Tech saksham

Case Study Report

Data analtic with power BI

“Real-Time Analysis of Bank Customers”

“Seared Heart College of Arts and Sicence Dindigul”

NM ID	NAME
F03D89A82929EEBABE7482809BEEFB62	T.G.SANTHALAKSHMI

R.UMAMAHESWARI	Trainer Name
R.UMAMAHESWARI	Master Trainer

ABSTRACT

The successful implementation of the "SUPPLY CHAIN OF ANALYTICS" project, powered by PowerBI, highlights the significant impact of data analytics in the sector. By analyzing customer data throughout the supply chain, valuable insights into customer behavior, product preferences, sales trends, discount strategies, and total revenue generation have been unearthed. The interactive dashboards and reports provided a comprehensive overview of customer data, enabling the identification of critical patterns and correlations. This not only streamlines data analysis but also empowers the organization to deliver personalized services, identify cross-selling and up-selling opportunities, and tailor products and services to meet specific customer needs. Additionally, this project contributes to the broader objective of digital transformation in the banking sector, fostering efficiency, innovation, and customer-centricity.

INDEX

CHAPTER 1

INTRODUCTION

- **Problem Statement**

In today's competitive banking landscape, understanding customer behavior and preferences is crucial for customer retention and revenue generation. However, banks often face challenges in analyzing customer data due to the sheer volume and velocity of data generated. Traditional data analysis methods are time-consuming and often fail to provide real-time insights. This lack of real-time analysis can lead to missed opportunities for customer engagement, cross-selling, and up-selling, impacting the bank's revenue generation and customer satisfaction. Furthermore, the complexity and diversity of customer data, which includes transaction history, customer feedback, and demographic data, pose additional challenges for data analysis.

- **Proposed Solution**

The proposed solution is to develop a PowerBI dashboard that can analyze and visualize real-time customer data. The dashboard will integrate data from various sources such as transaction history, customer feedback, and demographic data. It will provide a comprehensive view of customer behavior, preferences, and trends, enabling banks to make informed decisions. The dashboard will be interactive, user-friendly, and customizable, allowing banks to tailor it to their specific needs. The real-time analysis capability of the dashboard will enable banks to respond promptly to changes in customer behavior or preferences, identify opportunities for cross-selling and up-selling, and tailor their products and services to meet customer needs.

• 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.

• Advantages

- **Data-Driven Decisions**: Banks can make informed decisions based on real-time data analysis.
- **Improved Customer Engagement**: Understanding customer behavior and trends can help banks engage with their customers more effectively.
- **Increased Revenue**: By identifying opportunities for cross-selling and up-selling, banks can increase their revenue.

• 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:

- **PowerBI**: 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.

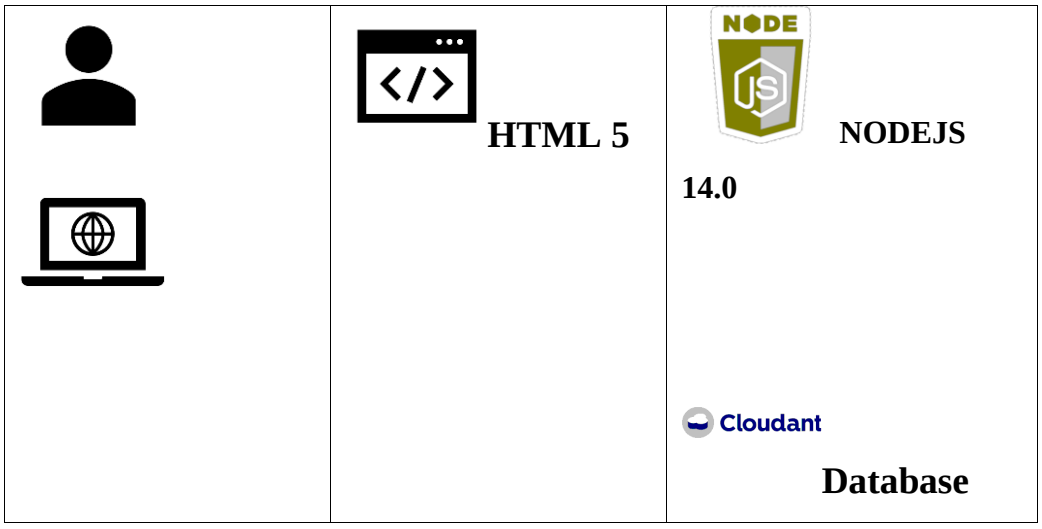
Software Requirements:

- **PowerBI Desktop**: This is a Windows application that you can use to create reports and publish them to PowerBI.
- **PowerBI Service**: This is an online SaaS (Software as a Service) service that you use to publish reports, create new dashboards, and share insights.
- **PowerBI 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:

- **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.
- **Data Storage**: The collected data is stored in a database for processing. Azure SQL Database or AWS RDS can be used for this purpose.
- **Data Processing**: The stored data is processed in real-time using services like Azure Stream Analytics or AWS Kinesis Data Analytics.
- **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.
- **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.
- **Data Access**: The dashboards created in PowerBI can be accessed through PowerBI Desktop, PowerBI Service (online), and PowerBI Mobile.

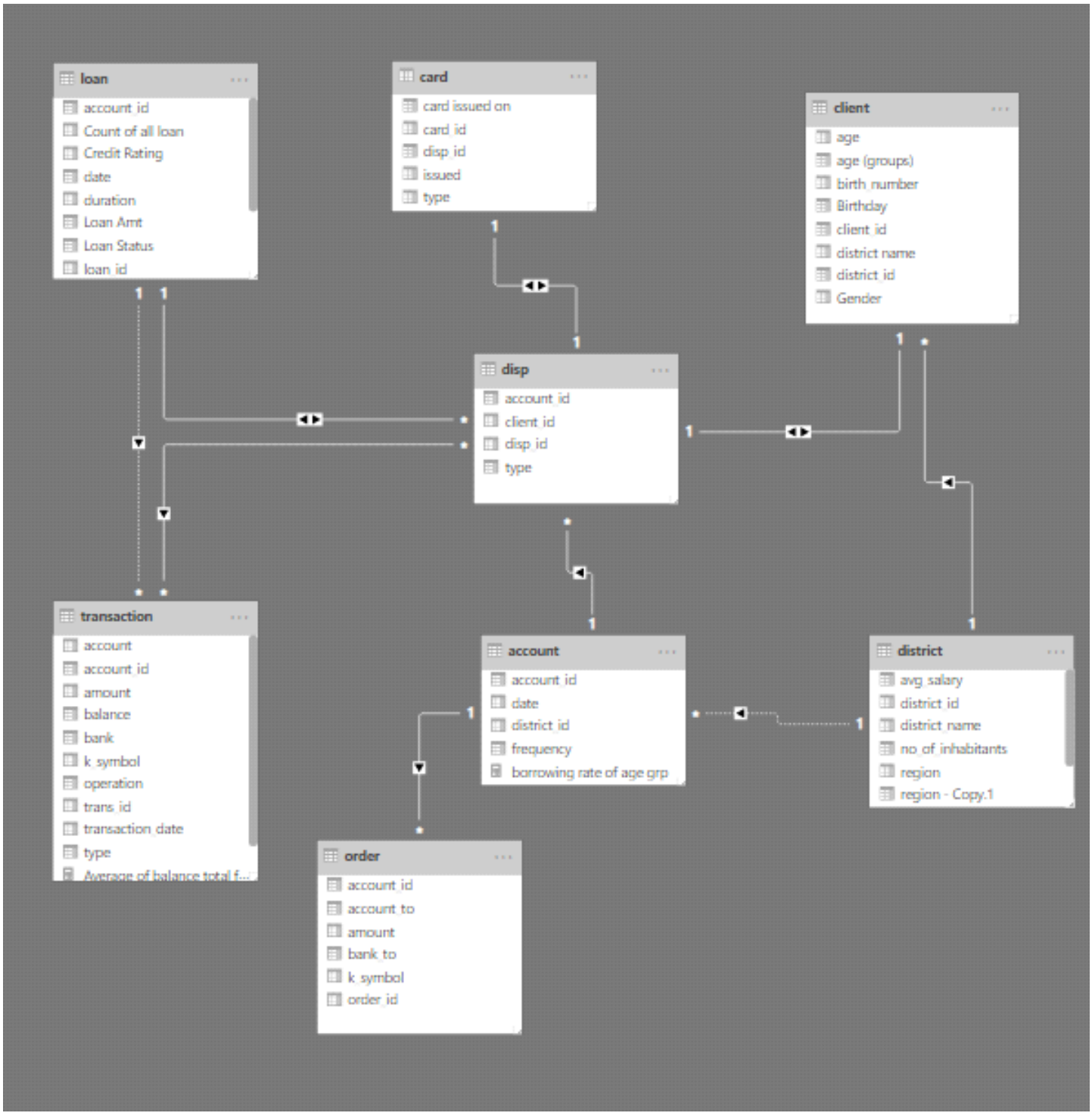
This architecture provides a comprehensive solution for real-time analysis of bank customers. However, it's important to note that the specific architecture may vary depending on the bank's existing infrastructure, specific requirements, and budget. It's also important to ensure that all tools and services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


loan
account_id
Count of all loan
Credit Rating
date
duration
Loan Amt
Loan Status
loan_id
card
card issued on
card_id
disp_id
issued
type
client
age
age (groups)
birth number
Birthday
client_id
district name
district_id
Gender
disp
account_id
client_id
disp_id
type
transaction
account
account_id
amount
balance
bank
k_symbol
operation
trans_id
transaction date
type
Average of balance total f...
account
account_id
date
district_id
frequency
borrowing rate of age grp
district
avg_salary
district_id
district name
no_of_inhabitants
region
region - Copy.1
order
account_id
account_to
amount
bank_to
k_symbol
order_id
1
*

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)

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
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 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

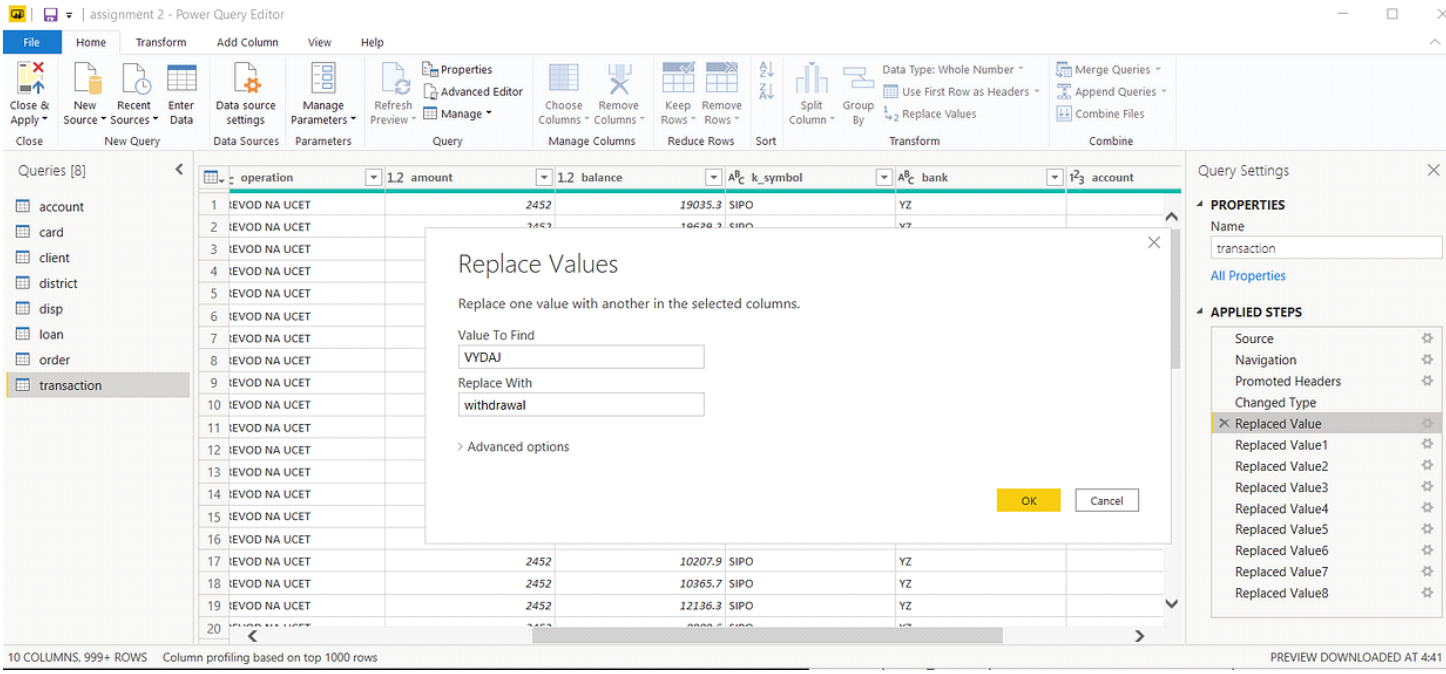
For Age, we shall assume it is year 1999 as explain previously and use it to minus from the birth year.

```
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.

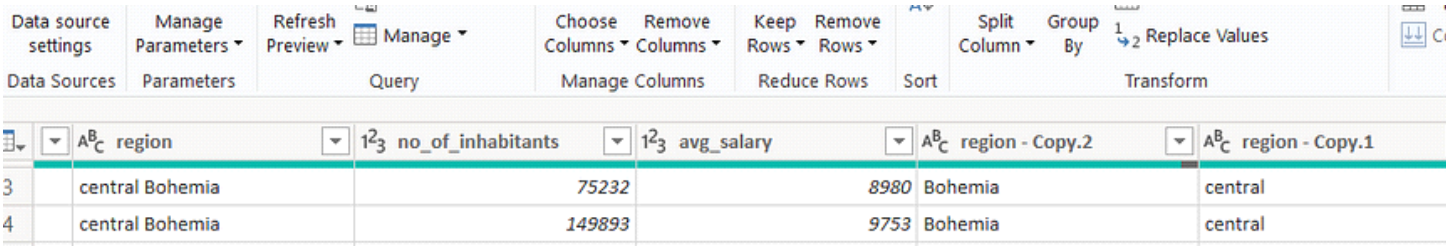


type	+/- transaction	"PRIJEM" stands for credit "VYDAJ" stands for withdrawal

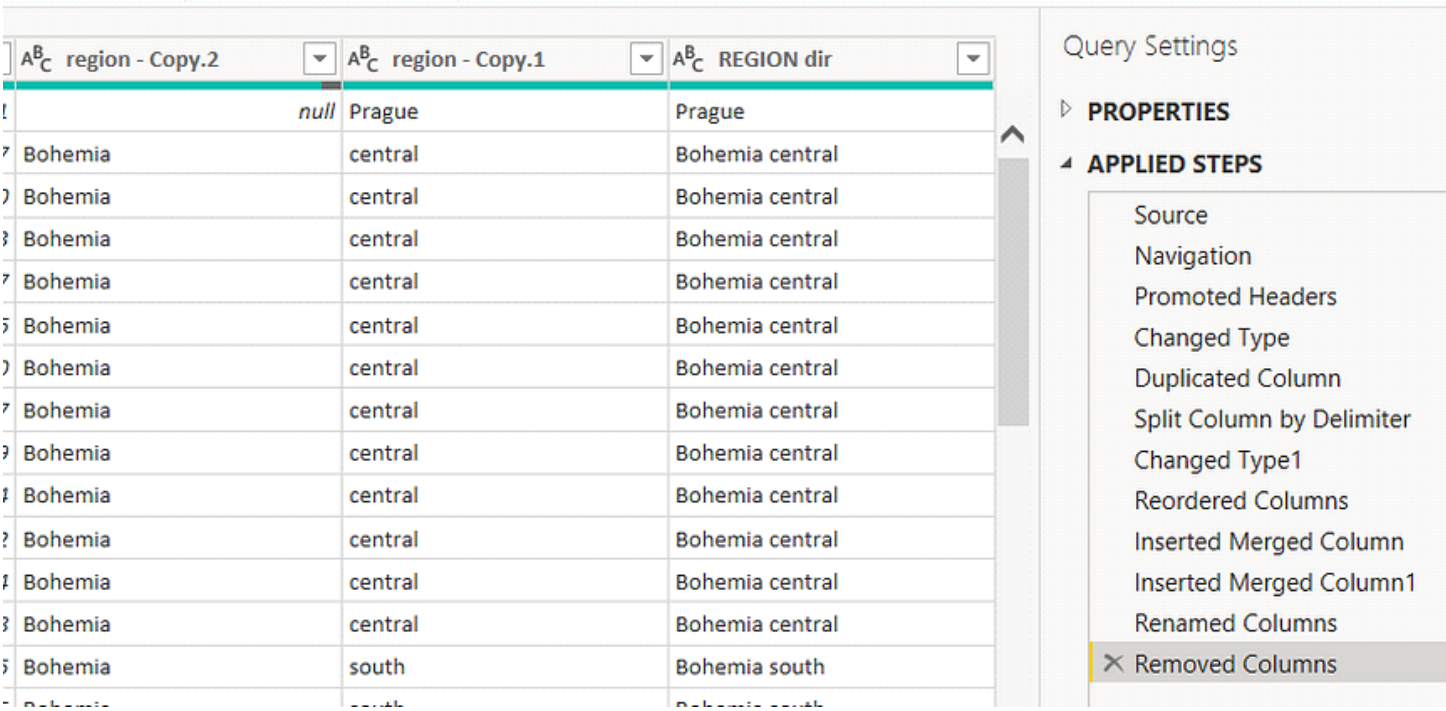
k_symbol	characterization of the transaction	"POJISTNE" stands for insurance payment "SLUZBY" stands for payment for statement "UROK" stands for interest credited "SANKC. UROK" sanction interest if negative balance "SIPO" stands for household "DUCHOD" stands for old-age pension "UVER" stands for loan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

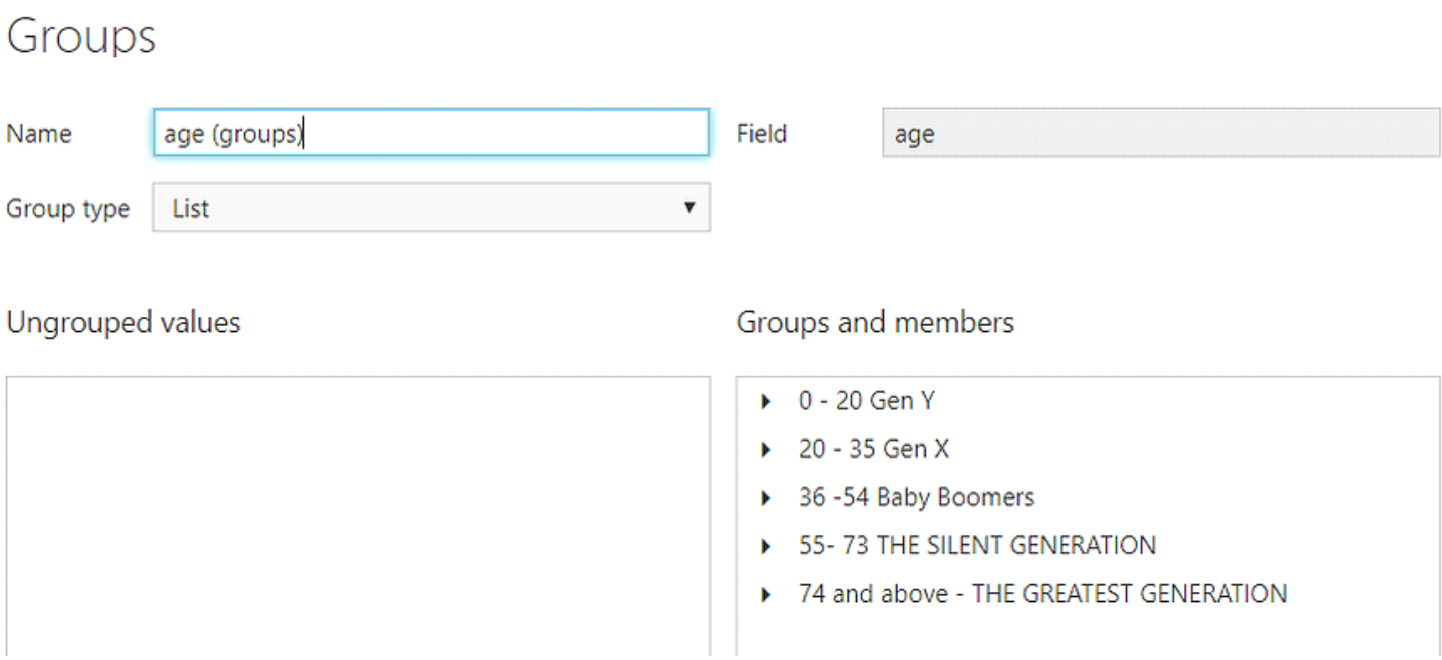
The Gen Y are youths,

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

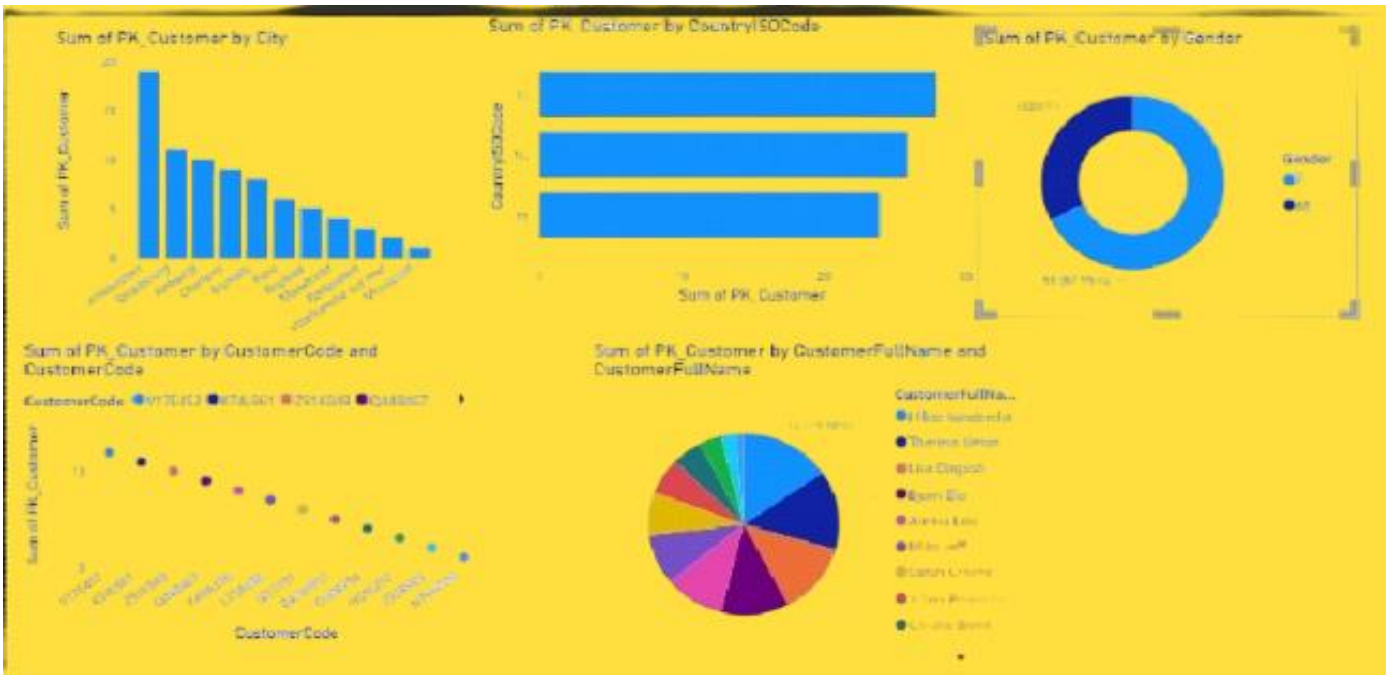
```
1 Credit Rating =
2 IF(loan[status]="A","GOOD",
3 IF(loan[status]="B","BAD",IF (loan[status]="c","GOOD","BAD" )))
```

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


Sum of PK_Customer by City
Sum of PK_Customer by CountryISOCode
Sum of PK_Customer
Sum of PK_Customer by Gender
Gender
Sum of PK_Customer by CustomerCode and CustomerCode
CustomerCode
Sum of PK_Customer by CustomerFullName and CustomerFullName

CONCLUSION

The successful execution of the "SUPPLY CHAIN OF ANALYTICS" project using PowerBI has effectively demonstrated the immense value of data analytics in the sector. By analyzing customer data across the supply chain, valuable insights into customer behavior, product preferences, sales trends, discount patterns, and total revenue have been gleaned. The interactive dashboards and reports have provided a comprehensive understanding of customer data, enabling the identification of patterns and correlations. This has not only optimized data analysis processes but has also bolstered the product's capacity to deliver personalized services to its customers. Moreover, the project has underscored the pivotal role of data visualization in simplifying complex data, making it more accessible and understandable. Leveraging PowerBI has enabled the presentation of data in visually appealing formats, facilitating better decision-making and enhancing product selling strategies.

FUTURE SCOPE

Moving forward, the project holds vast potential. As advanced analytics and machine learning continue to evolve, PowerBI can be utilized to forecast future trends based on historical data. By integrating predictive analytics, the bank can anticipate customer needs and offer proactive solutions. Moreover, PowerBI's ability to integrate with various data sources allows for a more comprehensive understanding of customers. To address growing concerns around data privacy and security, future iterations should focus on implementing rigorous data governance strategies. Additionally, exploring the integration of real-time data streams could provide even more timely insights, potentially transforming customer interactions and enhancing satisfaction and loyalty.

REFERENCES

https://medium.com/analytics-vidhya/analysis-of-bank-customers-using-dashboard-in-power-bi-a366f2b3e563

LINK

https://github.com/Santhaganesan/Santhaganesan.git