

# Creating Customer Churn Prediction Pipeline on Azure Cloud

---

**Kheem Dharmani 22I-0081**  
**Ahmed Ali 22I-0076**



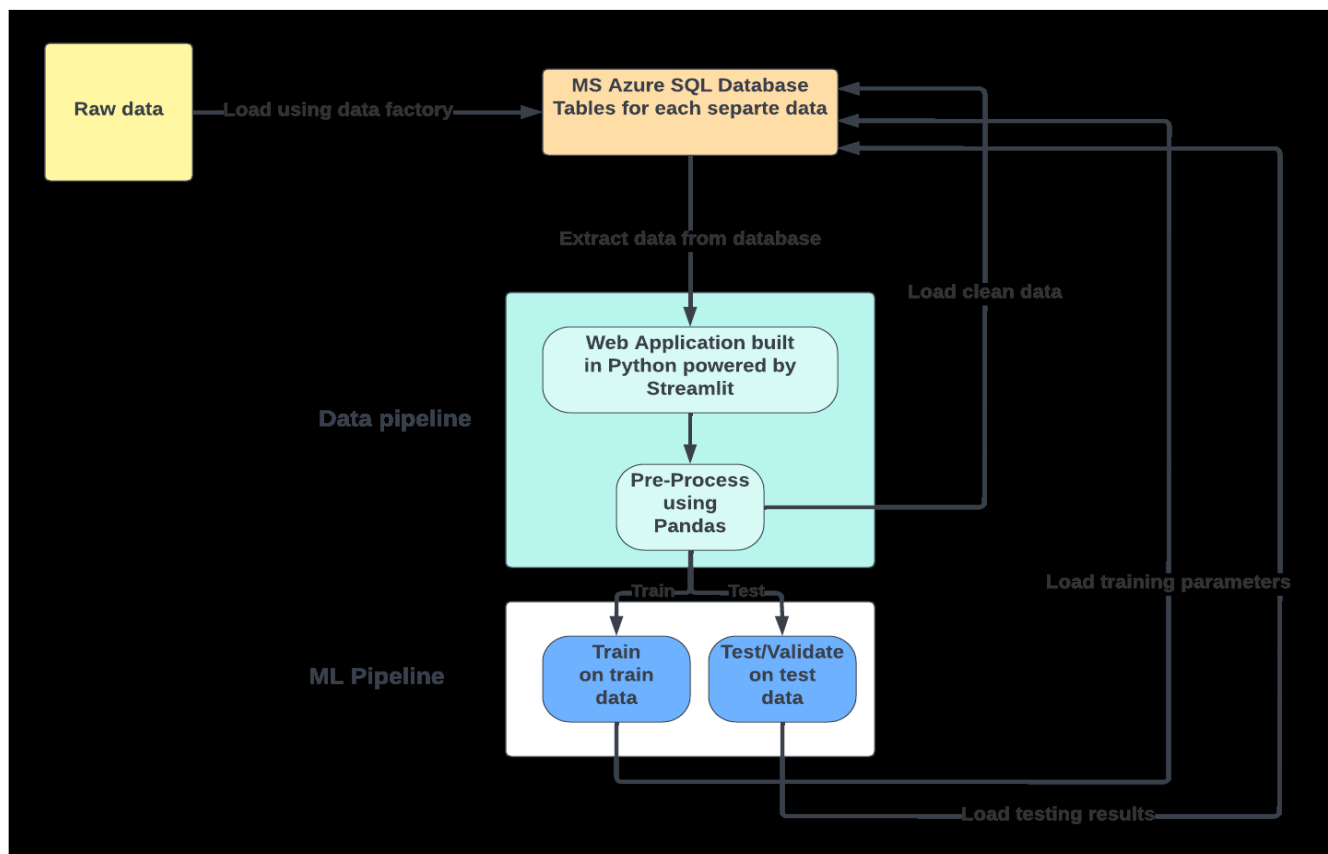
### Problem Statement:

As we stated in the project proposal, we aim to accomplish customer churn prediction by using a machine-learning algorithm. We used simple logistics regression for classifying customers to be churned or not churned. It's a binary classification problem.

Our project pipeline is:

Create database.

- Ingest data
- Run notebook.
- Connect data from SQL database.
- Store clean data in SQL database.
- Store results in SQL database.
- Move the local database to the cloud.
- change notebook connections to cloud SQL database.
- Read and validate the notebook using the data from the cloud SQL database.
- Create a Streamlit webpage.
- Show results on the webpage.



## STEP 1

### Microsoft Azure Account:

We created MS Azure free subscription account using SadaPay, but it expired and currently, the SadaPay debit card is not acceptable on the Azure portal. Therefore, we used an NU id for account creation, multiple times it shows that you are not eligible for a free account, after applying different NU emails it accepts one.

The screenshot shows the Microsoft Azure portal account summary page. At the top, there is a navigation bar with the text "resources, services, and docs (G+/)" and a user profile "khem.dharmani@hotmail...". Below the navigation bar, there is a "Summary" tab selected. The main content area displays two sections: "Amount due" and "Upcoming invoices".

**Amount due**

**\$0.00**

✓ No payment needed

[View invoices](#)

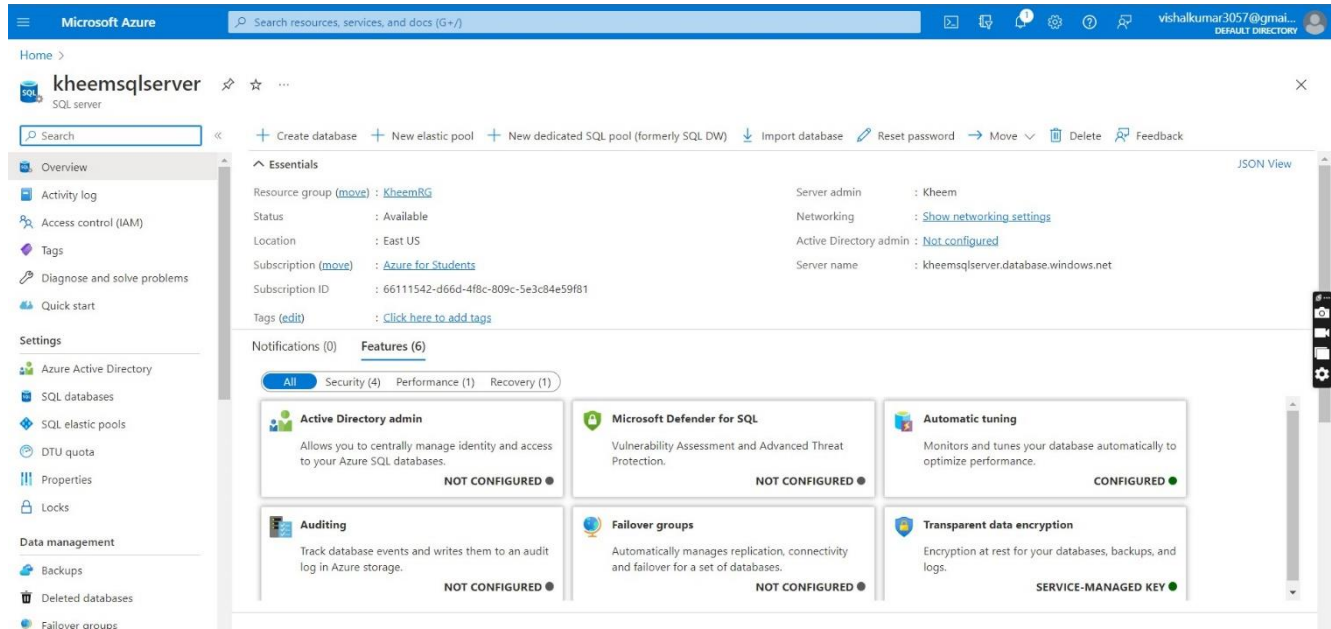
**Upcoming invoices**

Available on 12/9/2022	
Billing period 11/1/2022 - 11/30/2022	Pre-tax total so far ⓘ <b>\$0.00</b>
Available on 1/9/2023	
Billing period 12/1/2022 - 12/31/2022	Pre-tax total so far ⓘ <b>\$0.00</b>

## STEP 2

### Created MySQL server:

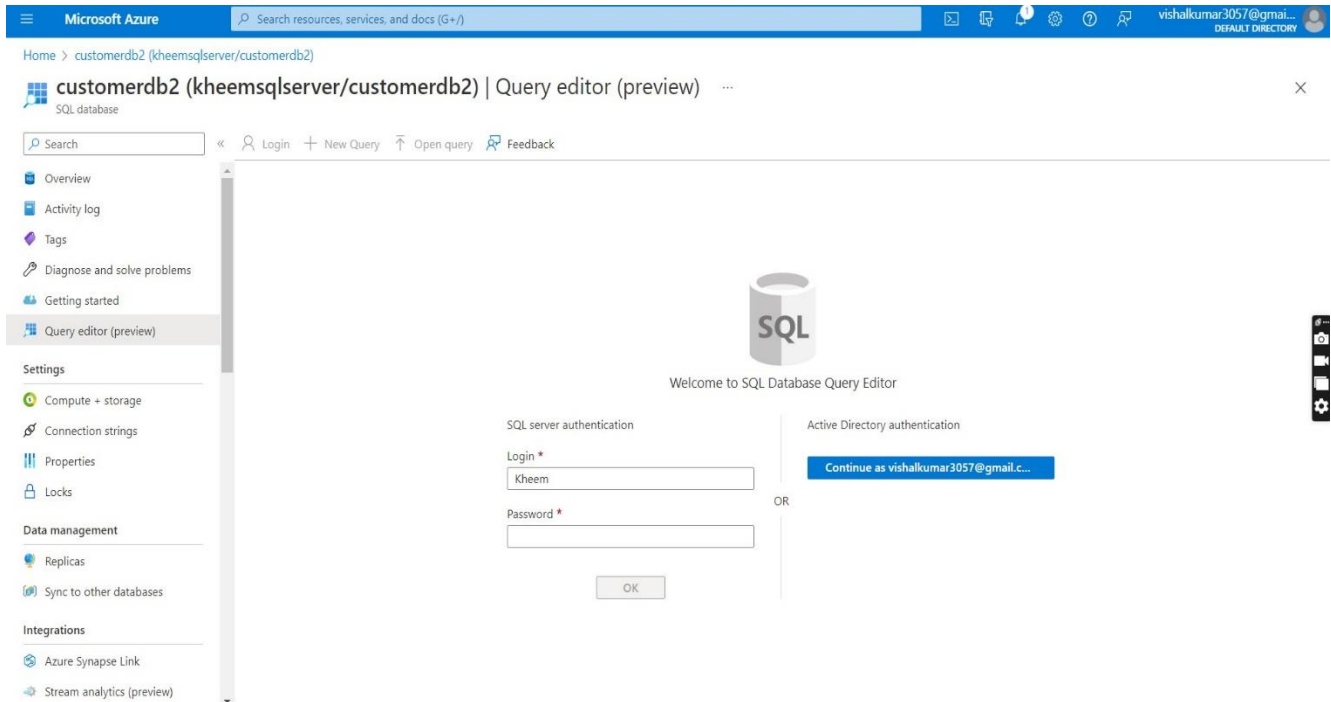
Select the SQL server (Logical) and named it, then created a new resource group for that, and choose the subscription. Rest settings are set as default.



## STEP 3

### Created MySQL Database:

Then we created SQL database in SQL server and ingest the data. We used bank customer data for churn prediction downloaded from Kaggle.




#### STEP 4

### Login To Database:

Here, we have four different tables:

1. Data (Original File)
2. Training Data (Split)
3. Test Data
4. Predication (The final results for Churned/Not Churned)

Home > customerdb2 (kheemsqlserver/customerdb2)

 **customerdb2 (kheemsqlserver/customerdb2)** | SQL database

Search

Overview  
Activity log  
Tags  
Diagnose and solve problems  
Getting started  
Query editor (preview)

Settings

Compute + storage  
Connection strings  
Properties  
Locks

Data management

Replicas  
Sync to other databases

Integrations

Azure Synapse Link

customerdb2 (Kheem)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

- > dbo.AzureBlobStorageFile
- > dbo.data
- > dbo.predictions
- > dbo.test\_data
- > dbo.training\_data

Views

Stored Procedures

## STEP 5

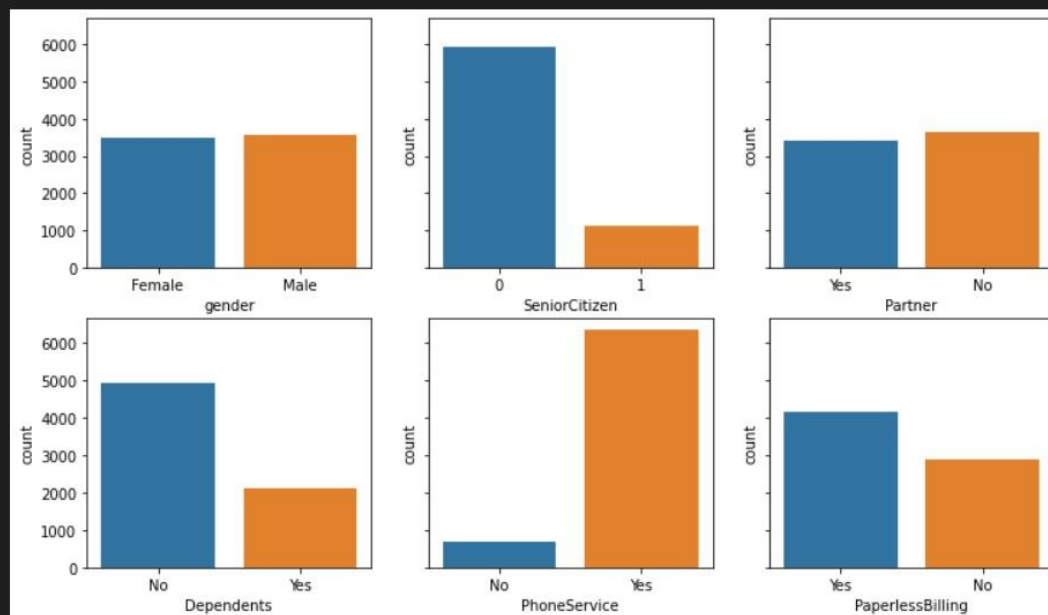
### Run Notebook:

Cleaned the data and did preprocessing. Also split the dataset in training and testing data frames and applied machine learning model Logistic Regression to predict the customer churn ratio.

Here the problem is, in the sandbox version the compute node is not supportable under the student subscriptions, therefore we train our classifier locally and then deployed it on the cloud.

```
fig, axes = plt.subplots(nrows =2 , ncols = 3, figsize =(12,7), sharey = True)
sns.countplot(data = df, x = "gender", ax = axes[0,0])
sns.countplot(data = df, x = 'SeniorCitizen', ax = axes[0,1])
sns.countplot(data = df, x = 'Partner', ax = axes[0,2])
sns.countplot(data = df, x = 'Dependents', ax = axes[1,0])
sns.countplot(data = df, x = 'PhoneService', ax = axes[1,1])
sns.countplot(data = df, x = 'PaperlessBilling', ax = axes[1,2])
plt.show()
```

[65]



```
> c = ['SeniorCitizen',
      'Partner',
      'Dependents',
      'PhoneService',
      'PaperlessBilling']

for i in c:
    print(df.groupby(i)["Churn"].mean())
```

[8]

SeniorCitizen  
0 0.236062  
1 0.416813  
Name: Churn, dtype: float64

Partner  
No 0.329580  
Yes 0.196649  
Name: Churn, dtype: float64

Dependents  
No 0.312791  
Yes 0.154502  
Name: Churn, dtype: float64

PhoneService  
No 0.249267  
Yes 0.267096  
Name: Churn, dtype: float64

PaperlessBilling  
No 0.163301  
Yes 0.335651  
Name: Churn, dtype: float64

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

[120]

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 42)
lg = LogisticRegression()
lg.fit(x_train, y_train)
```

[118]

...

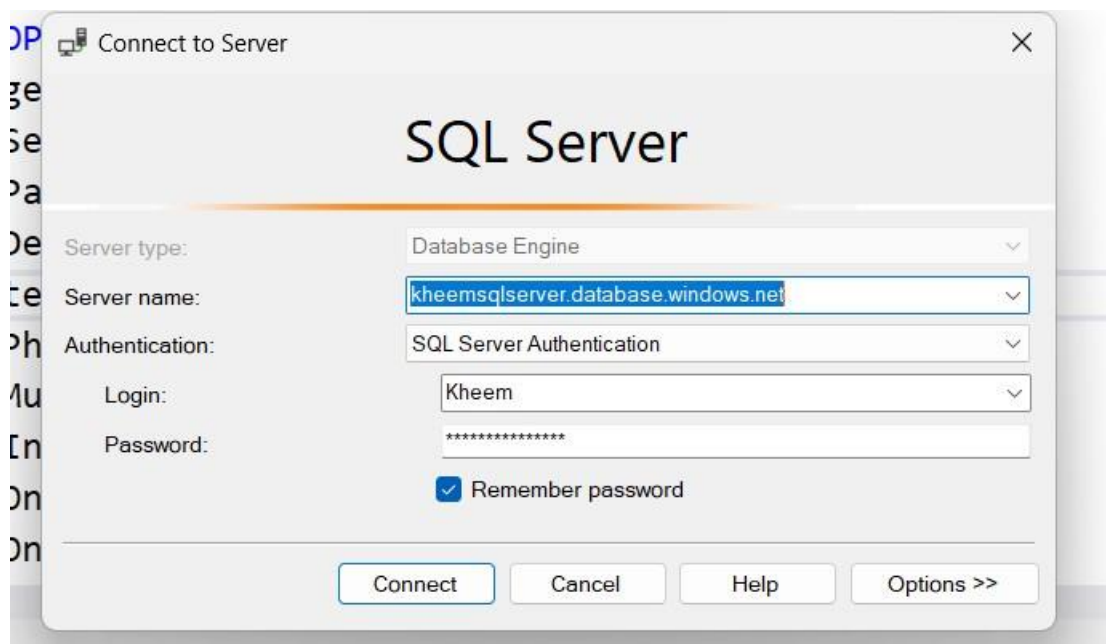
▼ LogisticRegression  
LogisticRegression()



## STEP 6

### Connect data from SQL database:

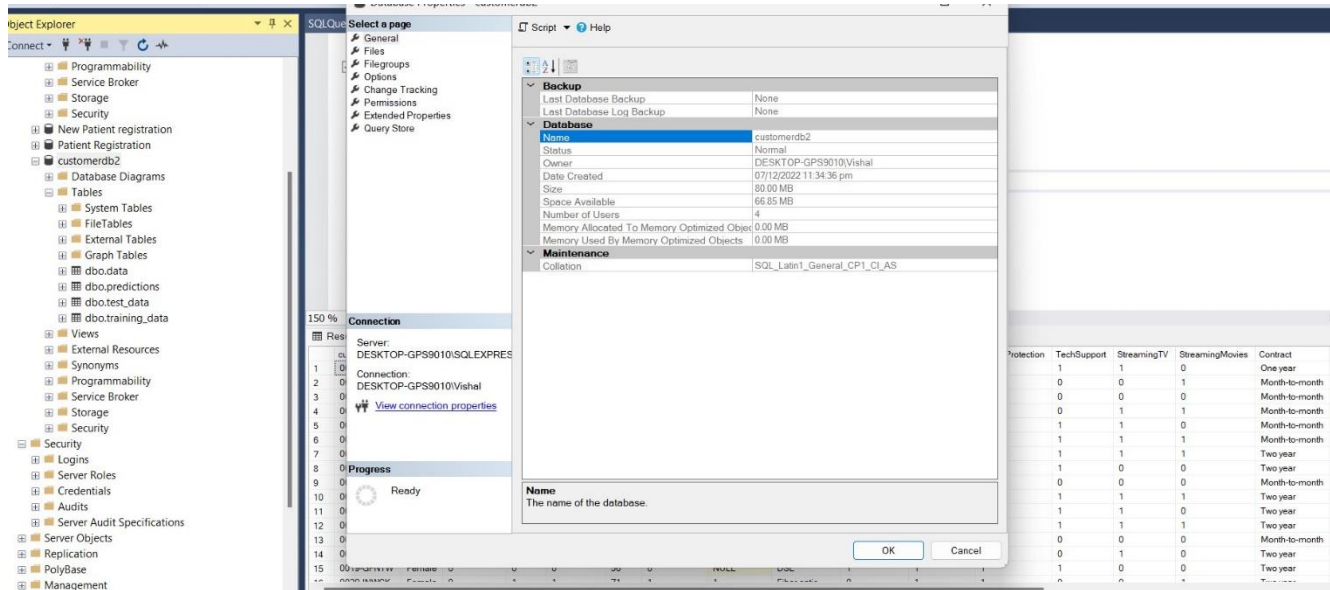
Used the SQL server authentication for connection.



## STEP 7

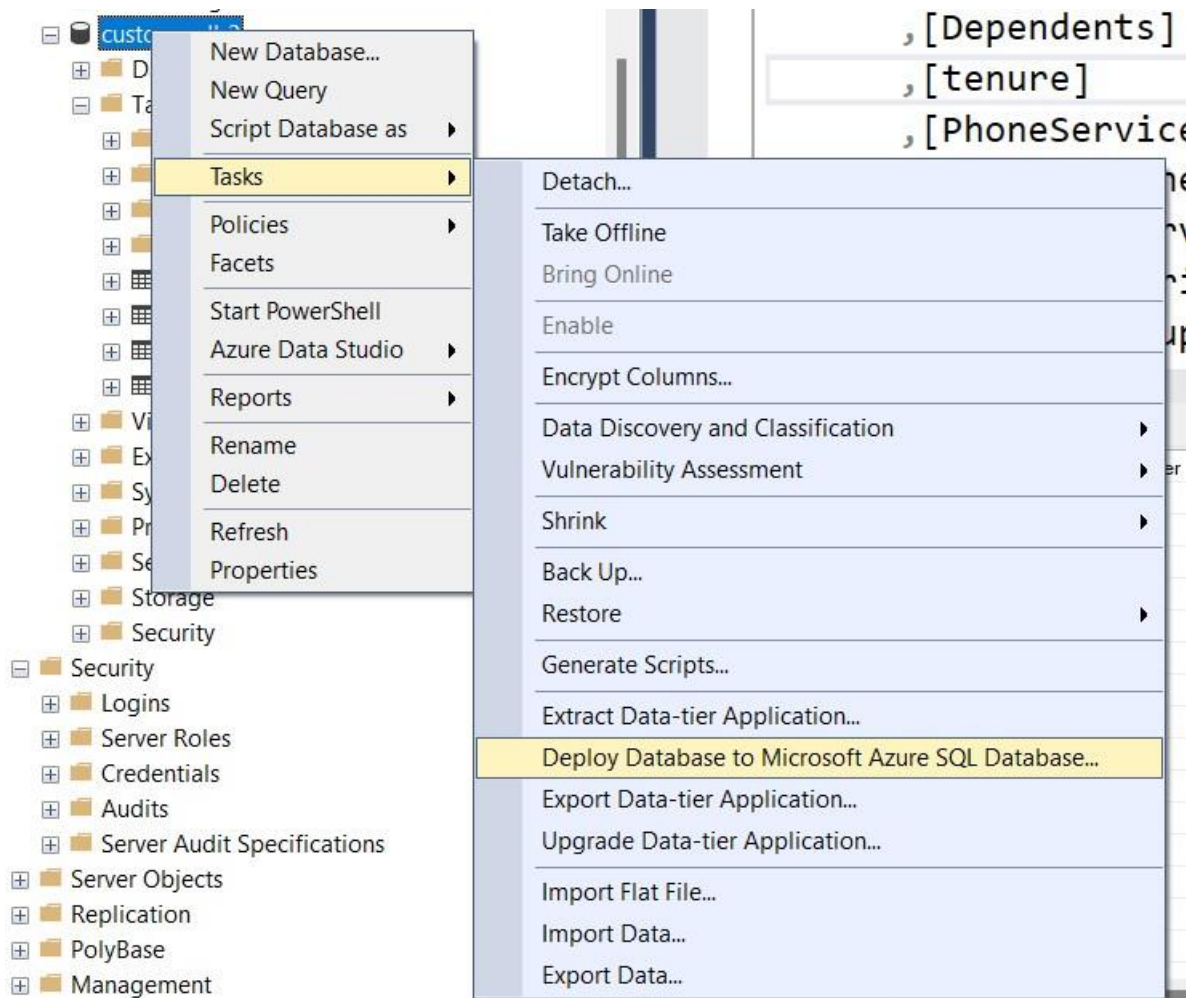
### Store clean data in SQL database:

Created the database and uploaded the tables (training, test, data, and prediction) in the database.




## STEP 8

### Move local database to cloud:



Deploy Database 'customerdb2'



Deployment Settings

Introduction

Deployment Settings

Summary

Results

Help

**Specify Target Connection**

Specify the name of the instance of SQL Server or the Microsoft Azure SQL Database server that will host the deployed database, name the new database, and then click Connect to login to the target server.

Server connection:

kheemsqlserver (Kheem)

Connect...

New database name:

customerdb2

Microsoft Azure SQL Database settings

Edition of Microsoft Azure SQL Database:

Basic

Maximum database size (GB):

2

Service Objective :

Basic

Other settings

Temporary file name:

C:\Users\dell\AppData\Local\Temp\customerdb2-20221208025929.bacpac

Browse...

< Previous

Next >

Cancel

## STEP 9

### Change notebook connections to cloud SQL database:

```
connection = pyodbc.connect('DRIVER='+driver+';SERVER='+server+';DATABASE='+database+';Trusted_Connection=yes')
cursor = connection.cursor()
return cursor

azure_server = 'kheemksqlserver.database.windows.net'
azure_database = 'bank'
azure_username = 'sqlserveradmin'
azure_password = '*****'
def create_azure_db_connection(azure_server, azure_database, azure_username, azure_password, driver)->pyodbc.Connection.cursor:
    connection = pyodbc.connect('DRIVER='+driver+';SERVER='+azure_server+';DATABASE='+azure_database+';UID='+azure_username+';PWD='+ azure_password)
    cursor = connection.cursor()
    return cursor
```

**Read and validate notebook using the data from cloud SQL database:**

Results Messages

🔍 Search to filter items...

Partner_No	Partner_Yes	Dependents_No	Dependents_Yes	MultipleLines_N
True	False	True	False	True
True	False	True	False	True
False	True	True	False	False
True	False	True	False	True

## Create Streamlit webpage:

```

60 def main():
61     server = 'localhost'
62     database = 'customerdb2'
63     driver = '{ODBC Driver 17 for SQL Server}'
64
65     azure_server = 'kheemsqlserver.database.windows.net'
66     azure_database = 'customerdb2'
67     azure_username = 'Kheem'
68     azure_password = 'MSazure@12'
69     azure_driver = '{ODBC Driver 17 for SQL Server}'
70
71     rows: int = 10
72     table_names: list() = ['dbo.data', 'dbo.training_data',
73                             'dbo.test_data', 'dbo.predictions']
74     query_main_data = query = f"SELECT * FROM {table_names[0]}"
75     query_training_data = query = f"SELECT * FROM {table_names[1]}"
76     query_test_data = query = f"SELECT * FROM {table_names[2]}"
77     query_predictions_data = query = f"SELECT * FROM {table_names[3]}"
78
79     df_main, df_train, df_test, df_predictions = (pd.DataFrame,)*4
80
81     # with return_sql_connection(server, database, driver) as connection:
82     #     df_main = pd.read_sql(query_main_data, connection)
83     #     df_train = pd.read_sql(query_training_data, connection)
84     #     df_test = pd.read_sql(query_test_data, connection)
85     #     df_predictions = pd.read_sql(query_predictions_data, connection)
86
87     with return_azure_connection(azure_server, azure_database, azure_driver, azure_username, azure_password) as connection:
88         df_main = pd.read_sql(query_main_data, connection)
89         df_train = pd.read_sql(query_training_data, connection)
90         df_test = pd.read_sql(query_test_data, connection)
91         df_predictions = pd.read_sql(query_predictions_data, connection)
92
93     data = (df_main, df_train, df_test, df_predictions)
94     set_page(data)
95

```

STEP 12

Show results in the webpage:

Customer Churn Prediction Web Application

Group Members

Kheem Dharmani 22I-0081

Ahmed Ali 22I-0076

Choose dataset

Predictions

Load data?

	.Yes	PaymentMethod_Bank_transfer_automatic	PaymentMethod_Credit_card_automatic	PaymentMethod_Electronic_check	PaymentMethod_Mailed_check	tenure	MonthlyCharges	SeniorCitizen	churn
0		<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.1528	0.5642	<input type="checkbox"/>	<input checked="" type="checkbox"/>
1		<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0139	0.5114	<input type="checkbox"/>	<input type="checkbox"/>
2		<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.8472	0.7547	<input type="checkbox"/>	<input type="checkbox"/>
3		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.3472	0.6129	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4		<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0278	0.5617	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.4583	0.6204	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
6		<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0694	0.6274	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.4306	0.6796	<input type="checkbox"/>	<input checked="" type="checkbox"/>

_No	PaperlessBilling_Yes	PaymentMethod_Bank_transfer_automatic	PaymentMethod_Credit_card_automatic	PaymentMethod_Electronic_check	PaymentMethod_Mailed_check	tenure	MonthlyCharges	SeniorCitizen	churn
67	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.1806	0.6463	<input type="checkbox"/>	<input type="checkbox"/>
68	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0.1667	0.7687	<input type="checkbox"/>	<input type="checkbox"/>
69	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0417	0.0647	<input type="checkbox"/>	<input checked="" type="checkbox"/>
70	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0556	0.3249	<input type="checkbox"/>	<input type="checkbox"/>
71	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.1944	0.5642	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
72	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0.0278	0.3542	<input type="checkbox"/>	<input checked="" type="checkbox"/>
73	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.1111	0.5721	<input checked="" type="checkbox"/>	<input type="checkbox"/>
74	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.9722	0.9682	<input type="checkbox"/>	<input checked="" type="checkbox"/>
75	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	0.0139	0.0164	<input type="checkbox"/>	<input checked="" type="checkbox"/>
76	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0139	0.7224	<input type="checkbox"/>	<input type="checkbox"/>
77	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.9167	0.9219	<input checked="" type="checkbox"/>	<input type="checkbox"/>
78	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0694	0.0667	<input type="checkbox"/>	<input checked="" type="checkbox"/>
79	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0833	0.7149	<input type="checkbox"/>	<input checked="" type="checkbox"/>
80	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.3056	0.5100	<input type="checkbox"/>	<input checked="" type="checkbox"/>
81	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.5139	0.8806	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
82	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0278	0.5786	<input type="checkbox"/>	<input checked="" type="checkbox"/>
83	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0278	0.6209	<input type="checkbox"/>	<input type="checkbox"/>
84	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	0.0139	0.6313	<input type="checkbox"/>	<input checked="" type="checkbox"/>