Title: *Amazon Review Sentiment Analysis*

Subtitle: *End-to-End Data Pipeline with Python, MySQL, Power BI, and Azure Simulation*

Name: Tanmay Sharma

Role: Data/Business-Analyst

Project Objective

- Analyse customer sentiment from product reviews.
- Classify reviews as **Positive**, **Negative**, or **Neutral**.
- Generate insights for **customer experience improvement**.

Key Goals:

- ✓ Sentiment Distribution
- ✓ Positive Review %
- ✓ Average Polarity & Rating
- ✓ Regional & Category Trends

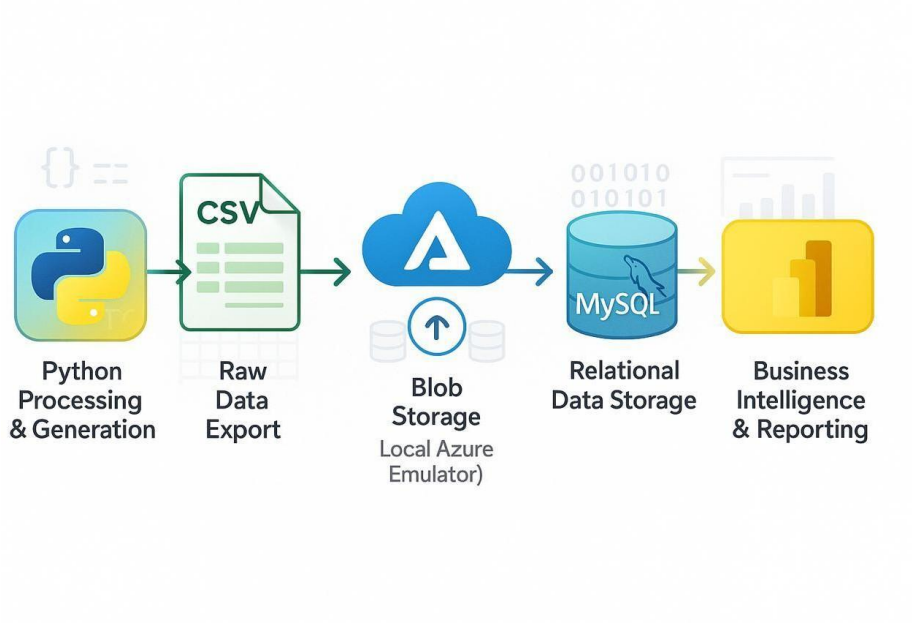
Tech Stack

Languages & Tools:

- **Python**: Data cleaning, sentiment analysis (TextBlob).
- **MySQL**: Data storage & relational modelling.
- **Power BI**: Visualization & interactive dashboard.
- **Azurite (Azure Simulation)**: Cloud integration concept.

: End-to-End Workflow Diagram

Python (Data Generation + NLP)
↓
MySQL (Data Storage & Queries)
↓
Power BI (Dashboard)
↓
Azurite (Cloud Simulation)



Dataset & Model

- **Dataset**: 500 Amazon product reviews (fake but realistic).
- **Fields**: Customer ID, Product ID, Review Text, Rating, Sentiment, Polarity.
- **Model**:
  - Sentiment: Positive / Negative / Neutral.
  - Polarity: Range (-1 to +1).

amazon

amazon-sentiments > Sentimental.py > ...

```python
# filepath: /amazon-sentiments/amazon-sentiments/Sentimental.py
# ------------------------------
# Amazon Review Sentiment Analysis
# ------------------------------

# ✅ Import Libraries
import pandas as pd
import numpy as np
from faker import Faker
import random
from textblob import TextBlob

# ------------------------------
# STEP 1: Generate Synthetic Amazon Review Dataset
# ------------------------------

fake = Faker()
Faker.seed(42)
random.seed(42)

# Define product categories for realism
categories = ['Electronics', 'Books', 'Clothing', 'Sports', 'Home']

# Number of reviews
num_reviews = 500

data = []
for i in range(1, num_reviews + 1):
    review_id = i
    customer_id = fake.uuid4()
    product_id = random.randint(1000, 9999)
    category = random.choice(categories)

    # Generate random review text (short sentences)
    review_text = fake.sentence(nb_words=random.randint(8, 20))

    # Simulate a rating (1-5 stars)
    rating = random.randint(1, 5)

    data.append([review_id, customer_id, product_id, category, review_text, rating])

# Create DataFrame
df = pd.DataFrame(data, columns=['review_id', 'customer_id', 'product_id', 'category', 'review_text', 'rating'])
```

```python
# ------------------------------
# STEP 2: Sentiment Analysis using TextBlob
# ------------------------------

def get_sentiment(text):
    analysis = TextBlob(text)
    polarity = analysis.sentiment.polarity  # range: -1 (negative) to +1 (positive)
    if polarity > 0.1:
        return 'Positive'
    elif polarity < -0.1:
        return 'Negative'
    else:
        return 'Neutral'

def get_polarity(text):
    return round(TextBlob(text).sentiment.polarity, 3)

# Apply sentiment analysis
df['sentiment'] = df['review_text'].apply(get_sentiment)
df['polarity'] = df['review_text'].apply(get_polarity)

# ------------------------------
# STEP 3: Save to CSV
# ------------------------------
# Unique Customers
customers_df = df[['customer_id']].drop_duplicates()
customers_df['name'] = [fake.name() for _ in range(len(customers_df))]
customers_df['region'] = [fake.state() for _ in range(len(customers_df))]
customers_df.to_csv('customers.csv', index=False)

# Unique Products
products_df = df[['product_id', 'category']].drop_duplicates()
products_df.to_csv('products.csv', index=False)

# Reviews (already have df)
df.to_csv('reviews.csv', index=False)
```

SQL Queries

- Total Reviews → SELECT COUNT(*) FROM amazon_reviews_sentiment;
- Avg Polarity by Category → SELECT category, ROUND(AVG(polarity), 2) ...
- Top 10 Products by Positive Reviews → SELECT product_id, COUNT(sentiment)...

✓ Show 2–3 queries with formatted SQL.

✅ ERD Structure

lua

Copy Edit

Relationships:

customers.customer_id → reviews.customer_id (1-to-many)

products.product_id → reviews.product_id (1-to-many)

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

review_id	customer_id	product_id	category	review_text	rating	sentiment	polarity
1	bdd640fb-0667-4ad1-9c80-317fa3b1799d	2824	Electronics	Each cause bill scientist nation opportunity all b...	3	Positive	0.2
2	571aa876-6c30-4511-b2b9-437a28df6ec4	5012	Books	Choice whatever from behavior benefit suggest...	1	Neutral	0
3	ec1b8ca1-f91e-4d4c-9ff4-9b7889463e85	9935	Electronics	Bad fall pick those gun court attorney product si...	4	Positive	0.158
4	a5e5a5ab-aefc-4ad8-afc8-9849b3aa7efe	1520	Electronics	Participant commercial rock clear here writer.	2	Neutral	0.05
5	3838b326-8e94-4239-b02b-61c4a3d70628	4811	Home	Successful simply director allow firm environmen...	1	Positive	0.2
6	6dadd6c7-95a7-4d79-bf3c-4c06434308bc	4257	Home	Human despite young meeting before another b...	2	Positive	0.152
7	c0398710-8976-4334-a281-7efdae849217	8359	Home	Shoulder future fall citizen about reveal rest will...	1	Neutral	0
8	a39231a7-d777-4477-8c66-e0a8a013ac6e	3615	Sports	Participant check several much single morning a ...	3	Neutral	-0.036
9	d0e6e660-7c69-4ee1-bb5e-4bcf15ed6269	3547	Books	Within mouth call process water close month pa...	3	Positive	0.2
10	a65e688e-abf3-4d39-bec2-1bbe66245bfa	2674	Electronics	Just military building different full open discover...	1	Neutral	0.062
11	7c441fe7-ab42-40a7-874a-493b3ceddf2d	6881	Clothing	Much rich think office drug list imagine behind pr...	3	Positive	0.258
12	6712303a-0f84-4fef-9931-e9eea56c0941	1711	Sports	Song body court movie cell contain leg themselv...	1	Positive	0.25
13	2a935d62-3c83-4dc0-9944-1fa5c0e9ab30	7201	Electronics	Magazine degree husband around her world ent...	3	Positive	0.136

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

customer_id	name	region
000c14a5-cf51-4203-a2d9-99bac146ed52	Samantha Navarro	Connecticut
00473139-185d-4528-b425-8e191758d47e	Michael Snow	Indiana
00e85ece-0b49-452d-86d4-83f3d450281c	Anthony Kelly	Maryland
0171f5a7-4148-4d11-9894-62e895801c3a	Ashley Simmons	New Mexico
01e06ad8-8880-4d4d-89b8-b7eec5f833cf	Brenda Raymond	Texas
0227b8d5-b070-4a53-90ce-857f75c72fb6	Karen Delgado	Minnesota
0351b09d-0dc5-4b06-8d58-09b7f740d87c	Bridget Hernandez	Ohio
03a2cee9-41c7-49b3-9b9c-9b7c4674a0eb	Laura Johnson	Massachusetts
047dd583-9344-4b2f-bba3-16688334e524	Laura Johnson	New Hampshire
04b40ac9-2831-4f0d-be90-63fb64b5d9a5	Jason Cole	New Mexico
04d07f29-13e6-45fd-9c99-d366c8b62c18	Anthony Miller	West Virginia
057fdd20-a50a-4a29-bcda-835f71a0cff8	Luis Long	New Hampshire

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

product_id	category
1025	Home
1043	Sports
1046	Clothing
1058	Clothing
1096	Clothing
1153	Sports
1207	Sports
1224	Home
1228	Clothing
1241	Electronics
1258	Books
1311	Clothing

Power BI Dashboard:

amazon

Region

All

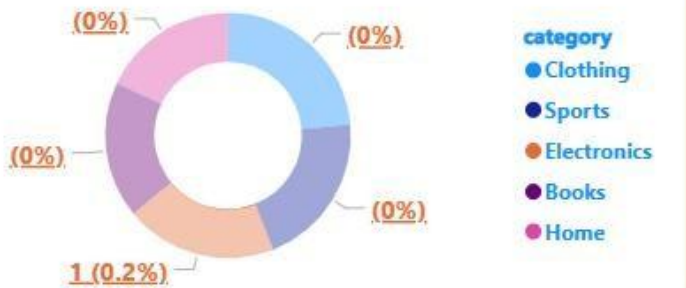
Amazon Review Sentiment Analysis

Rating by Category

All

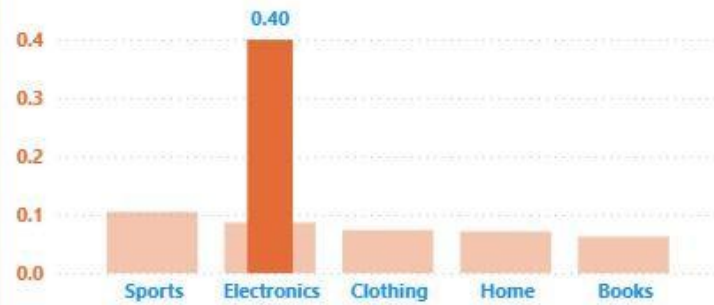
Positive Review %	Avg Polarity	Neutral Review %	Negative Review %	Avg Rating
100.00	0.40	(Blank)	(Blank)	2.00

Sentimental Distribution

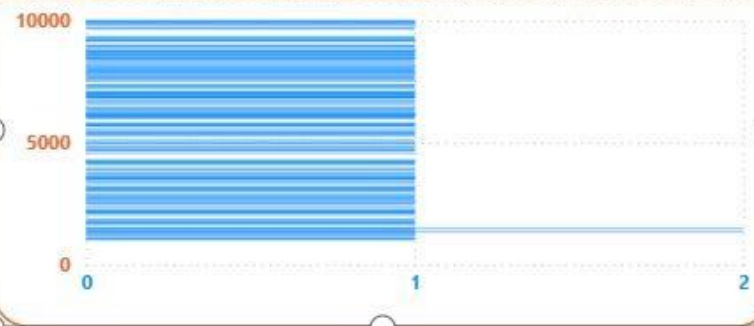


amazon

Average Polarity and Total Reviews by category

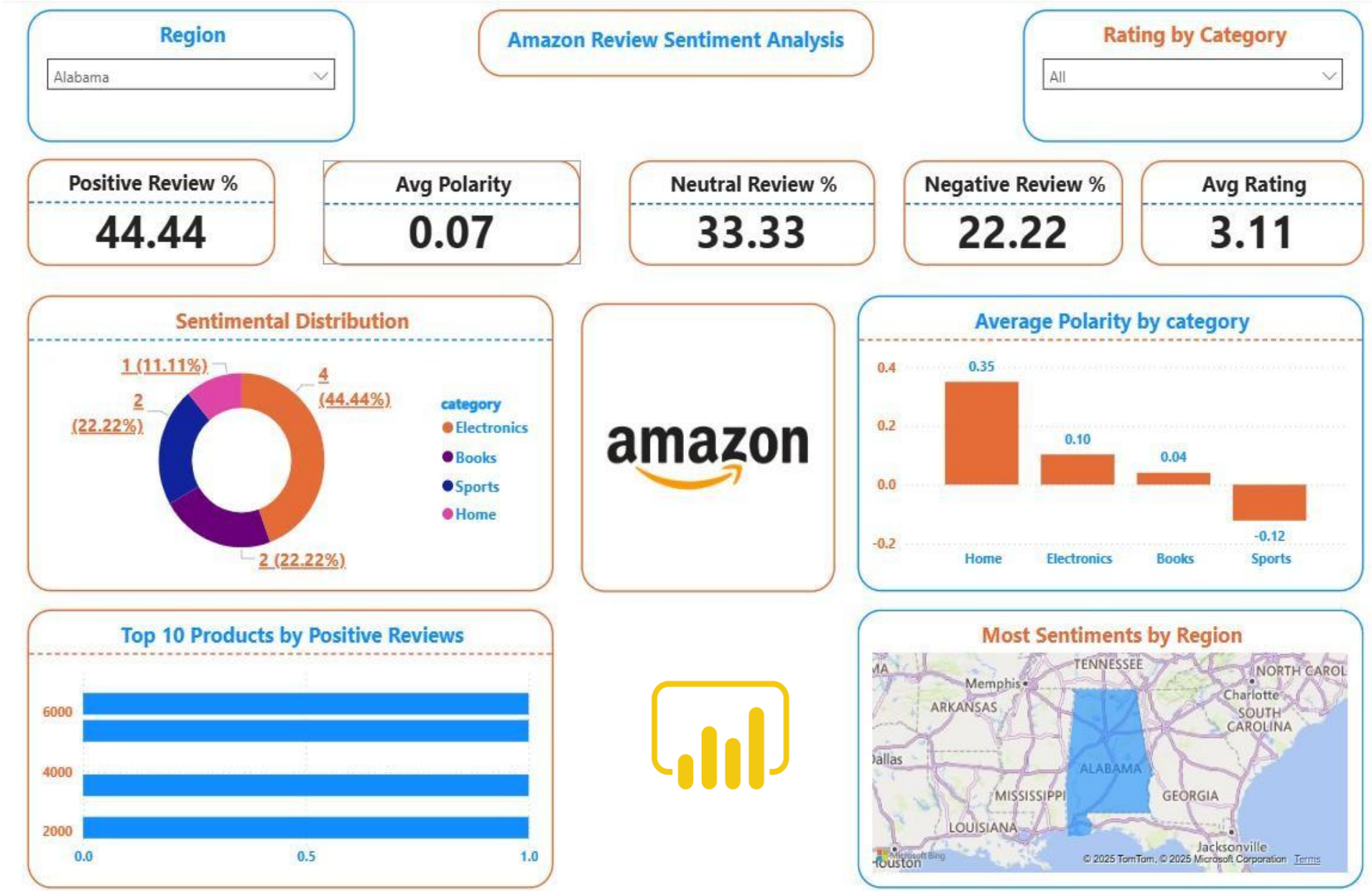


Top 10 Products by Positive Reviews



Most Sentiments by Region

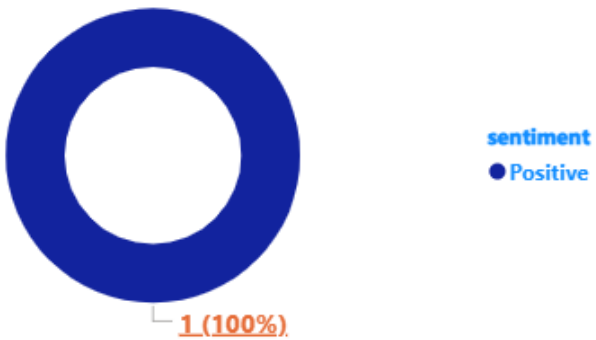
Slicers for Interactive Dashboard:



Drill-through Page:

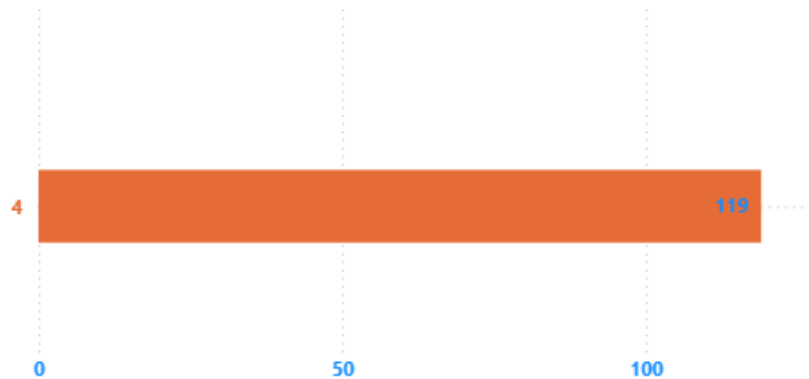
Positive Review %	Avg Polarity	Total Reviews	Avg Rating
100.00	0.43	1	4.00

Sentiment Breakdown



Sentiment vs Rating Breakdown

Rating Distribution



Average Polarity

Business Insights

✓ 37.4% Positive reviews overall.

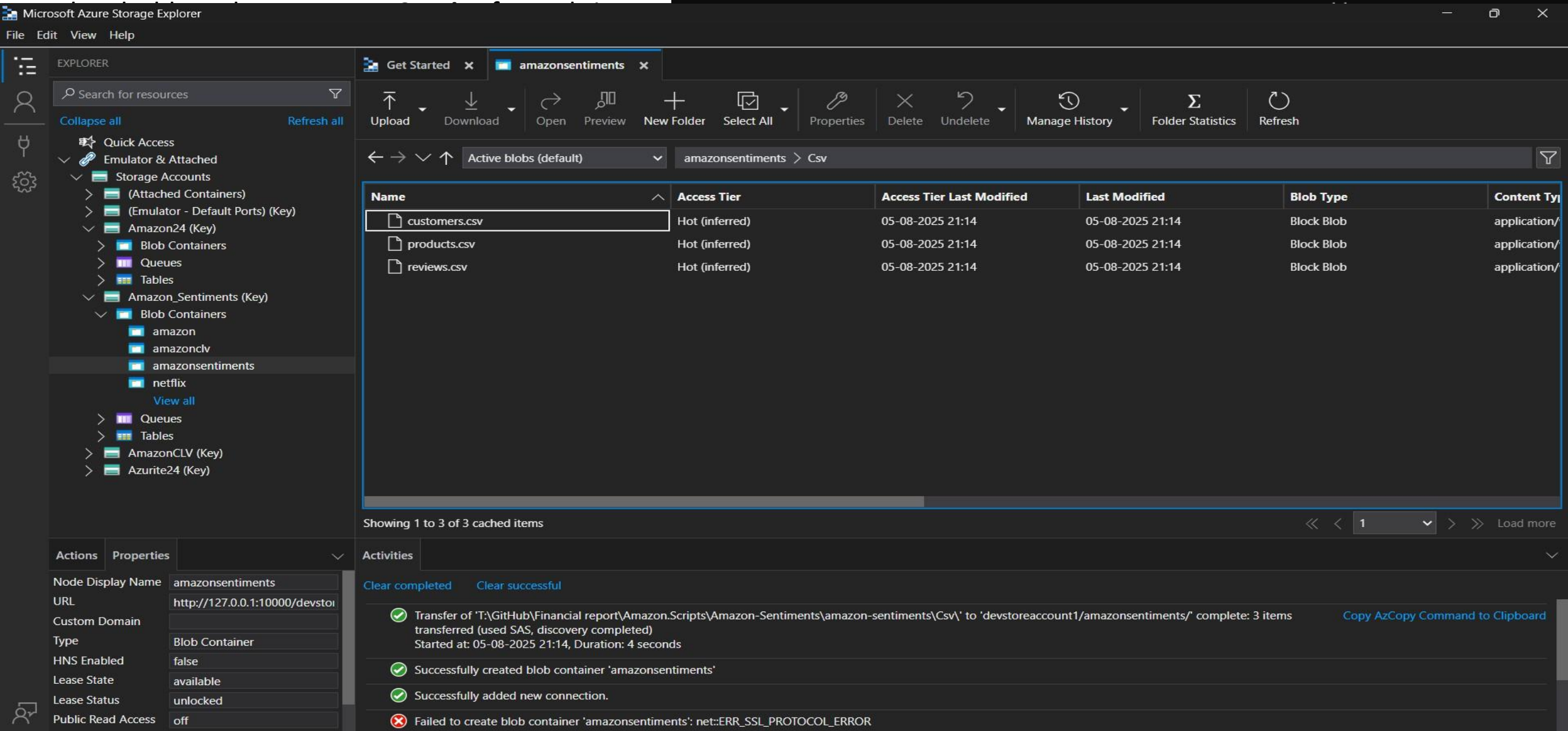
✓ Electronics category has highest polarity (0.40).

✓ Certain regions show higher negative sentiment → need improvement.

Future Enhancements

- Live connection with **Azure Blob Storage**.
- Use **Azure Cognitive Services** for advanced sentiment analysis.

```
Azurite Blob service is starting at http://127.0.0.1:10000
Azurite Blob service is successfully listening at http://127.0.0.1:10000
Azurite Queue service is starting at http://127.0.0.1:10001
Azurite Queue service is successfully listening at http://127.0.0.1:10001
Azurite Table service is starting at http://127.0.0.1:10002
```



Thank you/Let’s Connect

- LinkedIn: http://www.linkedin.com/in/
- Github: https://github.com/Tanu272004/Amazon_Sentiments.git