Tech Saksham

Capstone Project Report

“E- commerce SALES REPORT ANALYSIS”

COLLEGE NAME:GOVERNMENTCOLLEGE OF ENGINEERING- SALEM.

NM ID	NAME
au61772311t501	AAGASH G.V

ABSTRACT

A **sales analysis report** is a document that includes all of the most important data of your business's sales process and provides you with a complete overview of your sales trends, volume, and overall sales activities. Some of the metrics included in sales analysis reports are: Sales trends Lead conversion rate Number of leads in the sales pipeline

A forecasted vs. actual sales report can help your salespeople compare their progress against monthly and quarterly goals. These reports provide a quick way to analyze sales numbers and make adjustments as necessary.

INDEX

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

- ANALYZE SEASONAL VARIATION IN SALES PERFORMANCE TO ADJUST

MONTHLY TARGETS ACCORDINGLY AND TO COMPETITOR ACTION AND SHORTFALL OF MONTHLY SALES TARGETS ACROSS DIFFERENT PRODUCT CATEGORIES OR REGIONS AND EVEN TO MONITOR SALES METRICES AND TO UPSELLING BASED ON SALES DATA TO EVALUATE THE PROFITABLE PRODUCTS OR SERVICES CONTRIBUTING AND TO FORECAST AND INVESTIGATE ROOT CAUSES:

1.2 Proposed Solution

1. **Top-performing products/services**
2. **Underperforming products/services**
3. **Customer behavior and retention**
4. **New sales and market opportunities**
5. **The future outlook of your sales team**
6. **Monthly basis**: Overarching metrics like **net sales** and **deal size**.
7. **Shorter-term basis**: Specific metrics such as **calls** or **emails sent**.
8. Monitor **seasonal changes** and **year-over-year (YoY) metrics** as well.
9

9.1 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, customer satisfactions, 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.

9.2 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 sales engage with their customers more effectively.
- **Increased Revenue**: By identifying opportunities for cross-selling and up-selling, of products can increase their revenue.

9.3 Scope

10 1.PREDICTIVE ANALYTICS TO INCREASINGLY FOCUS ON PREDICTIVE MODELING.

11 2.TO DETECT ANOMALIES,IDENTIFY TRENDS AND GENERATE CUSTOMIZED INSIGHT TAILORED TO SPECIFIC NEEDS

12 3.REAL TIME ANALYSIS TO CHANGING MARKET CONDITION,AND MITIGATE RISKS

13 4.INTEGRATION WITH IOT DEVICES AND SUPPLY CHAIN DYNAMICS,INFORMED DECISION MAKING.

CHAPTER 2

SERVICES AND TOOLS REQUIRED

2.1 Services Used

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

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

- **Machine Learning Services**: Azure Machine Learning or data frames using scikitlearn matplotlib 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.

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

Software Requirements:

Microsoft Excel:

- **Type**: Spreadsheet software.
- **Usage**: Data wrangling and reporting.

Python:

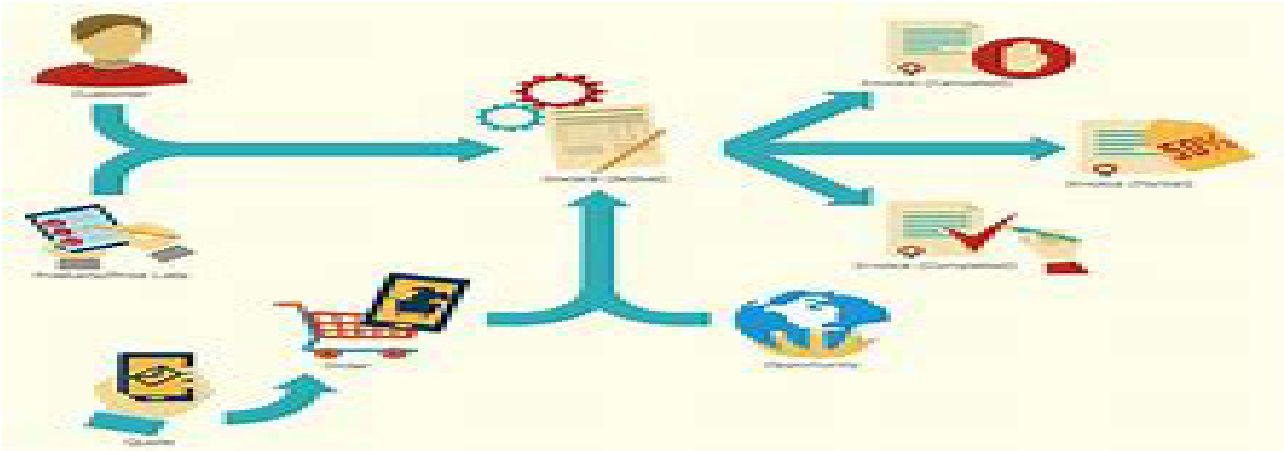
- **Type**: Programming language.
- **Usage**: Everything from data scraping to analysis and reporting
- Its extensive libraries (such as Pandas, NumPy, and Matplotlib) make it a go-to choice for data professionals.

Google colab: google colab allows you to create and share documents containing live code, equations, visualizations, and narrative text.

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 keras formulasis.
2. **Data Storage**: The collected data is stored in a database for processing.
3. Data analysis and optimize can be used for this purpose.
4. **Data Processing**: The stored data is processed in real-time using services like Azure Stream Analytics or AWS Kinesis Data Analytics.
5. **Machine Learning**: Predictive models are built based on processed data using Azure Machine Learning or process embassies. These models can help in predicting customer behavior, detecting fraud, etc.
6. **Data Visualization**: The processed data and the results from the predictive models are visualized in real-time using predict using matplotlib. Plot as pie chart forms allows you to create interactive dashboards that can provide valuable insights into the data.
7. **Data Access**: The dashboards created in google colab can be accessed through excel csv Desktop, Power Service (online), and using pandas libraries

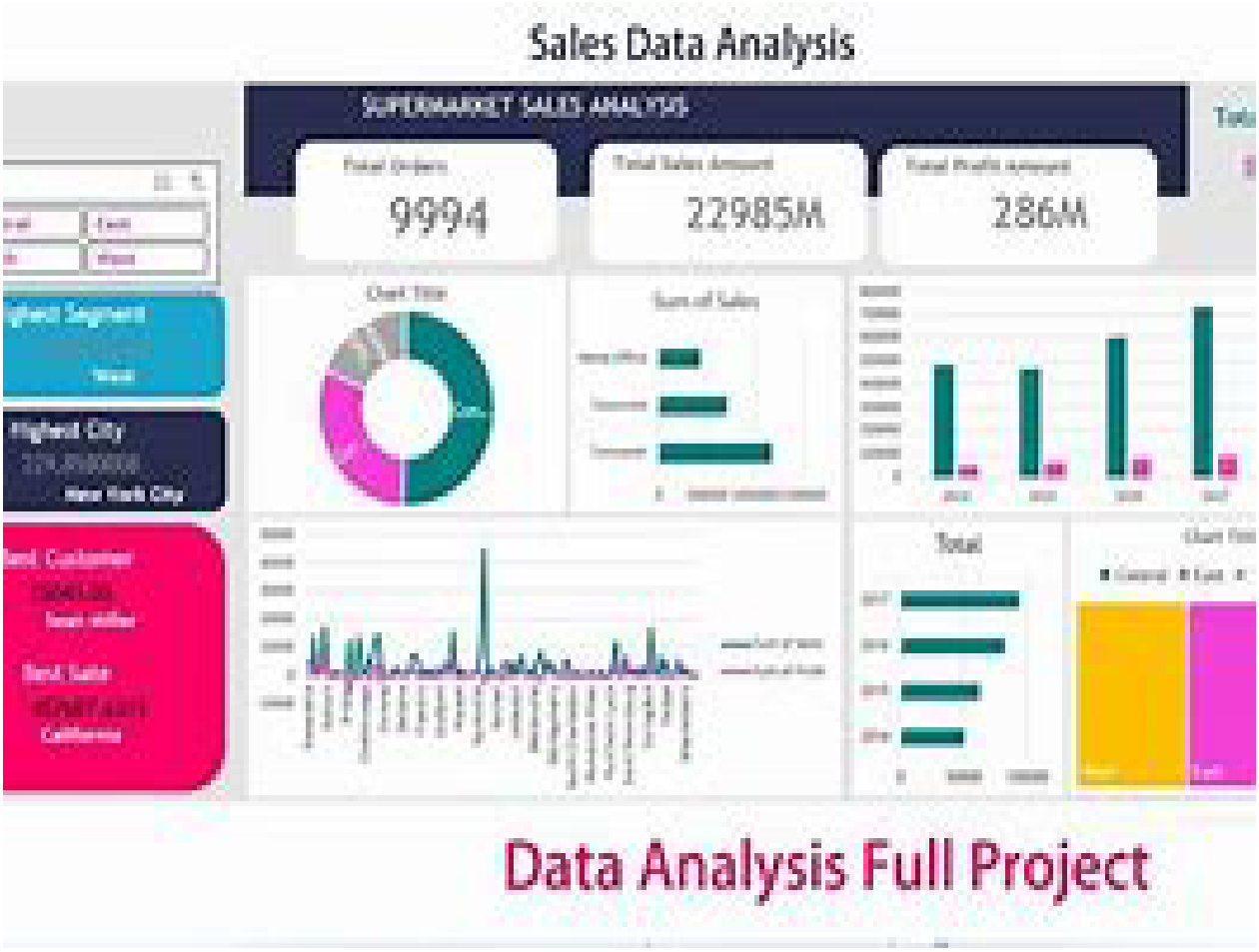
This architecture provides a comprehensive solution for real-time analysis of bank customers. However, it's important to note that the specific architecture may vary depending on the bank's existing infrastructure, specific requirements, and budget. It's also important to ensure that all tools and services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

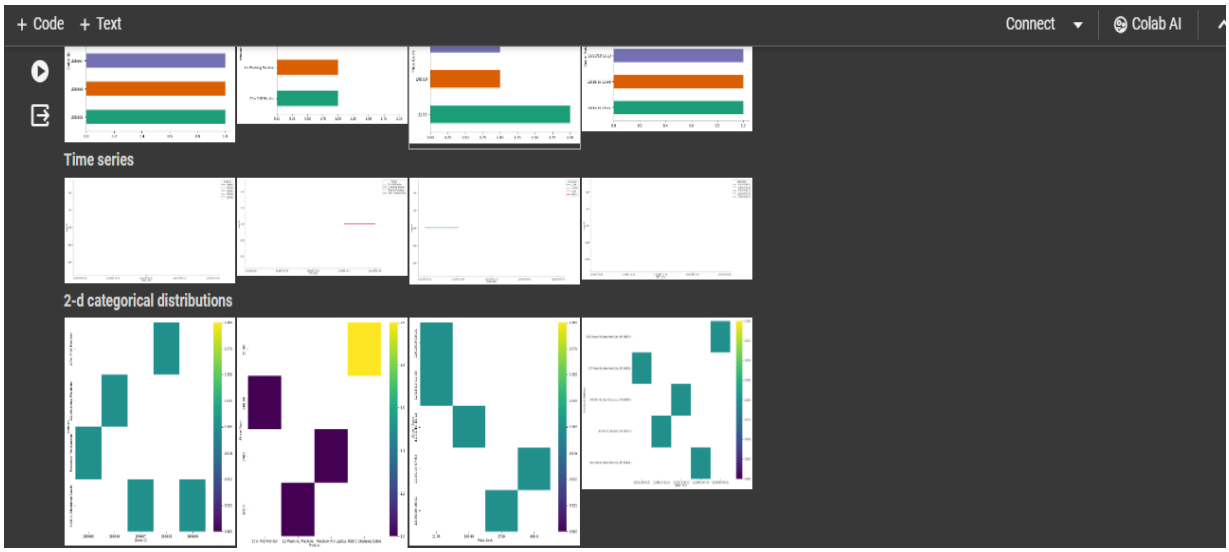
CHAPTER 4

MODELING AND PROJECT OUTCOME

Manage relationship

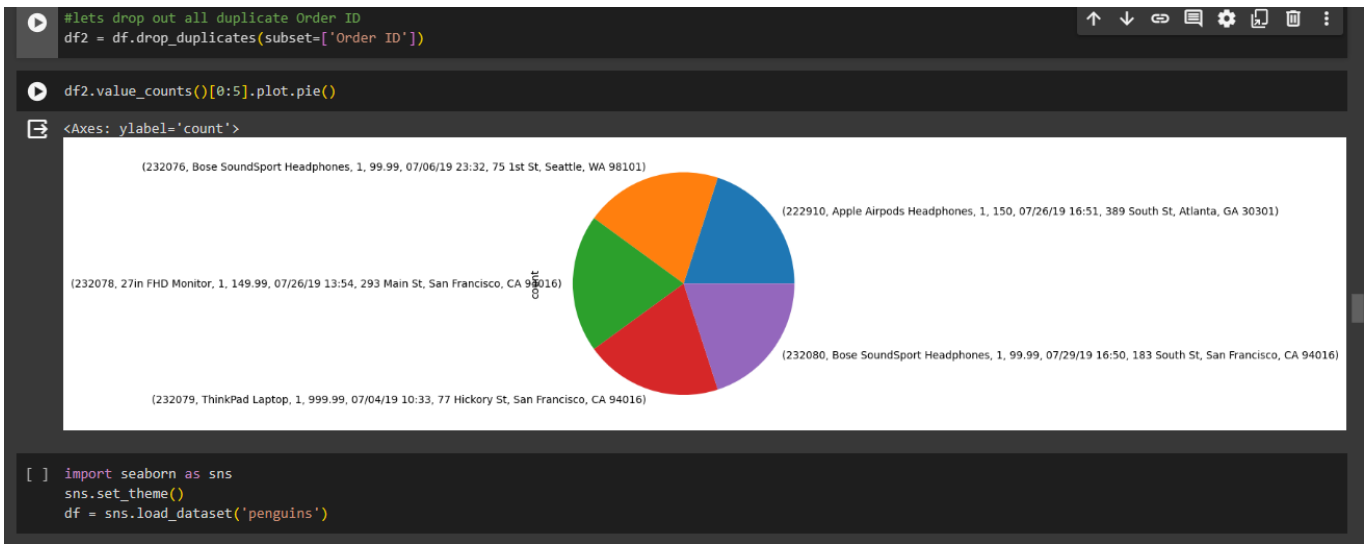
The “disp” file will be used as the main connector as it contains most key identifier (account id, client id and disp id) which can be use to relates the 8 data files together. The “district” file is use to link the client profile geographically with “district id”





Modelling for sales and division 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.



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

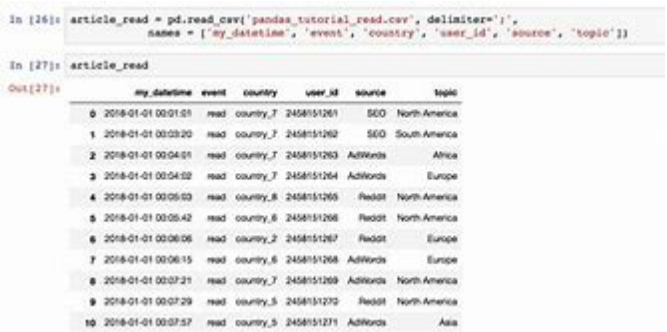
For Age, we shall assume it is year 1999 as explain previously and use it to minus from the birth year.

```
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

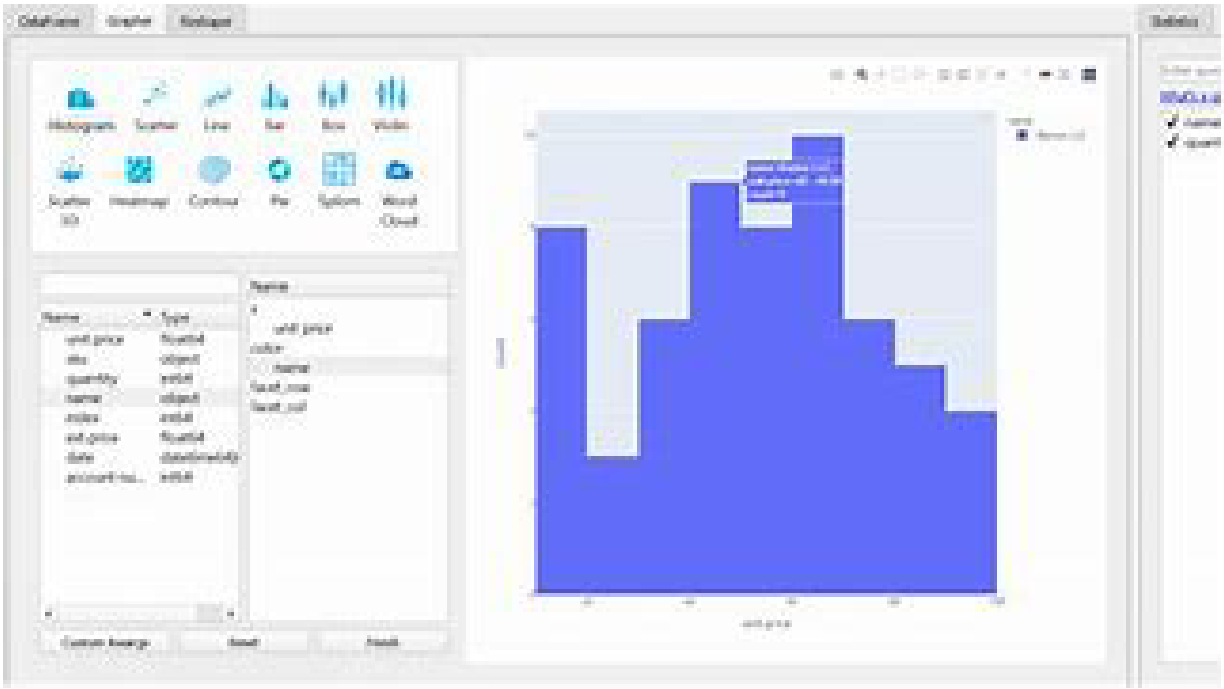
Replacing values

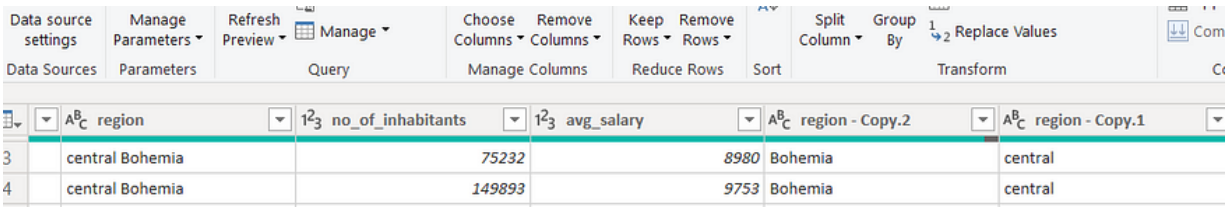
Set some fields to English for easy understanding, we replace values to English with the Power Query Editor.



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
3	central Bohemia	75232	8980	Bohemia	central
4	central Bohemia	149893	9753	Bohemia	central

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.

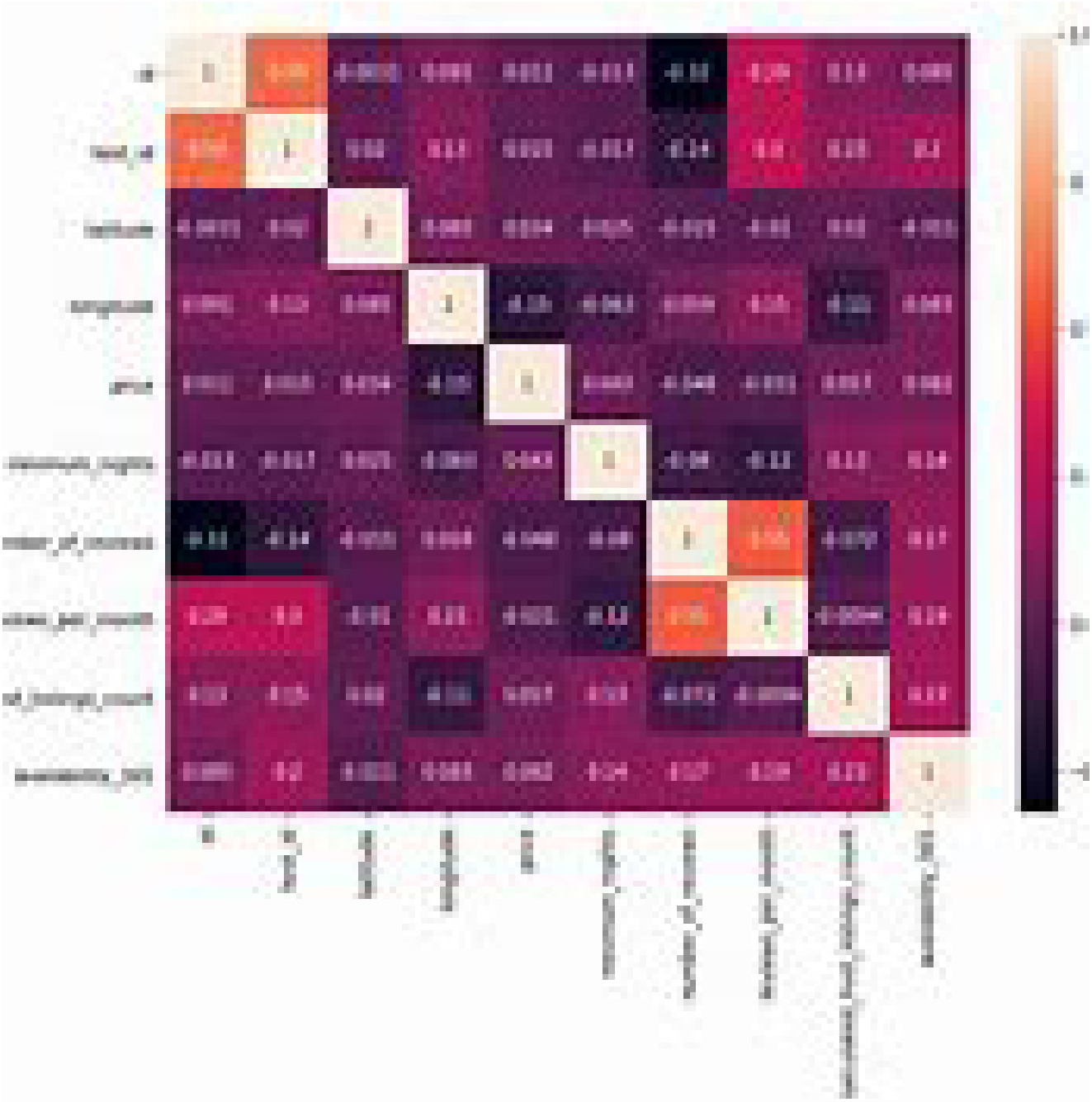
The Gen Y are youths,

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.



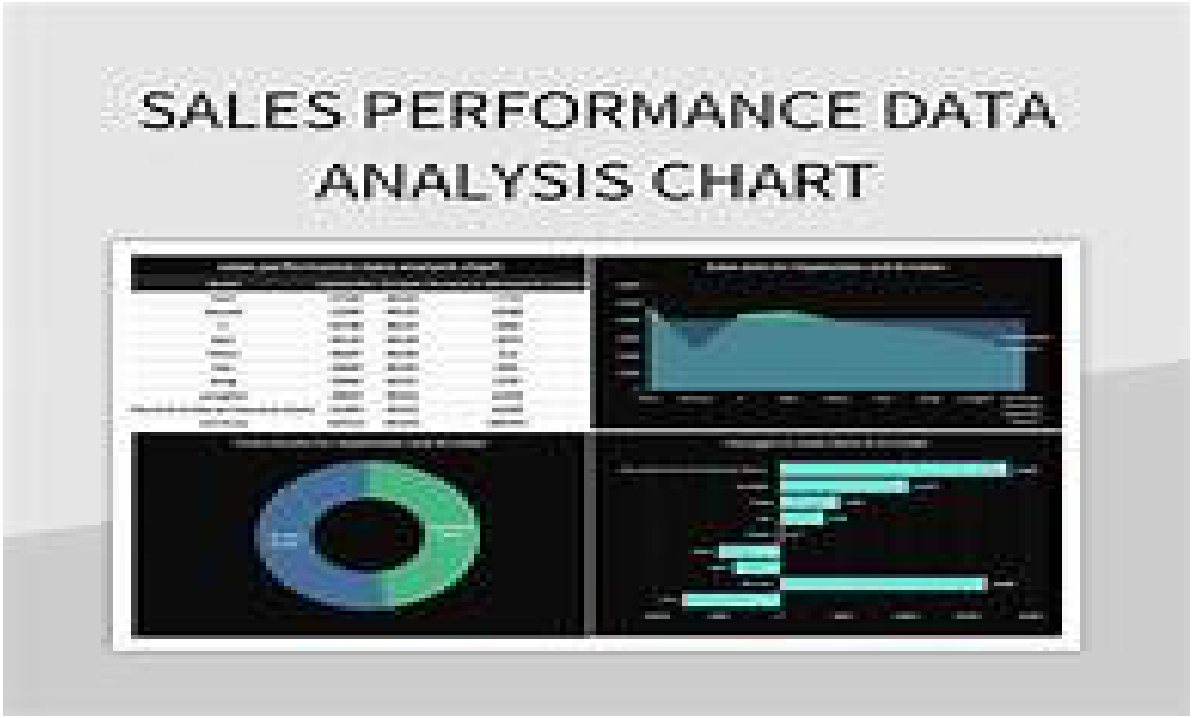
n look at a training **data** set with some **sales data** and gain some insights from it. Let's take a little look at the **data** as it looks in Excel. Upon initial inspection of the **data,** we can start thinking of some questions about it that we would want to answer

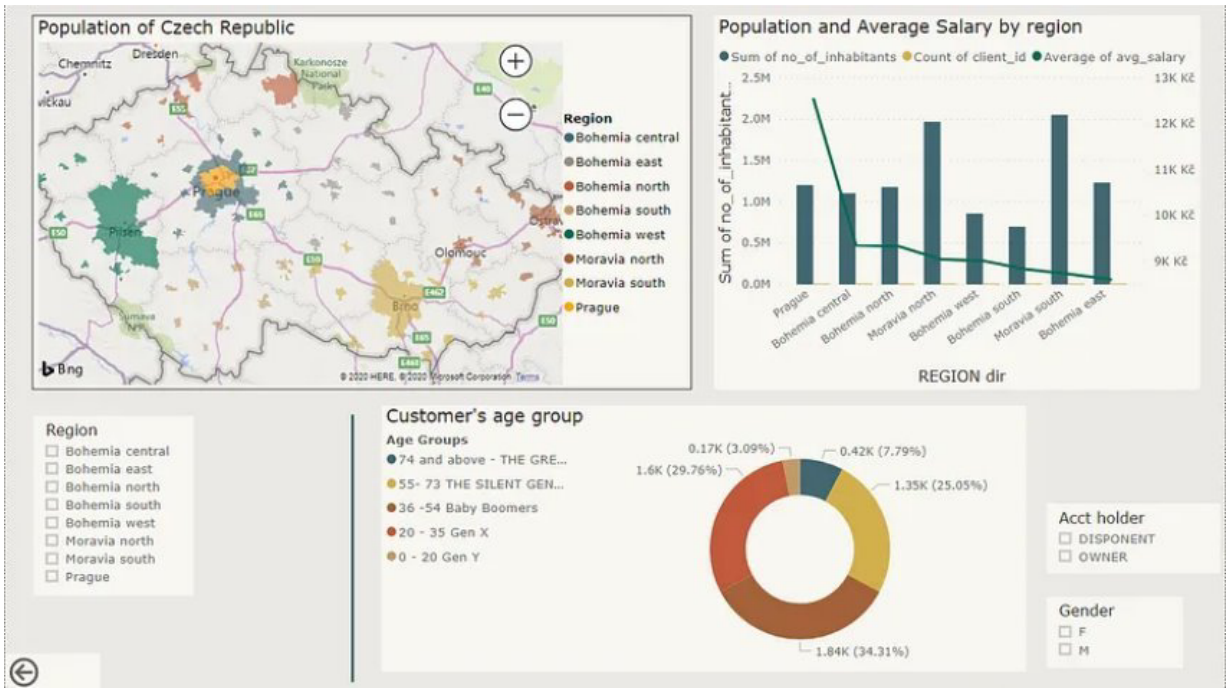
Values of such as "account Id" have also been set as Text.

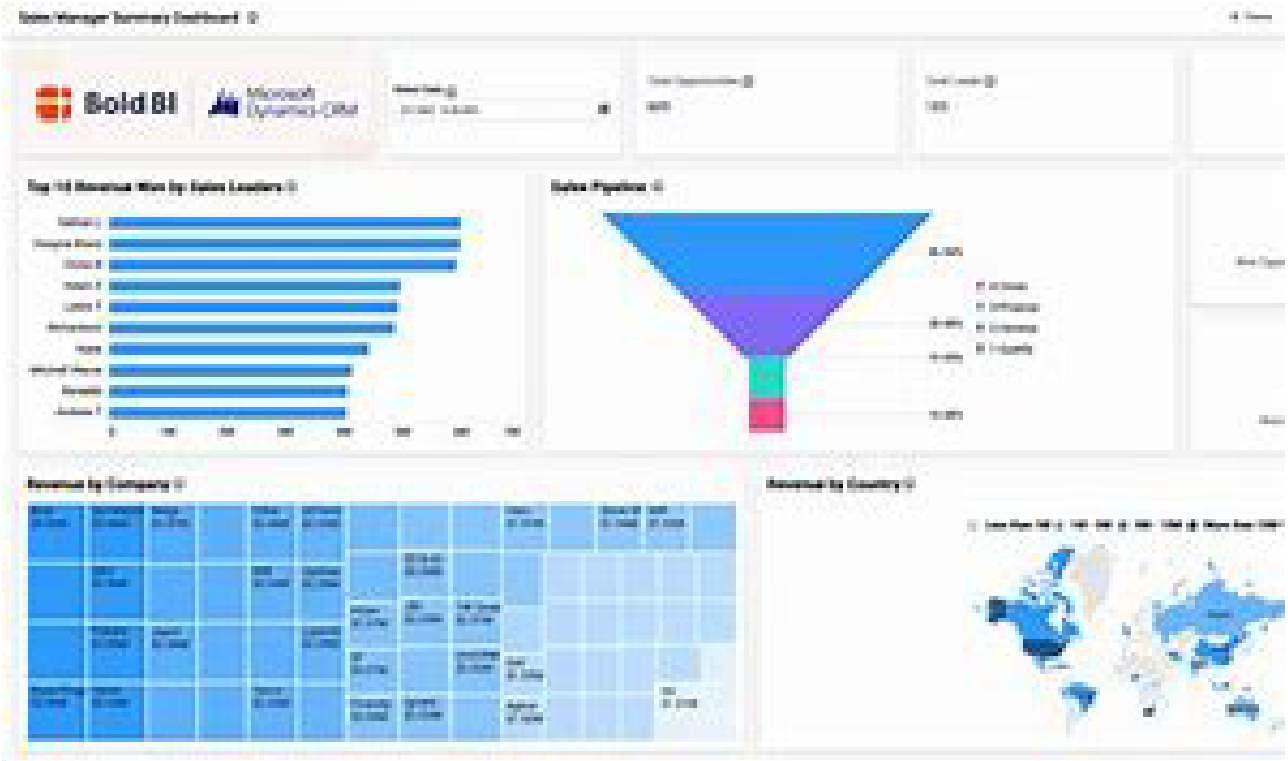
And District name have been categorized as place to be use for the map to show the sum of the inhabitants in each region.

	Date	SalesRep	Region	Product	Color	Units	Revenue
0	2015-11-06	Julie	East	Sunshine	Blue	4	78.8
1	2015-11-07	Adam	West	Bellen	Clear	4	123.0
2	2015-11-07	Julie	East	Aspen	Clear	1	26.0
3	2015-11-07	Nabil	South	Quad	Clear	2	69.0
4	2015-11-07	Julie	South	Aspen	Blue	2	51.0

Dashboard







CONCLUSION

Sales data analysis can reveal pricing trends and their impact on sales volumes. By examining price elasticity and competitive pricing, you can adjust your pricing strategies to maximize revenue. Data-driven pricing decisions can help you strike the right balance between attracting customers and maintaining.

Efficient inventory management, including accurate forecasting, is vital to controlling costs and ensuring you have the right products available at the right time. Analyzing sales data helps you optimize sales order management, preventing overstocking or stockouts. This optimization minimizes storage costs and lost sales opportunities, ultimately improving your bottom line.

Utilizing historical sales data and applying predictive analytics can help you forecast future sales trends and anticipate market shifts. This foresight enables you to proactively adjust your strategies, launch new products or services, and enter new markets with confidence. Predictive analysis empower

FUTURE SCOPE

The future scope of this project is vast. With the advent of advanced analytics and machine learning, PowerBI 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, PowerBI'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

Pull requests · witrioktafiani/Sales-Analysis-Excel (github.com)

- wesleyhsin (Wesley Hsin) (github.com)
- deepkorat / sales-data-cleaning-and-filterning

GITHUB LINK: