

Group No : 52

Group Member Names:

1. CHAGANTI S S V S LAKSHMINARAYANA(2023ad05097@wilp.bits-pilani.ac.in)- contribution -100%
2. PRADYOTHANA D P(2023ad05059@wilp.bits-pilani.ac.in) - contribution -100%
3. BRAJESH KUMAR SETHI(2023ac05335@wilp.bits-pilani.ac.in)- contribution -100%
4. BHOGE ANKIT ARUN MADHURI(2023dc04324@wilp.bits-pilani.ac.in)- contribution -100%

Here's a structured problem formulation for your customer churn prediction project:

1. Problem Formulation

Business Problem

Customer churn is a critical issue for subscription-based businesses. Losing customers impacts revenue, growth, and operational efficiency. The goal is to predict which customers are likely to churn so that proactive retention strategies can be implemented.

Key Business Objectives

- **Reduce customer churn rate** by identifying at-risk customers early.
- **Improve customer retention strategies** through targeted interventions.
- **Optimize marketing and customer support efforts** by understanding churn patterns.

Key Data Sources & Attributes

The dataset consists of customer-related attributes that influence churn. The key attributes include:

- **CustomerID**: Unique identifier for each customer.

- **Age:** Customer's age.
- **Gender:** Male or Female.
- **Tenure:** How long the customer has been subscribed (in months).
- **Usage Frequency:** How often the customer uses the service.
- **Support Calls:** Number of customer support interactions.
- **Payment Delay:** Delays in payment (in days).
- **Subscription Type:** Type of subscription (e.g., Basic, Premium).
- **Contract Length:** Duration of the customer's contract (e.g., Monthly, Annual).
- **Total Spend:** Total amount spent by the customer.
- **Last Interaction:** Time since last interaction (in days).
- **Churn:** Whether the customer has churned (Yes/No).

Expected Outputs from the Pipeline

1. Clean datasets for exploratory data analysis (EDA)

- Handle missing values and outliers.
- Convert categorical variables to numerical representations.

2. Transformed features for machine learning

- Normalize/scale numerical data.
- Engineer new features if necessary (e.g., interaction between variables).

3. A deployable model to predict customer churn

- Train a classification model (e.g., Decision Tree, Random Forest, XGBoost, or Neural Networks).
- Deploy the model for real-time predictions.

Measurable Evaluation Metrics

- **Accuracy:** Overall correctness of the model.
- **Precision & Recall:** Evaluate the model's ability to correctly identify churners.
- **F1-Score:** Balances precision and recall.

- **ROC-AUC Score:** Measures the ability of the model to distinguish between churners and non-churners.

Deliverables

- **A PDF/Markdown document** containing:
 - Business problem
 - Objectives
 - Data sources
 - Pipeline outputs
 - Evaluation metrics
- **A trained machine learning model** to predict customer churn.
- **Data processing and model training scripts** for reproducibility.

Install & Import Dependencies

```
In [1]: import os
import requests
import pandas as pd
import logging
import time
from datetime import datetime
import json

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
```

2. Data Ingestion

```
In [2]: # CSV file path
CSV_FILE_PATH = "~/Downloads/DMML_assignment/customer_churn_dataset-testing-master.csv"

# API URL
API_URL = "https://657165bcd61ba6fcc0126420.mockapi.io/customers"
```

Fetch Data from CSV

```
In [4]: logging.basicConfig(level=logging.INFO, format="%(asctime)s - %(levelname)s - %(message)s")
```

```
# Define log_message function
def log_message(level, message):
    logging.log(level, message)

def fetch_csv_data():
    try:
        df = pd.read_csv(CSV_FILE_PATH)
        log_message(logging.INFO, "Successfully ingested data from CSV file.")
        return df
    except Exception as e:
        log_message(logging.ERROR, f"Failed to fetch CSV data: {e}")
        return None

# Test CSV ingestion
csv_data = fetch_csv_data()
csv_data.head() if csv_data is not None else "CSV Data Fetch Failed"
```

```
2025-03-14 21:37:30,600 - INFO - Successfully ingested data from CSV file.
```

```
Out[4]:
```

	CustomerID	Age	Gender	Tenure	Usage Frequency	Support Calls	Payment Delay	Subscription Type	Contract Length	Total Spend	Last Interaction	Churn
0	1	22	Female	25	14	4	27	Basic	Monthly	598	9	1
1	2	41	Female	28	28	7	13	Standard	Monthly	584	20	0
2	3	47	Male	27	10	2	29	Premium	Annual	757	21	0
3	4	35	Male	9	12	5	17	Premium	Quarterly	232	18	0
4	5	53	Female	58	24	9	2	Standard	Annual	533	18	0

Fetch Data from API

```
In [5]: def fetch_api_data():
        try:
```

```

    response = requests.get(API_URL, timeout=10)
    response.raise_for_status()
    data = response.json()
    df = pd.DataFrame(data)
    log_message(logging.INFO, "Successfully ingested data from API.")
    return df
except requests.exceptions.RequestException as e:
    log_message(logging.ERROR, f"Failed to fetch API data: {e}")
    return None

# Test API ingestion
api_data = fetch_api_data()
api_data.head() if api_data is not None else "API Data Fetch Failed"

api_df = pd.DataFrame(api_data)
api_df.to_csv('api_data.csv', index=False)

```

2025-03-14 21:37:38,196 - INFO - Successfully ingested data from API.

Setup Logging

```

In [6]: # Define log file
log_file = "data_ingestion.log"

# Configure Logging
logging.basicConfig(filename=log_file, level=logging.INFO, format="%(asctime)s - %(levelname)s - %(message)s")

# Function to Log messages
def log_message(level, message):
    print(f"{datetime.now()} - {level} - {message}")
    logging.log(level, message)

```

3. Raw Data Storage

Combine and Store Data

```

In [7]: def ingest_data():
    csv_data = fetch_csv_data()

```

```

# api_df = fetch_api_data() # Fetch data from API and store as api_df

if csv_data is not None and api_df is not None:
    combined_data = pd.concat([csv_data, api_df], ignore_index=True) # Use api_df
    combined_data.to_csv("raw_ingested_data.csv", index=False)
    log_message(logging.INFO, "Successfully combined and stored ingested data.")
else:
    log_message(logging.WARNING, "Skipping data storage due to missing sources.")

# Run the ingestion process
ingest_data()

```

```

2025-03-14 21:37:42,553 - INFO - Successfully ingested data from CSV file.
2025-03-14 21:37:42,708 - INFO - Successfully combined and stored ingested data.
2025-03-14 21:37:42.552957 - 20 - Successfully ingested data from CSV file.
2025-03-14 21:37:42.708447 - 20 - Successfully combined and stored ingested data.

```

Automate Periodic Ingestion (Every Hour)

```

In [8]: # # Get the number of rows in CSV
# csv_data = fetch_csv_data()
# num_rows = len(csv_data) if csv_data is not None else 1 # Default to 1 if CSV is empty or missing

# # Run ingestion as many times as the number of rows
# for i in range(num_rows):
#     log_message(logging.INFO, f"Running ingestion {i+1}/{num_rows} times based on dataset size.")
#     ingest_data()
#     time.sleep(3600) # Wait for 1 hour

```

4. Data Validation

```

In [9]: def validate_data(df):
        """
        Performs data validation checks on a Pandas DataFrame.

        Args:
            df: The input DataFrame.

```

Returns:

 A dictionary containing data quality report.

"""

report = {}

Check for missing data

missing_data = df.isnull().sum()

report["missing_data"] = missing_data[missing_data > 0].to_dict()

Check for inconsistent data (example: checking for invalid gender values)

inconsistent_gender = df[~df["Gender"].isin(["Male", "Female"])]

report["inconsistent_gender"] = inconsistent_gender.to_dict()

Validate data types (example: check if 'Age' is numeric)

if not pd.api.types.is_numeric_dtype(df['Age']):

 report["invalid_age_type"] = "Age column is not numeric"

Validate data ranges (example: check for unrealistic age values)

invalid_age_range = df[(df["Age"] < 0) | (df["Age"] > 120)]

report["invalid_age_range"] = invalid_age_range.to_dict()

Identify duplicates

duplicates = df[df.duplicated()]

report["duplicates"] = duplicates.to_dict()

Identify anomalies (example: unusually high support calls)

anomalies = df[df["Support Calls"] > 10] # Example threshold

report["anomalies"] = anomalies.to_dict()

return report

df = pd.read_csv("~/Downloads/DMML_assignment/raw_ingested_data.csv")

data_quality_report = validate_data(df)

Print or save the report

print(data_quality_report)

Save to a CSV file

```
report_df = pd.DataFrame.from_dict(data_quality_report, orient="index")
report_df.to_csv("data_quality_report.csv")
```

```
{'missing_data': {}, 'inconsistent_gender': {'CustomerID': {}, 'Age': {}, 'Gender': {}, 'Tenure': {}, 'Usage Frequency': {}, 'Support Calls': {}, 'Payment Delay': {}, 'Subscription Type': {}, 'Contract Length': {}, 'Total Spend': {}, 'Last Interaction': {}}, 'invalid_age_range': {'CustomerID': {}, 'Age': {}, 'Gender': {}, 'Tenure': {}, 'Usage Frequency': {}, 'Support Calls': {}, 'Payment Delay': {}, 'Subscription Type': {}, 'Contract Length': {}, 'Total Spend': {}, 'Last Interaction': {}}, 'Churn': {}}, 'duplicates': {'CustomerID': {9: 9, 15: 14}, 'Age': {9: 36, 15: 42}, 'Gender': {9: 'Male', 15: 'Male'}, 'Tenure': {9: 20, 15: 46}, 'Usage Frequency': {9: 5, 15: 27}, 'Support Calls': {9: 10, 15: 5}, 'Payment Delay': {9: '8', 15: '8'}, 'Subscription Type': {9: 'Basic', 15: 'Premium'}, 'Contract Length': {9: 'Monthly', 15: 'Quarterly'}, 'Total Spend': {9: 687.0, 15: 438.0}, 'Last Interaction': {9: 8, 15: 30}, 'Churn': {9: 0, 15: 0}}, 'anomalies': {'CustomerID': {}, 'Age': {}, 'Gender': {}, 'Tenure': {}, 'Usage Frequency': {}, 'Support Calls': {}, 'Payment Delay': {}, 'Subscription Type': {}, 'Contract Length': {}, 'Total Spend': {}, 'Last Interaction': {}}, 'Churn': {}}}
```

5. Data Preparation

```
In [10]: import matplotlib.pyplot as plt
import seaborn as sns

# Load the combined dataset
df = pd.read_csv('raw_ingested_data.csv')

# Handling Missing Values
# Check for missing values
print(df.isnull().sum())

# Imputation (Example: fill missing 'Age' with the mean)
df['Age'].fillna(df['Age'].mean(), inplace=True)

# Removal (Example: remove rows with missing 'Total Spend')
df.dropna(subset=['Total Spend'], inplace=True)

# Standardize/Normalize Numerical Attributes
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
numerical_cols = ['Age', 'Tenure', 'Usage Frequency', 'Support Calls', 'Total Spend', 'Last Interaction']
df[numerical_cols] = scaler.fit_transform(df[numerical_cols])
```



```

# Encode Categorical Variables
# One-hot encoding
df = pd.get_dummies(df, columns=['Gender', 'Payment Delay', 'Subscription Type', 'Contract Length'], drop_first=True)

# Label encoding (example for 'Churn' - if it's not already numerical)
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['Churn'] = le.fit_transform(df['Churn'])

# Exploratory Data Analysis (EDA)
# Histograms
df.hist(figsize=(12, 8))
plt.show()

# Box plots
for col in numerical_cols:
    plt.figure(figsize=(6, 4))
    sns.boxplot(x=col, data=df)
    plt.title(f"Boxplot of {col}")
    plt.show()

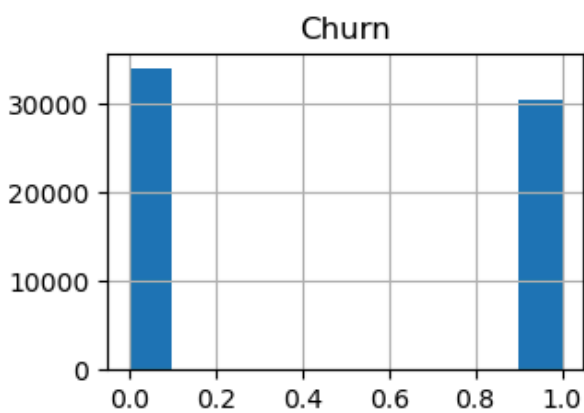
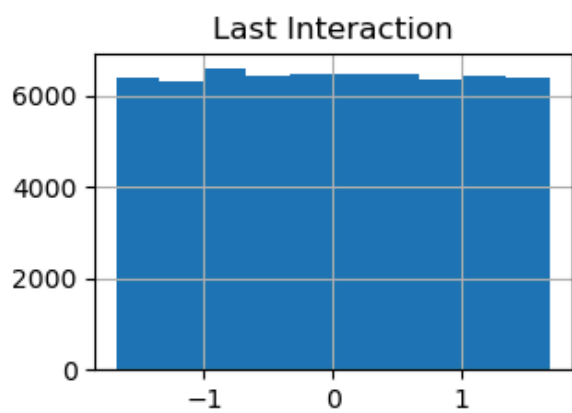
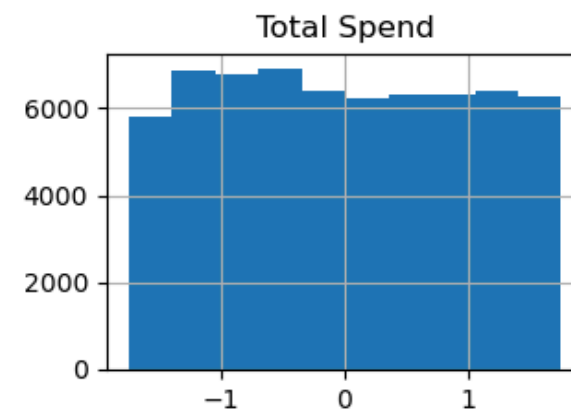
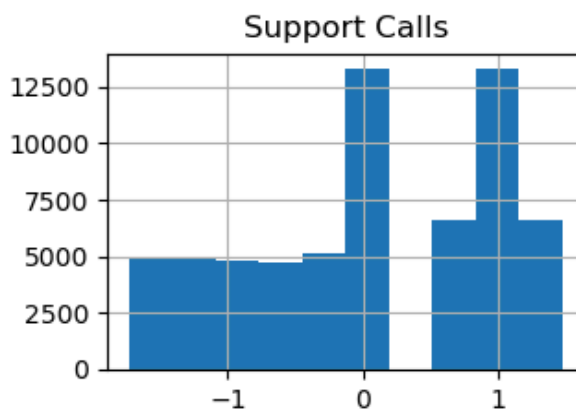
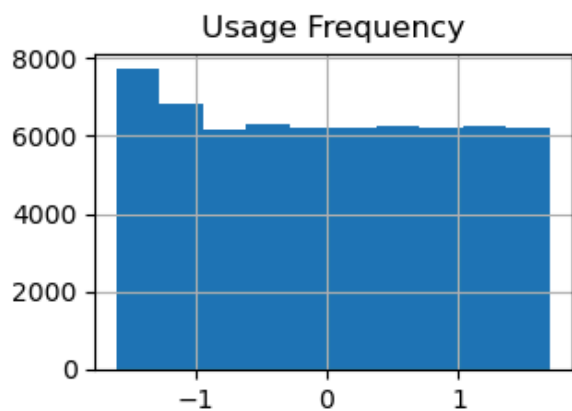
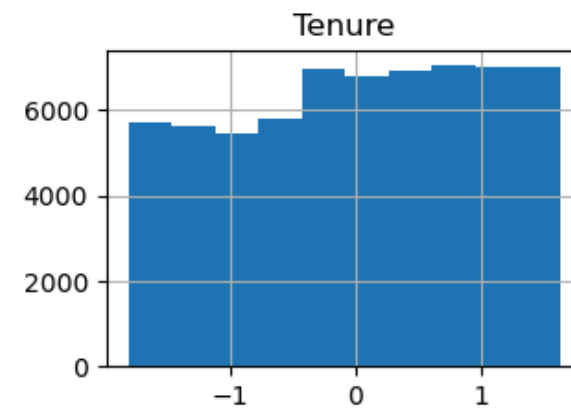
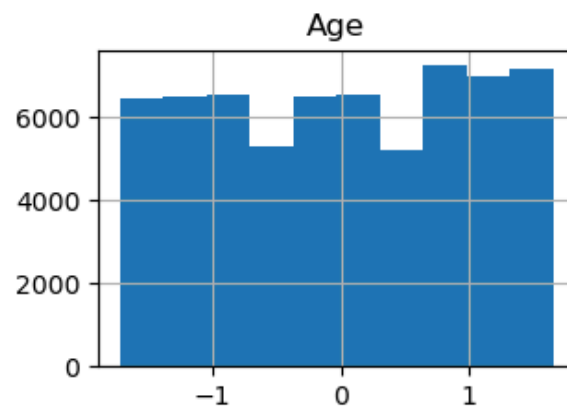
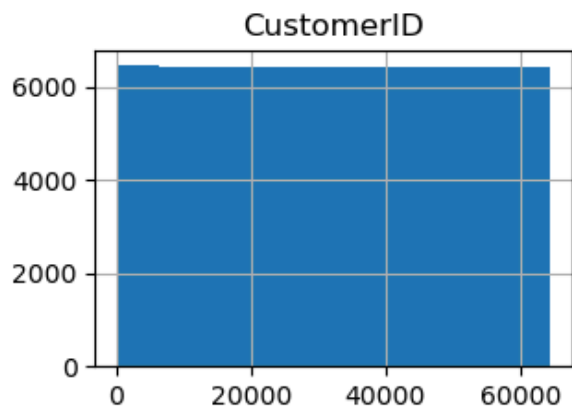
# Correlation Matrix
correlation_matrix = df.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Correlation Matrix")
plt.show()

# Outlier Detection (Example: using IQR for 'Total Spend')
Q1 = df['Total Spend'].quantile(0.25)
Q3 = df['Total Spend'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
df = df[(df['Total Spend'] >= lower_bound) & (df['Total Spend'] <= upper_bound)]

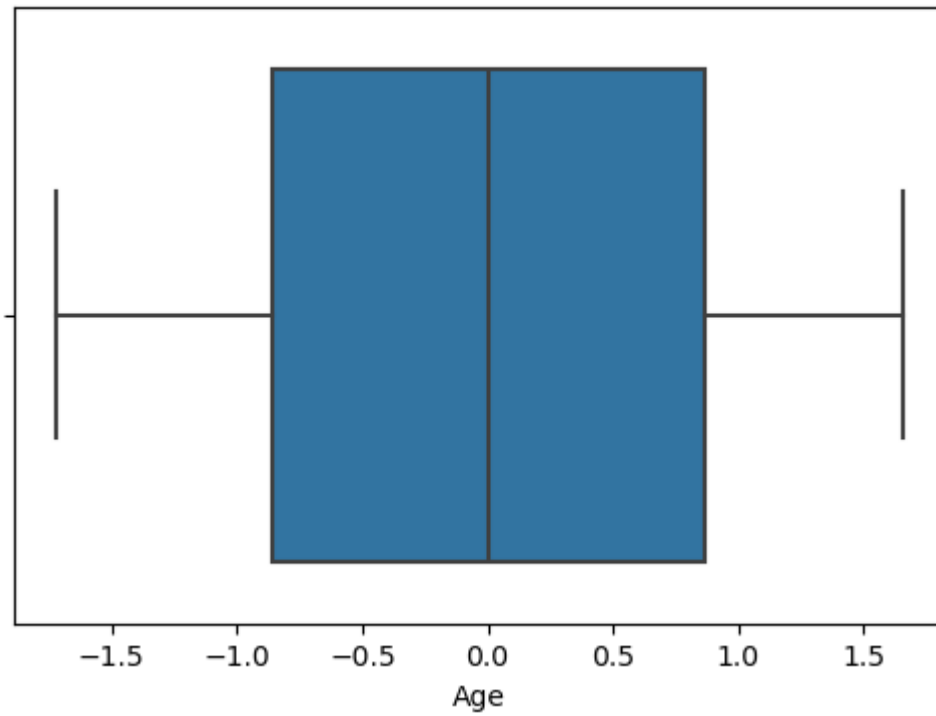
```

```
# Save the cleaned dataset  
df.to_csv("cleaned_data.csv", index=False)
```

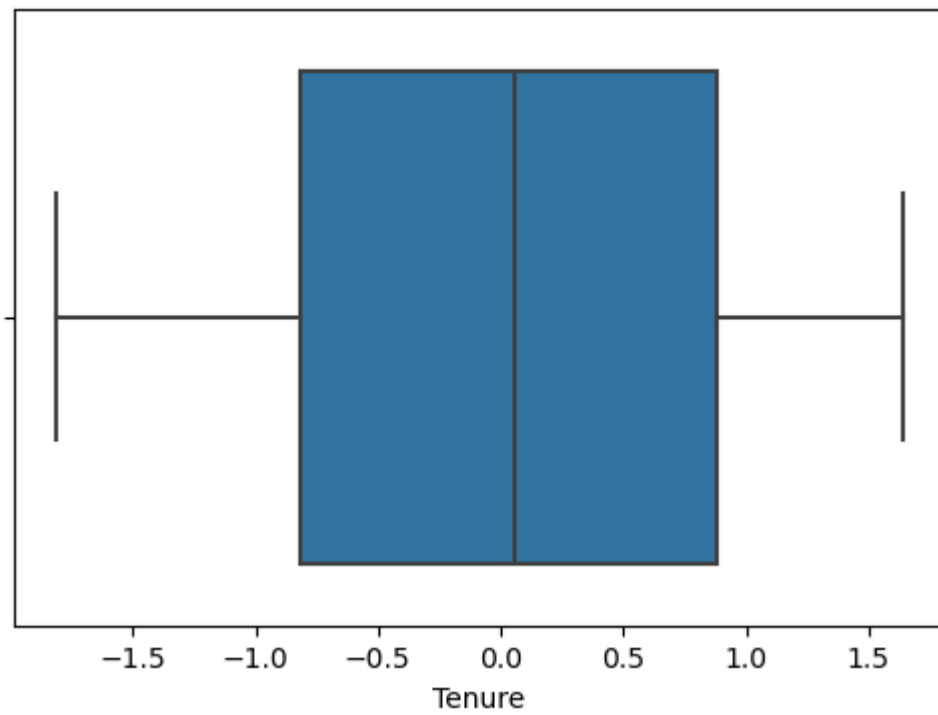
```
CustomerID      0  
Age             0  
Gender          0  
Tenure          0  
Usage Frequency 0  
Support Calls   0  
Payment Delay   0  
Subscription Type 0  
Contract Length 0  
Total Spend     0  
Last Interaction 0  
Churn           0  
dtype: int64
```



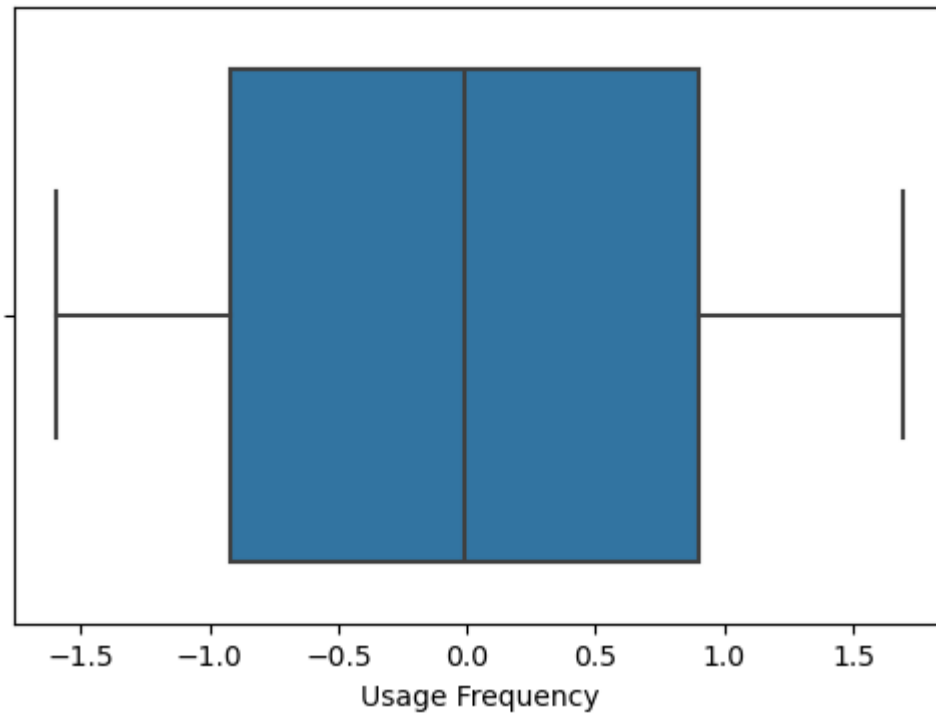
Boxplot of Age



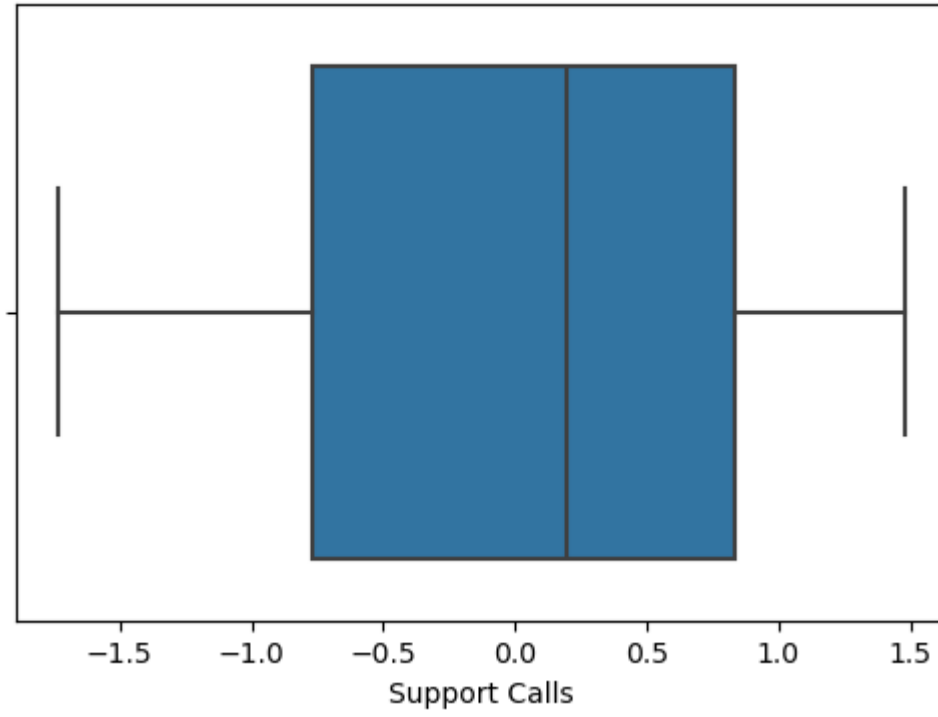
Boxplot of Tenure



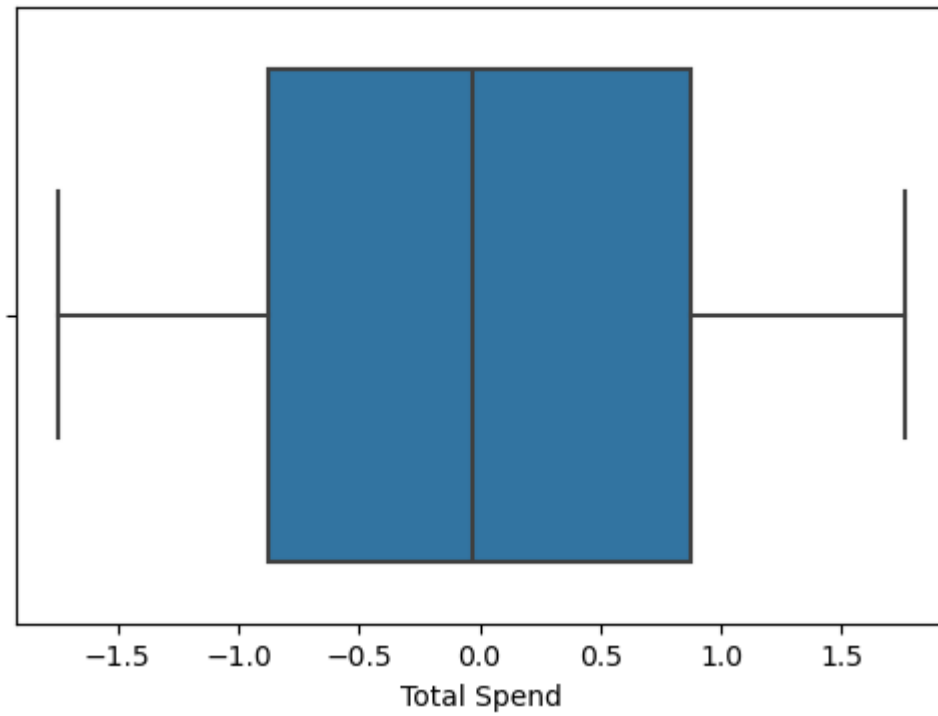
Boxplot of Usage Frequency



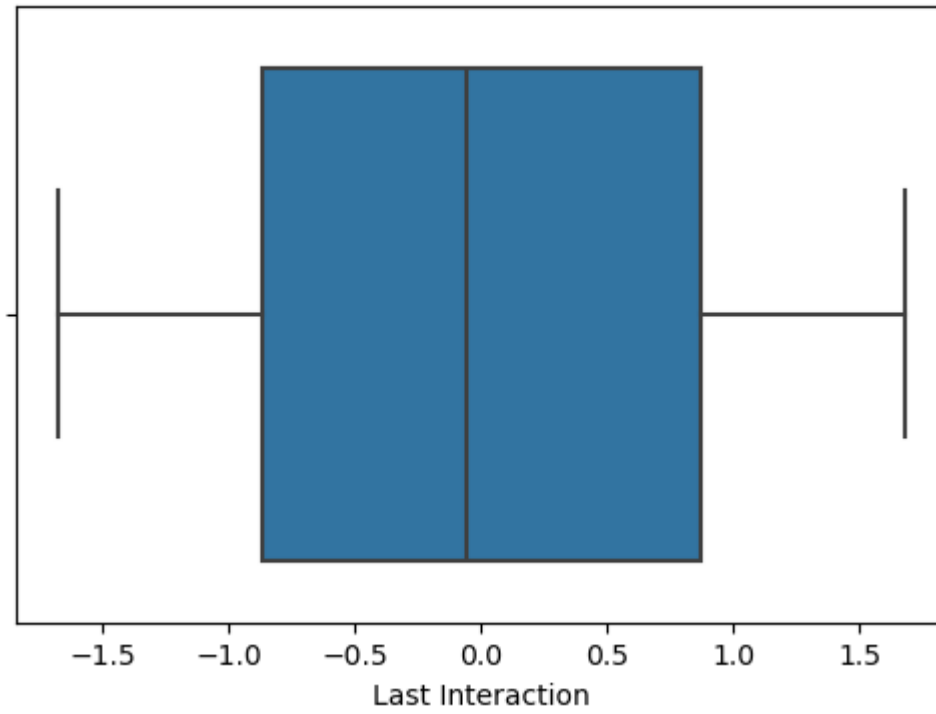
Boxplot of Support Calls

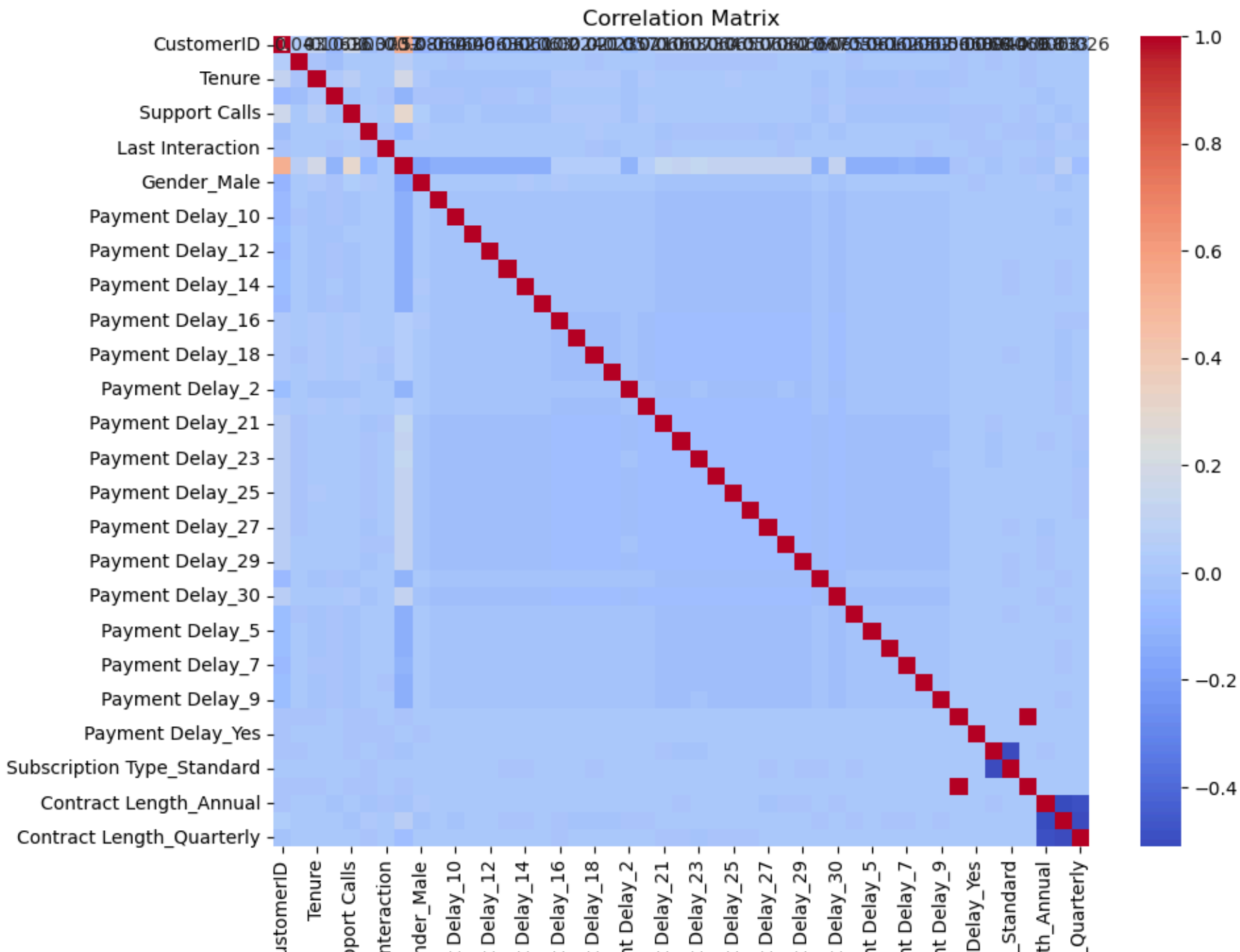


Boxplot of Total Spend



Boxplot of Last Interaction





Contract ID	Subscription Type	Contract Length	Contract Start Date	Contract End Date	Contract Value	Contract Status	Contract Owner	Contract Manager	Contract Description
1001	Subscription	12 Months	2023-01-01	2023-12-31	\$12,000	Active	John Doe	Jane Smith	Subscription for 12 months
1002	Subscription	6 Months	2023-02-01	2023-07-31	\$6,000	Active	John Doe	Jane Smith	Subscription for 6 months
1003	Subscription	24 Months	2023-03-01	2025-02-28	\$24,000	Active	John Doe	Jane Smith	Subscription for 24 months
1004	Subscription	3 Months	2023-04-01	2023-06-30	\$3,000	Active	John Doe	Jane Smith	Subscription for 3 months
1005	Subscription	18 Months	2023-05-01	2024-04-30	\$18,000	Active	John Doe	Jane Smith	Subscription for 18 months
1006	Subscription	9 Months	2023-06-01	2023-08-31	\$9,000	Active	John Doe	Jane Smith	Subscription for 9 months
1007	Subscription	15 Months	2023-07-01	2024-01-31	\$15,000	Active	John Doe	Jane Smith	Subscription for 15 months
1008	Subscription	7 Months	2023-08-01	2023-10-31	\$7,000	Active	John Doe	Jane Smith	Subscription for 7 months
1009	Subscription	11 Months	2023-09-01	2023-11-30	\$11,000	Active	John Doe	Jane Smith	Subscription for 11 months
1010	Subscription	5 Months	2023-10-01	2023-12-31	\$5,000	Active	John Doe	Jane Smith	Subscription for 5 months
1011	Subscription	13 Months	2023-11-01	2024-01-31	\$13,000	Active	John Doe	Jane Smith	Subscription for 13 months
1012	Subscription	8 Months	2023-12-01	2024-03-31	\$8,000	Active	John Doe	Jane Smith	Subscription for 8 months
1013	Subscription	10 Months	2024-01-01	2024-04-30	\$10,000	Active	John Doe	Jane Smith	Subscription for 10 months
1014	Subscription	4 Months	2024-02-01	2024-05-31	\$4,000	Active	John Doe	Jane Smith	Subscription for 4 months
1015	Subscription	16 Months	2024-03-01	2025-01-31	\$16,000	Active	John Doe	Jane Smith	Subscription for 16 months
1016	Subscription	6 Months	2024-04-01	2024-09-30	\$6,000	Active	John Doe	Jane Smith	Subscription for 6 months
1017	Subscription	20 Months	2024-05-01	2025-04-30	\$20,000	Active	John Doe	Jane Smith	Subscription for 20 months
1018	Subscription	14 Months	2024-06-01	2025-02-28	\$14,000	Active	John Doe	Jane Smith	Subscription for 14 months
1019	Subscription	17 Months	2024-07-01	2025-05-31	\$17,000	Active	John Doe	Jane Smith	Subscription for 17 months
1020	Subscription	19 Months	2024-08-01	2025-06-30	\$19,000	Active	John Doe	Jane Smith	Subscription for 19 months
1021	Subscription	12 Months	2024-09-01	2025-08-31	\$12,000	Active	John Doe	Jane Smith	Subscription for 12 months
1022	Subscription	10 Months	2024-10-01	2025-07-31	\$10,000	Active	John Doe	Jane Smith	Subscription for 10 months
1023	Subscription	8 Months	2024-11-01	2025-06-30	\$8,000	Active	John Doe	Jane Smith	Subscription for 8 months
1024	Subscription	6 Months	2024-12-01	2025-05-31	\$6,000	Active	John Doe	Jane Smith	Subscription for 6 months
1025	Subscription	4 Months	2025-01-01	2025-04-30	\$4,000	Active	John Doe	Jane Smith	Subscription for 4 months
1026	Subscription	3 Months	2025-02-01	2025-04-30	\$3,000	Active	John Doe	Jane Smith	Subscription for 3 months
1027	Subscription	2 Months	2025-03-01	2025-04-30	\$2,000	Active	John Doe	Jane Smith	Subscription for 2 months
1028	Subscription	1 Month	2025-04-01	2025-04-30	\$1,000	Active	John Doe	Jane Smith	Subscription for 1 month
1029	Subscription	12 Months	2025-05-01	2026-04-30	\$12,000	Active	John Doe	Jane Smith	Subscription for 12 months
1030	Subscription	18 Months	2025-06-01	2026-09-30	\$18,000	Active	John Doe	Jane Smith	Subscription for 18 months
1031	Subscription	24 Months	2025-07-01	2027-06-30	\$24,000	Active	John Doe	Jane Smith	Subscription for 24 months
1032	Subscription	30 Months	2025-08-01	2027-11-30	\$30,000	Active	John Doe	Jane Smith	Subscription for 30 months
1033	Subscription	36 Months	2025-09-01	2028-08-31	\$36,000	Active	John Doe	Jane Smith	Subscription for 36 months
1034	Subscription	42 Months	2025-10-01	2028-11-30	\$42,000	Active	John Doe	Jane Smith	Subscription for 42 months
1035	Subscription	48 Months	2025-11-01	2029-10-31	\$48,000	Active	John Doe	Jane Smith	Subscription for 48 months
1036	Subscription	54 Months	2025-12-01	2030-09-30	\$54,000	Active	John Doe	Jane Smith	Subscription for 54 months

Data Transformation and Storage

```
In [11]: import sqlite3

# Create a connection to the SQLite database (or your chosen database)
conn = sqlite3.connect('customer_data.db')
cursor = conn.cursor()

# SQL Schema Design (Example)
cursor.execute('''
    CREATE TABLE IF NOT EXISTS transformed_customer_data (
        CustomerID INTEGER PRIMARY KEY,
        TotalSpend REAL,
        CustomerTenure INTEGER,
        ActivityFrequency INTEGER,
        ScaledAge REAL,
        -- Add other transformed features here
        Churn INTEGER
    )
''')

# Data Transformation Logic
# Assuming 'df' is your cleaned DataFrame from the previous steps
for index, row in df.iterrows():
    # Create aggregated features
    total_spend = row['Total Spend']

    # Derive new features
    customer_tenure = row['Tenure']
    activity_frequency = row['Usage Frequency']
```

```

# Scale and normalize features
scaled_age = row['Age'] # Already scaled in the previous step

# Insert into the table
cursor.execute('''
    INSERT OR REPLACE INTO transformed_customer_data (CustomerID, TotalSpend, CustomerTenure, ActivityFrequency, ScaledAge
    VALUES (?, ?, ?, ?, ?, ?)
    ''', (row['CustomerID'], total_spend, customer_tenure, activity_frequency, scaled_age, row['Churn']))

# Commit changes and close the connection
conn.commit()
conn.close()

```

Feature store

```

In [17]: ##### 1. FEATURE STORE IMPLEMENTATION
# Using Feast (Feature Store for Machine Learning)

!pip install feast

from feast import FeatureStore
import pandas as pd
import numpy as np
import logging
import random
from datetime import datetime, timedelta

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
# Configure logging
logging.basicConfig(level=logging.INFO, format="%(asctime)s - %(levelname)s - %(message)s")

def log_message(level, message):
    logging.log(level, message)

# Create a larger dataset
num_samples = 100

```

```

random.seed(42)

data = {
    "customer_id": list(range(1, num_samples + 1)),
    "total_spend": [random.uniform(100, 1000) for _ in range(num_samples)],
    "tenure": [random.randint(1, 60) for _ in range(num_samples)],
    "churn": [random.choice([0, 1]) for _ in range(num_samples)],
    "last_purchase": [(datetime(2024, 1, 1) + timedelta(days=random.randint(1, 365))).strftime('%Y-%m-%d') for _ in range(num_
}]

feature_df = pd.DataFrame(data)

# Convert integer columns to float to avoid MLflow schema issue
feature_df = feature_df.astype({"total_spend": "float64", "tenure": "float64", "churn": "float64"})

# Save features to CSV (Custom Feature Store Alternative)
feature_df.to_csv("feature_store.csv", index=False)
log_message(logging.INFO, "Feature store updated successfully with 100 samples!")
print("Feature store updated successfully with 100 samples!") # Added print statement

df = pd.read_csv("feature_store.csv")
print(df.head())

retrieved_features = pd.read_csv("feature_store.csv")
print("Retrieved Feature Data:")
print(retrieved_features.head())

```

2025-03-14 21:49:39,426 - INFO - Feature store updated successfully with 100 samples!

Requirement already satisfied: feast in c:\users\laksh\anaconda3\lib\site-packages (0.47.0)

Requirement already satisfied: click<9.0.0,>=7.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (8.1.7)

Requirement already satisfied: colorama<1,>=0.3.9 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.4.6)

Requirement already satisfied: dill~0.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.3.7)

Requirement already satisfied: protobuf>=4.24.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (5.29.3)

Requirement already satisfied: Jinja2<4,>=2 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (3.1.3)

Requirement already satisfied: jsonschema in c:\users\laksh\anaconda3\lib\site-packages (from feast) (4.19.2)

Requirement already satisfied: mmh3 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (5.1.0)

Requirement already satisfied: numpy<2,>=1.22 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (1.26.4)

Requirement already satisfied: pandas<3,>=1.4.3 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (2.1.4)

Requirement already satisfied: pyarrow<18.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (14.0.2)

Requirement already satisfied: pydantic>=2.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (2.10.6)

Requirement already satisfied: pygments<3,>=2.12.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (2.15.1)

Requirement already satisfied: PyYAML<7,>=5.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (6.0.1)

Requirement already satisfied: requests in c:\users\laksh\anaconda3\lib\site-packages (from feast) (2.31.0)

Requirement already satisfied: SQLAlchemy>1 in c:\users\laksh\anaconda3\lib\site-packages (from SQLAlchemy[mypy]>1->feast) (2.0.25)

Requirement already satisfied: tabulate<1,>=0.8.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.9.0)

Requirement already satisfied: tenacity<9,>=7 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (8.2.2)

Requirement already satisfied: toml<1,>=0.10.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.10.2)

Requirement already satisfied: tqdm<5,>=4 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (4.65.0)

Requirement already satisfied: typeguard>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (4.4.2)

Requirement already satisfied: fastapi>=0.68.0 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.115.11)

Requirement already satisfied: uvicorn<1,>=0.14.0 in c:\users\laksh\anaconda3\lib\site-packages (from uvicorn[standard]<1,>=0.14.0->feast) (0.34.0)

Requirement already satisfied: uvicorn-worker in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.3.0)

Requirement already satisfied: dask>=2024.2.1 in c:\users\laksh\anaconda3\lib\site-packages (from dask[dataframe]>=2024.2.1->feast) (2025.2.0)

Requirement already satisfied: prometheus_client in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.21.1)

Requirement already satisfied: psutil in c:\users\laksh\anaconda3\lib\site-packages (from feast) (5.9.0)

Requirement already satisfied: bigtree>=0.19.2 in c:\users\laksh\anaconda3\lib\site-packages (from feast) (0.25.2)

Requirement already satisfied: pyjwt in c:\users\laksh\anaconda3\lib\site-packages (from feast) (2.4.0)

Requirement already satisfied: cloudpickle>=3.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from dask>=2024.2.1->dask[dataframe]>=2024.2.1->feast) (3.1.1)

Requirement already satisfied: fsspec>=2021.09.0 in c:\users\laksh\anaconda3\lib\site-packages (from dask>=2024.2.1->dask[dataframe]>=2024.2.1->feast) (2025.3.0)

Requirement already satisfied: packaging>=20.0 in c:\users\laksh\anaconda3\lib\site-packages (from dask>=2024.2.1->dask[dataframe]>=2024.2.1->feast) (23.1)

Requirement already satisfied: partd>=1.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from dask>=2024.2.1->dask[dataframe]>=2024.2.1->feast) (1.4.1)

Requirement already satisfied: toolz>=0.10.0 in c:\users\laksh\anaconda3\lib\site-packages (from dask>=2024.2.1->dask[dataframe]

e]>=2024.2.1->feast) (0.12.0)
Requirement already satisfied: importlib_metadata>=4.13.0 in c:\users\laksh\anaconda3\lib\site-packages (from dask>=2024.2.1->dask[dataframe]>=2024.2.1->feast) (7.0.1)
Requirement already satisfied: starlette<0.47.0,>=0.40.0 in c:\users\laksh\anaconda3\lib\site-packages (from fastapi>=0.68.0->fastapi) (0.46.1)
Requirement already satisfied: typing-extensions>=4.8.0 in c:\users\laksh\anaconda3\lib\site-packages (from fastapi>=0.68.0->fastapi) (4.12.2)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from Jinja2<4,>=2->feast) (2.1.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\laksh\anaconda3\lib\site-packages (from pandas<3,>=1.4.3->feast) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\laksh\anaconda3\lib\site-packages (from pandas<3,>=1.4.3->feast) (2025.1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\laksh\anaconda3\lib\site-packages (from pandas<3,>=1.4.3->feast) (2025.1)
Requirement already satisfied: annotated-types>=0.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic>=2.0.0->fastapi) (0.7.0)
Requirement already satisfied: pydantic-core==2.27.2 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic>=2.0.0->fastapi) (2.27.2)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\laksh\anaconda3\lib\site-packages (from SQLAlchemy>1->SQLAlchemy[mypy]>1->feast) (3.0.1)
Requirement already satisfied: mypy>=0.910 in c:\users\laksh\anaconda3\lib\site-packages (from SQLAlchemy[mypy]>1->feast) (1.8.0)
Requirement already satisfied: h11>=0.8 in c:\users\laksh\anaconda3\lib\site-packages (from uvicorn<1,>=0.14.0->uvicorn[standard]<1,>=0.14.0->feast) (0.14.0)
Requirement already satisfied: httptools>=0.6.3 in c:\users\laksh\anaconda3\lib\site-packages (from uvicorn[standard]<1,>=0.14.0->feast) (0.6.4)
Requirement already satisfied: python-dotenv>=0.13 in c:\users\laksh\anaconda3\lib\site-packages (from uvicorn[standard]<1,>=0.14.0->feast) (0.21.0)
Requirement already satisfied: watchfiles>=0.13 in c:\users\laksh\anaconda3\lib\site-packages (from uvicorn[standard]<1,>=0.14.0->feast) (1.0.4)
Requirement already satisfied: websockets>=10.4 in c:\users\laksh\anaconda3\lib\site-packages (from uvicorn[standard]<1,>=0.14.0->feast) (15.0.1)
Requirement already satisfied: attrs>=22.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema->feast) (23.1.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema->feast) (2023.7.1)
Requirement already satisfied: referencing>=0.28.4 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema->feast) (0.30.2)
Requirement already satisfied: rpds-py>=0.7.1 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema->feast) (0.10.6)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\laksh\anaconda3\lib\site-packages (from requests->feast) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\laksh\anaconda3\lib\site-packages (from requests->feast) (3.4)

Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\laksh\anaconda3\lib\site-packages (from requests->feast) (2.0.7)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\laksh\anaconda3\lib\site-packages (from requests->feast) (2024.7.4)

Requirement already satisfied: gunicorn>=20.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from uvicorn-worker->feast) (23.0.0)

Requirement already satisfied: zipp>=0.5 in c:\users\laksh\anaconda3\lib\site-packages (from importlib_metadata>=4.13.0->dask>=2024.2.1->dask[dataframe]>=2024.2.1->feast) (3.17.0)

Requirement already satisfied: mypy_extensions>=1.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from mypy>=0.910->SQLAlchemy[mypy]>1->feast) (1.0.0)

Requirement already satisfied: locket in c:\users\laksh\anaconda3\lib\site-packages (from partd>=1.4.0->dask>=2024.2.1->dask[dataframe]>=2024.2.1->feast) (1.0.0)

Requirement already satisfied: six>=1.5 in c:\users\laksh\anaconda3\lib\site-packages (from python_dateutil>=2.8.2->pandas<3,>=1.4.3->feast) (1.16.0)

Requirement already satisfied: anyio<5,>=3.6.2 in c:\users\laksh\anaconda3\lib\site-packages (from starlette<0.47.0,>=0.40.0->fastapi>=0.68.0->feast) (4.8.0)

Requirement already satisfied: sniffio>=1.1 in c:\users\laksh\anaconda3\lib\site-packages (from anyio<5,>=3.6.2->starlette<0.47.0,>=0.40.0->fastapi>=0.68.0->feast) (1.3.0)

Feature store updated successfully with 100 samples!

	customer_id	total_spend	tenure	churn	last_purchase
0	1	675.484119	1.0	0.0	2024-07-22
1	2	122.509680	44.0	0.0	2024-03-09
2	3	347.526387	47.0	0.0	2024-12-10
3	4	300.889664	8.0	1.0	2024-11-27
4	5	762.824093	44.0	0.0	2024-06-03

Retrieved Feature Data:

	customer_id	total_spend	tenure	churn	last_purchase
0	1	675.484119	1.0	0.0	2024-07-22
1	2	122.509680	44.0	0.0	2024-03-09
2	3	347.526387	47.0	0.0	2024-12-10
3	4	300.889664	8.0	1.0	2024-11-27
4	5	762.824093	44.0	0.0	2024-06-03

Data Versioning

```
In [13]: ##### 2. DATA VERSIONING
# Using DVC (Data Version Control) for tracking datasets
# Install DVC (only needed once)
!pip install dvc
```



```
# Change directory using Python (not cd)
import os

# Use raw string or double backslashes
os.chdir(r"C:\Users\laksh\Downloads\DMML_assignment")

# Verify the current working directory
print("Current Directory:", os.getcwd())

# Initialize DVC (force re-init if needed)
!dvc init -f

# Add feature store CSV to DVC tracking
!dvc add feature_store.csv

# Track the DVC file with Git
!git add feature_store.csv.dvc .gitignore .dvc
!git commit -m "Added feature store tracking"
```

Requirement already satisfied: dvc in c:\users\laksh\anaconda3\lib\site-packages (3.59.1)
Requirement already satisfied: attrs>=22.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (23.1.0)
Requirement already satisfied: celery in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (5.4.0)
Requirement already satisfied: colorama>=0.3.9 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.4.6)
Requirement already satisfied: configobj>=5.0.9 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (5.0.9)
Requirement already satisfied: distro>=1.3 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (1.8.0)
Requirement already satisfied: dpath<3,>=2.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (2.2.0)
Requirement already satisfied: dulwich in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.22.8)
Requirement already satisfied: dvc-data<3.17,>=3.16.2 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (3.16.9)
Requirement already satisfied: dvc-http>=2.29.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (2.32.0)
Requirement already satisfied: dvc-objects in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (5.1.0)
Requirement already satisfied: dvc-render<2,>=1.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (1.0.2)
Requirement already satisfied: dvc-studio-client<1,>=0.21 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.21.0)
Requirement already satisfied: dvc-task<1,>=0.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.40.2)
Requirement already satisfied: flatten_dict<1,>=0.4.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.4.2)
Requirement already satisfied: fluff.lock<9,>=8.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (8.1.0)
Requirement already satisfied: fsspec>=2024.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (2025.3.0)
Requirement already satisfied: funcy>=1.14 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (2.0)
Requirement already satisfied: grandalf<1,>=0.7 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.8)
Requirement already satisfied: gto<2,>=1.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (1.7.2)
Requirement already satisfied: hydra-core>=1.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (1.3.2)
Requirement already satisfied: iterative-telemetry>=0.0.7 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.0.10)
Requirement already satisfied: kombu in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (5.5.0)
Requirement already satisfied: networkx>=2.5 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (3.1)
Requirement already satisfied: omegaconf in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (2.3.0)
Requirement already satisfied: packaging>=19 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (23.1)
Requirement already satisfied: pathspec>=0.10.3 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.10.3)
Requirement already satisfied: platformdirs<5,>=3.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (3.10.0)
Requirement already satisfied: psutil>=5.8 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (5.9.0)
Requirement already satisfied: pydot>=1.2.4 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (3.0.4)
Requirement already satisfied: pygtrie>=2.3.2 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (2.5.0)
Requirement already satisfied: pyparsing>=2.4.7 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (3.0.9)
Requirement already satisfied: requests>=2.22 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (2.31.0)
Requirement already satisfied: rich>=12 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (13.3.5)
Requirement already satisfied: ruamel.yaml>=0.17.11 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.17.21)
Requirement already satisfied: scmrepo<4,>=3.3.8 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (3.3.10)
Requirement already satisfied: shortuuid>=0.5 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (1.0.13)
Requirement already satisfied: shtab<2,>=1.3.4 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (1.7.1)
Requirement already satisfied: tabulate>=0.8.7 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.9.0)
Requirement already satisfied: tomlkit>=0.11.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.11.1)
Requirement already satisfied: tqdm<5,>=4.63.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (4.65.0)

Requirement already satisfied: voluptuous>=0.11.7 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (0.15.2)

Requirement already satisfied: zc.lockfile>=1.2.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc) (3.0.post1)

Requirement already satisfied: dictdiffer>=0.8.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc-data<3.17,>=3.16.2->dvc) (0.9.0)

Requirement already satisfied: diskcache>=5.2.1 in c:\users\laksh\anaconda3\lib\site-packages (from dvc-data<3.17,>=3.16.2->dvc) (5.6.3)

Requirement already satisfied: sqltrie<1,>=0.11.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc-data<3.17,>=3.16.2->dvc) (0.11.2)

Requirement already satisfied: orjson<4,>=3 in c:\users\laksh\anaconda3\lib\site-packages (from dvc-data<3.17,>=3.16.2->dvc) (3.10.15)

Requirement already satisfied: aiohttp-retry>=2.5.0 in c:\users\laksh\anaconda3\lib\site-packages (from dvc-http>=2.29.0->dvc) (2.9.1)

Requirement already satisfied: pywin32>=225 in c:\users\laksh\anaconda3\lib\site-packages (from dvc-task<1,>=0.3.0->dvc) (305.1)

Requirement already satisfied: billiard<5.0,>=4.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (4.2.1)

Requirement already satisfied: vine<6.0,>=5.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (5.1.0)

Requirement already satisfied: click<9.0,>=8.1.2 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (8.1.7)

Requirement already satisfied: click-didyoumean>=0.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (0.3.1)

Requirement already satisfied: click-repl>=0.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (0.3.0)

Requirement already satisfied: click-plugins>=1.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (1.1.1)

Requirement already satisfied: tzdata>=2022.7 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (2025.1)

Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\laksh\anaconda3\lib\site-packages (from celery->dvc) (2.8.2)

Requirement already satisfied: six<2.0,>=1.12 in c:\users\laksh\anaconda3\lib\site-packages (from flatten_dict<1,>=0.4.1->dvc) (1.16.0)

Requirement already satisfied: atpublic in c:\users\laksh\anaconda3\lib\site-packages (from fluf1.lock<9,>=8.1.0->dvc) (5.1)

Requirement already satisfied: entrypoints in c:\users\laksh\anaconda3\lib\site-packages (from gto<2,>=1.6.0->dvc) (0.4)

Requirement already satisfied: pydantic!=2.0.0,<3,>=1.9.0 in c:\users\laksh\anaconda3\lib\site-packages (from gto<2,>=1.6.0->dvc) (2.10.6)

Requirement already satisfied: semver>=2.13.0 in c:\users\laksh\anaconda3\lib\site-packages (from gto<2,>=1.6.0->dvc) (2.13.0)

Requirement already satisfied: typer>=0.4.1 in c:\users\laksh\anaconda3\lib\site-packages (from gto<2,>=1.6.0->dvc) (0.15.2)

Requirement already satisfied: antlr4-python3-runtime==4.9.* in c:\users\laksh\anaconda3\lib\site-packages (from hydra-core>=1.1->dvc) (4.9.3)

Requirement already satisfied: appdirs in c:\users\laksh\anaconda3\lib\site-packages (from iterative-telemetry>=0.0.7->dvc) (1.4.4)

Requirement already satisfied: filelock in c:\users\laksh\anaconda3\lib\site-packages (from iterative-telemetry>=0.0.7->dvc) (3.13.1)

Requirement already satisfied: amqp<6.0.0,>=5.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from kombu->dvc) (5.3.1)

Requirement already satisfied: PyYAML>=5.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from omegaconf->dvc) (6.0.1)

Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\laksh\anaconda3\lib\site-packages (from requests>=2.22->dvc) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in c:\users\laksh\anaconda3\lib\site-packages (from requests>=2.22->dvc) (3.4)

Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\laksh\anaconda3\lib\site-packages (from requests>=2.22->dvc) (2.0.7)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\laksh\anaconda3\lib\site-packages (from requests>=2.22->dvc) (2024.7.4)

Requirement already satisfied: markdown-it-py<3.0.0,>=2.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from rich>=12->dvc) (2.2.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\users\laksh\anaconda3\lib\site-packages (from rich>=12->dvc) (2.15.1)

Requirement already satisfied: gitpython>3 in c:\users\laksh\anaconda3\lib\site-packages (from scmrepo<4,>=3.3.8->dvc) (3.1.37)

Requirement already satisfied: pygit2>=1.14.0 in c:\users\laksh\anaconda3\lib\site-packages (from scmrepo<4,>=3.3.8->dvc) (1.17.0)

Requirement already satisfied: asyncssh<3,>=2.13.1 in c:\users\laksh\anaconda3\lib\site-packages (from scmrepo<4,>=3.3.8->dvc) (2.20.0)

Requirement already satisfied: setuptools in c:\users\laksh\anaconda3\lib\site-packages (from zc.lockfile>=1.2.1->dvc) (68.2.2)

Requirement already satisfied: aiohttp in c:\users\laksh\anaconda3\lib\site-packages (from aiohttp-retry>=2.5.0->dvc-http>=2.29.0->dvc) (3.9.3)

Requirement already satisfied: cryptography>=39.0 in c:\users\laksh\anaconda3\lib\site-packages (from asyncssh<3,>=2.13.1->scmrepo<4,>=3.3.8->dvc) (42.0.2)

Requirement already satisfied: typing_extensions>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from asyncssh<3,>=2.13.1->scmrepo<4,>=3.3.8->dvc) (4.12.2)

Requirement already satisfied: prompt-toolkit>=3.0.36 in c:\users\laksh\anaconda3\lib\site-packages (from click-repl>=0.2.0->celery->dvc) (3.0.43)

Requirement already satisfied: gitdb<5,>=4.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from gitpython>3->scmrepo<4,>=3.3.8->dvc) (4.0.7)

Requirement already satisfied: mdurl~=0.1 in c:\users\laksh\anaconda3\lib\site-packages (from markdown-it-py<3.0.0,>=2.2.0->rich>=12->dvc) (0.1.0)

Requirement already satisfied: annotated-types>=0.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic!=2.0.0,<3,>=1.9.0->gto<2,>=1.6.0->dvc) (0.7.0)

Requirement already satisfied: pydantic-core==2.27.2 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic!=2.0.0,<3,>=1.9.0->gto<2,>=1.6.0->dvc) (2.27.2)

Requirement already satisfied: cffi>=1.17.0 in c:\users\laksh\anaconda3\lib\site-packages (from pygit2>=1.14.0->scmrepo<4,>=3.3.8->dvc) (1.17.1)

Requirement already satisfied: shellingham>=1.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from typer>=0.4.1->gto<2,>=1.6.0->dvc) (1.5.4)

Requirement already satisfied: aiosignal>=1.1.2 in c:\users\laksh\anaconda3\lib\site-packages (from aiohttp->aiohttp-retry>=2.5.0->dvc-http>=2.29.0->dvc) (1.2.0)

Requirement already satisfied: frozenlist>=1.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from aiohttp->aiohttp-retry>=2.5.0->dvc-http>=2.29.0->dvc) (1.4.0)

Requirement already satisfied: multidict<7.0,>=4.5 in c:\users\laksh\anaconda3\lib\site-packages (from aiohttp->aiohttp-retry>=2.5.0->dvc-http>=2.29.0->dvc) (6.0.4)

Requirement already satisfied: yarl<2.0,>=1.0 in c:\users\laksh\anaconda3\lib\site-packages (from aiohttp->aiohttp-retry>=2.5.0->dvc-http>=2.29.0->dvc) (1.9.7)

```
->dvc-http>=2.29.0->dvc) (1.9.3)
Requirement already satisfied: pycparser in c:\users\laksh\anaconda3\lib\site-packages (from cffi>=1.17.0->pygit2>=1.14.0->scmr
epo<4,>=3.3.8->dvc) (2.21)
Requirement already satisfied: smmap<5,>=3.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from gitdb<5,>=4.0.1->gitpython>3
->scmrepo<4,>=3.3.8->dvc) (4.0.0)
Requirement already satisfied: wcwidth in c:\users\laksh\anaconda3\lib\site-packages (from prompt-toolkit>=3.0.36->click-repl>=
0.2.0->celery->dvc) (0.2.5)
Current Directory: C:\Users\laksh\Downloads\DMML_assignment
Initialized DVC repository.
```

You can now commit the changes to git.

```
+-----+
|
|      DVC has enabled anonymous aggregate usage analytics.
|      Read the analytics documentation (and how to opt-out) here:
|      <https://dvc.org/doc/user-guide/analytics>
|
+-----+
```

What's next?

- Check out the documentation: <<https://dvc.org/doc>>
- Get help and share ideas: <<https://dvc.org/chat>>
- Star us on GitHub: <<https://github.com/iterative/dvc>>

To track the changes with git, run:

```
git add feature_store.csv.dvc
```

To enable auto staging, run:

```
dvc config core.autostage true
```

\u280b Checking graph

On branch master

Your branch is up to date with 'origin/master'.

Changes not staged for commit:

(use "git add <file>..." to update what will be committed)

(use "git restore <file>..." to discard changes in working directory)

modified: DMML_assignment_group52.ipynb

modified: raw_ingested_data.csv

Untracked files:

(use "git add <file>..." to include in what will be committed)

.ipynb_checkpoints/DMML_Assignment_Group52-checkpoint.ipynb

cleaned_data.csv

customer_data.db

data_quality_report.csv

no changes added to commit (use "git add" and/or "git commit -a")

Model Building

```
In [18]: ##### 3. MODEL BUILDING & TRACKING WITH MLflow
# Install necessary libraries

def fetch_csv_data():
    try:
        df = pd.read_csv("feature_store.csv")
        log_message(logging.INFO, "Successfully ingested data from CSV file.")
        return df
    except Exception as e:
        log_message(logging.ERROR, f"Failed to fetch CSV data: {e}")
        return None

##### 3. MODEL BUILDING & TRACKING WITH MLflow
# Install necessary libraries
!pip install mlflow scikit-learn

import mlflow
import mlflow.sklearn
```

```

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

def log_message(level, message):
    logging.log(level, message)
# Load dataset
feature_df = fetch_csv_data()
if feature_df is None:
    raise ValueError("Error loading feature store. Check CSV file.")

X = feature_df[["total_spend", "tenure"]]
y = feature_df["churn"].astype(float) # Ensure labels are float to prevent MLflow warning

test_size = 0.2
# Disable stratify if dataset is too small
stratify_option = y if len(y.unique()) > 1 and len(y) * test_size >= len(y.unique()) else None

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, random_state=42, stratify=stratify_option)

# Train model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
log_message(logging.INFO, f"Model trained with accuracy: {accuracy}")

# Log model with MLflow
mlflow.set_experiment("churn_prediction")
with mlflow.start_run():
    mlflow.log_param("n_estimators", 100)
    mlflow.log_metric("accuracy", accuracy)

# Provide an example input with proper feature names
input_example = pd.DataFrame(X_test.iloc[:1], columns=X_train.columns)
mlflow.sklearn.log_model(model, "churn_model", input_example=input_example)

```

```
print(f"Model trained and logged with accuracy: {accuracy}")
```

```
2025-03-14 21:50:04,946 - INFO - Successfully ingested data from CSV file.
```

```
2025-03-14 21:50:05,038 - INFO - Model trained with accuracy: 0.45
```


Requirement already satisfied: mlflow in c:\users\laksh\anaconda3\lib\site-packages (2.20.4)
Requirement already satisfied: scikit-learn in c:\users\laksh\anaconda3\lib\site-packages (1.2.2)
Requirement already satisfied: mlflow-skinny==2.20.4 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (2.20.4)
Requirement already satisfied: Flask<4 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (2.2.5)
Requirement already satisfied: Jinja2<4,>=3.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (3.1.3)
Requirement already satisfied: alembic!=1.10.0,<2 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (1.15.1)
Requirement already satisfied: docker<8,>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (7.1.0)
Requirement already satisfied: graphene<4 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (3.4.3)
Requirement already satisfied: markdown<4,>=3.3 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (3.4.1)
Requirement already satisfied: matplotlib<4 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (3.8.0)
Requirement already satisfied: numpy<3 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (1.26.4)
Requirement already satisfied: pandas<3 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (2.1.4)
Requirement already satisfied: pyarrow<20,>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (14.0.2)
Requirement already satisfied: scipy<2 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (1.11.4)
Requirement already satisfied: sqlalchemy<3,>=1.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (2.0.25)
Requirement already satisfied: waitress<4 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow) (3.0.2)
Requirement already satisfied: cachetools<6,>=5.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (5.5.2)
Requirement already satisfied: click<9,>=7.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (8.1.7)
Requirement already satisfied: cloudpickle<4 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (3.1.1)
Requirement already satisfied: databricks-sdk<1,>=0.20.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (0.46.0)
Requirement already satisfied: gitpython<4,>=3.1.9 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (3.1.37)
Requirement already satisfied: importlib_metadata!=4.7.0,<9,>=3.7.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (7.0.1)
Requirement already satisfied: opentelemetry-api<3,>=1.9.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (1.31.0)
Requirement already satisfied: opentelemetry-sdk<3,>=1.9.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (1.31.0)
Requirement already satisfied: packaging<25 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (23.1)
Requirement already satisfied: protobuf<6,>=3.12.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (5.29.3)
Requirement already satisfied: pydantic<3,>=1.10.8 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (2.10.6)
Requirement already satisfied: pyyaml<7,>=5.1 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (6.0.1)
Requirement already satisfied: requests<3,>=2.17.3 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (2.31.0)

lflow) (2.31.0)
Requirement already satisfied: sqlparse<1,>=0.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (0.5.3)
Requirement already satisfied: typing-extensions<5,>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from mlflow-skinny==2.20.4->mlflow) (4.12.2)
Requirement already satisfied: joblib>=1.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: Mako in c:\users\laksh\anaconda3\lib\site-packages (from alembic!=1.10.0,<2->mlflow) (1.3.9)
Requirement already satisfied: pywin32>=304 in c:\users\laksh\anaconda3\lib\site-packages (from docker<8,>=4.0.0->mlflow) (305.1)
Requirement already satisfied: urllib3>=1.26.0 in c:\users\laksh\anaconda3\lib\site-packages (from docker<8,>=4.0.0->mlflow) (2.0.7)
Requirement already satisfied: Werkzeug>=2.2.2 in c:\users\laksh\anaconda3\lib\site-packages (from Flask<4->mlflow) (2.2.3)
Requirement already satisfied: itsdangerous>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from Flask<4->mlflow) (2.0.1)
Requirement already satisfied: graphql-core<3.3,>=3.1 in c:\users\laksh\anaconda3\lib\site-packages (from graphene<4->mlflow) (3.2.6)
Requirement already satisfied: graphql-relay<3.3,>=3.1 in c:\users\laksh\anaconda3\lib\site-packages (from graphene<4->mlflow) (3.2.0)
Requirement already satisfied: python-dateutil<3,>=2.7.0 in c:\users\laksh\anaconda3\lib\site-packages (from graphene<4->mlflow) (2.8.2)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from Jinja2<4,>=3.0->mlflow) (2.1.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from matplotlib<4->mlflow) (1.2.0)
Requirement already satisfied: cyclor>=0.10 in c:\users\laksh\anaconda3\lib\site-packages (from matplotlib<4->mlflow) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\laksh\anaconda3\lib\site-packages (from matplotlib<4->mlflow) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from matplotlib<4->mlflow) (1.4.4)
Requirement already satisfied: pillow>=6.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from matplotlib<4->mlflow) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\laksh\anaconda3\lib\site-packages (from matplotlib<4->mlflow) (3.0.9)
Requirement already satisfied: pytz>=2020.1 in c:\users\laksh\anaconda3\lib\site-packages (from pandas<3->mlflow) (2025.1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\laksh\anaconda3\lib\site-packages (from pandas<3->mlflow) (2025.1)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\laksh\anaconda3\lib\site-packages (from sqlalchemy<3,>=1.4.0->mlflow) (3.0.1)
Requirement already satisfied: colorama in c:\users\laksh\anaconda3\lib\site-packages (from click<9,>=7.0->mlflow-skinny==2.20.4->mlflow) (0.4.6)
Requirement already satisfied: google-auth~=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from databricks-sdk<1,>=0.20.0->mlflow-skinny==2.20.4->mlflow) (2.38.0)
Requirement already satisfied: gitdb<5,>=4.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from gitpython<4,>=3.1.9->mlflow-

```

skinny==2.20.4->mlflow) (4.0.7)
Requirement already satisfied: zipp>=0.5 in c:\users\laksh\anaconda3\lib\site-packages (from importlib_metadata!=4.7.0,<9,>=3.7.0->mlflow-skinny==2.20.4->mlflow) (3.17.0)
Requirement already satisfied: deprecated>=1.2.6 in c:\users\laksh\anaconda3\lib\site-packages (from opentelemetry-api<3,>=1.9.0->mlflow-skinny==2.20.4->mlflow) (1.2.18)
Requirement already satisfied: opentelemetry-semantic-conventions==0.52b0 in c:\users\laksh\anaconda3\lib\site-packages (from opentelemetry-sdk<3,>=1.9.0->mlflow-skinny==2.20.4->mlflow) (0.52b0)
Requirement already satisfied: annotated-types>=0.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic<3,>=1.10.8->mlflow-skinny==2.20.4->mlflow) (0.7.0)
Requirement already satisfied: pydantic-core==2.27.2 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic<3,>=1.10.8->mlflow-skinny==2.20.4->mlflow) (2.27.2)
Requirement already satisfied: six>=1.5 in c:\users\laksh\anaconda3\lib\site-packages (from python-dateutil<3,>=2.7.0->graphene<4->mlflow) (1.16.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\laksh\anaconda3\lib\site-packages (from requests<3,>=2.17.3->mlflow-skinny==2.20.4->mlflow) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\laksh\anaconda3\lib\site-packages (from requests<3,>=2.17.3->mlflow-skinny==2.20.4->mlflow) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\laksh\anaconda3\lib\site-packages (from requests<3,>=2.17.3->mlflow-skinny==2.20.4->mlflow) (2024.7.4)
Requirement already satisfied: wrapt<2,>=1.10 in c:\users\laksh\anaconda3\lib\site-packages (from deprecated>=1.2.6->opentelemetry-api<3,>=1.9.0->mlflow-skinny==2.20.4->mlflow) (1.14.1)
Requirement already satisfied: smmap<5,>=3.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from gitdb<5,>=4.0.1->gitpython<4,>=3.1.9->mlflow-skinny==2.20.4->mlflow) (4.0.0)
Requirement already satisfied: pyasn1-modules>=0.2.1 in c:\users\laksh\anaconda3\lib\site-packages (from google-auth~=2.0->databricks-sdk<1,>=0.20.0->mlflow-skinny==2.20.4->mlflow) (0.2.8)
Requirement already satisfied: rsa<5,>=3.1.4 in c:\users\laksh\anaconda3\lib\site-packages (from google-auth~=2.0->databricks-sdk<1,>=0.20.0->mlflow-skinny==2.20.4->mlflow) (4.9)
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in c:\users\laksh\anaconda3\lib\site-packages (from pyasn1-modules>=0.2.1->google-auth~=2.0->databricks-sdk<1,>=0.20.0->mlflow-skinny==2.20.4->mlflow) (0.4.8)
Downloading artifacts: 0%|          | 0/7 [00:00<?, ?it/s]
Model trained and logged with accuracy: 0.45

```

Orchestrating the Data Pipeline

Install Dependencies

```
In [19]: !pip install prefect
```

```
!pip install pandas scikit-learn prefect
```

Requirement already satisfied: prefect in c:\users\laksh\anaconda3\lib\site-packages (3.2.12)

Requirement already satisfied:aiosqlite<1.0.0,>=0.17.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.18.0)

Requirement already satisfied:alembic<2.0.0,>=1.7.5 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.15.1)

Requirement already satisfied:anyio<5.0.0,>=4.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.8.0)

Requirement already satisfied:apprise<2.0.0,>=1.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.9.2)

Requirement already satisfied:asgi-lifespan<3.0,>=1.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.1.0)

Requirement already satisfied:asyncpg<1.0.0,>=0.23 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.30.0)

Requirement already satisfied:cachetools<6.0,>=5.3 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (5.5.2)

Requirement already satisfied:click<8.2,>=8.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (8.1.7)

Requirement already satisfied:cloudpickle<4.0,>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.1.1)

Requirement already satisfied:coolname<3.0.0,>=1.0.4 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.2.0)

Requirement already satisfied:cryptography>=36.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (42.0.2)

Requirement already satisfied:dateparser<2.0.0,>=1.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.2.1)

Requirement already satisfied:docker<8.0,>=4.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (7.1.0)

Requirement already satisfied:exceptiongroup>=1.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.2.2)

Requirement already satisfied:fastapi<1.0.0,>=0.111.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.115.11)

Requirement already satisfied:fsspec>=2022.5.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2025.3.0)

Requirement already satisfied:graphviz>=0.20.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.20.3)

Requirement already satisfied:griffe<2.0.0,>=0.49.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.6.0)

Requirement already satisfied:httpcore<2.0.0,>=1.0.5 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.0.7)

Requirement already satisfied:httpx!=0.23.2,>=0.23 in c:\users\laksh\anaconda3\lib\site-packages (from httpx[http2]!=0.23.2,>=0.23->prefect) (0.26.0)

Requirement already satisfied:humanize<5.0.0,>=4.9.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.12.1)

Requirement already satisfied:jinja2-humanize-extension>=0.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.4.0)

Requirement already satisfied:jinja2<4.0.0,>=3.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.1.3)

Requirement already satisfied:jsonpatch<2.0,>=1.32 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.32)

Requirement already satisfied:jsonschema<5.0.0,>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.19.2)

Requirement already satisfied:opentelemetry-api<2.0.0,>=1.27.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.31.0)

Requirement already satisfied:orjson<4.0,>=3.7 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.10.15)

Requirement already satisfied:packaging<24.3,>=21.3 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (23.1)

Requirement already satisfied:pathspec>=0.8.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.10.3)

Requirement already satisfied:pendulum<4,>=3.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.0.0)

Requirement already satisfied:prometheus-client>=0.20.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.21.1)

Requirement already satisfied:pydantic!=2.10.0,<3.0.0,>=2.9 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.10.6)

Requirement already satisfied:pydantic-core<3.0.0,>=2.12.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.27.2)

Requirement already satisfied:pydantic-extra-types<3.0.0,>=2.8.2 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.10.3)

Requirement already satisfied: pydantic-settings>2.2.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.8.1)

Requirement already satisfied: python-dateutil<3.0.0,>=2.8.2 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.8.2)

Requirement already satisfied: python-slugify<9.0,>=5.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (5.0.2)

Requirement already satisfied: python-socks<3.0,>=2.5.3 in c:\users\laksh\anaconda3\lib\site-packages (from python-socks[asyncio]<3.0,>=2.5.3->prefect) (2.7.1)

Requirement already satisfied: pytz<2026,>=2021.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2025.1)

Requirement already satisfied: pyyaml<7.0.0,>=5.4.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (6.0.1)

Requirement already satisfied: readchar<5.0.0,>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.2.1)

Requirement already satisfied: rfc3339-validator<0.2.0,>=0.1.4 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.1.4)

Requirement already satisfied: rich<14.0,>=11.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (13.3.5)

Requirement already satisfied: ruamel-yaml>=0.17.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.17.21)

Requirement already satisfied: sniffio<2.0.0,>=1.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.3.0)

Requirement already satisfied: sqlalchemy<3.0.0,>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from sqlalchemy[asyncio]<3.0.0,>=2.0->prefect) (2.0.25)

Requirement already satisfied: toml>=0.10.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.10.2)

Requirement already satisfied: typer!=0.12.2,<0.16.0,>=0.12.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.15.2)

Requirement already satisfied: typing-extensions<5.0.0,>=4.5.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.12.2)

Requirement already satisfied: ujson<6.0.0,>=5.8.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (5.10.0)

Requirement already satisfied: uv>=0.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.6.6)

Requirement already satisfied: uvicorn!=0.29.0,>=0.14.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.34.0)

Requirement already satisfied: websockets<16.0,>=13.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (15.0.1)

Requirement already satisfied: Mako in c:\users\laksh\anaconda3\lib\site-packages (from alembic<2.0.0,>=1.7.5->prefect) (1.3.9)

Requirement already satisfied: idna>=2.8 in c:\users\laksh\anaconda3\lib\site-packages (from anyio<5.0.0,>=4.4.0->prefect) (3.4)

Requirement already satisfied: certifi in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (2024.7.4)

Requirement already satisfied: requests in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (2.31.0)

Requirement already satisfied: requests-oauthlib in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (2.0.0)

Requirement already satisfied: markdown in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (3.4.1)

Requirement already satisfied: colorama in c:\users\laksh\anaconda3\lib\site-packages (from click<8.2,>=8.0->prefect) (0.4.6)

Requirement already satisfied: cffi>=1.12 in c:\users\laksh\anaconda3\lib\site-packages (from cryptography>=36.0.1->prefect) (1.17.1)

Requirement already satisfied: regex!=2019.02.19,!2021.8.27,>=2015.06.24 in c:\users\laksh\anaconda3\lib\site-packages (from dateparser<2.0.0,>=1.1.1->prefect) (2023.10.3)

Requirement already satisfied: tzlocal>=0.2 in c:\users\laksh\anaconda3\lib\site-packages (from dateparser<2.0.0,>=1.1.1->prefect) (2.1)

Requirement already satisfied: pywin32>=304 in c:\users\laksh\anaconda3\lib\site-packages (from docker<8.0,>=4.0->prefect) (305.1)

Requirement already satisfied: urllib3>=1.26.0 in c:\users\laksh\anaconda3\lib\site-packages (from docker<8.0,>=4.0->prefect) (2.0.7)

Requirement already satisfied: starlette<0.47.0,>=0.40.0 in c:\users\laksh\anaconda3\lib\site-packages (from fastapi<1.0.0,>=0.111.0->prefect) (0.46.1)

Requirement already satisfied: h11<0.15,>=0.13 in c:\users\laksh\anaconda3\lib\site-packages (from httpcore<2.0.0,>=1.0.5->prefect) (0.14.0)

Requirement already satisfied: h2<5,>=3 in c:\users\laksh\anaconda3\lib\site-packages (from httpx[http2]!=0.23.2,>=0.23->prefect) (4.2.0)

Requirement already satisfied: MarkupSafe>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from jinja2<4.0.0,>=3.0.0->prefect) (2.1.3)

Requirement already satisfied: jsonpointer>=1.9 in c:\users\laksh\anaconda3\lib\site-packages (from jsonpatch<2.0,>=1.32->prefect) (2.1)

Requirement already satisfied: attrs>=22.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (23.1.0)

Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (2023.7.1)

Requirement already satisfied: referencing>=0.28.4 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (0.30.2)

Requirement already satisfied: rpds-py>=0.7.1 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (0.10.6)

Requirement already satisfied: deprecated>=1.2.6 in c:\users\laksh\anaconda3\lib\site-packages (from opentelemetry-api<2.0.0,>=1.27.0->prefect) (1.2.18)

Requirement already satisfied: importlib-metadata<8.7.0,>=6.0 in c:\users\laksh\anaconda3\lib\site-packages (from opentelemetry-api<2.0.0,>=1.27.0->prefect) (7.0.1)

Requirement already satisfied: tzdata>=2020.1 in c:\users\laksh\anaconda3\lib\site-packages (from pendulum<4,>=3.0.0->prefect) (2025.1)

Requirement already satisfied: time-machine>=2.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from pendulum<4,>=3.0.0->prefect) (2.16.0)

Requirement already satisfied: annotated-types>=0.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic!=2.10.0,<3.0.0,>=2.9->prefect) (0.7.0)

Requirement already satisfied: python-dotenv>=0.21.0 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic-settings>2.2.1->prefect) (0.21.0)

Requirement already satisfied: six>=1.5 in c:\users\laksh\anaconda3\lib\site-packages (from python-dateutil<3.0.0,>=2.8.2->prefect) (1.16.0)

Requirement already satisfied: text-unidecode>=1.3 in c:\users\laksh\anaconda3\lib\site-packages (from python-slugify<9.0,>=5.0->prefect) (1.3)

Requirement already satisfied: markdown-it-py<3.0.0,>=2.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from rich<14.0,>=11.

0->prefect) (2.2.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\users\laksh\anaconda3\lib\site-packages (from rich<14.0,>=11.0->prefect) (2.15.1)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\laksh\anaconda3\lib\site-packages (from sqlalchemy<3.0.0,>=2.0->sqlalchemy[asyncio]<3.0.0,>=2.0->prefect) (3.0.1)
Requirement already satisfied: shellingham>=1.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from typer!=0.12.2,<0.16.0,>=0.12.0->prefect) (1.5.4)
Requirement already satisfied: pycparser in c:\users\laksh\anaconda3\lib\site-packages (from cffi>=1.12->cryptography>=36.0.1->prefect) (2.21)
Requirement already satisfied: wrapt<2,>=1.10 in c:\users\laksh\anaconda3\lib\site-packages (from deprecated>=1.2.6->opentelemetry-api<2.0.0,>=1.27.0->prefect) (1.14.1)
Requirement already satisfied: hyperframe<7,>=6.1 in c:\users\laksh\anaconda3\lib\site-packages (from h2<5,>=3->httpx[http2]!=0.23.2,>=0.23->prefect) (6.1.0)
Requirement already satisfied: hpack<5,>=4.1 in c:\users\laksh\anaconda3\lib\site-packages (from h2<5,>=3->httpx[http2]!=0.23.2,>=0.23->prefect) (4.1.0)
Requirement already satisfied: zipp>=0.5 in c:\users\laksh\anaconda3\lib\site-packages (from importlib-metadata<8.7.0,>=6.0->opentelemetry-api<2.0.0,>=1.27.0->prefect) (3.17.0)
Requirement already satisfied: mdurl~=0.1 in c:\users\laksh\anaconda3\lib\site-packages (from markdown-it-py<3.0.0,>=2.2.0->rich<14.0,>=11.0->prefect) (0.1.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\laksh\anaconda3\lib\site-packages (from requests->apprise<2.0.0,>=1.1.0->prefect) (2.0.4)
Requirement already satisfied: oauthlib>=3.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from requests-oauthlib->apprise<2.0.0,>=1.1.0->prefect) (3.2.2)
Requirement already satisfied: pandas in c:\users\laksh\anaconda3\lib\site-packages (2.1.4)
Requirement already satisfied: scikit-learn in c:\users\laksh\anaconda3\lib\site-packages (1.2.2)
Requirement already satisfied: prefect in c:\users\laksh\anaconda3\lib\site-packages (3.2.12)
Requirement already satisfied: numpy<2,>=1.23.2 in c:\users\laksh\anaconda3\lib\site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\laksh\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\laksh\anaconda3\lib\site-packages (from pandas) (2025.1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\laksh\anaconda3\lib\site-packages (from pandas) (2025.1)
Requirement already satisfied: scipy>=1.3.2 in c:\users\laksh\anaconda3\lib\site-packages (from scikit-learn) (1.11.4)
Requirement already satisfied: joblib>=1.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: aiosqlite<1.0.0,>=0.17.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.18.0)
Requirement already satisfied: alembic<2.0.0,>=1.7.5 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.15.1)
Requirement already satisfied: anyio<5.0.0,>=4.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.8.0)
Requirement already satisfied: apprise<2.0.0,>=1.1.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.9.2)
Requirement already satisfied: asgi-lifespan<3.0,>=1.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.1.0)
Requirement already satisfied: asyncpg<1.0.0,>=0.23 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.30.0)
Requirement already satisfied: cachetools<6.0,>=5.3 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (5.5.2)
Requirement already satisfied: click<8.2,>=8.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (8.1.7)

Requirement already satisfied: cloudpickle<4.0,>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.1.1)

Requirement already satisfied: coolname<3.0.0,>=1.0.4 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.2.0)

Requirement already satisfied: cryptography>=36.0.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (42.0.2)

Requirement already satisfied: dateparser<2.0.0,>=1.1.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.2.1)

Requirement already satisfied: docker<8.0,>=4.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (7.1.0)

Requirement already satisfied: exceptiongroup>=1.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.2.2)

Requirement already satisfied: fastapi<1.0.0,>=0.111.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.115.11)

Requirement already satisfied: fsspec>=2022.5.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2025.3.0)

Requirement already satisfied: graphviz>=0.20.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.20.3)

Requirement already satisfied: griffe<2.0.0,>=0.49.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.6.0)

Requirement already satisfied: httpcore<2.0.0,>=1.0.5 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.0.7)

Requirement already satisfied: httpx!=0.23.2,>=0.23 in c:\users\laksh\anaconda3\lib\site-packages (from httpx[http2]!=0.23.2,>=0.23->prefect) (0.26.0)

Requirement already satisfied: humanize<5.0.0,>=4.9.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.12.1)

Requirement already satisfied: jinja2-humanize-extension>=0.4.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.4.0)

Requirement already satisfied: jinja2<4.0.0,>=3.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.1.3)

Requirement already satisfied: jsonpatch<2.0,>=1.32 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.32)

Requirement already satisfied: jsonschema<5.0.0,>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.19.2)

Requirement already satisfied: opentelemetry-api<2.0.0,>=1.27.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.31.0)

Requirement already satisfied: orjson<4.0,>=3.7 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.10.15)

Requirement already satisfied: packaging<24.3,>=21.3 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (23.1)

Requirement already satisfied: pathspec>=0.8.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.10.3)

Requirement already satisfied: pendulum<4,>=3.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (3.0.0)

Requirement already satisfied: prometheus-client>=0.20.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.21.1)

Requirement already satisfied: pydantic!=2.10.0,<3.0.0,>=2.9 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.10.6)

Requirement already satisfied: pydantic-core<3.0.0,>=2.12.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.27.2)

Requirement already satisfied: pydantic-extra-types<3.0.0,>=2.8.2 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.10.3)

Requirement already satisfied: pydantic-settings>2.2.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (2.8.1)

Requirement already satisfied: python-slugify<9.0,>=5.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (5.0.2)

Requirement already satisfied: python-socks<3.0,>=2.5.3 in c:\users\laksh\anaconda3\lib\site-packages (from python-socks[asyncio]<3.0,>=2.5.3->prefect) (2.7.1)

Requirement already satisfied: pyyaml<7.0.0,>=5.4.1 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (6.0.1)

Requirement already satisfied: readchar<5.0.0,>=4.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.2.1)

Requirement already satisfied: rfc3339-validator<0.2.0,>=0.1.4 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.1.4)

Requirement already satisfied: rich<14.0,>=11.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (13.3.5)

Requirement already satisfied: ruamel-yaml>=0.17.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.17.21)

Requirement already satisfied: sniffio<2.0.0,>=1.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (1.3.0)

Requirement already satisfied: sqlalchemy<3.0.0,>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from sqlalchemy[asyncio]<3.0.0,>=2.0->prefect) (2.0.25)

Requirement already satisfied: toml>=0.10.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.10.2)

Requirement already satisfied: typer!=0.12.2,<0.16.0,>=0.12.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.15.2)

Requirement already satisfied: typing-extensions<5.0.0,>=4.5.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (4.12.2)

Requirement already satisfied: ujson<6.0.0,>=5.8.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (5.10.0)

Requirement already satisfied: uv>=0.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.6.6)

Requirement already satisfied: uvicorn!=0.29.0,>=0.14.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (0.34.0)

Requirement already satisfied: websockets<16.0,>=13.0 in c:\users\laksh\anaconda3\lib\site-packages (from prefect) (15.0.1)

Requirement already satisfied: Mako in c:\users\laksh\anaconda3\lib\site-packages (from alembic<2.0.0,>=1.7.5->prefect) (1.3.9)

Requirement already satisfied: idna>=2.8 in c:\users\laksh\anaconda3\lib\site-packages (from anyio<5.0.0,>=4.4.0->prefect) (3.4)

Requirement already satisfied: certifi in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (2024.7.4)

Requirement already satisfied: requests in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (2.31.0)

Requirement already satisfied: requests-oauthlib in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (2.0.0)

Requirement already satisfied: markdown in c:\users\laksh\anaconda3\lib\site-packages (from apprise<2.0.0,>=1.1.0->prefect) (3.4.1)

Requirement already satisfied: colorama in c:\users\laksh\anaconda3\lib\site-packages (from click<8.2,>=8.0->prefect) (0.4.6)

Requirement already satisfied: cffi>=1.12 in c:\users\laksh\anaconda3\lib\site-packages (from cryptography>=36.0.1->prefect) (1.17.1)

Requirement already satisfied: regex!=2019.02.19,!2021.8.27,>=2015.06.24 in c:\users\laksh\anaconda3\lib\site-packages (from dateparser<2.0.0,>=1.1.1->prefect) (2023.10.3)

Requirement already satisfied: tzlocal>=0.2 in c:\users\laksh\anaconda3\lib\site-packages (from dateparser<2.0.0,>=1.1.1->prefect) (2.1)

Requirement already satisfied: pywin32>=304 in c:\users\laksh\anaconda3\lib\site-packages (from docker<8.0,>=4.0->prefect) (305.1)

Requirement already satisfied: urllib3>=1.26.0 in c:\users\laksh\anaconda3\lib\site-packages (from docker<8.0,>=4.0->prefect) (2.0.7)

Requirement already satisfied: starlette<0.47.0,>=0.40.0 in c:\users\laksh\anaconda3\lib\site-packages (from fastapi<1.0.0,>=0.111.0->prefect) (0.46.1)

Requirement already satisfied: h11<0.15,>=0.13 in c:\users\laksh\anaconda3\lib\site-packages (from httpcore<2.0.0,>=1.0.5->prefect) (0.14.0)

Requirement already satisfied: h2<5,>=3 in c:\users\laksh\anaconda3\lib\site-packages (from httpx[http2]!=0.23.2,>=0.23->prefect) (4.2.0)

Requirement already satisfied: MarkupSafe>=2.0 in c:\users\laksh\anaconda3\lib\site-packages (from jinja2<4.0.0,>=3.0.0->prefect) (2.1.3)

Requirement already satisfied: jsonpointer>=1.9 in c:\users\laksh\anaconda3\lib\site-packages (from jsonpatch<2.0,>=1.32->prefect) (2.1)

Requirement already satisfied: attrs>=22.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (23.1.0)

Requirement already satisfied: jsonschema-specifications>=2023.03.6 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (2023.7.1)

Requirement already satisfied: referencing>=0.28.4 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (0.30.2)

Requirement already satisfied: rpds-py>=0.7.1 in c:\users\laksh\anaconda3\lib\site-packages (from jsonschema<5.0.0,>=4.0.0->prefect) (0.10.6)

Requirement already satisfied: deprecated>=1.2.6 in c:\users\laksh\anaconda3\lib\site-packages (from opentelemetry-api<2.0.0,>=1.27.0->prefect) (1.2.18)

Requirement already satisfied: importlib-metadata<8.7.0,>=6.0 in c:\users\laksh\anaconda3\lib\site-packages (from opentelemetry-api<2.0.0,>=1.27.0->prefect) (7.0.1)

Requirement already satisfied: time-machine>=2.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from pendulum<4,>=3.0.0->prefect) (2.16.0)

Requirement already satisfied: annotated-types>=0.6.0 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic!=2.10.0,<3.0.0,>=2.9->prefect) (0.7.0)

Requirement already satisfied: python-dotenv>=0.21.0 in c:\users\laksh\anaconda3\lib\site-packages (from pydantic-settings>2.2.1->prefect) (0.21.0)

Requirement already satisfied: six>=1.5 in c:\users\laksh\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

Requirement already satisfied: text-unidecode>=1.3 in c:\users\laksh\anaconda3\lib\site-packages (from python-slugify<9.0,>=5.0->prefect) (1.3)

Requirement already satisfied: markdown-it-py<3.0.0,>=2.2.0 in c:\users\laksh\anaconda3\lib\site-packages (from rich<14.0,>=11.0->prefect) (2.2.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\users\laksh\anaconda3\lib\site-packages (from rich<14.0,>=11.0->prefect) (2.15.1)

Requirement already satisfied: greenlet!=0.4.17 in c:\users\laksh\anaconda3\lib\site-packages (from sqlalchemy<3.0.0,>=2.0->sqlalchemy[asyncio]<3.0.0,>=2.0->prefect) (3.0.1)

Requirement already satisfied: shellingham>=1.3.0 in c:\users\laksh\anaconda3\lib\site-packages (from typer!=0.12.2,<0.16.0,>=0.12.0->prefect) (1.5.4)

Requirement already satisfied: pycparser in c:\users\laksh\anaconda3\lib\site-packages (from cffi>=1.12->cryptography>=36.0.1->prefect) (2.21)

Requirement already satisfied: wrapt<2,>=1.10 in c:\users\laksh\anaconda3\lib\site-packages (from deprecated>=1.2.6->opentelemetry-api<2.0.0,>=1.27.0->prefect) (1.14.1)

Requirement already satisfied: hyperframe<7,>=6.1 in c:\users\laksh\anaconda3\lib\site-packages (from h2<5,>=3->httpx[http2]!=0.23.2,>=0.23->prefect) (6.1.0)

Requirement already satisfied: hpack<5,>=4.1 in c:\users\laksh\anaconda3\lib\site-packages (from h2<5,>=3->httpx[http2]!=0.23.

```
2,>=0.23->prefect) (4.1.0)
Requirement already satisfied: zipp>=0.5 in c:\users\laksh\anaconda3\lib\site-packages (from importlib-metadata<8.7.0,>=6.0->opentelemetry-api<2.0.0,>=1.27.0->prefect) (3.17.0)
Requirement already satisfied: mdurl~=0.1 in c:\users\laksh\anaconda3\lib\site-packages (from markdown-it-py<3.0.0,>=2.2.0->rich<14.0,>=11.0->prefect) (0.1.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\laksh\anaconda3\lib\site-packages (from requests->apprise<2.0.0,>=1.1.0->prefect) (2.0.4)
Requirement already satisfied: oauthlib>=3.0.0 in c:\users\laksh\anaconda3\lib\site-packages (from requests-oauthlib->apprise<2.0.0,>=1.1.0->prefect) (3.2.2)
```

Import Required Libraries

```
In [20]: from prefect import task, flow
import pandas as pd
import time
import random
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

Data Ingestion Task

```
In [21]: @task
def ingest_data():
    print("Ingesting data...")
    time.sleep(2) # Simulate data fetching
    df = pd.DataFrame({
        "customer_id": range(1, 101),
        "tenure": random.choices(range(1, 36), k=100),
        "monthly_charges": random.choices(range(20, 120), k=100),
        "churn": random.choices([0, 1], k=100),
    })
    return df
```

Data Validation Task

```
In [22]: @task
def validate_data(df):
    print("Validating data...")
    assert not df.isnull().values.any(), "Data contains missing values!"
    return df
```

Data Transformation Task

```
In [23]: @task
def transform_data(df):
    print("Transforming data...")
    df["tenure_squared"] = df["tenure"] ** 2
    return df
```

Feature Engineering Task

```
In [24]: @task
def feature_engineering(df):
    print("Performing feature engineering...")
    features = df[["tenure", "tenure_squared", "monthly_charges"]]
    labels = df["churn"]
    return features, labels
```

Model Training Task

```
In [25]: @task
def train_model(features, labels):
    print("Training model...")
    X_train, X_test, y_train, y_test = train_test_split(features, labels, test_size=0.2, random_state=42)
    model = RandomForestClassifier(n_estimators=50, random_state=42)
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f"Model Accuracy: {accuracy:.2f}")
    return model
```

Define Prefect Flow

```
In [26]: @flow(name="Customer Churn Prediction Pipeline")
def churn_pipeline():
    df = ingest_data()
    validated_df = validate_data(df)
    transformed_df = transform_data(validated_df)
    features, labels = feature_engineering(transformed_df)
    model = train_model(features, labels)
```

Execute the Pipeline

```
In [27]: if __name__ == "__main__":
    churn_pipeline()
```

```
21:50:50.823 | INFO      | prefect - Starting temporary server on http://127.0.0.1:8587
See https://docs.prefect.io/3.0/manage/self-host#self-host-a-prefect-server for more information on running a dedicated Prefect server.
21:50:57.318 | INFO      | Flow run 'granite-seahorse' - Beginning flow run 'granite-seahorse' for flow 'Customer Churn Prediction Pipeline'
Ingesting data...
21:50:59.645 | INFO      | Task run 'ingest_data-9b5' - Finished in state Completed()
Validating data...
21:50:59.911 | INFO      | Task run 'validate_data-09e' - Finished in state Completed()
Transforming data...
21:51:00.149 | INFO      | Task run 'transform_data-c2c' - Finished in state Completed()
Performing feature engineering...
21:51:00.381 | INFO      | Task run 'feature_engineering-639' - Finished in state Completed()
Training model...
Model Accuracy: 0.80
21:51:00.760 | INFO      | Task run 'train_model-cc7' - Finished in state Completed()
21:51:00.834 | INFO      | Flow run 'granite-seahorse' - Finished in state Completed()
```