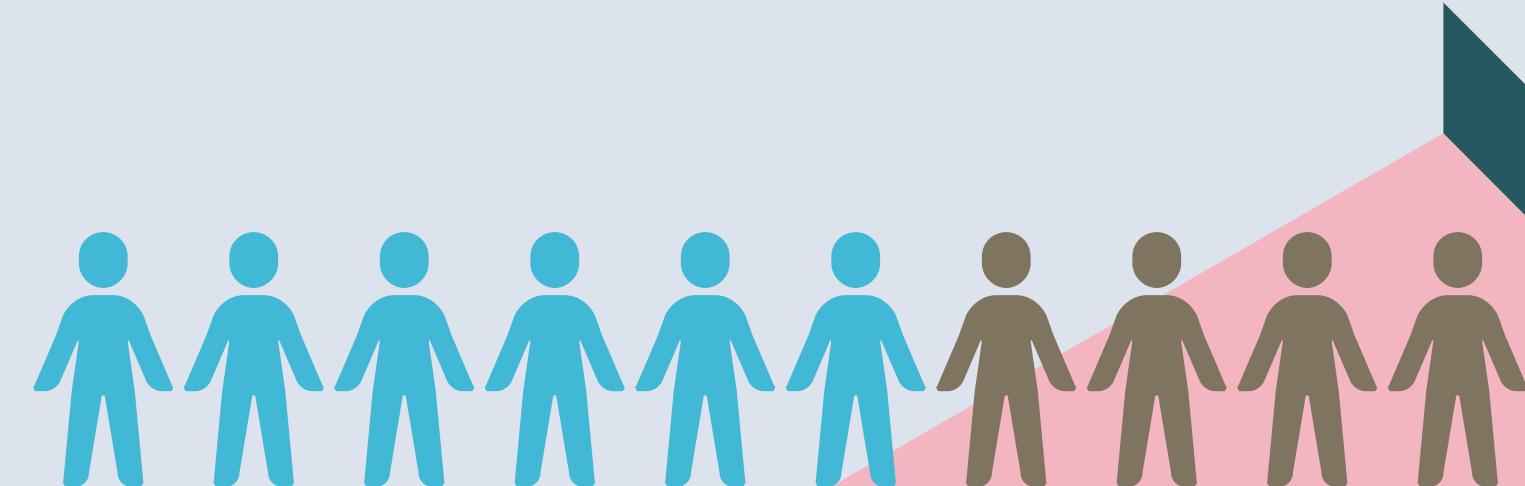


CUSTOMER DATA MANAGEMENT AND ANALYSIS

PRESENTED BY:

| | |
|-----------------------|------------|
| Menna Allah Sayed Ali | 1112149599 |
| Basant Waleed Yehia | 1112144832 |
| Nadine Ibrahim | 1112131247 |
| Yazied Mohamed Hassan | 1112131350 |
| Ahmed Mohamed Yousef | 1112140141 |



Agenda

Introduction

Project Objectives

Week-by-Week
Breakdown

Key Insights

Project benefits

Conclusion

INTRODUCTION

Today's companies face significant challenges in managing the variety and variety of data they want. This project aims to address these challenges by providing innovation and customer data analysis solutions.

This project aims to design and implement an integrated system for managing and analyzing customer data, through creating a specialized database, developing a data warehouse, and building predictive models.

Project Objectives



1. Building a robust infrastructure for customer data management and analysis:

- Designing and implementing a comprehensive SQL database: To store all customer-related data.
- Constructing an SQL data warehouse: For aggregating and analyzing large volumes of data.
- Utilizing advanced data analysis tools: Such as Python and Azure Machine Learning.

2. Developing predictive models for enhanced decision-making:^{*}

- Creating models to predict customer behavior: Such as churn probability or product purchase likelihood.
- Leveraging models for personalized recommendations: And improving customer experience

3. Integrating AI into business operations:

- Implementing MLOps principles: For efficient management of the machine learning model lifecycle.
- Deploying developed models: Into a production environment.

Week-by-Week Breakdown

week 1

week 2

week 3

week 4

week 1

Tools

Overview

Deliverables

toools

- Microsoft SQL Server: It is the database management system used to operate the database.
- SQL Management Studio: It is a software tool for interacting with SQL Server. With this tool, you can write SQL tags, manage the database, and design it.

OVERVIEW

- Data Sourcing: Utilized datasets sourced from Kaggle to obtain real-world customer data.
- Database Design and schema : Designed a robust schema to manage customer data.

Tables Created:

- telco_customer_data: Contains comprehensive customer information.
- customer_trans: Tracks transaction details for each customer.
- customer_inter: Captures customer interaction records.
- Implementation: Successfully created and populated the database using Microsoft SQL Server, ensuring data integrity and accessibility.

DELIVERABLES

Delivered a well-structured database schema and developed SQL queries for efficient data extraction and analysis.

```
-- استعلام انتخاب اقل البيانات
SELECT *
FROM dbo.customer_inter;
-- استعلام انتخاب اقل البيانات
SELECT Customer_ID,
       [Satisfaction Score],
       [Churn_Label],
       CLTV,
       [Customer_Status]
FROM dbo.customer_inter;
-- استعلام انتخاب اقل البيانات درجة رضا أقل من 3
SELECT *
FROM dbo.customer_inter
WHERE [Satisfaction Score] > 3; -- او قيمة ما ذكرتها
-- استعلام انتخاب اقل البيانات لم يتبرع لها
SELECT *
FROM dbo.customer_inter
WHERE [Customer_Status] = 'Churned'; -- لبيانه هو
-- استعلام تعداد اصحاب هذه المدفوعة من فئة
SELECT [Churn Category],
       COUNT(Customer_ID) AS Customer_Count
FROM dbo.customer_inter
GROUP BY [Churn Category]
ORDER BY Customer_Count DESC;
-- استعلام ببيانات telco_customer_data
SELECT *
FROM dbo.telco_customer_data;
-- استعلام انتخاب اقل البيانات
SELECT *
FROM dbo.customer_trans;
-- استعلام تعداد اصحاب بين 18 و 65 سنة
SELECT Customer_ID,
       [Count],
       [Gender],
       [Age],
       [Under 30];
-- استعلام انتخاب اقل البيانات
```

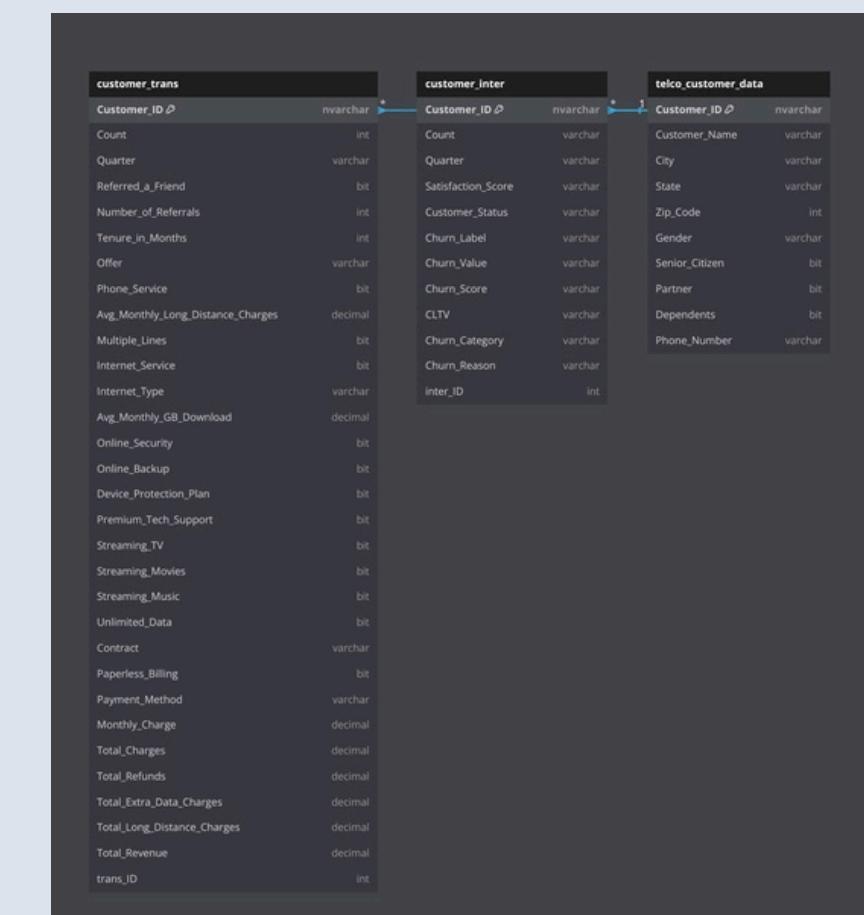
Results

| Customer_ID | Monthly_Charge | Total_Revenue | Contract |
|-------------|----------------|---------------|----------|
| 8 | 3318-NMQXL | 111.1 | 313.6 |
| 9 | 1022-RKODR | 24.85 | 2301.31 |
| 10 | 2361-FJWNO | 36 | 1512.9 |
| 11 | 2272-UOINI | 95.7 | 594.43 |
| 12 | 8232-UTFOZ | 19.95 | 1894.77 |
| 13 | 3750-YHRYO | 20.65 | 297.63 |
| 14 | 6637-KYRCV | 37.4 | 167.2 |
| 15 | 5668-MEISB | 106.1 | 8118.92 |

| Avg_Monthly_Charge | Total_Revenue |
|--------------------|---------------|
| 71.911543 | 21558123 |

| Total_Customers | Total_Referrals |
|-----------------|-----------------|
| 3222 | 13747 |

| Quarter | Customer_Count |
|---------|----------------|
| | |



week 2

Tools

Overview

Deliverables

toools

- Microsoft SQL Data Warehouse: A powerful platform for creating and managing data warehouses.
- Python: A flexible and powerful programming language for data managing and analysis.
Pandas: A Python library for statistical data analysis.
SQLAlchemy: A Python library for interacting with various databases, including SQL Server.

OVERVIEW

- Data Warehouse Implementation:

Built a SQL Data Warehouse to aggregate and manage large volumes of customer data for analytical insights.

- Schema Overview:

Fact Table: `customer_trans`

Primary Key: `trans_ID` (Surrogate Key)

Business Key: `trans_Customer_ID`

Dimension Tables:

- `dbo.customer_inter`

Primary Key: `interaction_ID`

- `dbo.telco_customer_data`

Primary Key: `Customer_Telco_ID`

- Importance: This structure allows for efficient querying and analysis, enabling better insights into customer behavior.

OVERVIEW

Schema Overview

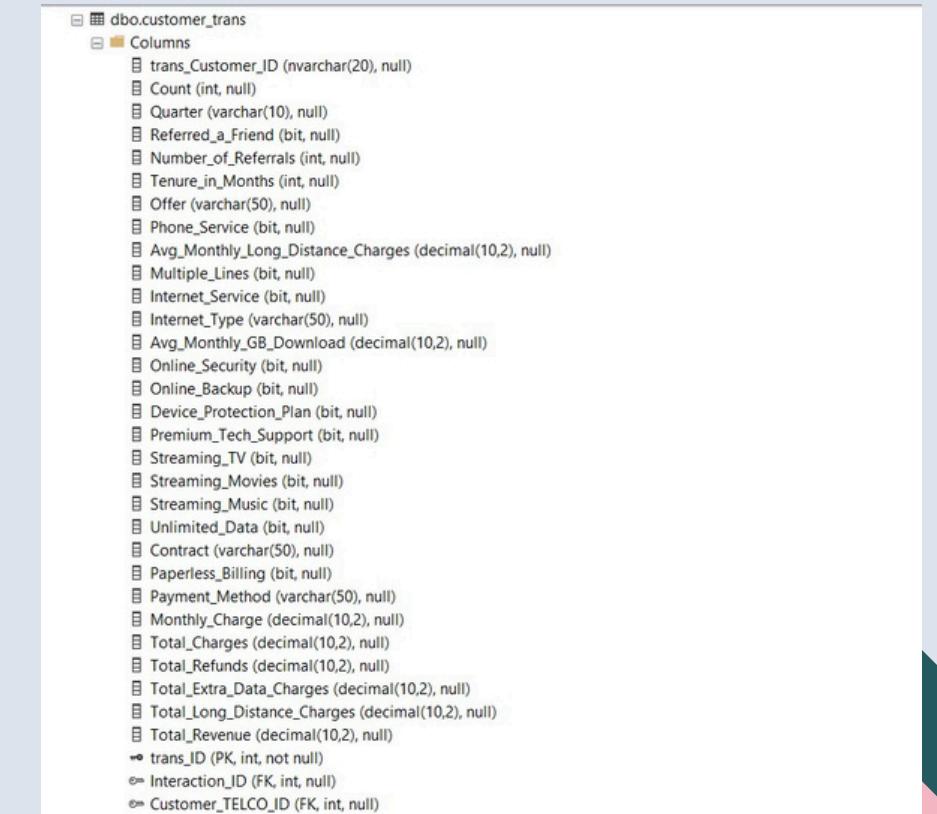
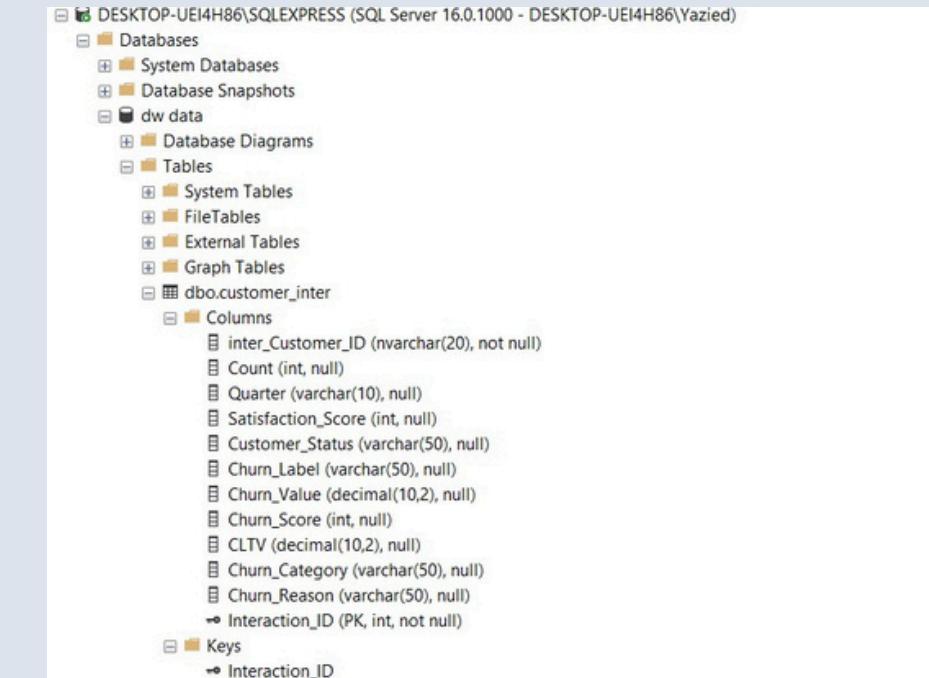
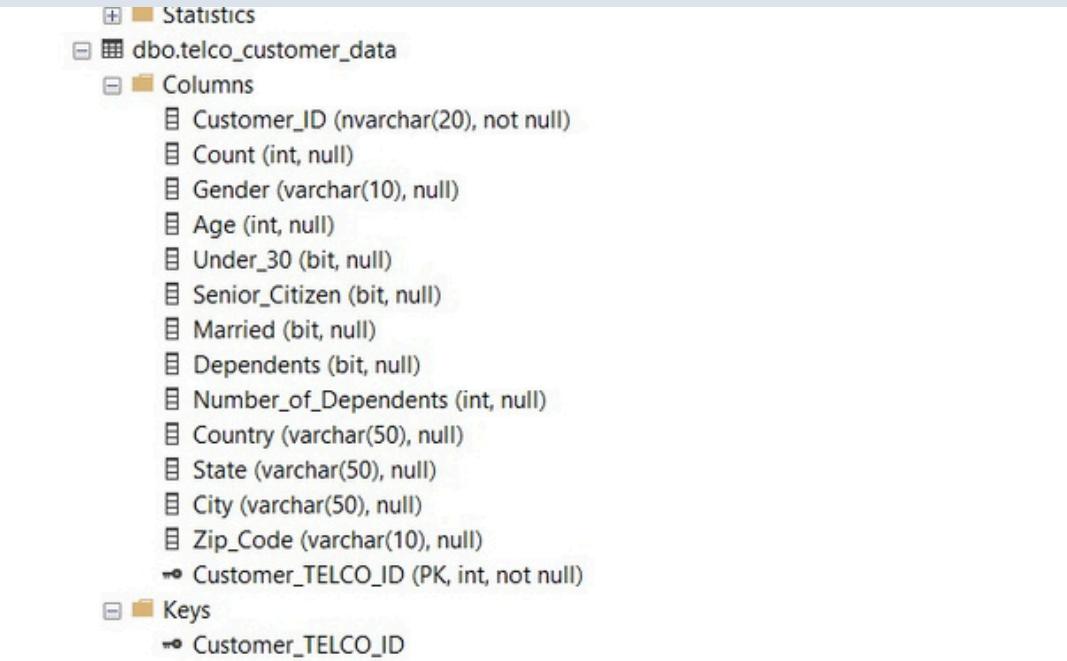
- Fact Table: customer_trans
 - Primary Key: trans_ID (Surrogate Key)
 - Business Key: trans_Customer_ID
 - Foreign Keys:
 - inter_Customer_ID → References dbo.customer_inter(inter_Customer_ID)
 - Customer_Telco_ID → References dbo.telco_customer_data(Customer_Telco_ID)
- Dimension Table 1: dbo.customer_inter
 - Primary Key: interaction_ID (Surrogate Key)
 - Business Key: inter_Customer_ID
- Dimension Table 2: dbo.telco_customer_data
 - Primary Key: Customer_Telco_ID (Surrogate Key)
 - Business Key: Customer_ID

Keys Overview

- Surrogate Keys:
 - trans_ID (Primary Key for customer_trans)
 - interaction_ID (Primary Key for dbo.customer_inter)
 - Customer_Telco_ID (Primary Key for dbo.telco_customer_data)
- Business Keys:
 - trans_Customer_ID (Business Key in customer_trans)
 - inter_Customer_ID (Business Key in dbo.customer_inter)
 - Customer_ID (Business Key in dbo.telco_customer_data)

DELIVERABLES

A fully functional SQL Data Warehouse along with Python scripts for data extraction and preparation.



week 3

Tools

Overview

Deliverables

tools

-  Python: A programming language for data analysis and predictive forecasting.
- Scikit-Learn: A library in Python that provides a wide range of learning algorithms.
- Matplotlib: Another Python library for weathering and data visualization.
-  Azure Data Studio: A tool for creating and exploring data in Azure services.
-  Azure Machine Learning: Microsoft's cloud service for deploying machine learning models.

OVERVIEW

- Data Science with Python:
we focused on the integration of data science and Azure services. We explored data science with Python, conducting comprehensive data analysis and predictive modeling
- Azure Integration:
Leveraged Azure services : Additionally, we leveraged Azure services for efficient data management and analytics..
- Model Development:
Developed and evaluated machine learning models to predict customer behavior based on transaction and interaction data.

DATA SCIENCE WITH PYTHON

- Churn prediction to identify at-risk customers:

Churn prediction was a key aspect of our data science work. By analyzing transaction and interaction data, we were able to identify customers who were likely to churn. This allowed us to proactively implement retention strategies and minimize customer attrition.

- Conducting comprehensive data analysis and predictive modeling:

We conducted in-depth data analysis and built predictive models using Python. One notable application was churn prediction, which helped us identify at-risk customers. These models provided valuable insights into customer behavior and enabled targeted retention strategies.

AZURE INTEGRATION

Leveraging Azure services for data management and analytics: We seamlessly integrated Azure services into our data science workflow. Azure provided robust solutions for data management and analytics, streamlining our processes and enhancing our capabilities. By leveraging Azure, we were able to efficiently store, process, and analyze large datasets.

DELIVERABLES

Generated a detailed analysis report outlining insights and presented predictive models to stakeholders.

1) Read the Three CSV Tables

Read the paths of our three csv tables

+ Code + Markdown

```
1 import pandas as pd
2 import numpy as np
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 import warnings
6 %matplotlib inline
7 warnings.filterwarnings("ignore")
8 from azureml.core import Dataset, Workspace
9
10 # Connect to your Azure ML workspace
11 ws = Workspace.from_config()
12
13 # Get the dataset (the folder you've uploaded in Data Assets)
14 dataset = Dataset.File.from_files(path=(ws.datastores['workspaceblobstore'], 'UI/2024-10-17_090040_UTC/dw data/'))
15
16 # Display the file paths in the dataset
17 file_paths = dataset.to_path()
18 print(file_paths)
```

```
1 from IPython.display import display
2 # Iterate over the file paths to read each CSV file
3 dfs = []
4
5 # Make the static path or the folder path that will be the same in the three files
6 static_path = "azureml://subscriptions/0146ae78-9468-4285-8cab-97231deb201d/resourcegroups/Final_Project/workspaces/CustomerChurnPredict"
7
8 for file_path in file_paths:
9     if file_path.endswith('.csv'):
10         # Read the CSV file directly from the blob storage into a pandas DataFrame
11         df = pd.read_csv(static_path.format(file_path))
12         dfs.append(df)
13
14 # Check the number of dataframes loaded
15 print(f"Successfully Loaded {len(dfs)} CSV files.")
16
17 # Display the first few rows of each DataFrame to verify
18 print("\ncustomer_inter.csv:")
19 display(dfs[0].head())
20 print(dfs[0].shape)
21
22 print("\ncustomer_trans.csv:")
23 display(dfs[1].head())
24 print(dfs[1].shape)
25
26 print("\ntelco_customer_data.csv:")
27 display(dfs[2].head())
28 print(dfs[2].shape)
```

| | Missing_Number | Missing_Percent |
|-----------------------------|----------------|-----------------|
| Offer | 3877 | 0.550476 |
| Internet_Type | 1526 | 0.216669 |
| Satisfaction_Score | 0 | 0.000000 |
| Total_Long_Distance_Charges | 0 | 0.000000 |
| Contract | 0 | 0.000000 |
| Paperless_Billing | 0 | 0.000000 |
| Payment_Method | 0 | 0.000000 |
| Monthly_Charge | 0 | 0.000000 |
| Total_Charges | 0 | 0.000000 |
| Total_Refunds | 0 | 0.000000 |

week 4

Tools

Overview

Deliverables

tools

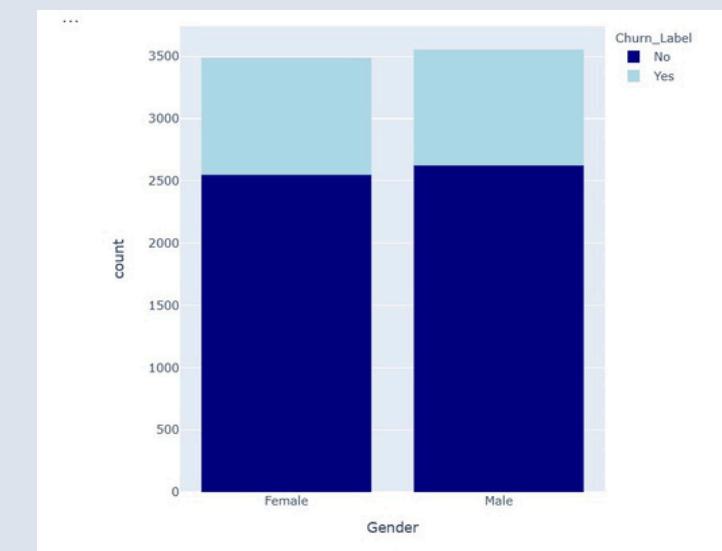
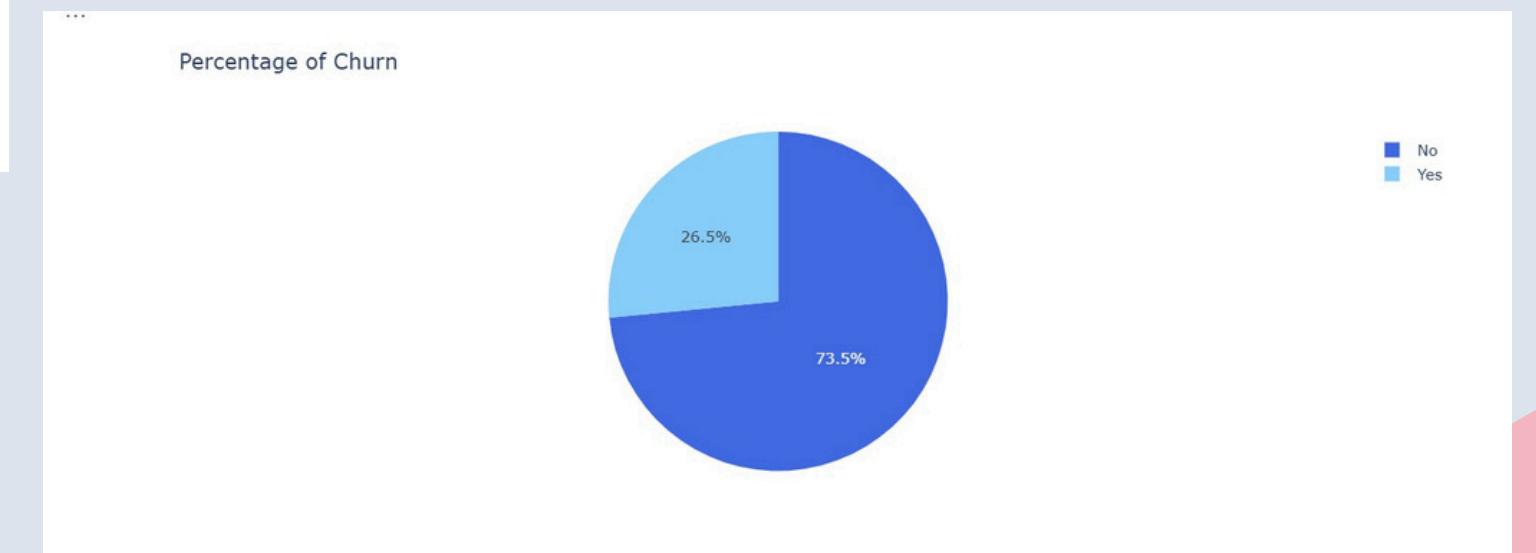
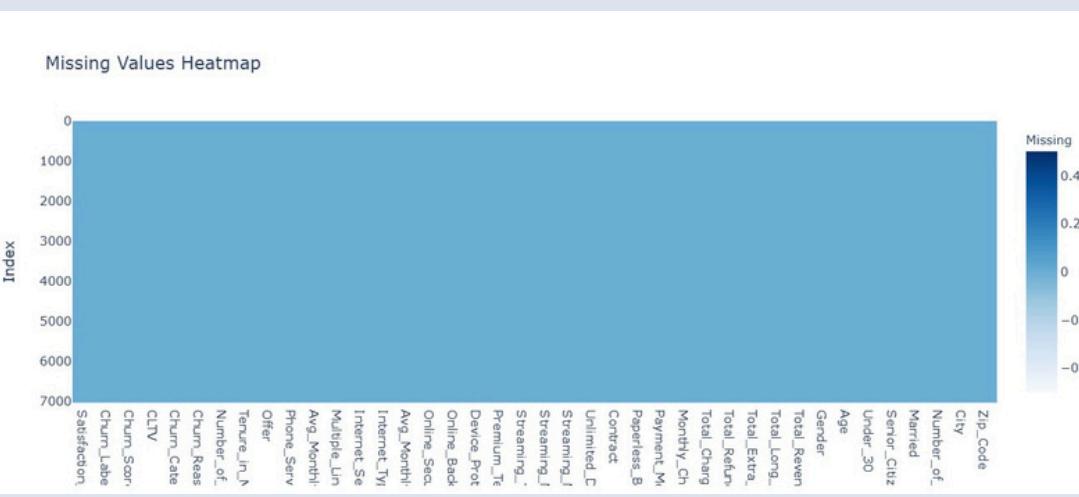
- MLflow: This resource usage utility uses an integrated, shared learning lifecycle of link tracking and link publishing.
- A Azure Service: Note this point to the wide range of services that Microsoft offers on the cloud, which can be used typically, such as Azure Machine Learning or Azure Container Instances.
 - Web frameworks (such as Flask, Streamlit): These frameworks are used to create simple web applications and through detection can provide predictions using the model.

OVERVIEW

- MLOps Implementation:
Integrating MLflow for tracking experiments and version control: We implemented MLOps practices by integrating MLflow into our workflow. MLflow allowed us to track machine learning experiments, manage model versions, and efficiently collaborate with team members. This ensured reproducibility and traceability throughout the project.
- Deployment:
Deploying ML models using Azure or web applications (Flask or Streamlit): We deployed our machine learning models using Azure services or web applications such as Flask or Streamlit. This allowed us to provide user-friendly access to our predictive insights. By deploying our models, we made our solutions accessible and actionable for stakeholders.

DELIVERABLES

Our final deliverables included the deployment of our machine learning models or web applications. Alongside that, we provided a comprehensive final project report summarizing our methodology, findings, and recommendations.



Key Insights

- A_ The importance of SQL for structured data management.
- B_ Power of Python in data extraction and model development.
- C_ Efficiency of cloud services (Azure) for data integration and machine learning.
- D_ Highlight important findings derived from the data analysis.
- E_ Discuss the value added to decision-making processes and operational efficiency.
- F_ Automation of customer data management significantly reduced manual processing time.
- G_ Advanced machine learning techniques improved customer churn prediction capabilities, enabling targeted retention strategies.
- H_ Azure cloud services streamlined data storage and analysis processes, enhancing scalability and performance.

Project benefits

- Enable customer understanding: By analyzing data, a company can understand more about customer behavior and needs.
- Make Better Decisions: Predictive reporting will help the company make more integrated decisions with marketing, forecasts, and products.
- Experience Clients: Clients can specialize in specialists resulting in them being introduced to us and them.
- Increase profits: By better understanding customers of marketing, we can increase sales for the company and its partners.
- Improved strategies for customer retention based on data-driven insights.
- Established a scalable infrastructure capable of handling large datasets, enabling future growth and analysis.

CONCLUSION

In conclusion, we successfully developed a comprehensive solution for customer data management and predictive analytics. Our deployed machine learning models provided valuable insights into customer behavior, enabling effective business decisions and retention strategies. Through the integration of data science and Azure services, we enhanced our data management and predictive analytics capabilities, empowering informed decision-making.

THANK YOU

