Tech Saksham

Case Study Report

Data Analytics with Power BI

“Anlayics of Unemployment in Republic of india”

“Using power bi”

“Govt arts and science college”

NM ID	NAME
A31A9276F2EDF8EA7054 5507BFEF3B59	S.Sarathi

ABSTRACT

The latest data indicates a glimmer of hope, as India's unemployment rate has recently declined. According to the the unemployment rate for individuals aged 15 years and above in urban areas decreased to 6.8 percent during January-March 2023 from 8.2 percent a year ago. This positive development suggests a potential turnaround in the job market amidst the prevailing economic complexities. However, continued vigilance and effective policy measures remain crucial to foster sustainable job growth and secure the nation's future prosperity.

The Current Unemployment Rate in India The unemployment rate witnessed a sharp decrease in January 2024. According to the latest data from the independent think-tank, the unemployment rate in India stood at 6.8 percent in January. The unemployment rate in India saw a decrease of 1.9 percent in a month, as it stood at 8.7 percent in December last year.

INDEX

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

The unemployment rate in India, amidst lockdown and restrictions on mobility, is 12.81% as of June 8th 2021 based on the data provided by the CMIE. Earlier, the unemployment rate in India shot up from 6.5 per cent in March 2021 to 8 per cent in April 2021, to 14.7% by May end, while the employment rate fell from 37.6 per cent in March to 36.8 per cent in April, says the report of CMIE – Centre For Monitoring Indian Economy.

In 2020, the unemployment rate in India fell to 7% in September 2020 from the record high of 29% since the country went into lockdown from March 2020, However, it later increased to 9.1% in December 2020.

The unemployment rate again declined to 6.5 per cent in January 2021 from 9.1 per cent in December 2020, while the employment rate surged to 37.9 per cent as compared to 36.9 per cent.

The lockdown to contain the coronavirus outbreak has forced many industries to shut down thus increasing unemployment across the country.

1.2Proposed Solution

Unemployment means a person willing to work but unable to find a qualified job. Our country is facing many problems but one of the serious problem is of unemployment. Many graduates, doctors, engineers, scientist are unemployed or working underemployed. Due to unemployment we are wasting our country's human resource.

The unemployed rate in between age group 15- 29 has been increased since 2009-2010. According to the Global Employment Trends 2014 the unemployment rate has raised to 3.8%, last year.

1.3 Feature

- **Real-Time Analysis**: The dashboard will provide real-time analysis of customer data.
- **Customer Segmentation**: It will segment customers based on various parameters like age, income, transaction behavior, etc.
- **Trend Analysis**: The dashboard will identify and display trends in customer behavior.
- **Predictive Analysis**: It will use historical data to predict future customer behavior.

1.4 Advantages

To provide unemployment benefits, various countries have policies and programs in place. The US enacted policies qua unemployment insurance as far back as 1935. The basic unemployment insurance program is run by various states which provide most of the funding and pay for the actual benefits given to workers. A person is from the state where the person was last employed at, even if the person now lives in a different state.

1.5 Scope

The scope of this project extends to all banking institutions that aim to leverage data for decision-making and customer engagement. The project can be further extended to incorporate more data sources and advanced analytics techniques, such as machine learning and artificial intelligence, to provide more sophisticated insights into customer behavior. The project also has the potential to be adapted for other sectors, such as retail, healthcare, and telecommunications, where understanding customer behavior is crucial. Furthermore, the project contributes to the broader goal of digital transformation in the banking sector, promoting efficiency, innovation, and customer-centricity.

CHAPTER 2

SERVICES AND TOOLS REQUIRED

2.1 Services Used

- **Data Collection and Storage Services**: Banks need to collect and store customer data in real-time. This could be achieved through services like Azure Data Factory, Azure Event Hubs, or AWS Kinesis for real-time data collection, and Azure SQL Database or AWS RDS for data storage.

- **Data Processing Services**: Services like Azure Stream Analytics or AWS Kinesis Data Analytics can be used to process the real-time data.

- **Machine Learning Services**: Azure Machine Learning or AWS SageMaker can be used to build predictive models based on historical data.

2.2 Tools and Software used

Tools:

- **Power BI**: The main tool for this project is PowerBI, which will be used to create interactive dashboards for real-time data visualization.

- **Power Query**: This is a data connection technology that enables you to discover, connect, combine, and refine data across a wide variety of sources.

2.3 Software Requirements:

- **Power BI Desktop**: This is a Windows application that you can use to create reports and publish them to Power BI.

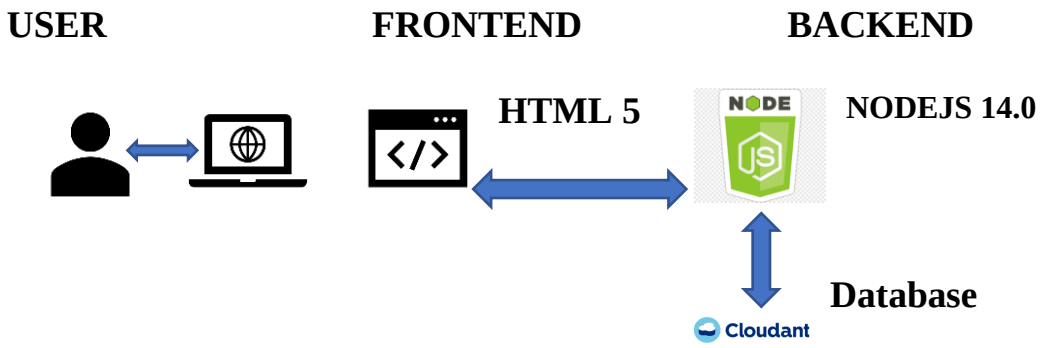
- **Power BI Service**: This is an online SaaS (Software as a Service) service that you use to publish reports, create new dashboards, and share insights.

- **Power BI Mobile**: This is a mobile application that you can use to access your reports and dashboards on the go.

CHAPTER 3

PROJECT ARCHITECTURE

3.1 Architecture



Here's a high-level architecture for the project:

1. **Data Collection**: Real-time customer data is collected from various sources like bank transactions, customer interactions, etc. This could be achieved using services like Azure Event Hubs or AWS Kinesis.
2. **Data Storage**: The collected data is stored in a database for processing. Azure SQL Database or AWS RDS can be used for this purpose.
3. **Data Processing**: The stored data is processed in real-time using services like Azure Stream Analytics or AWS Kinesis Data Analytics.
4. **Machine Learning**: Predictive models are built based on processed data using Azure Machine Learning or AWS Sage Maker. These models can help in predicting customer behavior, detecting fraud, etc.

5. **Data Visualization**: The processed data and the results from the predictive models are visualized in real-time using Power BI. Power BI allows you to create interactive dashboards that can provide valuable insights into the data.
6. **Data Access**: The dashboards created in Power BI can be accessed through Power BI Desktop, Power BI Service (online), and Power BI Mobile.

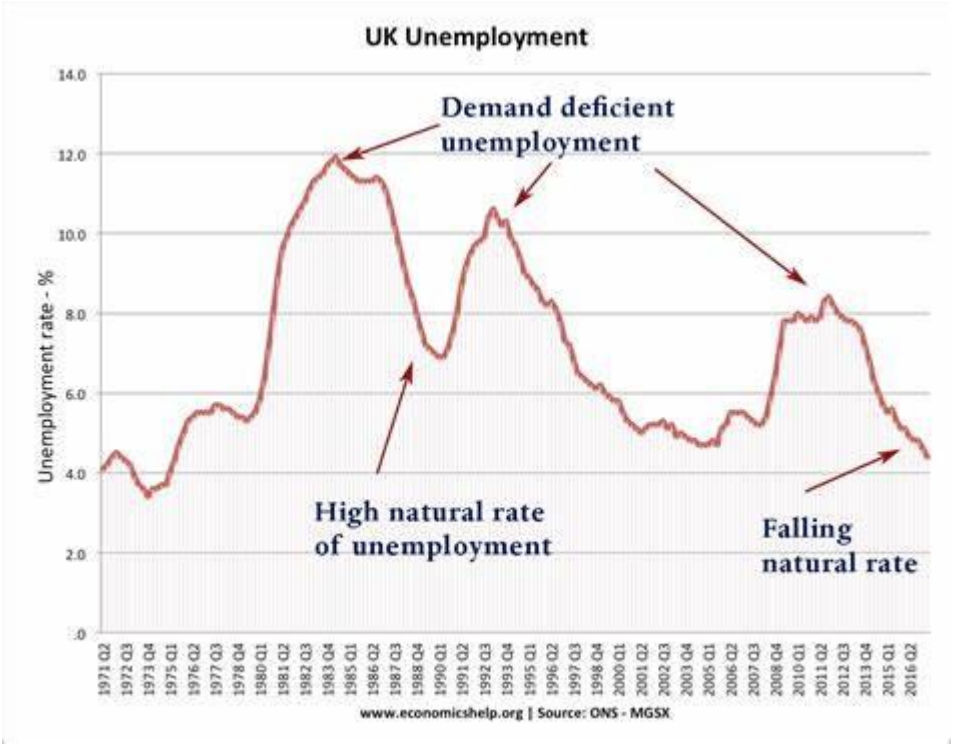
This architecture provides a comprehensive solution for real-time analysis of bank customers. However, it's important to note that the specific architecture may vary depending on the bank's existing infrastructure, specific requirements, and budget. It's also important to ensure that all tools and services comply with relevant data privacy and security regulations.

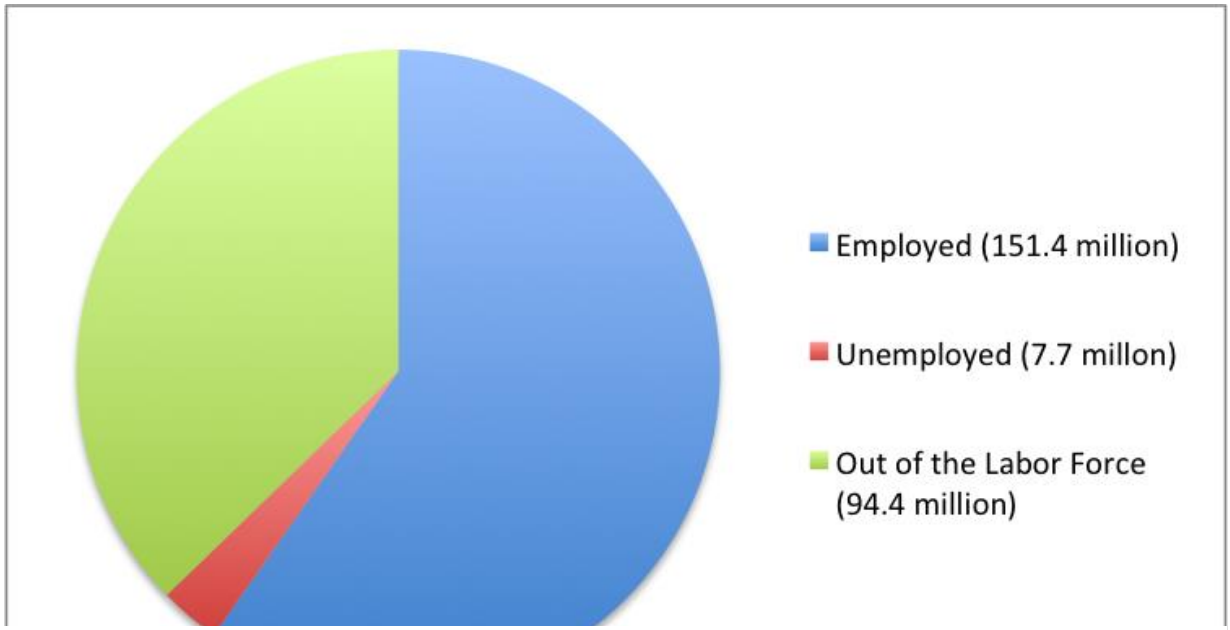
CHAPTER 4

MODELING AND RESULT

4.1 Manage relationship

The “unemployment” file will be used as the main connector as it contains most key identifier (account id, client id and unemployment id) which can be to relate the 8 data files together. The “district” file is use to link the client profile geographically with “district id”





Manage relationships

Active	From: Table (Column)	To: Table (Column)
☑	card (disp_id)	disp (disp_id)
☑	client (district_id)	district (district_id)
☑	disp (account_id)	account (account_id)
☑	disp (account_id)	loan (account_id)
☑	disp (client_id)	client (client_id)
☑	order (account_id)	account (account_id)
☑	transaction (account_id)	disp (account_id)
☐	account (district_id)	district (district_id)
☐	transaction (account_id)	loan (account_id)


```
Gender =
VAR stringDate = FORMAT(client[birth_number],"General Number")
VAR month = VALUE(MID(stringDate,3,2))
RETURN IF(month > 50,"F","M")
```

client_id	birth_number	district_id	Gender	Birthday	age
3428	875927	42	F	27/09/1987	13
4354	860813	28	M	13/08/1986	14
3417	855318	35	F	18/03/1985	15
10201	851019	13	M	19/10/1985	15

2013	8.5%	8.1%	7.6%	7.1%	7.3%	7.8%	7.7%	7.3%	7.0%	7.0%	6.6%	6.5%
2012	8.8%	8.7%	8.4%	7.7%	7.9%	8.4%	8.6%	8.2%	7.6%	7.5%	7.4%	7.6%
2011	9.8%	9.5%	9.2%	8.7%	8.7%	9.3%	9.3%	9.1%	8.8%	8.5%	8.2%	8.3%
2010	10.6%	10.4%	10.2%	9.5%	9.3%	9.6%	9.7%	9.5%	9.2%	9.0%	9.3%	9.1%

the date format to 20/05/2004 adding 365 to the year.

```
Birthday =
VAR stringDate = FORMAT(client[birth_number],"General Number")
VAR stringMonth = VALUE(MID(stringDate,3,2))
VAR mth = IF(stringMonth > 50, stringMonth - 50,stringMonth)
VAR year = VALUE(MID(stringDate,1,2))
VAR day = VALUE(MID(stringDate,5,2))
RETURN FORMAT(DATE(year+1900,mth,day),"DD/MM/YYYY")
```

client_id	birth_number	district_id	Gender	Birthday	age
3428	875927	42	F	27/09/1987	13
4354	860813	28	M	13/08/1986	14
3417	855318	35	F	18/03/1985	15
10201	851019	13	M	19/10/1985	15

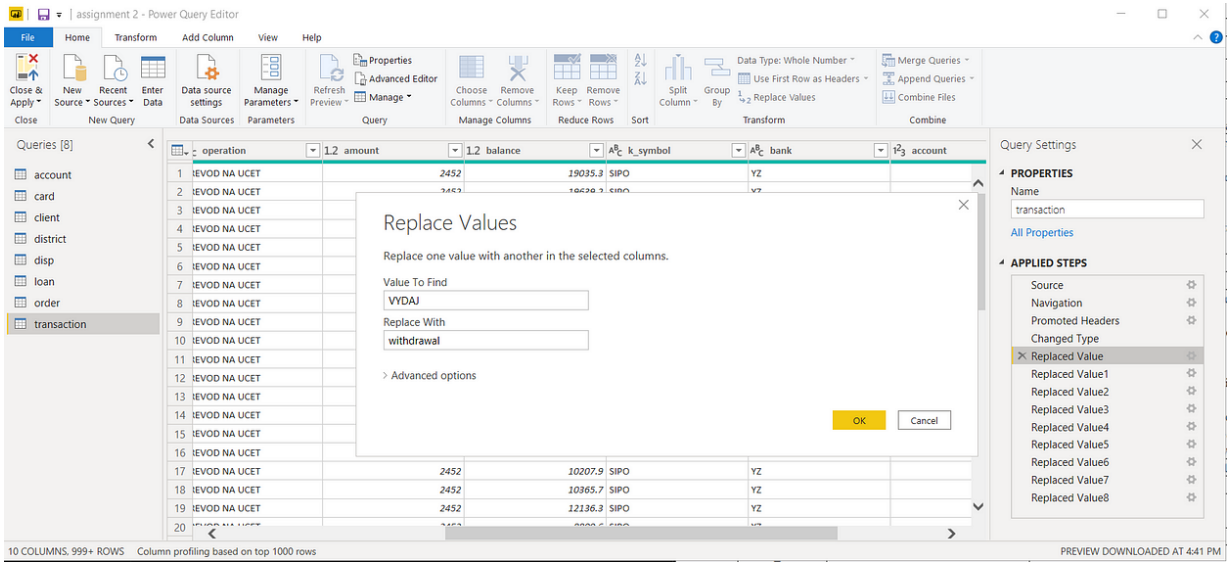
For Age, we shall assume it is year 1999 as explain previously and use it to minus from the birth year.

```
1  age = 1999 -RIGHT(client[Birthday],4)
```

client_id	birth_number	district_id	Gender	Birthday	age	age (groups)
2	450204	1	M	04/02/1945	54	36 -54 Baby Boomers

4.2 Replacing values

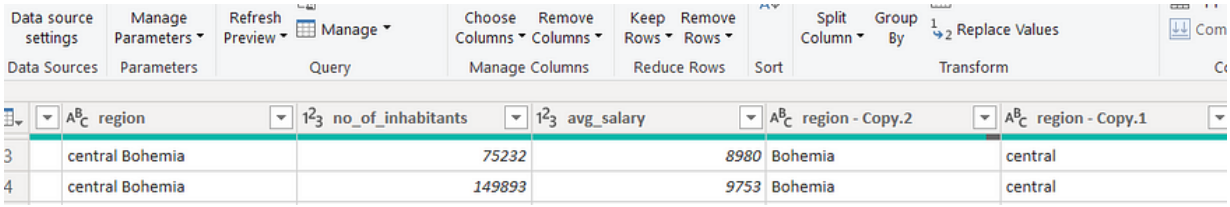
Set some fields to English for easy understanding, we replace values to English with the Power Query Editor.



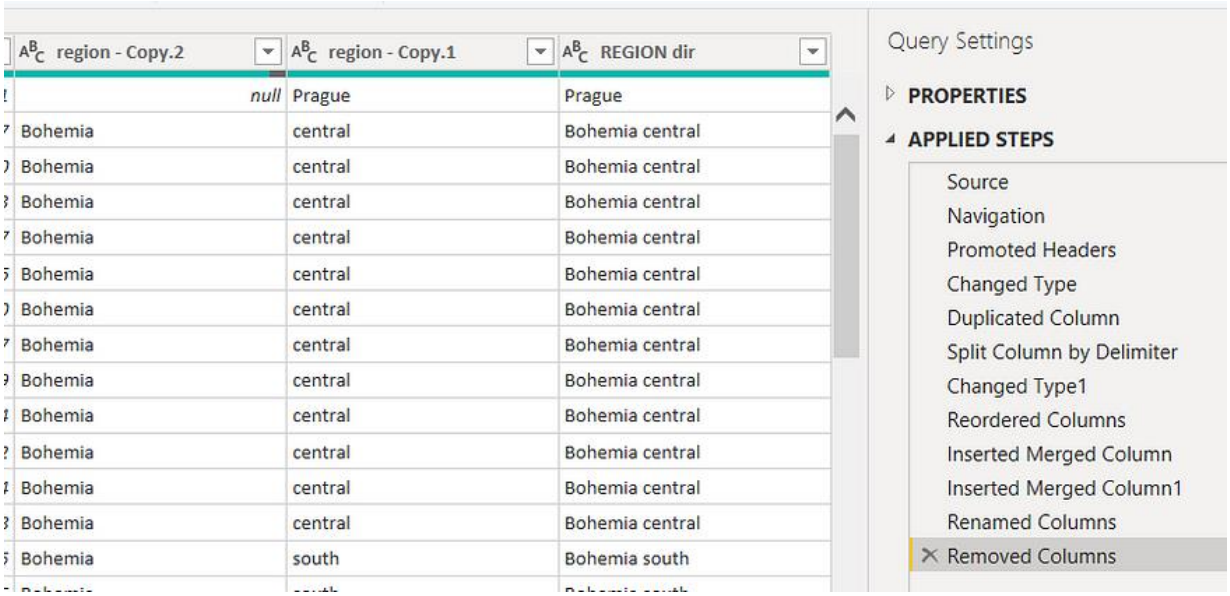
type	+/- transaction	"PRIJEM" stands for credit "VYDAJ" stands for withdrawal
k_symbol	characterization of the transaction	"POJISTNE" stands for insurance payment "SLUZBY" stands for payment for statement "UROK" stands for interest credited "SANKC. UROK" sanction interest if negative balance "SIPO" stands for household "DUCHOD" stands for old-age pension "UVER" stands for loan payment

Changing the order of Region name at Power Query

Duplicate the “district /region” then split column using space as delimiter.



Then merge column by Region and direction. Refer to applied steps for details.



4.3 Grouping of age by ranges

As the customers' age ranges from 12 to 88, we shall group them into different generation age range for easier profiling we will group the ages into 5 groups.

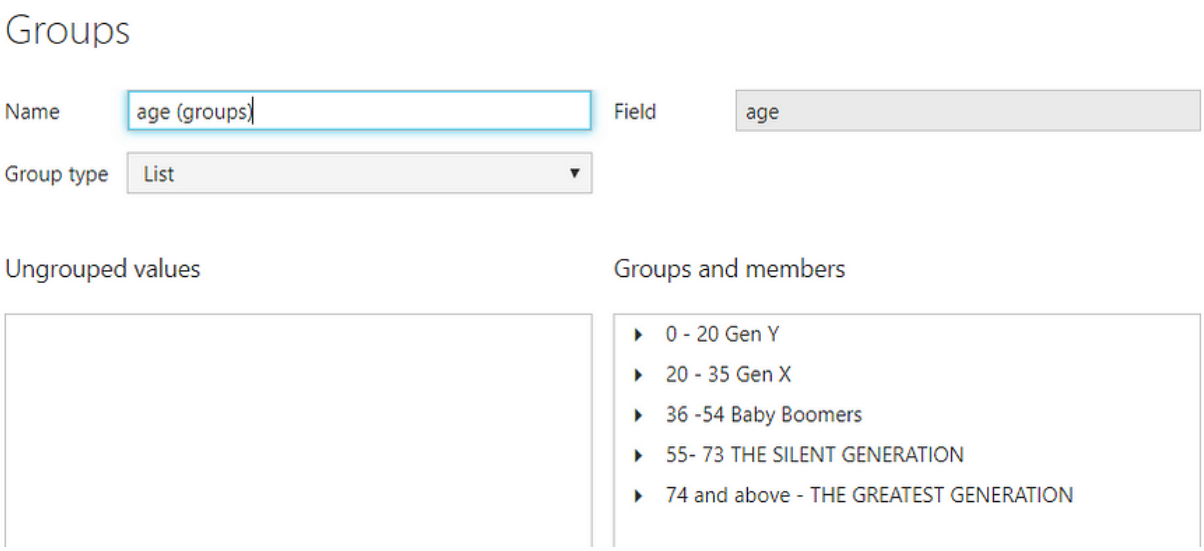
The Gen youths,

Gen are young working adults, some starting their families

Baby Boomer working adults with are families.

The silent Generations some are working and retired, living on pensions.

The greatest Generation, retired elderly living on pensions.



4.4 Credit Rating and Loan Status

As the Loan status uses A, B, C, D which are not reader friendly. We can add a column to represent what it stands for, we also simplify the classification of those with late or default on payment as bad credit, refer to the table below for details on the new columns added.

Status in "loan" data	New column "loan status"	New column "credit rating"
'A' stands for contract finished no problems	Fully Repaid	Good
'B' stands for contract finished loan not payed	Default	Bad
'C' stands for running contract OK so far	Timely Payment	Good
'D' stands for running contract client in debt	Late payment	Bad

```
1 Loan Status =
2 IF(loan[status]="A","Repaid Full",
3 IF(loan[status]="B","Default",IF (loan[status]="c","Timely payment","Late payment" )))
```

loan_id	account_id	date	Loan Amt	duration	payments	status	Credit Rating	Loan Status
6059	5196	971228	79,824 Kč	12	6652	A	GOOD	Repaid Full
6727	8505	971210	42,840 Kč	12	3570	A	GOOD	Repaid Full

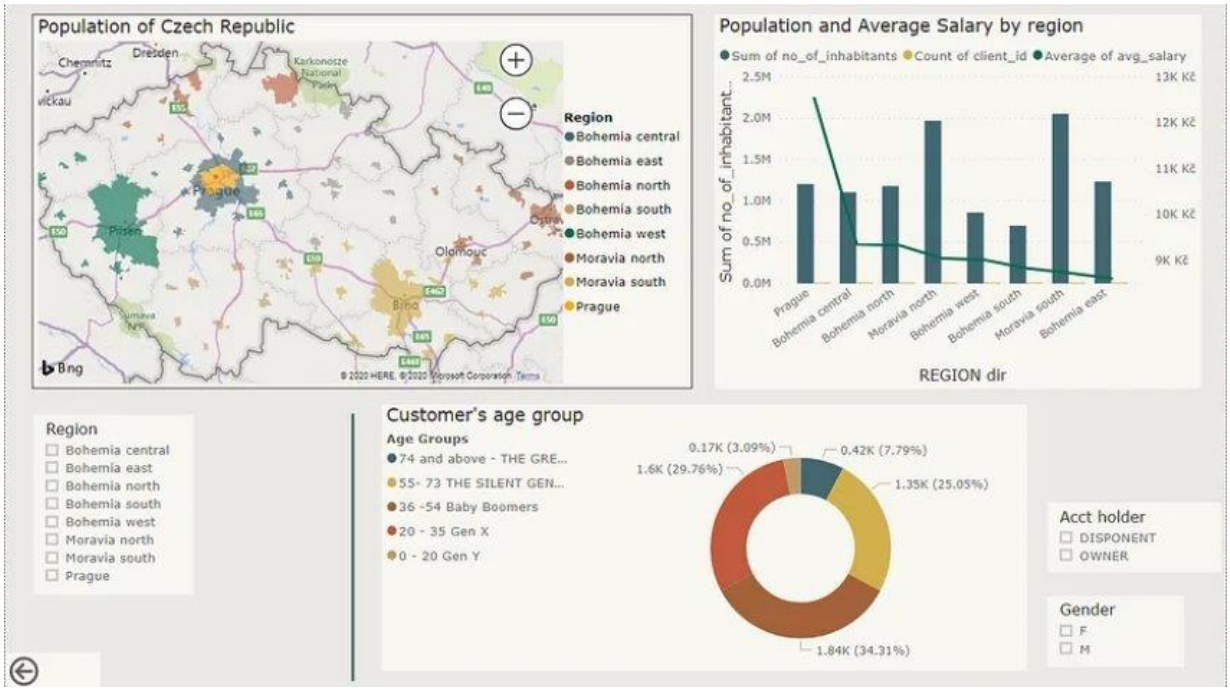
```
1 Credit Rating =
2 IF(loan[status]="A","GOOD",
3 IF(loan[status]="B","BAD",IF (loan[status]="c","GOOD","BAD" )))
```

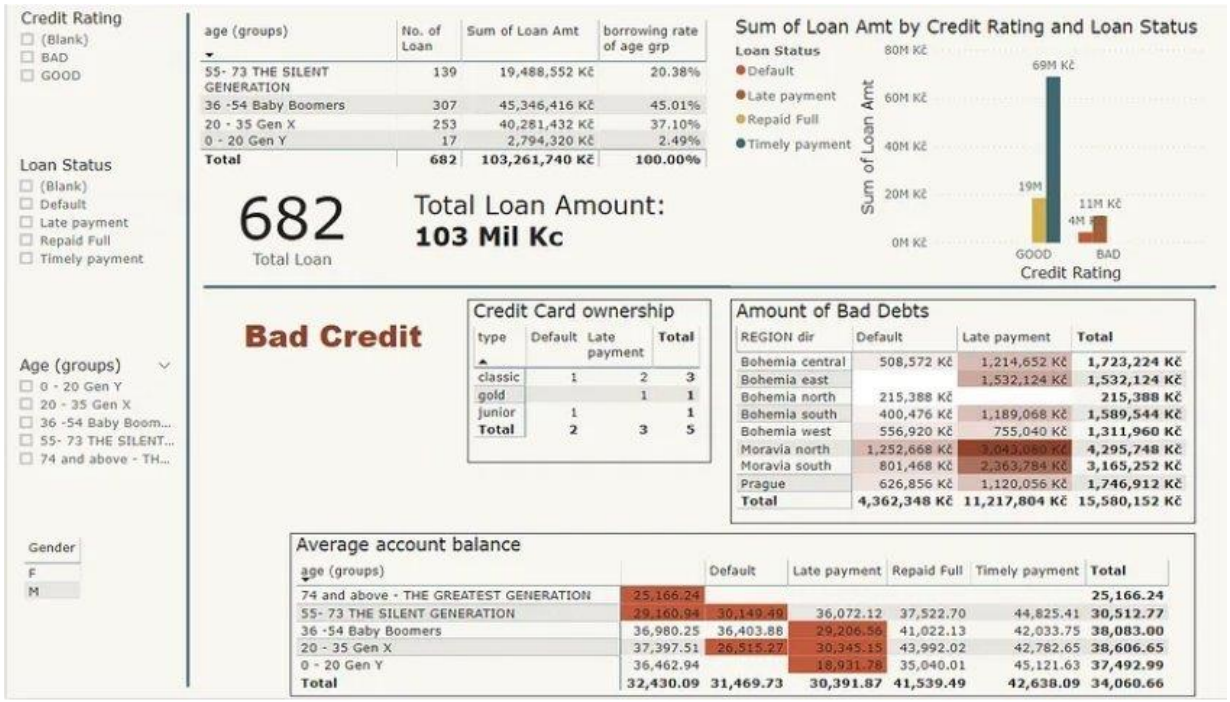
loan_id	account_id	date	Loan Amt	duration	payments	status	Credit Rating	Loan Status
5221	1284	981205	52,512 Kč	12	4376	C	GOOD	Timely payment
5841	4268	981104	41,988 Kč	12	3499	C	GOOD	Timely payment

Values of such as “account Id” have also been set as Text.

And District name have been categorized as place to be use for the map to show the sum of the inhabitants in each region.

4.5 Dashboard





age (groups)	No. of Loan	Sum of Loan Amt	borrowing rate of age grp
55- 73 THE SILENT GENERATION	139	19,488,552 Kč	20.38%
36 -54 Baby Boomers	307	45,346,416 Kč	45.01%
20 - 35 Gen X	253	40,281,432 Kč	37.10%
0 - 20 Gen Y	17	2,794,320 Kč	2.49%
Total	**682**	**103,261,740 Kč**	**100.00%**

type	Default	Late payment	Total
classic	1	2	3
gold		1	1
junior	1		1
Total	**2**	**3**	**5**

REGION dir	Default	Late payment	Total
Bohemia central	508,572 Kč	1,214,652 Kč	1,723,224 Kč
Bohemia east		1,532,124 Kč	1,532,124 Kč
Bohemia north	215,388 Kč		215,388 Kč
Bohemia south	400,476 Kč	1,189,068 Kč	1,589,544 Kč
Bohemia west	556,920 Kč	755,040 Kč	1,311,960 Kč
Moravia north	1,252,668 Kč	3,043,080 Kč	4,295,748 Kč
Moravia south	801,468 Kč	2,363,784 Kč	3,165,252 Kč
Prague	626,856 Kč	1,120,056 Kč	1,746,912 Kč
Total	**4,362,348 Kč**	**11,217,804 Kč**	**15,580,152 Kč**

age (groups)	Default	Late payment	Repaid Full	Timely payment	Total
74 and above - THE GREATEST GENERATION	25,166.24				25,166.24
55- 73 THE SILENT GENERATION	29,160.94	30,149.40	36,072.12	37,522.70	44,825.41	30,512.77
36 -54 Baby Boomers	36,980.25	36,403.88	29,206.56	41,022.13	42,033.75	38,083.00
20 - 35 Gen X	37,397.51	26,515.27	30,345.15	43,992.02	42,782.65	38,606.65
0 - 20 Gen Y	36,462.94		18,931.78	35,040.01	45,121.63	37,492.99
Total	**32,430.09**	**31,469.73**	**30,391.87**	**41,539.49**	**42,638.09**	**34,060.66**

CONCLUSION

The project real time analysis of “unemployment” using Power Bi has successfully demonstrated the potential of data analysis in the unemployment sector. The real time analysis of unemployment data has provided valuable insights into unemployment preference and trends thereby foliating informed decision making the interactive dashboards and reports have offered a comprehensive view of unemployment data enabling the identification of patterns correlations this has not only improved the efficiency of data analysis but also enhanced unemployment ability to provided unemployed.

FUTURE SCOPE

The future scope of this project is vast. With the advent of advanced analytics and machine learning, Power BI can be leveraged to predict future trends based on historical data. Integrating these predictive analytics into the project could enable the bank to anticipate customer needs and proactively offer solutions. Furthermore, Power BI's capability to integrate with various data sources opens up the possibility of incorporating more diverse datasets for a more holistic view of customers. As data privacy and security become increasingly important, future iterations of this project should focus on implementing robust data governance strategies. This would ensure the secure handling of sensitive customer data while complying with data protection regulations. Additionally, the project could explore the integration of real-time data streams to provide even more timely and relevant insights. This could potentially transform the way banks interact with their customers, leading to improved customer satisfaction and loyalty.

REFERENCES

https://github.com/sara-sara123/sarathi20/upload

LINK

https://github.com/githubtraining/hellogitworld.git

