

# BANK LOAN CASE STUDY

# Hello!

I am Preety Mohanta

I am here to give presentations **Bank Loan Case Study.**

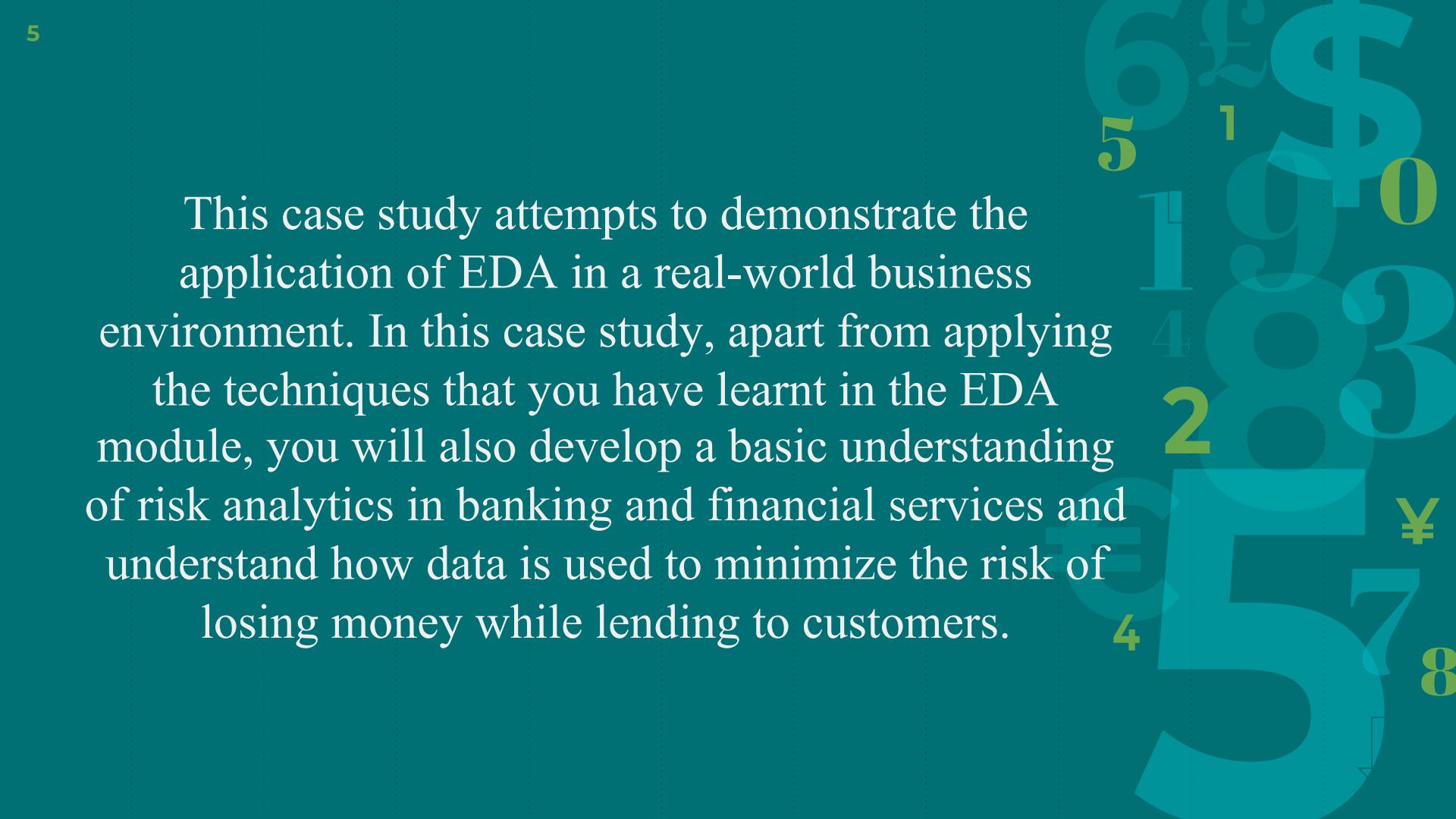


# CONTENT

- ❖ **Introduction**
- ❖ **Approach**
- ❖ **Tech Stack Used**
- ❖ **Tasks**
- ❖ **Insight**
- ❖ **Summary**
- ❖ **Source**



# PROJECT DESCRIPTION



This case study attempts to demonstrate the application of EDA in a real-world business environment. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

# APPROACH. . .

## Here are the steps we will follow for executing our project:

- Download the data set
- Then clean the data by removing unwanted information, duplicate, handle the missing data.
- After cleaning the data convert the data types and make it ready to use.
- Now understand the data and use the excel and perform the tasks
- With the help of excel we will perform Outliers, Data Imbalance, Univariate, Segmented Univariate, Bivariate and Correlations.
- And at the end we get insight from it.



# TECH STACK

USED . . .



Excel



PowerPoint



Presentation

# → DATA ANALYTICS TASKS:

## 1<sup>st</sup> Task

- Identify Missing Data and Deal with it Appropriately

## 2<sup>nd</sup> Task

- Identify Outliers in the Dataset

## 3<sup>rd</sup> Task

- Analyze Data Imbalance

## 4<sup>th</sup> Task

- Perform Univariate, Segmented Univariate, and Bivariate Analysis

## 5<sup>th</sup> Task

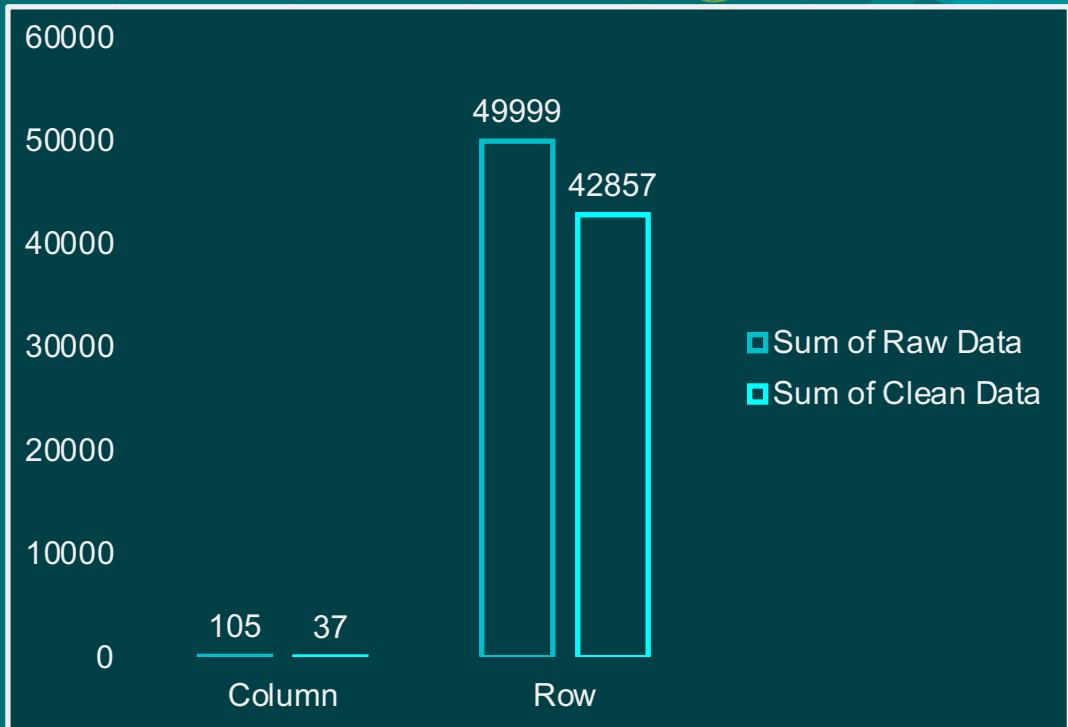
- Identify Top Correlations for Different Scenarios

# INSIGHT



# <sup>12</sup> → Identify Missing Data and Deal with it Appropriately

	Sum of Raw Data	Sum of Clean Data
Column	105	37
Row	49999	42857
Grand Total	50104	42894



## RESULT

After clean the data using functions like COUNTA and COUNTBLANK, after removing the null and missing values we get 49,999 rows and 105 columns to 42,857 rows and 37 columns. This will significantly reduce the missing values or null values.

This process likely involve removing rows and columns with more than 30% of null value in the dataset. With this cleaned dataset, further analysis can be conducted with greater confidence in the data's reliability.

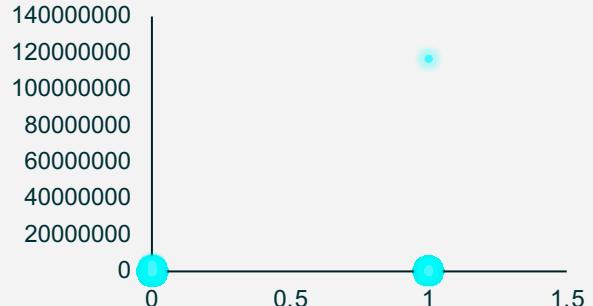


# → Identify Outliers in the Dataset

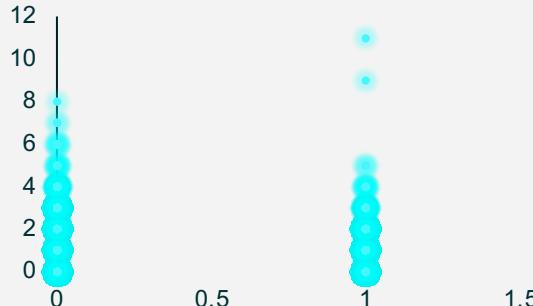
**Days Employed(YRS)**



**Amt\_Income\_Total**



**CNT\_Children**



**Quartile - 1**

**2.652054795**

**Quartile - 3**

**202500**

**Inter Quartile Range**

**202497.3479**

**UPPER LIMIT**

**506246.0219**

**Lower Limit**

**-303743.3699**

**Quartile - 1**

**112500**

**Quartile - 3**

**202500**

**Inter Quartile Range**

**90000**

**UPPER LIMIT**

**337500**

**Lower Limit**

**-22500**

**Quartile - 1**

**0**

**Quartile - 3**

**1**

**Inter Quartile Range**

**4**

**UPPER LIMIT**

**2.5**

**Lower Limit**

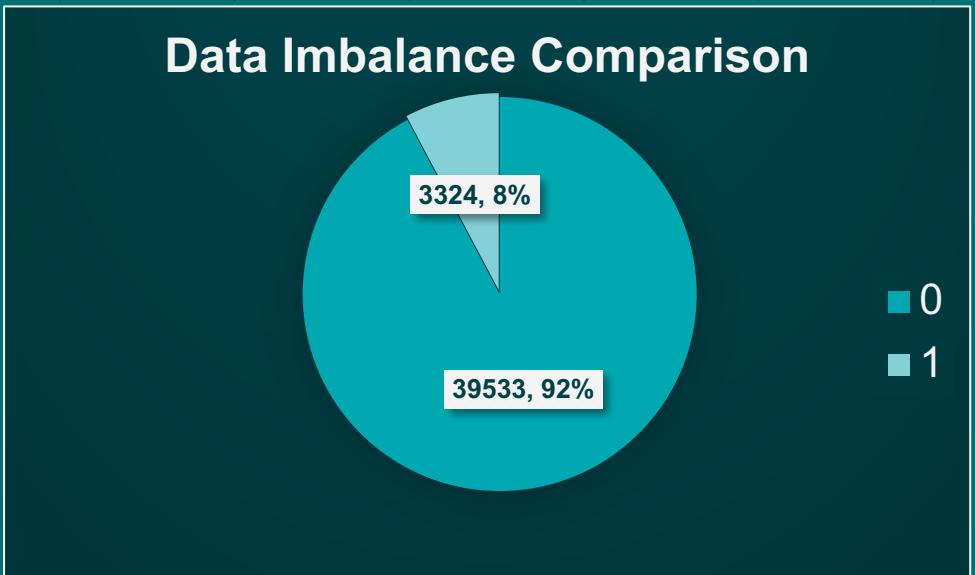
**-1.5**

# RESULT

- ❖ Quartiles statistics provide, among other things, an idea of the distribution and central tendency of the data that can help us identify whether there are outliers or biases and can help with decisions such as thresholding or comparing stocks.
- ❖ The "Amt\_Income\_Total" data set shows the average income is 173,367.59, with a mean of 157,500 and a mode of 572,899.79. The standard deviation of 572,899.79 indicates a large variation. Of 42,857 observations, the income ranged from 27,000 to 117,000,000, amounting to 7,430,014,715.
- ❖ The "Days Employed(yrs)" data set shows that the average length of employment is 182.29 years, with a median of 6.25 years and a mode of 379.08 years with a standard deviation of 379.08 showing high variability. The 42,855 observations range from 0 to 1000.67 years, for a total of 7,812,010.90 years.
- ❖ The "CNT\_Children" data set shows a significant absence of children, with an average of 173,367 but moderate and 0. There is a moderate difference, reflected in a standard deviation of 0.73, with children 0 to 11 in 42,857 observations This indicates that a childless population is common, but large families Some examples.

# Analyze Data Imbalance

Row Labels	Count of TARGET	RATIO	CONTRIBUTION
0	39533	11.89320096	92.24
1	3324		7.76
Grand Total	42857		100.00

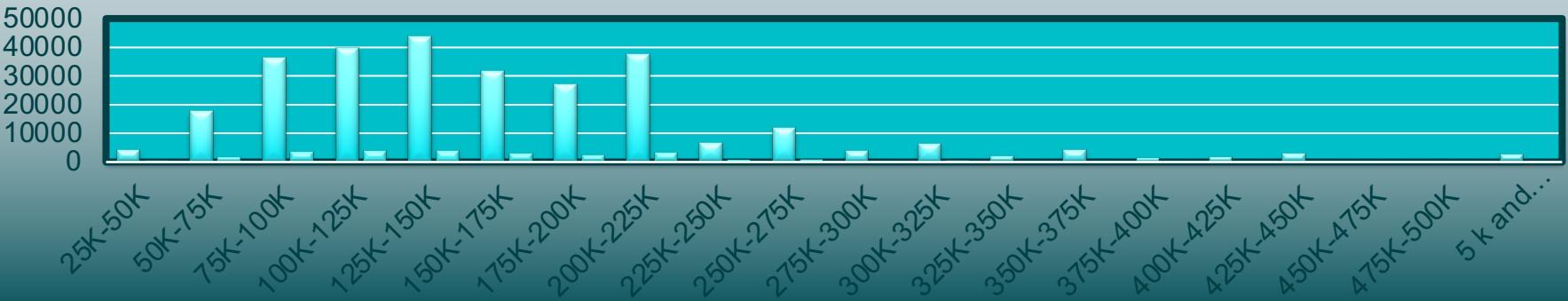


# RESULT

- **Row Labels:** This column in all likelihood represents the types or groups for your dataset. In this example, it seems to be the target values, with "zero" and "1" indicating the two viable values for the goal variable.
- **Count of TARGET:** This column shows the number of occurrences of every goal value on your dataset. For instance, there are 39,533 instances wherein the goal value is zero, and three,324 times in which the goal price is 1.
- **RATIO:** This column represents the ratio of each goal value to the whole remember. For instance, the ratio for the target price zero is about 11.89%, meaning that about eleven.89% of the full instances have a target value of zero. Similarly, the ratio for the goal fee 1 is ready 7.Seventy six%.
- **CONTRIBUTION:** This column probable indicates the contribution of each goal price to the entire in percentage terms. In this situation, the contribution of the target value 0 is ninety two.24%, indicating that about 92.24% of the overall instances have a goal price of zero. The contribution of the target value 1 is 7.Seventy six%.
- Value 0: accounts for about 92.24% of the total sample.
- Value 1: accounts for about 7.76% of the total sample.
- This figure provides a clear comparison of the objective value distributions, showing that the objective value is 0 in most cases, while the objective value is 1 for a subset

18 → Perform Univariate, Segmented Univariate, and Bivariate Analysis

TARGET APPLICANTS PER INCOME BINS



APPLICANTS PER CREDIT BINS



AVERAGE CREDIT AMOUNT PER INCOME BIN



## RESULT

- o Customers have largely chosen cash and consumer loans. The majority of our clients are repeat customers.
- o The majority of current loan applicants are individuals who applied for loans less than ten months ago. More loans have been requested for consumer gadgets.
- o Customers who applied for more than Rs. 350,000 will most likely be denied. The majority of loans sought for through Credit and Cash agencies are cancelled.
- o Clients who apply for another loan within 10 months of their previous loan are more likely to have it cancelled. Walk-in loans have a higher refusal rate.



20

# Identify Top Correlations for Different Scenarios

		TARGET 0						
<b>CORRELATION FOR TIMELY PAYMENTS</b>								
<i>CNT Of Children</i>	1	0.009300239	0.004117726	-0.025043214	-9.7283E-07	0.002368016	0.003949962	0.026733198
<i>AMT_Income_Total</i>	0.009300239	1	0.063511987	0.026543706	-0.000502928	-0.003429977	0.003573112	-0.035127419
<i>AMT_Credit</i>	0.004117726	0.063511987	1	0.098803701	-0.004437401	-0.000222036	-0.004349401	-0.102897278
<i>Region_Population_Relative</i>	-0.025043214	0.026543706	0.098803701	1	-0.003303866	-0.006023773	0.000844092	-0.527253308
<i>Days_Birth(Yrs)</i>	-9.7283E-07	-0.000502928	-0.004437401	-0.003303866	1	0.523388736	0.50850239	0.003170646
<i>Days_Employed(YRS)</i>	0.002368016	-0.003429977	-0.000222036	-0.006023773	0.523388736	1	0.296604311	0.001543287
<i>Days_ID_Publish(YRS)</i>	0.003949962	0.003573112	-0.004349401	0.000844092	0.50850239	0.296604311	1	0.003088812
<i>Region_Rating_Client</i>	0.026733198	-0.035127419	-0.102897278	-0.527253308	0.003170646	0.001543287	0.003088812	1
	<i>CNT Of Children</i>	<i>AMT_Income_Total</i>	<i>AMT_Credit</i>	<i>Region_Population_Relative</i>	<i>Days_Birth(Yrs)</i>	<i>Days_Employed(YRS)</i>	<i>Days_ID_Publish(YRS)</i>	<i>Region_Rating_Client</i>

		TARGET 1						
<b>Correlation For Timely Payments</b>								
<i>CNT Of Children</i>	1	0.009300239	0.004117726	-0.025043214	-9.7283E-07	0.002368016	0.003949962	0.026733198
<i>AMT_Income_Total</i>	0.009300239	1	0.063511987	0.026543706	-0.000502928	-0.003429977	0.003573112	-0.035127419
<i>AMT_Credit</i>	0.004117726	0.063511987	1	0.098803701	-0.004437401	-0.000222036	-0.004349401	-0.102897278
<i>Region_Population_Relative</i>	-0.025043214	0.026543706	0.098803701	1	-0.003303866	-0.006023773	0.000844092	-0.527253308
<i>Days_Birth(Yrs)</i>	-9.7283E-07	-0.000502928	-0.004437401	-0.003303866	1	0.523388736	0.50850239	0.003170646
<i>Days_Employed(YRS)</i>	0.002368016	-0.003429977	-0.000222036	-0.006023773	0.523388736	1	0.296604311	0.001543287
<i>Days_ID_Publish(YRS)</i>	0.003949962	0.003573112	-0.004349401	0.000844092	0.50850239	0.296604311	1	0.003088812
<i>Region_Rating_Client</i>	0.026733198	-0.035127419	-0.102897278	-0.527253308	0.003170646	0.001543287	0.003088812	1
	<i>CNT Of Children</i>	<i>AMT_Income_Total</i>	<i>AMT_Credit</i>	<i>Region_Population_Relative</i>	<i>Days_Birth(Yrs)</i>	<i>Days_Employed(YRS)</i>	<i>Days_ID_Publish(YRS)</i>	<i>Region_Rating_Client</i>

# RESULT

## Target 0

- **CNT Of Children:** Weak negative correlation with region population and region rating client.
- **AMT\_Income\_Total:** Weak positive correlation with region population.
- **AMT\_Credit:** Moderate positive correlation with region population.
- **Region Population Relative:** Weak negative correlation with days employed.
- **Days\_Birth(Yrs), Days\_Employed(YRS), Days\_ID\_Publish(YRS):** Moderate positive correlation with region rating client.
- **Region\_Rating\_Client:** Moderate positive correlation with days employed.

## Target 1

- The number of children has weak correlations with income and credit amount, and a weak negative correlation with region population.
- Total income weakly correlates positively with credit amount and negatively with region rating.
- Credit amount moderately correlates positively with region population and negatively with region rating.
- Region population has a moderate negative correlation with region rating.
- Age-related variables moderately correlate positively with each other and with region rating.
- Region rating moderately correlates negatively with region population and credit amount.

# SUMMARY

- ❖ As the Age and Years of Experience Increase, the chances of defaulting decrease. So the bank should prioritize older And Experienced Clients.
- ❖ Educated Clients tend to default lesser compared to clients with lower education such as Lower Secondary & Secondary education
- ❖ Male clients tend to default more than female clients.
- ❖ Corporate clients are a safer bet compared to labour class clients.
- ❖ People belonging to Region Rating 3 have the highest % of defaulters, bank could formulate a stricter loan policy m the clients from these region. Clients from Region 1 are the safest bet
- ❖ We also notice that as the Age increases, the amount taken by the clients is considerably higher and since the default percentage with higher age is lower these should be least risky and highly profitable clients for the bank.

# LINKS



[EXCEL Sheet LINK](#)



[LOOM Video LINK](#)



# Thanks!

