaulting on that particular loan.
- We assume by the term **'Lending patterns'** that how the amount sanctioned by the bank differs from the amount applied for the loan by the customer and the time duration set by the bank for that particular loan.
- **'Maturity date original'** means the last date set by the bank to repay the whole loan after restructuring the loan (if applicable).
- **'Maturity date last'** means the date followed by the customer to repay the whole loan.
- By the term **loan restructuring**, we assume that the time for repayment of loan is increased by the banks to decrease the probability of default.

KEY INSIGHTS FROM THE DATA



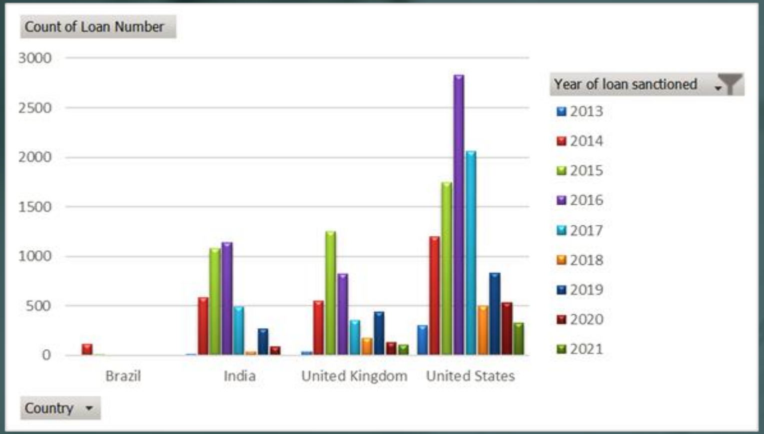
Females take 3 types of more loans than men ie, Personal, Payday and Auto loans

- Student Loan has the least frequency and Small Business Loan has the most frequency across all countries and age groups

So, banks giving loans to customers having previous pending loans (>1) and high probability of default is very improbable.

So, those who have number of previous loans, the company gives loan to only those who have low probability of default.

NO EVERGREENING OF LOANS!



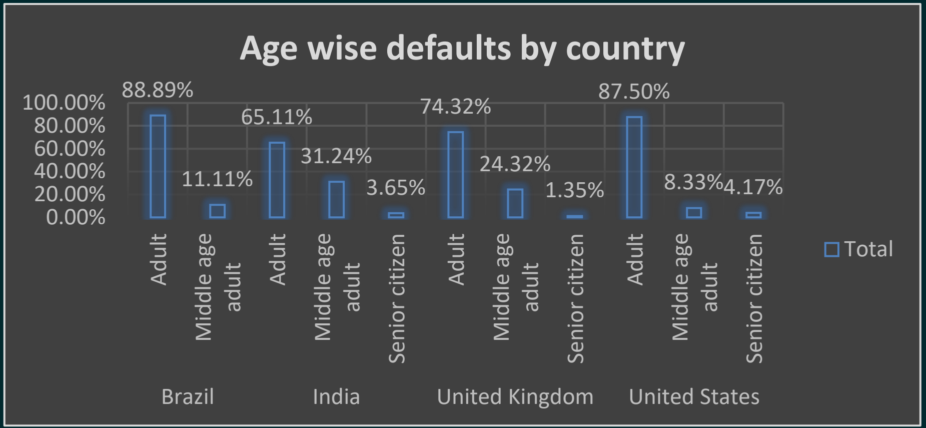
Acc. to the data, number of loans given in Brazil are very less.

There is an unusual spike in the lending of loans in USA for the year 2016. It may be because of the very LOW INTEREST RATES(1.8%) in USA that year.

- Maximum number of loans are given through Verification Type 4 (7879) and Minimum number of loans are given through Verification Type 2 (49)

Parameters affecting Probability of Default

Age Group in every Country



In all the countries, **Adults (17- 40 Yrs)** have the **highest probability of Default**

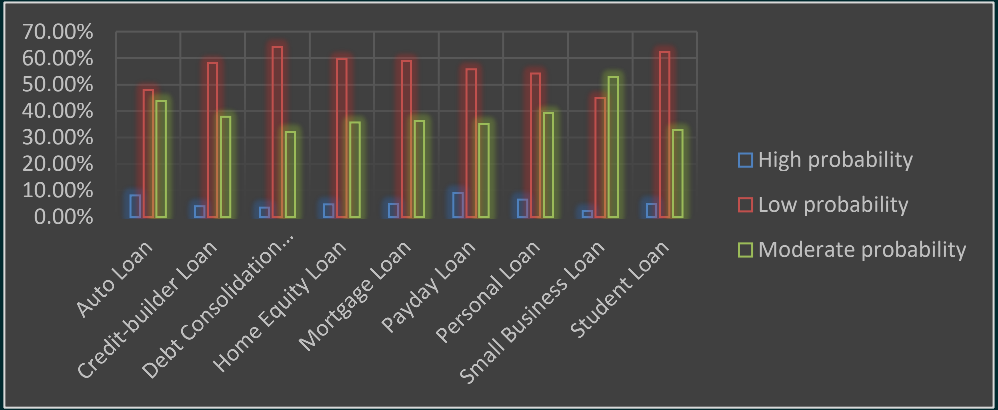
New Credit Customer

Row Labels	Total	Adhered customers	Percentage adhering
0	6852	4264	62%
1	11155	8213	74%

0 -	Old customer
1 -	New customer

So, from the above table, we can say that **old customers** do not adhere to maturity date as much as the new customers do

Use of Loan



So, we can infer that loan taken for **debt consolidation** use have a very **low and moderate probability of defaults**

Work Experience

Row Labels	High probability	Low probability	Moderate probability
10To15Years	5.48%	56.77%	37.75%
15To25Years	4.82%	62.72%	32.47%
2To5Years	6.54%	46.41%	47.06%
5To10Years	6.34%	53.61%	40.05%
LessThan2Years	6.16%	49.61%	44.22%
MoreThan25Years	4.27%	69.76%	25.97%
No record found	1.07%	43.90%	55.04%

So, we can infer that there are highest number of customers who are **highly experienced** and having **low probability of default**

Parameters affecting Lending Patterns

Country

Row Labels	Total	Full sanctioned loans	Percentage
Brazil	115	109	94.78
India	3697	2632	71.19
United Kingdom	3866	3325	86.01
United States	10329	9550	92.46

So it can be seen that in India, full loan are not given to many people. What can be the reason for it??

Defaults by Country

Defaulting_on_loan	High probability

Row Labels	Count of Loan Number
Brazil	9
India	685
United Kingdom	74
United States	48

So, we can see that customers with high probability of default is maximum in India. Hence, we can assume that this is the reason for not giving full loans to customers in India

Use of Loan

Row Labels	Count	Full loans	Percentage
Auto Loan	568	482	84.86
Credit-builder Loan	848	661	77.95
Debt Consolidation Loan	3826	3222	84.21
Home Equity Loan	2332	1992	85.42
Mortgage Loan	1084	921	84.96
Payday Loan	714	614	85.99
Personal Loan	3930	3307	84.15
Small Business Loan	4321	4107	95.05
Student Loan	384	310	80.73

So, for Small Business Loan, percentage of full loan sanctioned is highest and for Credit Builder Loan, it is the least

Rating

Row Labels	Count	Full loans sanctioned	Percentage
A	656	640	98%
AA	363	362	100%
B	2122	1997	94%
C	3140	2887	92%
D	3212	2909	91%
E	2900	2533	87%
F	2068	1779	86%
G	3546	2509	71%

We can infer that number of full loans sanctioned are the least for rating G and highest for rating AA

Country where Dunder Mifflin Bank should shut its operations

Restructured	TRUE

Row Labels	Count of Loan Number	Adhered customers	Percentage
Brazil	36	14	39%
India	700	158	23%
United Kingdom	1361	427	31%
United States	3912	950	24%

So, from the above table, we can infer that India has the highest number of customers for whom loan date was restructured and they still couldn't repay the loan on time

Row Labels	Average of amt owed
India	5796.032453
United Kingdom	3837.190277
United States	2134.587547

From the above table, we can say that average of total amount (principal balance + Interest balance) owed by the customers in India is the highest.

Count of Loan Number	Column Labels		
Row Labels	High probability	Low probability	Moderate probability
Brazil	7.83%	19.13%	73.04%
India	18.53%	19.31%	62.16%
United Kingdom	1.91%	59.18%	38.90%
United States	0.46%	67.14%	32.39%

Here also we can see that in India, the customers having high probability of default is the maximum among all countries.

So, from the above observations, we can say there is maximum negligence on the part of the customers in India for loan repayment and other issues and the company should plan to reduce/stop their operations in INDIA as it is affecting their overall efficiency.

CONCLUSION

- There are various factors affecting Probability of default, some of the prominent one being **Age group, New or Old customer, Use of loan and Work experience**. **Senior citizens and highly experienced customers** have the lowest probability of default. **Old customers** do not adhere to maturity dates and have a higher probability of defaulting on loans.
- Primary factors that contribute to differences in lending pattern from one customer to other are **Country, Use of loan, and Rating**. In **India**, lending patterns are different as full loans are given to very few people as compared to other countries. Lending patterns are favorable for uses like **small business loans** but not favorable for uses like **credit builder loans**. The percentage of full loans sanctioned decreases from **rating AA to G**. (We can plot them on a rating scale starting from the highest AA to lowest G).
- **Data analytics** is an integral part for every company and it helps in identifying key parameters which would help in lending the loans to right customers and monitor their collections.

SUGGESTIONS

- The Bank should consider **reducing** its operations in **India** because the customers there, have a **high probability of default** despite **restructuring** their loans.
- Bank should focus more on the **Adult** age group as it has a **higher probability of default** in all the countries. It should be more vigilant while sanctioning loans to them.
- Bank should focus more on the customers having work experience of **less than 5 yrs** and the process of lending loans to them should be more rigorous and prudent as they have a **higher probability of default**.
- If a **new customer** comes with the same defaulting parameters as the old ones, then the process of lending loans should be more **scrutinizing and watchful**.
- Bank should start analyzing the **Credit Score** of customers using data analytics, before lending the loans in order to even reduce **NPAs** and **increase its profitability**.

THANK YOU!

TEAM MEMBERS

- ISHAN SONI
- SHIVANI ARYA