

Tech Saksham

Case Study Report

Data Analytics with Power BI

“An Analysis of Unemployment in republic of India”

“Shrimati Indira Gandhi College Trichy”

NM ID	NAME
70C4DEB840E7D59876145E164B2CD9E1	R.RAJARAJESHWARI

Trainer Name : Umamaheshwari

ABSTRACT

Unemployment remains a critical issue affecting the socioeconomic landscape of the Republic of India. This project aims to conduct an in-depth analysis of unemployment trends in India, exploring its various dimensions, causes, and implications. By utilizing a multidisciplinary approach, encompassing economic, sociological, and policy perspectives, this study seeks to offer valuable insights into the dynamics of unemployment in India. The project will begin by examining the historical trends of unemployment in India, tracing its evolution over time and identifying key turning points. It will analyse the different types of unemployment prevalent in the country, including structural, cyclical, and frictional unemployment, and their respective impacts on the labour market. Further more, this study will delve into the underlying causes of unemployment in India, considering factors such as rapid population growth, technological advancements, skill mismatches, and labour market rigidities. Special attention will be paid to the impact of the COVID -19 pandemic on unemployment, including its short -term disruptions and long -term repercussions on employment patterns.

INDEX

Sr. No.	Table of Contents	Page No.
1	Chapter 1: Introduction	4
2	Chapter 2: Services and Tools Required	6
3	Chapter 3: Project Architecture	7
4	Chapter 4: Modeling and Result	9
5	Conclusion	18
6	Future Scope	19
7	References	20
8	Links	21

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Despite being one of the fastest-growing economies globally, India continues to grapple with persistent unemployment challenges. The problem statement for this study revolves around understanding the multifaceted nature of unemployment in India, including its historical trends, causes, demographic disparities, and the effectiveness of existing government policies. By addressing these issues, the study aims to provide insights that can inform evidence-based policy interventions to mitigate unemployment and foster inclusive economic growth in the Republic of India.

1.2 Proposed Solution

To address the unemployment challenges in India, a multifaceted approach is necessary. Firstly, there needs to be a focus on enhancing skill development programs to bridge the gap between the skills demanded by the labour market and those possessed by the workforce. This can be achieved through collaboration between educational institutions, industry stakeholders, and government bodies. Secondly, targeted employment generation schemes should be implemented, especially in sectors with high growth potential such as renewable energy, infrastructure development, and technology-driven industries. These initiatives can create job opportunities while also addressing pressing societal needs. Thirdly, there is a need for labour market reforms aimed at reducing regulatory barriers and promoting entrepreneurship. Simplifying bureaucratic processes for starting and running businesses can encourage job creation and promote a more dynamic and resilient labour market.

1.3 Feature

- **Skill Development Programs:** Collaborative initiatives between educational institutions, industries, and government bodies.
- **Employment Generation Schemes:** Targeted investment in sectors with high growth potential.
- **Labor Market Reforms:** Simplification of bureaucratic processes for starting and running businesses.

1.4 Advantages

- **Economic Growth:** Effective measures to address unemployment can stimulate economic growth by harnessing the untapped potential of the workforce, leading to increased productivity and innovation.
- **Poverty Reduction:** Creating employment opportunities enables individuals and families to generate income, lifting them out of poverty and contributing to overall socioeconomic development.
- **Social Stability:** Lower unemployment rates foster social stability by reducing income inequality, mitigating social tensions, and enhancing community well-being.

1.5 Scope

The scope for implementing solutions to address unemployment challenges in India is extensive and multifaceted. It involves the development of a comprehensive national policy framework that encompasses various dimensions of unemployment, including skill development, entrepreneurship promotion, labor market reforms, and social inclusion. Additionally, there is a need for regional integration to tailor solutions according to regional disparities in employment opportunities, leveraging local resources and industries. Public-private partnerships are essential for fostering collaboration between the government, private sector, civil society organizations, and academia to design and implement effective employment generation programs and initiatives.

CHAPTER 2

SERVICES AND TOOLS REQUIRED

2.1 Services Used

- **Data Sources:** Power BI can connect to a wide range of data sources, including databases, Excel spreadsheets, cloud-based services (e.g., Salesforce, Google Analytics), and onpremises data sources.
- **Power Query Editor:** This component of Power BI allows you to import, transform, and clean data from various sources.
- **Data Modeling:** Power BI Desktop provides a data modeling environment where you can create relationships between different tables or data sources, enabling you to analyze unemployment data from multiple perspectives.

2.2 Tools and Software used Tools:

- **Power BI Desktop:** This is the primary software tool used for data modeling, data transformation, and creating interactive reports and visualizations. Power BI Desktop is a free application that can be installed on your local machine.
- **Power Query Editor:** This is a component within Power BI Desktop that allows you to import, transform, and clean data from various sources. It provides a user-friendly interface for data preparation tasks.
- **Power BI Service:** This is the cloud-based service offered by Microsoft that allows you to publish your Power BI reports and dashboards, and share them with others. It also provides collaboration features and integration with other Microsoft

Software Equipment:

- **Power BI Desktop (Latest Version):** This is the primary software tool required for creating data models, transforming data, and building interactive reports and

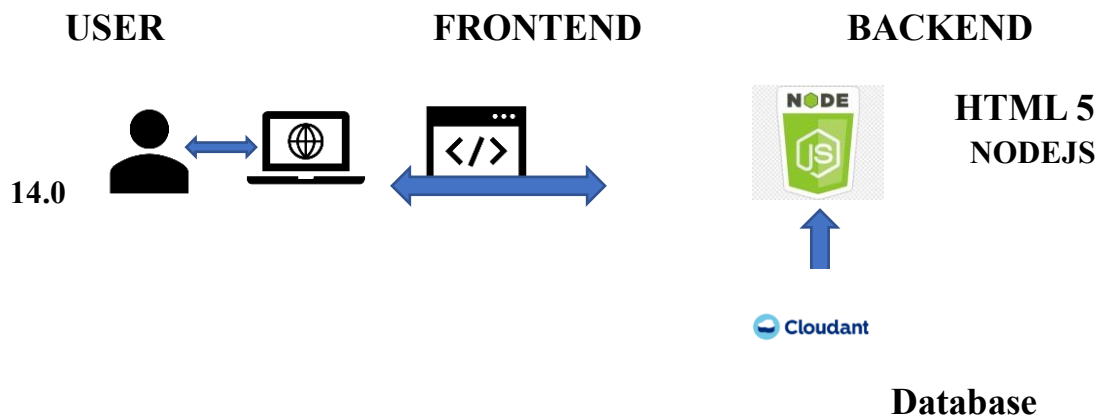
visualizations. Power BI Desktop is available as a free download from the Microsoft website.

- **Operating System:** Power BI Desktop is supported on Windows 10 or later versions. For Mac users, Power BI Desktop can be run on a Windows virtual machine or using cloud-based solutions like Power BI Service.
- **Microsoft Office (Optional):** If you plan to use Excel files as data sources, you may need Microsoft Office (Excel) installed on your machine.

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 PowerBI can be accessed through PowerBI Desktop, PowerBI Service (online), and PowerBI Mobile.

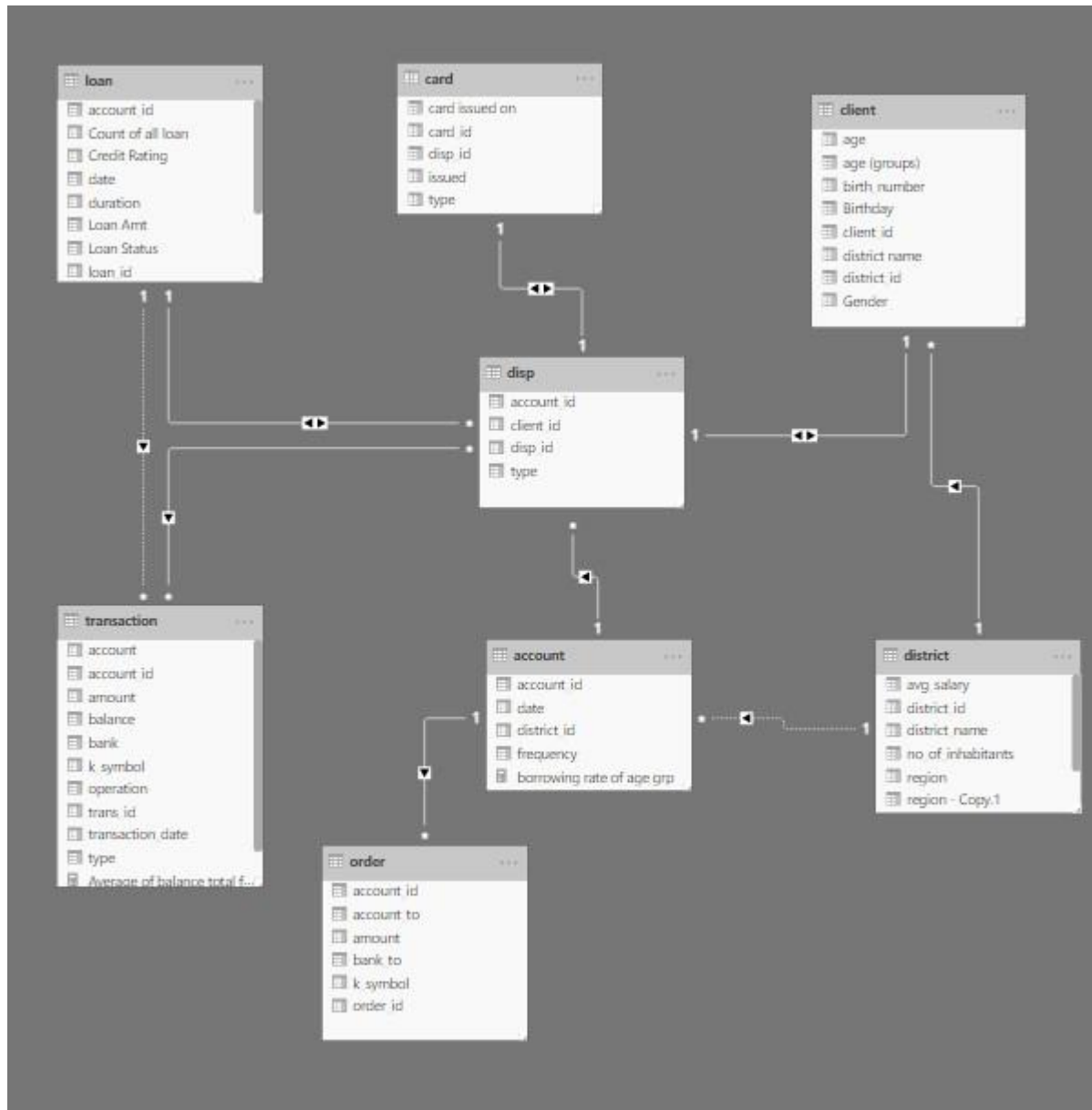
This architecture provides a comprehensive solution for real-time analysis of bank customers. However, it's important to note that the specific architecture may vary depending on the bank's existing infrastructure, specific requirements, and budget. It's also important to ensure that all tools and services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

CHAPTER 4

MODELING AND RESULT

Manage relationship

The “disp” file will be used as the main connector as it contains most key identifier (account id, client id and disp id) which can be use to relates 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)
<input checked="" type="checkbox"/>		card (disp_id)	disp (disp_id)
<input checked="" type="checkbox"/>		client (district_id)	district (district_id)
<input checked="" type="checkbox"/>		disp (account_id)	account (account_id)
<input checked="" type="checkbox"/>		disp (account_id)	loan (account_id)
<input checked="" type="checkbox"/>		disp (client_id)	client (client_id)
<input checked="" type="checkbox"/>		order (account_id)	account (account_id)
<input checked="" type="checkbox"/>		transaction (account_id)	disp (account_id)
<input type="checkbox"/>		account (district_id)	district (district_id)
<input type="checkbox"/>		transaction (account_id)	loan (account_id)

Edit relationship

Select tables and columns that are related.

card ▼

card_id	disp_id	type	issued	card issued on
1005	9285	classic	931107	Sunday, 7 November 1993
104	588	classic	940119	Wednesday, 19 January 1994
747	4915	classic	940205	Saturday, 5 February 1994

disp ▼

disp_id	client_id	account_id	type
1	1	1	OWNER
2	2	2	OWNER
4	4	3	OWNER

Cardinality

One to one (1:1) ▼

Cross filter direction

Both

- ☒ Make this relationship active
- ☐ Apply security filter in both directions
- ☐ Assume referential integrity

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
724	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 DD/MM/YYYY adding 1900 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.

<div> <div>✕</div> <div>✓</div> </div>		<pre> 1 age = 1999 -RIGHT(client[Birthday],4) </pre>				
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.

region	no_of_inhabitants	avg_salary	region - Copy.2	region - Copy.1
central Bohemia	75232	8980	Bohemia	central
central Bohemia	149893	9753	Bohemia	central

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

A ^B _C region - Copy.2	A ^B _C region - Copy.1	A ^B _C REGION dir
	Prague	Prague
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	central	Bohemia central
Bohemia	south	Bohemia south

Query Settings

- PROPERTIES
- APPLIED STEPS
 - Source
 - Navigation
 - Promoted Headers
 - Changed Type
 - Duplicated Column
 - Split Column by Delimiter
 - Changed Type1
 - Reordered Columns
 - Inserted Merged Column
 - Inserted Merged Column1
 - Renamed Columns
 - Removed Columns

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.

Groups

Name	age (groups)	Field	age
Group type	List		

Ungrouped values

Groups and members

- ▶ 0 - 20 Gen Y
- ▶ 20 - 35 Gen X
- ▶ 36 - 54 Baby Boomers
- ▶ 55- 73 THE SILENT GENERATION
- ▶ 74 and above - THE GREATEST GENERATION

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

X
✓

```

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

X
✓

```

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

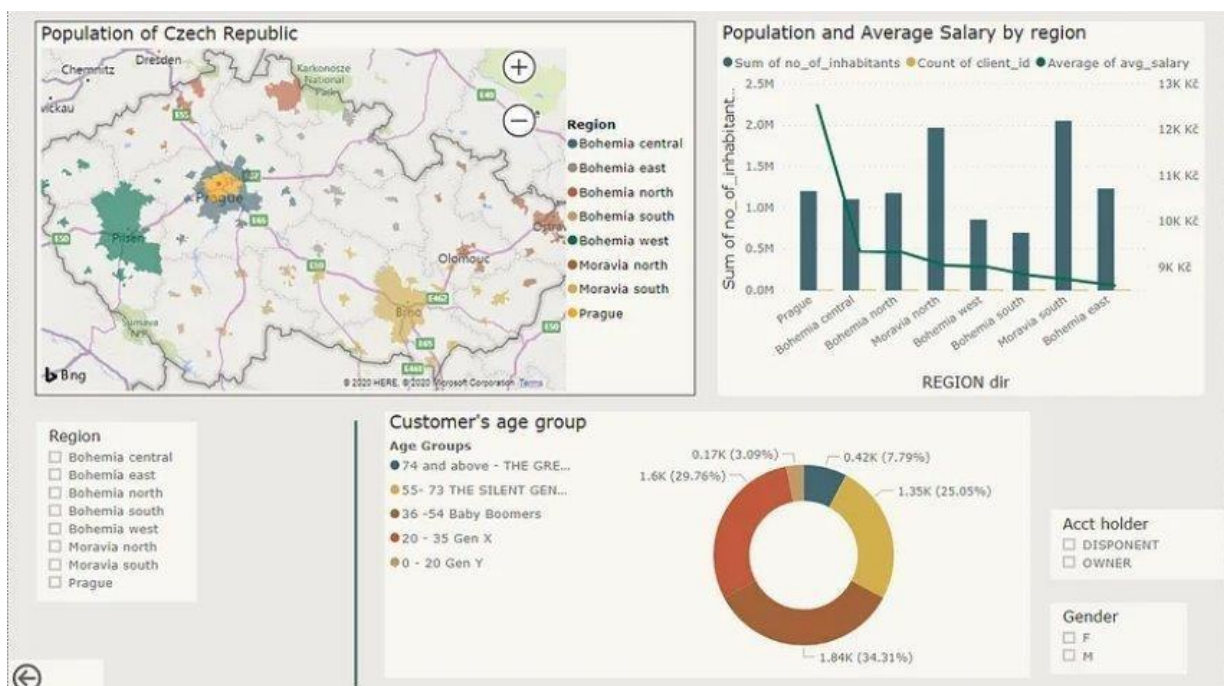
```

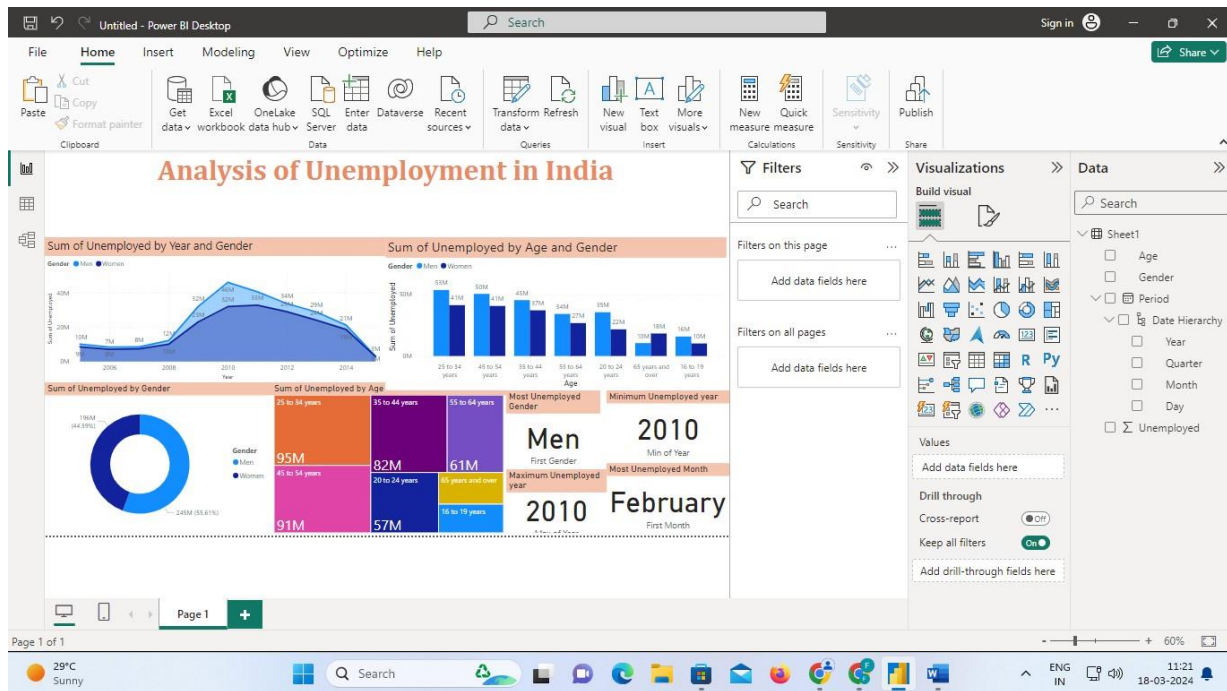
loan_id	account_id	date	Loan Amt	duration	payments	status	Credit Rating	Loan Status
5221	1284	981205	52,512 Kč	12	4376	C	GOOD	Timely payment
5841	4268	981104	41,988 Kč	12	3499	C	GOOD	Timely payment

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





CONCLUSION

An unemployment analysis in India using Power BI could uncover valuable insights into overall trends, regional variations, demographic factors impacting unemployment rates, and industry or sector vulnerabilities. Visualizing the data may reveal patterns over time, areas with high unemployment needing targeted interventions, and demographic groups disproportionately affected - informing policy decisions. The analysis could also highlight potential skill mismatches between education levels and job market demands, allowing reforms to vocational training. By incorporating economic indicators and policy changes, Power BI's modeling capabilities enable evaluating the effectiveness of government initiatives and forecasting future unemployment scenarios. Benchmarking India's situation against other countries could identify best practices to emulate.

FUTURE SCOPE

The future scope of an unemployment analysis in India using Power BI is vast and promising. As more data becomes available from government agencies, surveys, and other sources, the analysis can be extended to incorporate additional factors influencing unemployment, such as socioeconomic indicators, migration patterns, and changing workforce demographics. Power BI's integration with advanced analytics tools like Azure Machine Learning could enable predictive modeling and forecasting of unemployment trends based on various economic scenarios. The analysis could be expanded to explore the gig economy and its impact on traditional employment metrics. Furthermore, Power BI's natural language capabilities can be leveraged to create conversational interfaces, allowing stakeholders to interact with the unemployment data using intuitive voice or text queries.