رواد مصر الرقمية

Customer Feedback Analysis and Improvement

Project Overview

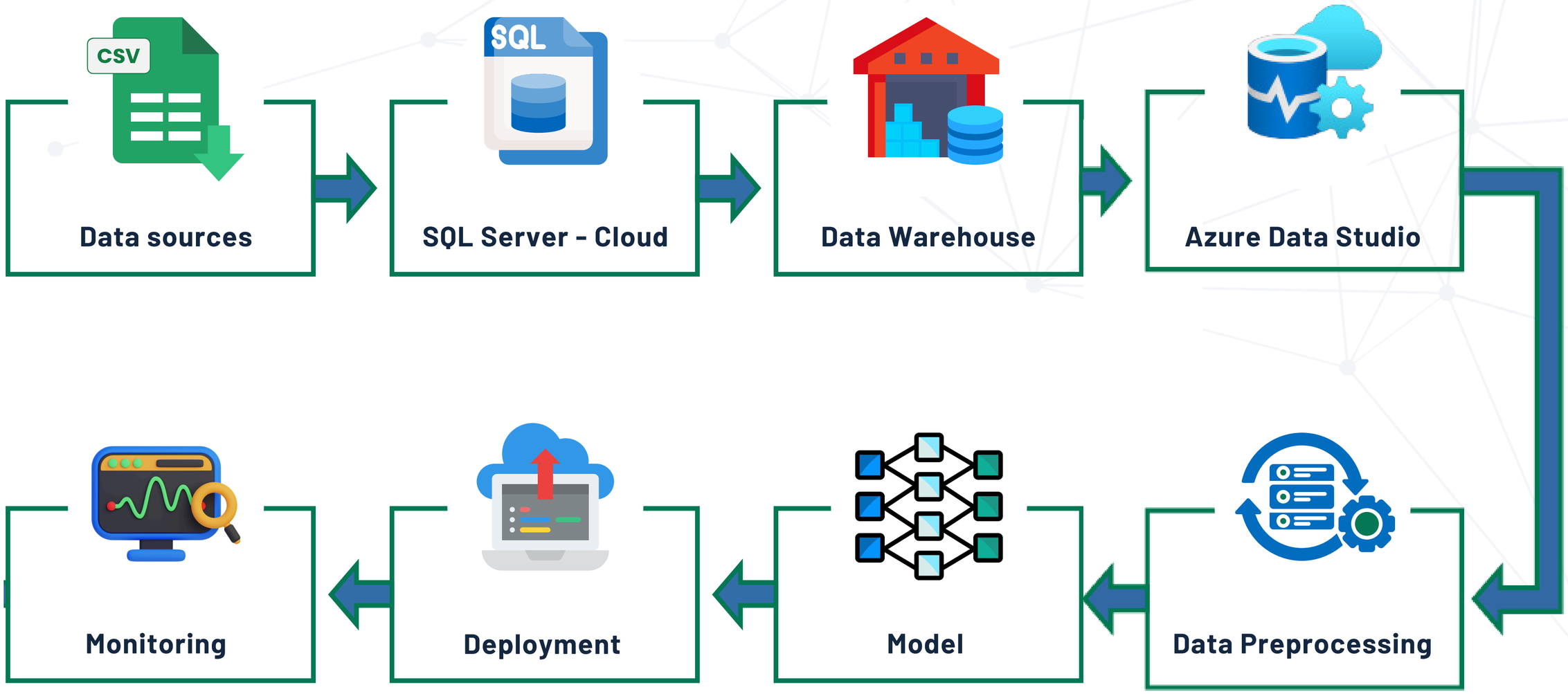
Objective

- The primary goal of this project was to analyze customer feedback to identify trends, classify sentiment, and provide actionable insights to improve customer experience.

Approach

- Data Collection: We gathered and organized historical customer feedback in a SQL database.
- Data Warehousing: A structured data warehouse was designed to store and manage this feedback for efficient querying and analysis.
- Sentiment Analysis: We developed a Deep learning model to classify customer feedback as positive or negative.
- Deployment: The final model was deployed as a web application for real-time feedback insights.

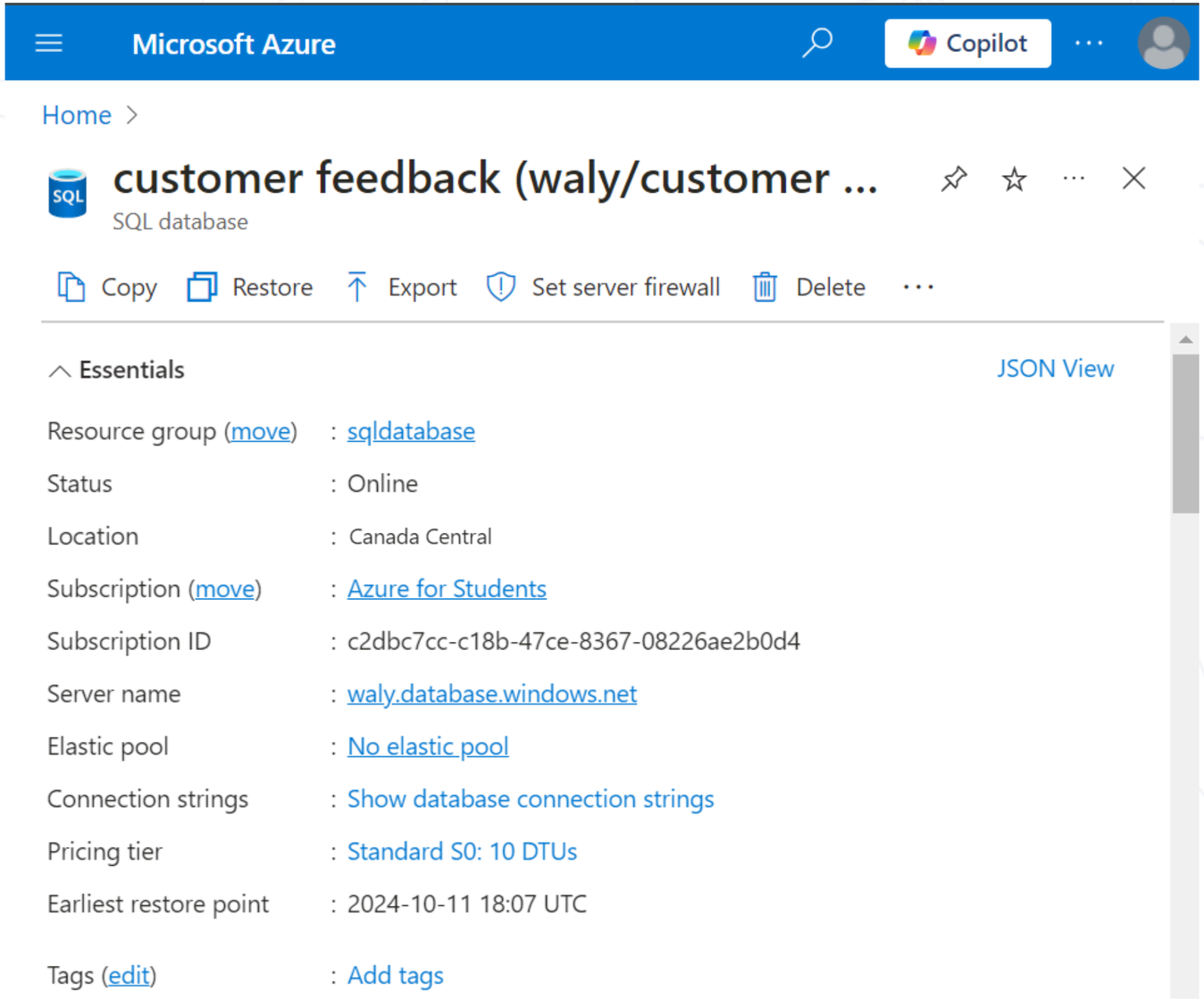
Comprehensive Data Pipeline



This pipeline automates feedback collection, storage, cleaning, and sentiment analysis.
Data is processed in SQL Server, analyzed with Python, and deployed for insights.

SQL Server on Azure

We have uploaded the data to SQL Server on Azure and connected it with Microsoft SQL Server. This setup ensures that our team stays updated and allows us to collaborate more efficiently, making it easier to access and manage the data.

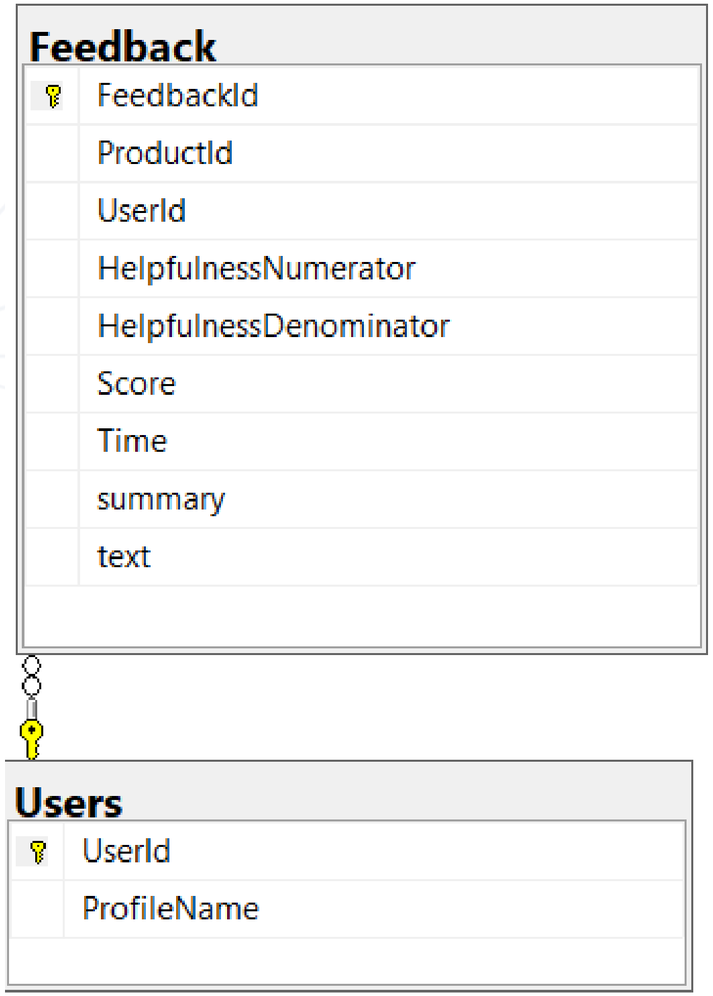


Customer Feedback Database Schema

This database schema efficiently stores and organizes customer feedback data, allowing for structured and meaningful analysis of feedback related to products and users. It includes two main tables: Users and Feedback.

Users Table The Users table holds user-specific information. Each row in this table represents a unique user who provides feedback on products.

Feedback Table
The Feedback table stores individual feedback entries provided by users about products. This table captures both numerical ratings and text-based feedback, as well as helpfulness ratings given by other users.



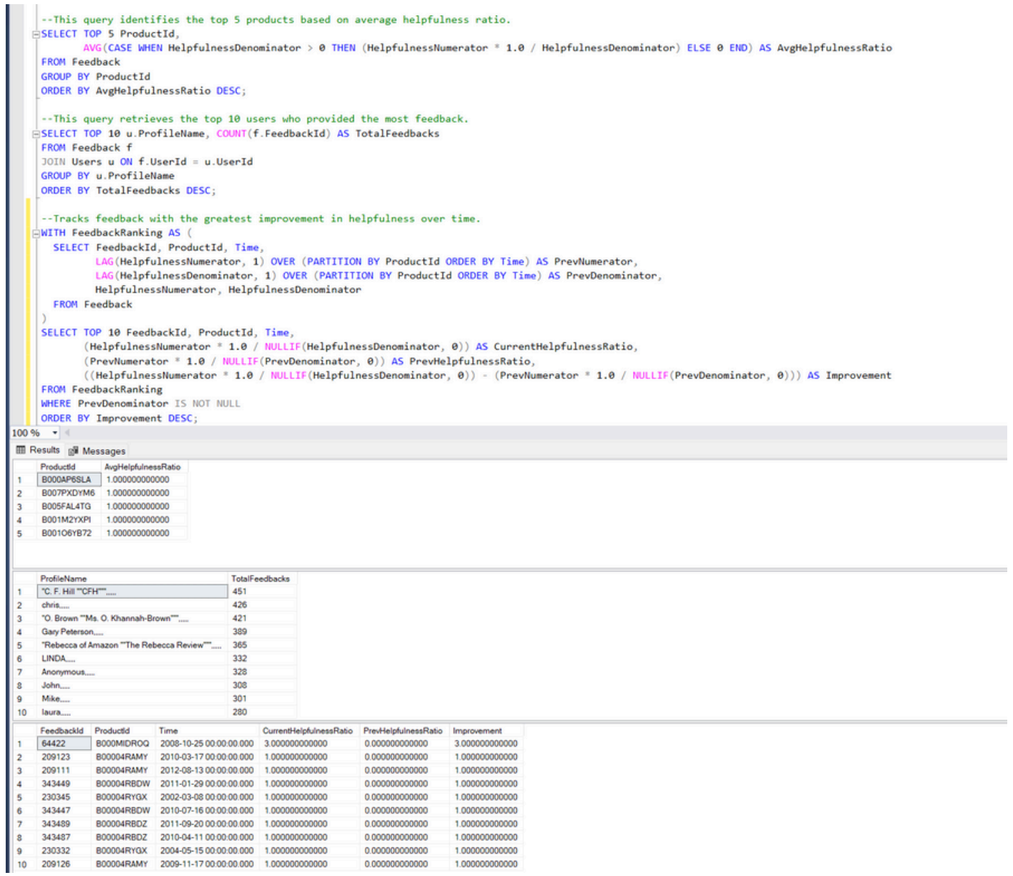
SQL Queries for Feedback Insights

These SQL queries are designed to analyze customer feedback data, efficiently extracting insights on user contributions, product performance, and feedback helpfulness. By leveraging this structured approach, we can make data-driven decisions that enhance product offerings and improve customer satisfaction. Below are three sample queries that illustrate this analysis

- **Top Products by Helpfulness Ratio:** his query identifies the top 5 products based on their average helpfulness ratio, providing insights into which products are perceived as most helpful by users.

- **Top Users by Feedback Contribution:** This query retrieves the top 10 users who have provided the most feedback, allowing us to recognize our most engaged customers.

- **Improvement in Helpfulness Over Time:** This query tracks feedback entries with the greatest improvement in helpfulness over time, helping us identify areas of progress and success.



```sql
--This query identifies the top 5 products based on average helpfulness ratio.
SELECT TOP 5 ProductId,
       AVG(CASE WHEN HelpfulnessDenominator > 0 THEN (HelpfulnessNumerator * 1.0 / HelpfulnessDenominator) ELSE 0 END) AS AvgHelpfulnessRatio
FROM Feedback
GROUP BY ProductId
ORDER BY AvgHelpfulnessRatio DESC;

--This query retrieves the top 10 users who provided the most feedback.
SELECT TOP 10 u.ProfileName, COUNT(f.FeedbackId) AS TotalFeedbacks
FROM Feedback f
JOIN Users u ON f.UserId = u.UserId
GROUP BY u.ProfileName
ORDER BY TotalFeedbacks DESC;

--Tracks feedback with the greatest improvement in helpfulness over time.
WITH FeedbackRanking AS (
  SELECT FeedbackId, ProductId, Time,
         LAG(HelpfulnessNumerator, 1) OVER (PARTITION BY ProductId ORDER BY Time) AS PrevNumerator,
         LAG(HelpfulnessDenominator, 1) OVER (PARTITION BY ProductId ORDER BY Time) AS PrevDenominator,
         HelpfulnessNumerator, HelpfulnessDenominator
  FROM Feedback
)
SELECT TOP 10 FeedbackId, ProductId, Time,
       (HelpfulnessNumerator * 1.0 / NULLIF(HelpfulnessDenominator, 0)) AS CurrentHelpfulnessRatio,
       (PrevNumerator * 1.0 / NULLIF(PrevDenominator, 0)) AS PrevHelpfulnessRatio,
       ((HelpfulnessNumerator * 1.0 / NULLIF(HelpfulnessDenominator, 0)) - (PrevNumerator * 1.0 / NULLIF(PrevDenominator, 0))) AS Improvement
FROM FeedbackRanking
WHERE PrevDenominator IS NOT NULL
ORDER BY Improvement DESC;
```

100 %

Results | Messages

	ProductId	AvgHelpfulnessRatio
1	B000AP6SLA	1.000000000000
2	B007PXDYM6	1.000000000000
3	B005FAL4TG	1.000000000000
4	B001M2YXPI	1.000000000000
5	B001O6YB72	1.000000000000

	ProfileName	TotalFeedbacks
1	"C. F. Hill ""CFH"".....	451
2	chris,....	426
3	"O. Brown ""Ms. O. Khannah-Brown"".....	421
4	Gary Peterson,....	389
5	"Rebecca of Amazon ""The Rebecca Review"".....	365
6	LINDA,....	332
7	Anonymous,....	328
8	John,....	308
9	Mike,....	301
10	laura,....	280

	FeedbackId	ProductId	Time	CurrentHelpfulnessRatio	PrevHelpfulnessRatio	Improvement
1	64422	B000MIDROQ	2008-10-25 00:00:00.000	3.000000000000	0.000000000000	3.000000000000
2	209123	B00004RAMY	2010-03-17 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
3	209111	B00004RAMY	2012-08-13 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
4	343449	B00004RBDW	2011-01-29 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
5	230345	B00004RYGX	2002-03-08 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
6	343447	B00004RBDW	2010-07-16 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
7	343489	B00004RBDZ	2011-09-20 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
8	343487	B00004RBDZ	2010-04-11 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
9	230332	B00004RYGX	2004-05-15 00:00:00.000	1.000000000000	0.000000000000	1.000000000000
10	209126	B00004RAMY	2009-11-17 00:00:00.000	1.000000000000	0.000000000000	1.000000000000

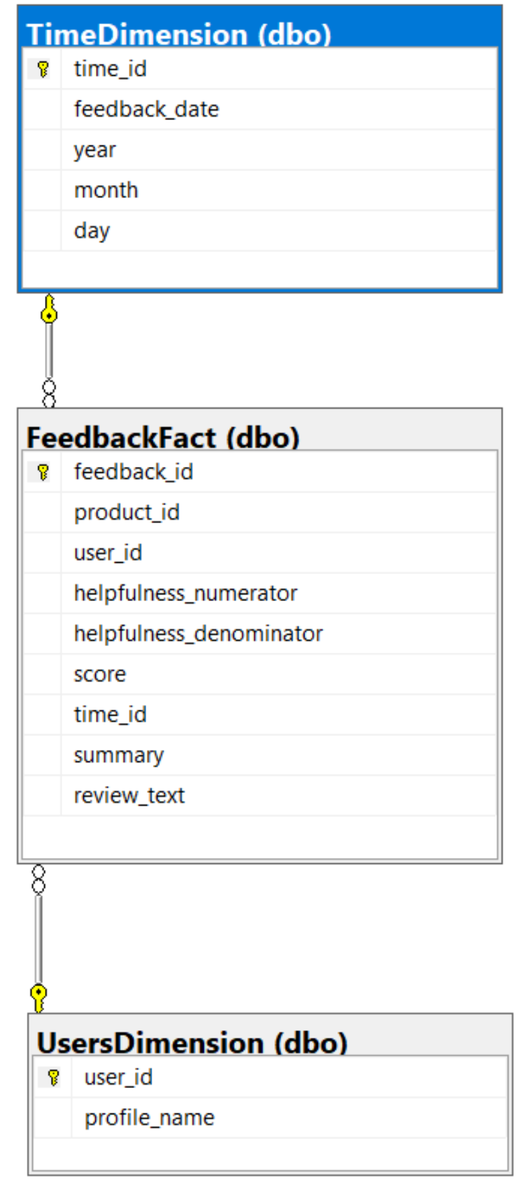
Customer Feedback Data Warehouse Schema

This is a data warehouse to efficiently store and analyze customer feedback. The warehouse uses a star schema to enable quick querying and reporting, optimized for analyzing large amounts of customer review data. This schema includes one fact table: FeedbackFact and two dimension tables : UsersDimension and TimeDimension.

The TimeDimension table stores structured time information. This helps us break down and analyze feedback by year, month, and day, which is essential for time-series analysis, trends, and reporting.

The UsersDimension table stores user-related information, which allows to analyze feedback at the user level to look at user behavior, track which users are providing feedback, and segment by profile name.

Finally, FeedbackFact table is the central fact table that stores the actual feedback data, which includes product reviews, user IDs, helpfulness scores, and ratings.This table allows to perform analysis on feedback across different dimensions (time and user).



Data Processing with Python

Column Dropping: Remove columns which aren't relevant to the next steps in the analysis.

Null Check: Identify columns with missing values for potential handling.

Text Cleaning: Remove unwanted characters such as extra quotes that can interfere with analysis or model training. This ensures the text is clean and ready for processing.

Text Normalization: Converts text to lowercase for consistency in future text processing tasks.

BEFORE

```
[7]  1  df.head(15)
```

Python

	FeedbackId	ProductId	UserId	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	summary	text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	1	1	5	2011-04-27	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	5	B006K2ZZ7K	A1UQRSCLF8GW1T	0	0	5	2012-10-21	Great taffy	"Great taffy at a great price. There was a wi...
2	7	B006K2ZZ7K	A1SP2KVKFXXRU1	0	0	5	2012-06-20	Great! Just as good as the expensive brands!	"This saltwater taffy had great flavors and wa...
3	8	B006K2ZZ7K	A3JRGQVEQN31IQ	0	0	5	2012-05-03	"Wonderful, tasty taffy"	This taffy is so good. It is very soft and ch...
4	9	B000E7L2R4	A1MZYO9TZK0BBI	1	1	5	2011-11-23	Yay Barley	Right now I'm mostly just sprouting this so my...
5	10	B00171APVA	A21BT40VZCCYT4	0	0	5	2012-10-26	Healthy Dog Food	This is a very healthy dog food. Good for thei...
6	11	B0001PB9FE	A3HDKO7OW0QNK4	1	1	5	2005-02-08	The Best Hot Sauce in the World	"I don't know if it's the cactus or the tequil...
7	12	B0009XLVG0	A2725IB4YY9JEB	4	4	5	2010-08-27	"My cats LOVE this ""diet"" food better than t...	"One of my boys needed to lose some weight and...

AFTER

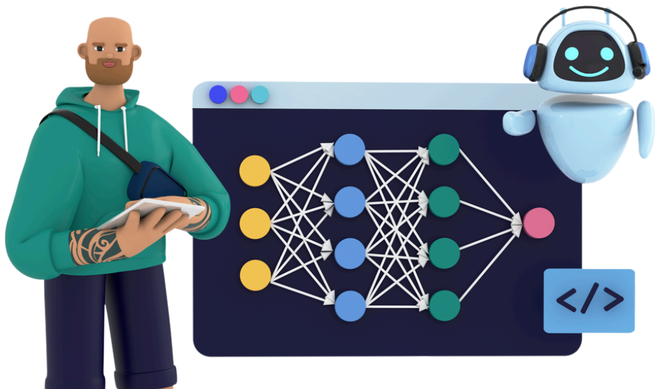
```
[13]  1  df
```

Python

	HelpfulnessNumerator	Score	summary	text
0	1	5	good quality dog food	i have bought several of the vitality canned d...
1	0	5	great taffy	great taffy at a great price. there was a wid...
2	0	5	great! just as good as the expensive brands!	this saltwater taffy had great flavors and was...
3	0	5	wonderful, tasty taffy	this taffy is so good. it is very soft and ch...
4	1	5	yay barley	right now i'm mostly just sprouting this so my...
...	...	...	...	...
162032	0	1	tastes horrible!	i just bought this soup today at my local groc...
162033	0	2	not so good	this soup is mostly broth. although it has a k...
162034	0	2	where's the tortellini?	it is mostly broth, with the advertised 3/4 cu...
162035	0	2	mixed wrong	i had ordered some of these a few months back ...

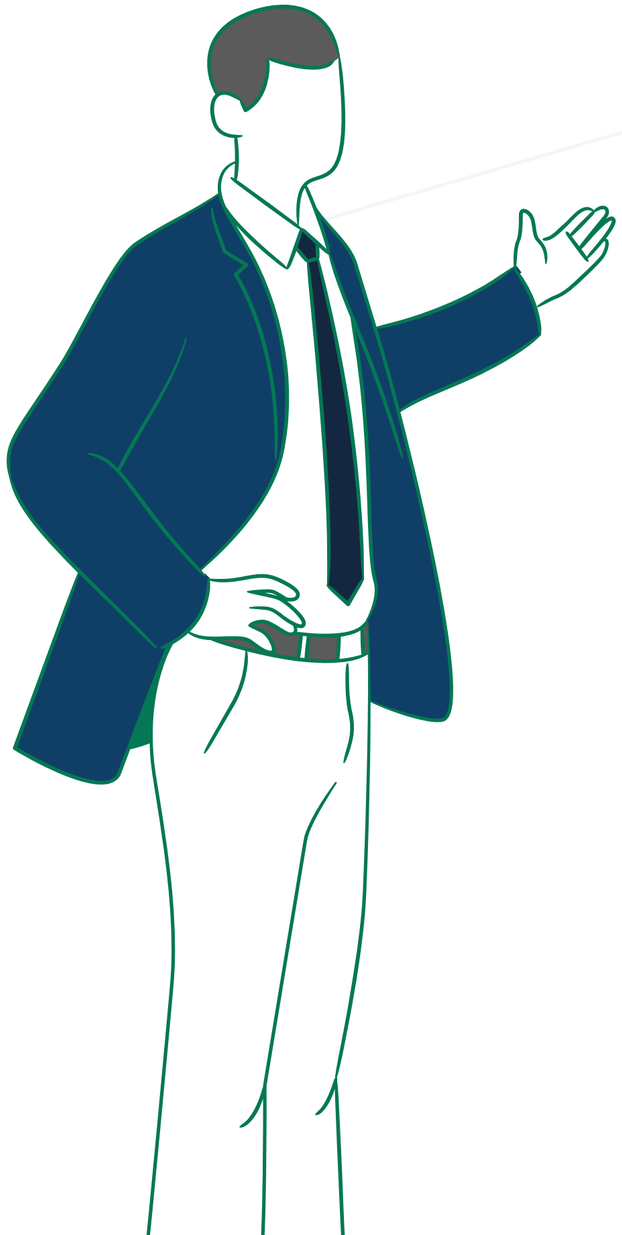
Sentiment Analysis Model (LSTM)

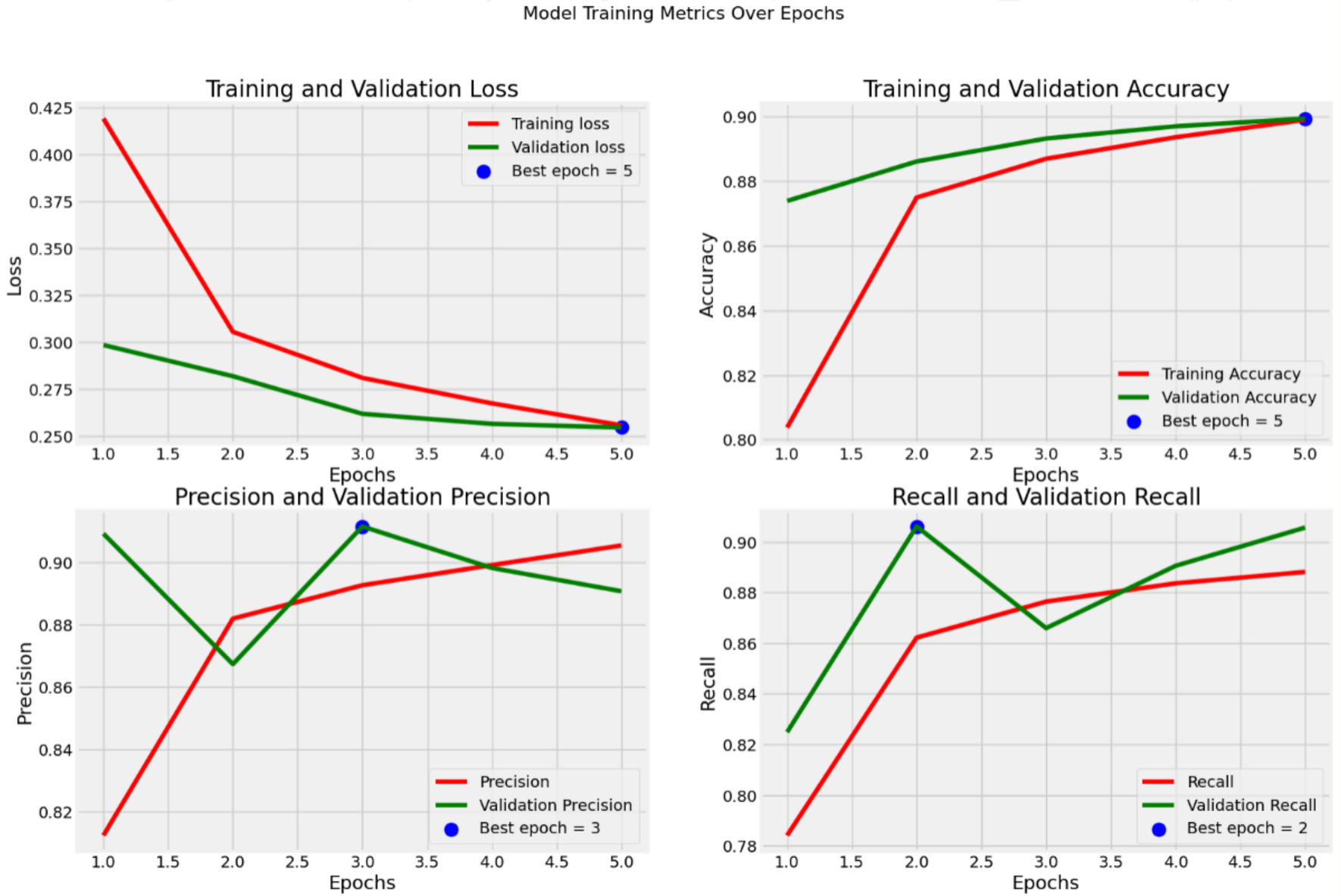
The LSTM model was trained on 110,000 samples from a dataset of 162,000. Over 5 epochs, the model achieved the following performance metrics: Training Loss: 0.2556, Validation Loss: 0.2544, with the Best Epoch at 5. The Training Accuracy reached 89.89%, while Validation Accuracy was 89.93%. Precision scores were 90.55% (training) and 89.08% (validation). Recall scores were 88.81% (training) and 90.56% (validation). These results show consistent improvement across epochs, indicating a well-performing model.




```
Shape of y: (162037,)
After training split:
X_rem shape: (52037, 100)
y_rem shape: (52037,)
Shapes after validation and test
X_train shape: (110000, 100)
y_train shape: (110000,)
X_valid shape: (27037, 100)
y_valid shape: (27037,)
X_test shape: (25000, 100)
y_test shape: (25000,)
```

LSTM Model Performance Metrics Overview

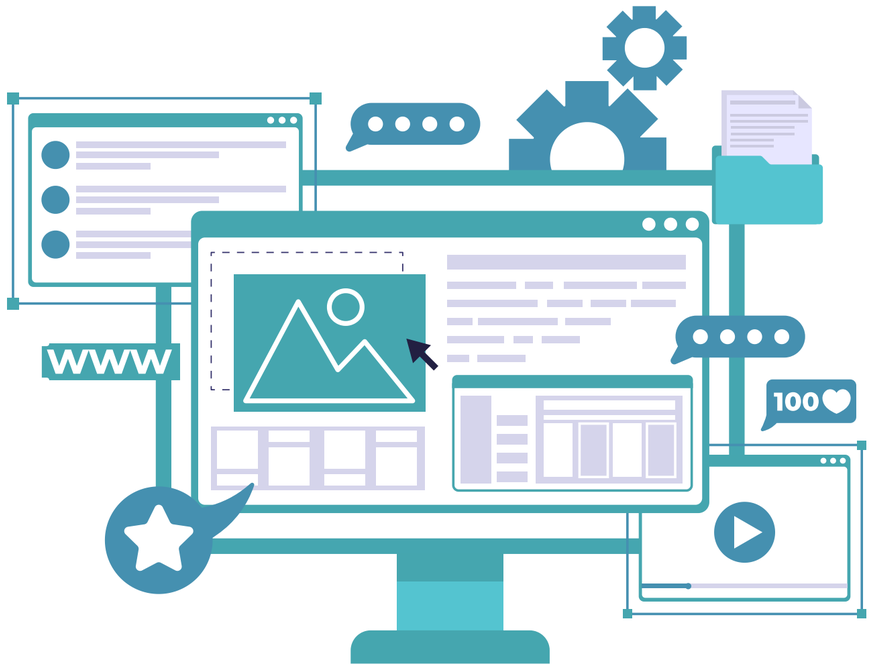


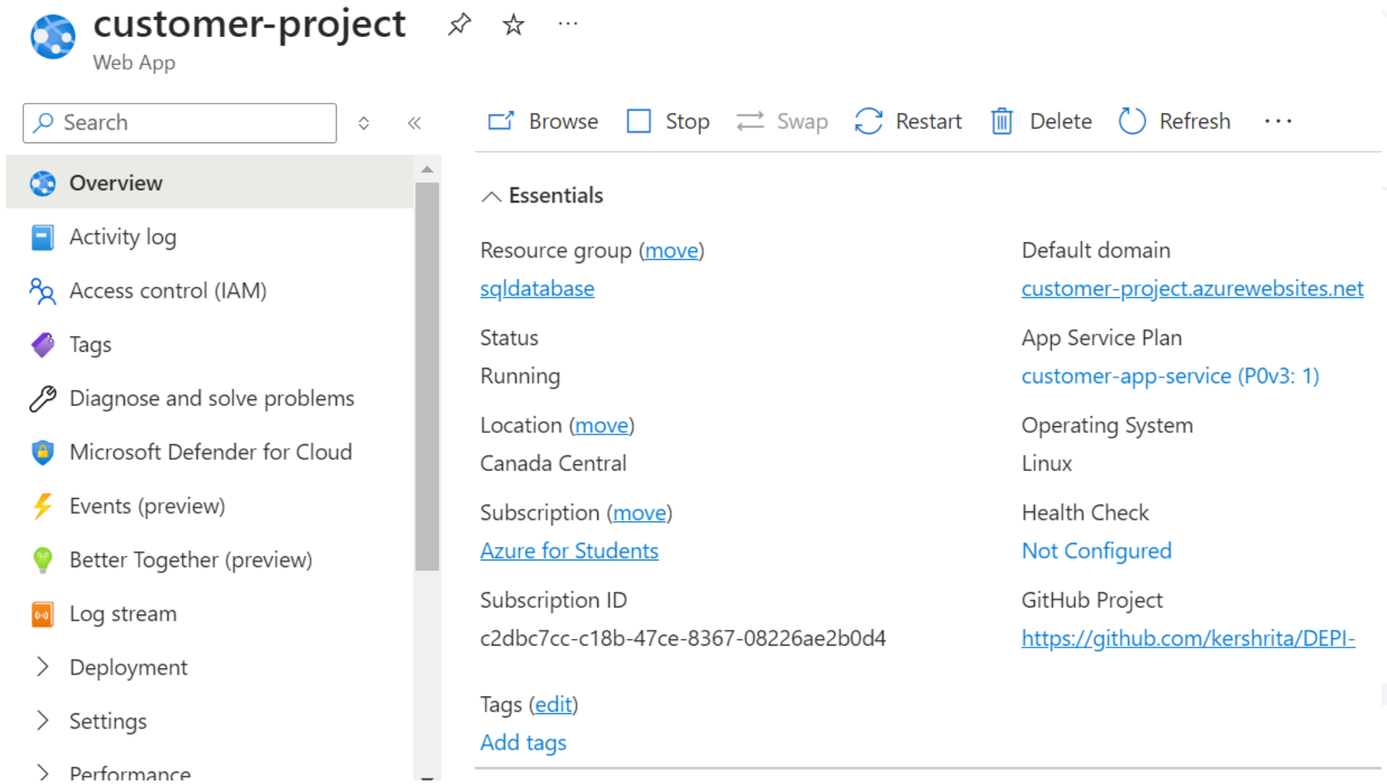


Model Deployment

The project involved deploying a sentiment analysis model using Streamlit, with the model and tokenizer hosted on GitHub. Azure was integrated with the GitHub repository for seamless deployment.The live application allows users to input text and receive a sentiment prediction (positive or negative) along with the confidence percentage.

Access the web page here:
customer-project.azurewebsites.net





Sentiment Analysis Web Page

Start analyzing sentiments

Sentiment Analysis

Enter text for sentiment analysis:

This thing is a piece of crap and a waste of money. DO NOT BUY. It's not useful at all and has more issues

The sentiment of the text is: **Negative** 😠

Confidence percentage is: **0.82%**

Sentiment Analysis

Enter text for sentiment analysis:

better and give me more data. And the green color is a nice change for the always boring smart watches.

The sentiment of the text is: **Positive** 😀

Confidence percentage is: **97.94%**

Shortcomings of the Project

Fixed Response Categories

The analysis only offers binary sentiment results (e.g., positive or negative), neglecting a neutral or more nuanced classification.

Single Input Analysis

The page only allows for one text input at a time, limiting bulk analysis.

Limited Explanation

The confidence percentage alone does not explain why a particular sentiment was assigned, reducing user understanding.

Limited Language Support

The current implementation does not support multiple languages, limiting its usability for a broader audience.

Marketing Strategy

Overview: Provide sentiment analysis as a Software as a Service (SaaS) to attract a wide range of users without the need for significant upfront investment.

Subscription Plans

Implement tiered subscription models (basic, professional, enterprise) to cater to different user needs.

User-Friendly Platform

Develop an intuitive web-based interface for easy text input and analysis.

Targeted Marketing

Utilize online advertising and social media to reach potential users in various industries and create engaging educational content to demonstrate the service's value.

Free Trials

Offer a free trial period (e.g., 14 or 30 days) to allow users to experience the service before committing.

Scan Me

WEBSITE

customer-project.azurewebsites.net

GITHUB

github.com/A7md-Waly/sentiment_analysis

Meet Our Best Team

Ahmed Gamel

Alaa Awad

Rowyda Abdelrahem

Ganna Zayed

Hams Tamer

Salma Mohsen

رواد مصر الرقمية

Thank You

For Your Attention