Tech Saksham

Case Study Report

Data Analytics with Power BI

“AN ANALYSIS OF UNEMPLOYMENT IN REPUBLIC OF INDIA”

“Govt arts and science college”

NM ID	NAME
7D98109492F013F85CCF2557F945460	V. Mariyammal

ABSTRACT

The word unemployment belongs to a state in which a respective actively seeks employment but is unsuccessful. It is said to be one of the critical measures of the economy's strength. The unemployment rate is the most generally used method to arbitrate a country's unemployment rate. This can be found by honestly dividing the number of people without jobs by the total population covered in a nation's labor force. National and local governments often effort to offer employment convenience to secure people who meet the acceptability criteria set by them. Commonly, work is availed for groups of particular upon a fixed minimum wage sufficient for bare continuance and provides further chances for them to find permanent jobs. These attempts are made to develop the country's growths and cut down the overall unemployment rate. The rate of unemployment in India has been expanding over the years. The current paper purpose to interpret the element leading to unemployment and its impact on the Indian economy. The study focal point on how employment rate performance a vital role in overall advancement of the economy. The study makes use of secondary data sources and point out on the present outline of unemployment in rural and urban areas. The paper interprets how an increase in population, poverty, illiteracy, inflation and lack of full employment can lead to a downturn in the hike of the economy. The paper deliberate the issue faced by the economy due to high rate of unemployment and prescribe strategies to improve the current status of employment in the country.

INDEX

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Unemployment is the greatest challenge that India is facing today. Unemployment is a situation where the person willing to work, fails to find a job that earns him a living. The reasons for this unemployment situation are high population, defective education system, excessive burden on agriculture, low productivity in agricultural sector combined with lack of alternative opportunities for agricultural workers, unskilled workforce, etc. Though millions of students are pouring out of educational institutions every year, they are jobless. The supply of jobs is much less than the huge demand. The reason for this huge demand is the increase in population by leaps and bounds. This again gets directly linked up to poverty which aggravates the situation. Due to poverty, illiteracy has spread its harmful tentacles and in the course of time pushed the Indian youth into anti-social activities. Even if they become educated, the failure of getting job according to their capacity makes them hostile and thus results in strikes and protests. Big cities like Mumbai, Kolkata, Chennai all are pressurised with the influx of rural folk who come here in the hope of better prospects and as a result create pressure on the civic amenities.

1.2 Proposed Solution

The proposed solution is to develop a Power BI dashboard that can analyze and visualize unemployment data. Invest in public infrastructure projects that can create jobs in construction, maintenance , and related industries. These projects can stimulate economic activity and create employment opportunities. Support for small and medium-sized SMES often drive job creation.

1.3 Feature

- **Analysis of Unemployment**: The dashboard will provide unemployment data.
- **Unemployed Segmentation**: It will segment unemployment people based on various parameters like age, gender, period, etc.
- **Trend Analysis**: The dashboard will identify and display trends in people's unemployment.
- **Predictive Analysis**: It will use historical data.

1.4 Advantages

Unemployment advantages enable workers to maintain consumption while spending more time searching for a job fitting their skills. Unemployment benefits provide additional support workers during recessions, without large negative side effects.

1.5 Scope

The scope of this project extends to unemployment of India . Analyzing unemployment in India involves examining various factors such as the types of unemployment (structural, frictional, cyclical), regional disparities, demographic trends, education and skill levels, government policies, labor force participation rates, informal sector employment, and the impact of technological advancements. Additionally, understanding the effects of globalization, economic growth rates, and social factors on employment dynamics is crucial.

CHAPTER 2

SERVICES AND TOOLS REQUIRED

2.1 Services Used

- **Data Collection and Storage Services**: Unemployment need to collect and store 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 Sage Maker 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 Power BI, 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.

Software Requirements:

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

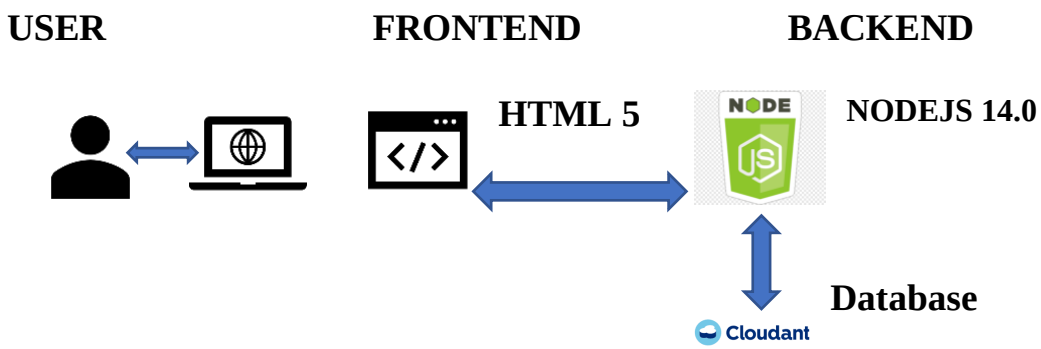
- **Power BI Service**: This is an online (Software as a Service) service that you use to publish reports, create new dashboards, and share insights.

- **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 SageMaker. 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 PowerBI. PowerBI 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

Manage relationship:

If you are losing respect for an unemployed husband or are feeling unemployed wife resentment, it's time to reconsider your thoughts.

Is there anything you can do to help your spouse? Yes!

- You can lovingly help them look for jobs that interest them.
- You can look over their resume to make sure they are presenting themselves in the best way possible
- You can give them personal space to deal with their job loss grief Sometimes all your unemployed spouse needs to hear is that your there for them. They don't need you to find them a new job or solve all their problems. They just need to know that you're there whenever they need to talk.

Modelling for Gender and Age data

Notice that the Gender and age of the client are missing from the data. These can be formulated from the birth number YYMMDD where at months (the 3rd and 4th digits) greater than 50 means that client is a Female. We can create a column for Gender.

```
1  Gender =
2  VAR stringDate = FORMAT(client[birth_number],"General Number")
3  VAR month = VALUE(MID(stringDate,3,2))
4  RETURN IF(month > 50,"F","M")
5
```

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
734	855114	46	F	14/01/1985	15

For birthday, we need to reduce the birth month of the female by 50 and then change the date format to 21/04/2004 adding 365 to the year.


```
1  Birthday =
2  VAR stringDate = FORMAT(client[birth_number],"General Number")
3  VAR stringMonth = VALUE(MID(stringDate,3,2))
4  VAR mth = IF(stringMonth > 50, stringMonth - 50,stringMonth)
5  VAR year = VALUE(MID(stringDate,1,2))
6  VAR day = VALUE(MID(stringDate,5,2))
7  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

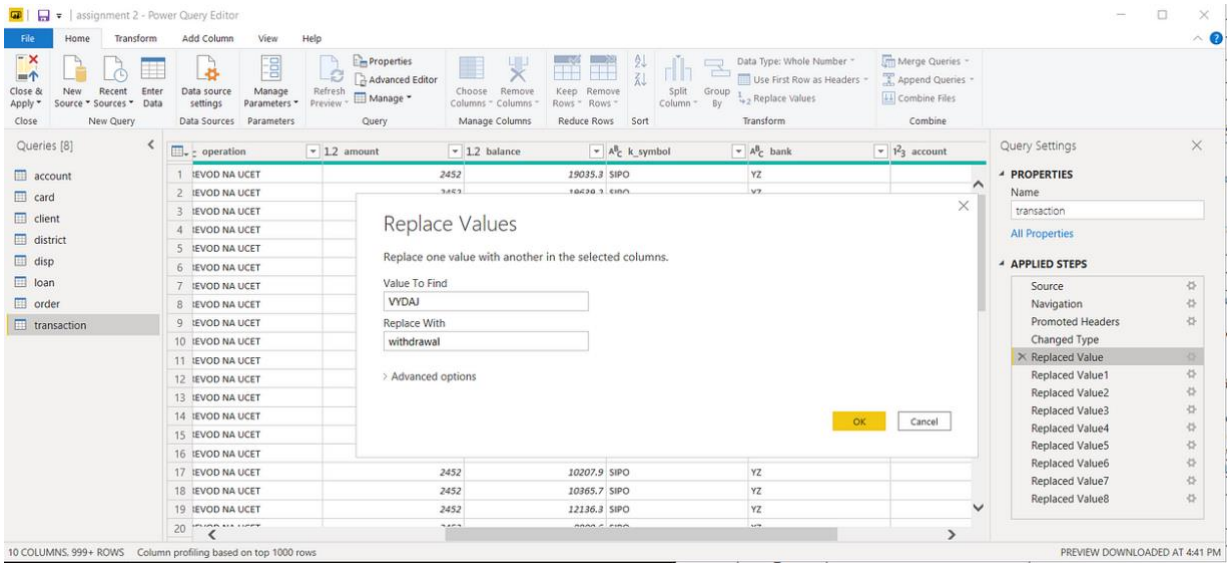
For Age, we shall assume it is year 1999 as explain previously and use it to minus from the birth year.

```
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

Replacing values

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

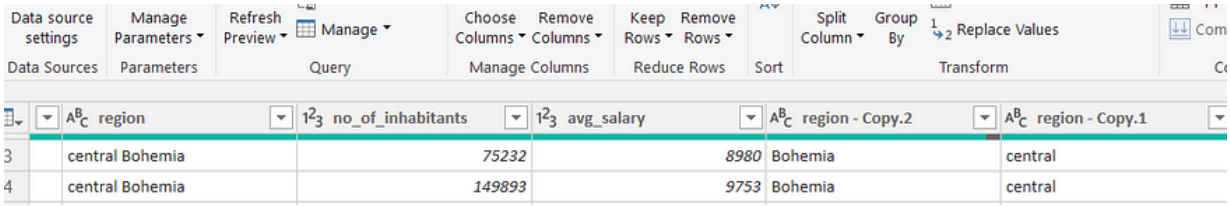


type	+/- transaction	"PRIJEM" stands for credit "VYDAJ" stands for withdrawal

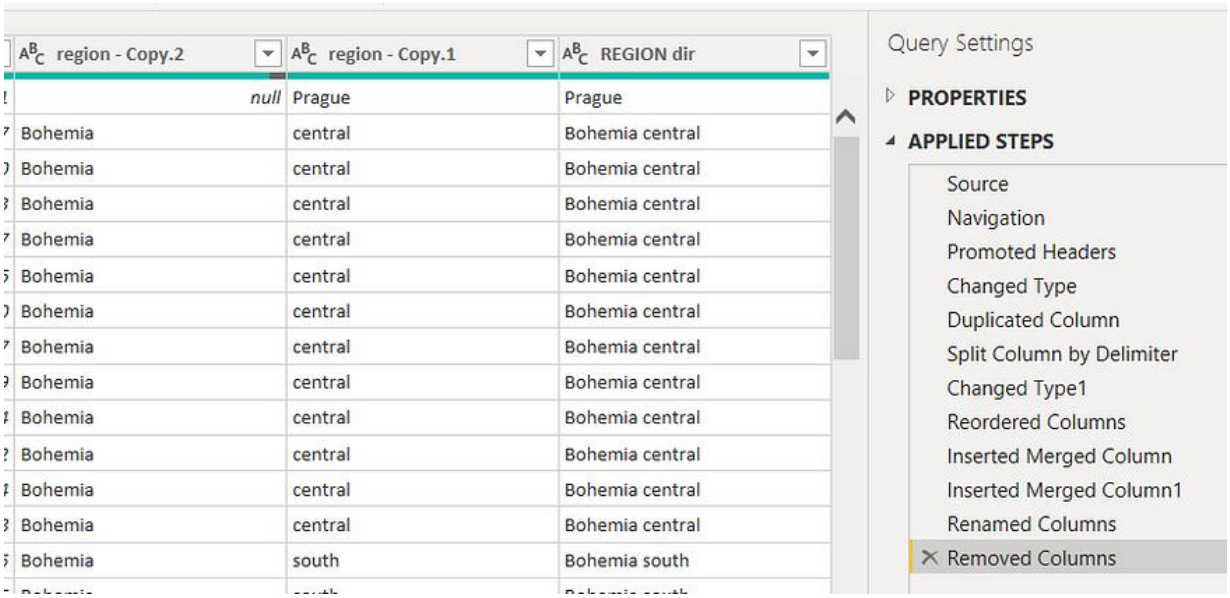
k_symbol	characterization of the transaction	"POJISTNE" stands for insurance payment "SLUZBY" stands for payment for statement "UROK" stands for interest credited "SANKC. UROK" sanction interest if negative balance "SIPO" stands for household "DUCHOD" stands for old-age pension "UVER" stands for loan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



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.

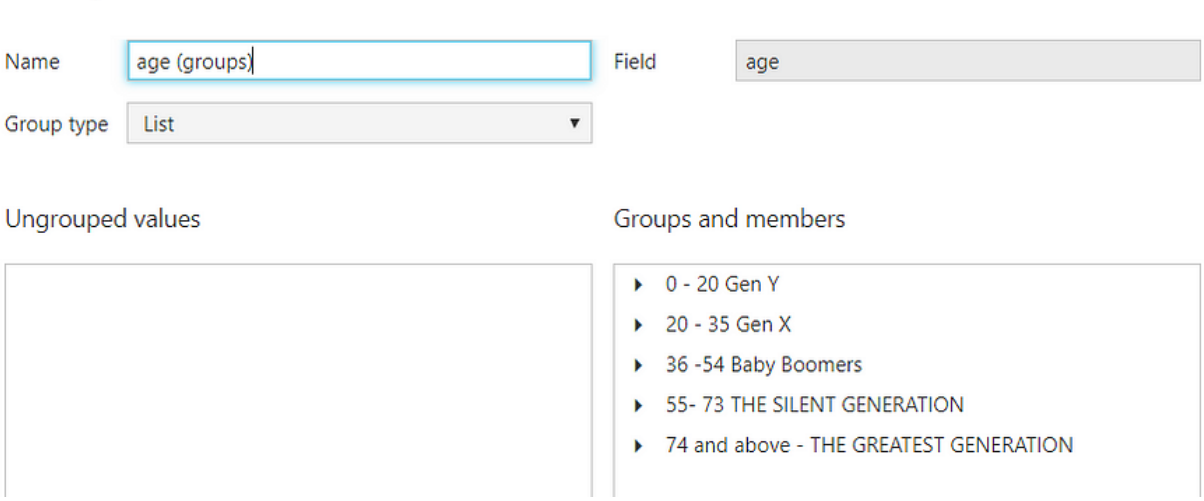
The Gen Y are youths,

Gen X are young working adults, some starting their families

Baby Boomer are working adults with families.

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

The greatest Generation, retired elderly living on pensions.



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

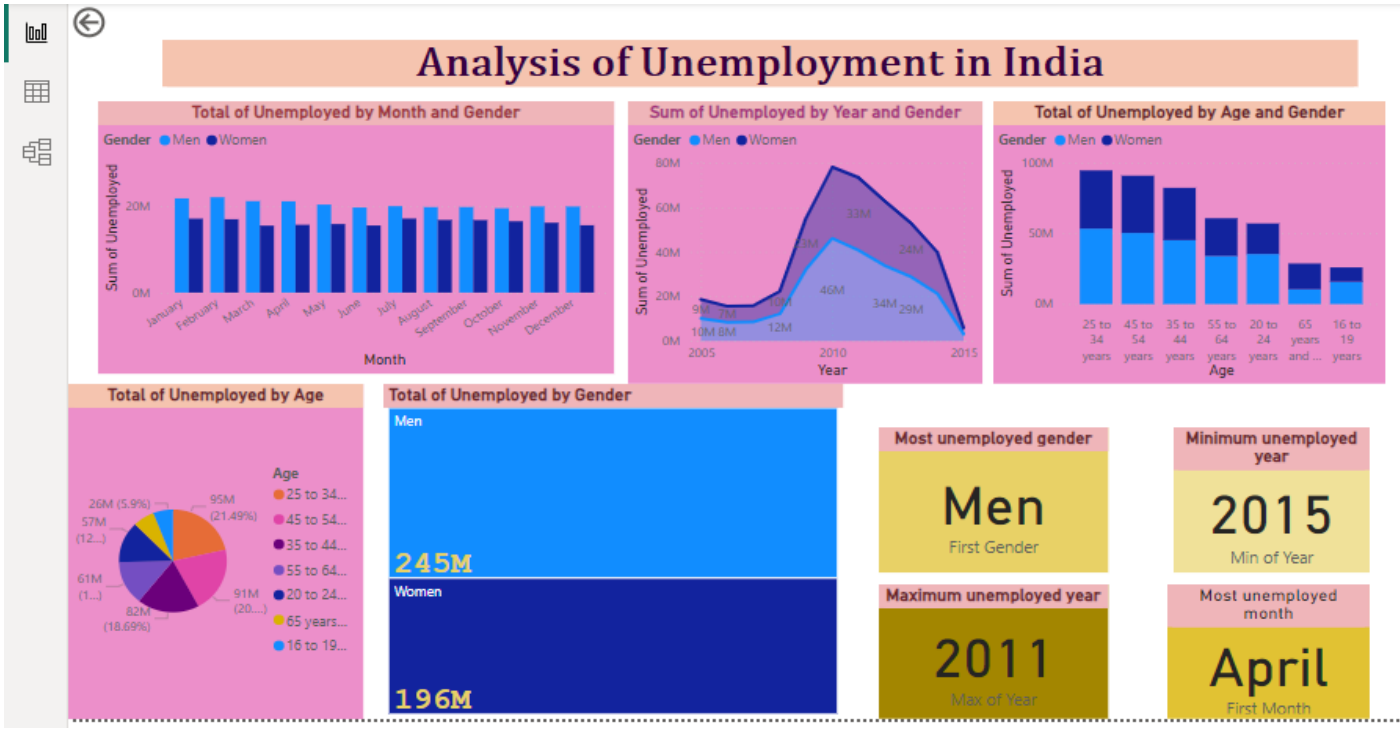
```
1 Credit Rating =
2 IF(loan[status]="A","GOOD",
3 IF(loan[status]="B","BAD",IF (loan[status]="c","GOOD","BAD" )))
```

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.

Dashboard



Labor Data for Aridia

Year	2010	2011	2012
Adult population	2,000	3,000	3,200
Number of employed	1,400	1,300	1,600
Number of unemployed	200	600	200

CONCLUSION

The unemployment rate in India inched higher to 7.45% in February 2023, taking the total number of unemployed in the country to 33 million. India needs to make dedicated efforts in order to decrease the number of unemployed people.

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 .

REFERENCES

https://www.scribd.com/document/287189229/184415093-Project-About-Unemployment-in-India

LINK

https://github.com/Mariyammal2004/Mariyammal

