•Title: **Amazon Customer Lifetime Value Analysis**
•Subtitle: *End-to-End Analytics with Python | MySQL | Power BI | Azure*
•Name: Tanmay Sharma
•Role: Data/Business-Analyst(“Aspiring Fresher”)

Project Objective:

- **Goal:**

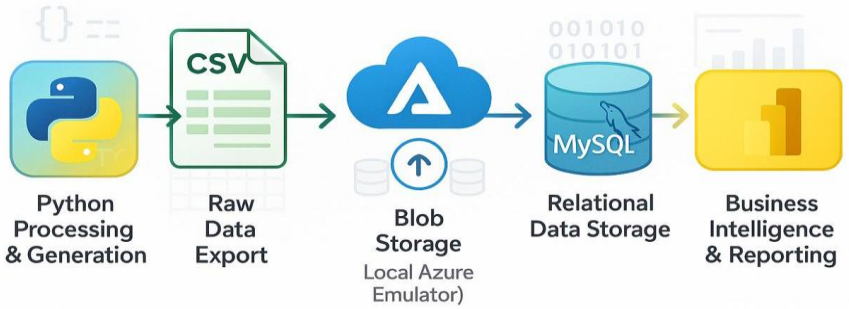
Analyze CLV, identify high-value customers, revenue patterns, and category performance.

- **Business Impact:**
- Improve retention strategy.
- Maximize revenue from top customers.

Tech Stack & Workflow:

- **Tech Used:** Python, MySQL, Power BI, Azure.
- **Workflow Diagram:**

Python → MySQL → Power BI → Azure



Data Pipeline

•Steps:

- **Python:** Data generation & CLV calculation.
- **MySQL:** Data storage, schema, queries.
- **Power BI:** Dashboard & insights.
- **Azure:** Container simulation (Azurite).

```python
import pandas as pd
import random
from faker import Faker
from datetime import datetime
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error


# ------------------------------
# STEP 4: Calculate CLV
# ------------------------------
agg_orders = orders_df.groupby('customer_id').agg({
    'quantity': 'sum',
    'price': 'mean',
    'order_id': 'count'
}).rename(columns={'price': 'avg_price', 'order_id': 'frequency'})

agg_orders['monetary'] = agg_orders['avg_price'] * agg_orders['quantity']
current_date = datetime.now()
customers_df['lifespan_years'] = customers_df['signup_date'].apply(
    lambda x: (current_date - datetime.strptime(x, '%Y-%m-%d')).days / 365
)
```

Result Grid | Filter Rows: | Edit: | Export/Im

customer_id	name	region	signup_date
1	Gloria Mitchell	Peru	2024-10-22
2	Jessica Ross	Malta	2023-05-15
3	Franklin Walker	Anguilla	2023-08-06
4	Michael Huff	Greece	2025-05-26
5	April Reyes	Sao Tome and Principe	2023-05-23
6	Tammy Callahan	Solomon Islands	2024-01-03
7	Priscilla Garcia	Zambia	2022-10-10
8	Lisa Hudson	Honduras	2024-09-12
9	Rodney Johnson	Japan	2025-05-27
10	Joseph Carson	Kuwait	2023-06-20
11	Michele Smith	Congo	2025-02-25
12	Shannon Jordan	Saint Vincent and the ...	2022-11-23
13	Kendra Young	Liberia	2023-10-23

customers 1

Result Grid | Filter Rows:

product_id	category	price
1	Electronics	55.66
2	Home	31.66
3	Sports	437.91
4	Books	310.64
5	Electronics	234.83
6	Clothing	93.43
7	Clothing	199.03
8	Home	84.35
9	Books	380.90
10	Home	453.28
11	Electronics	32.84
12	Sports	317.10
13	Books	308.95

products 1

amazon

MySQL

order_id	customer_id	product_id	order_date	quantity	price
1	149	10	2024-10-10	2	453.28
2	32	22	2024-11-24	1	206.02
3	453	40	2025-01-12	2	178.45
4	205	4	2025-04-14	5	310.64
5	309	39	2023-12-25	1	356.55
6	293	3	2025-04-14	1	437.91
7	19	34	2024-04-07	1	383.91
8	396	43	2024-10-02	5	466.22
9	7	31	2025-04-21	4	423.58
10	36	18	2024-08-16	3	416.98
11	403	22	2025-01-06	3	206.02
12	274	34	2025-06-21	4	383.91
13	185	17	2024-09-24	5	236.40

orders 1

customer_id	avg_order_value	frequency	lifespan_years	clv	predicted_clv
1	944.99	10	0.78	7378.65	8093.84
2	1058.98	9	2.22	21176.77	19809.14
3	683.19	7	1.99	9538.52	10958.52
4	965.42	18	0.19	3285.07	13219.73
5	860.19	14	2.20	26493.93	23182.61
6	505.05	8	1.58	6398.24	6679.43
7	672.13	10	2.82	18930.13	20986.46
8	1015.02	12	0.89	10845.46	12232.88
9	801.96	10	0.19	1494.06	1581.77
10	676.58	4	2.12	5746.31	8305.31
11	663.53	14	0.44	4046.61	6665.82
12	749.07	8	2.70	16155.26	18572.50
13	470.12	10	1.78	8372.03	10224.14

customer_clv 1

- Query Part–
- 1): Overall Average CLV Select Round(avg(clv),
- 2) as Avg_CustomerCLVFrom customer_clv;-- Top 10 High-Value CustomersSELECT c.customer_id, c.name, customer_clv.clvFROM customer_clv JOIN customers c ON c.customer_id = customer_clv.customer_idORDER BY customer_clv.clv DESCLIMIT 10;-- Revenue by regionSELECT c.region, ROUND(SUM(o.quantity * o.price), 2) AS total_revenueFROM orders oJOIN customers c ON o.customer_id = c.customer_idGROUP BY c.regionORDER BY total_revenue DESC;

- 3) Customer Segmentation by CLVSELECT c.customer_id, c.name, Round((customer_clv.clv),2) as CLVFROM customer_clv JOIN customers c ON c.customer_id = customer_clv.customer_idORDER BY CLV DESCLIMIT 10;-- Avg order value by category;SELECT p.category, ROUND(AVG(o.price * o.quantity), 2) AS avg_order_valueFROM orders oJOIN products p ON o.product_id = p.product_idGROUP BY p.categoryORDER BY avg_order_value DESC;

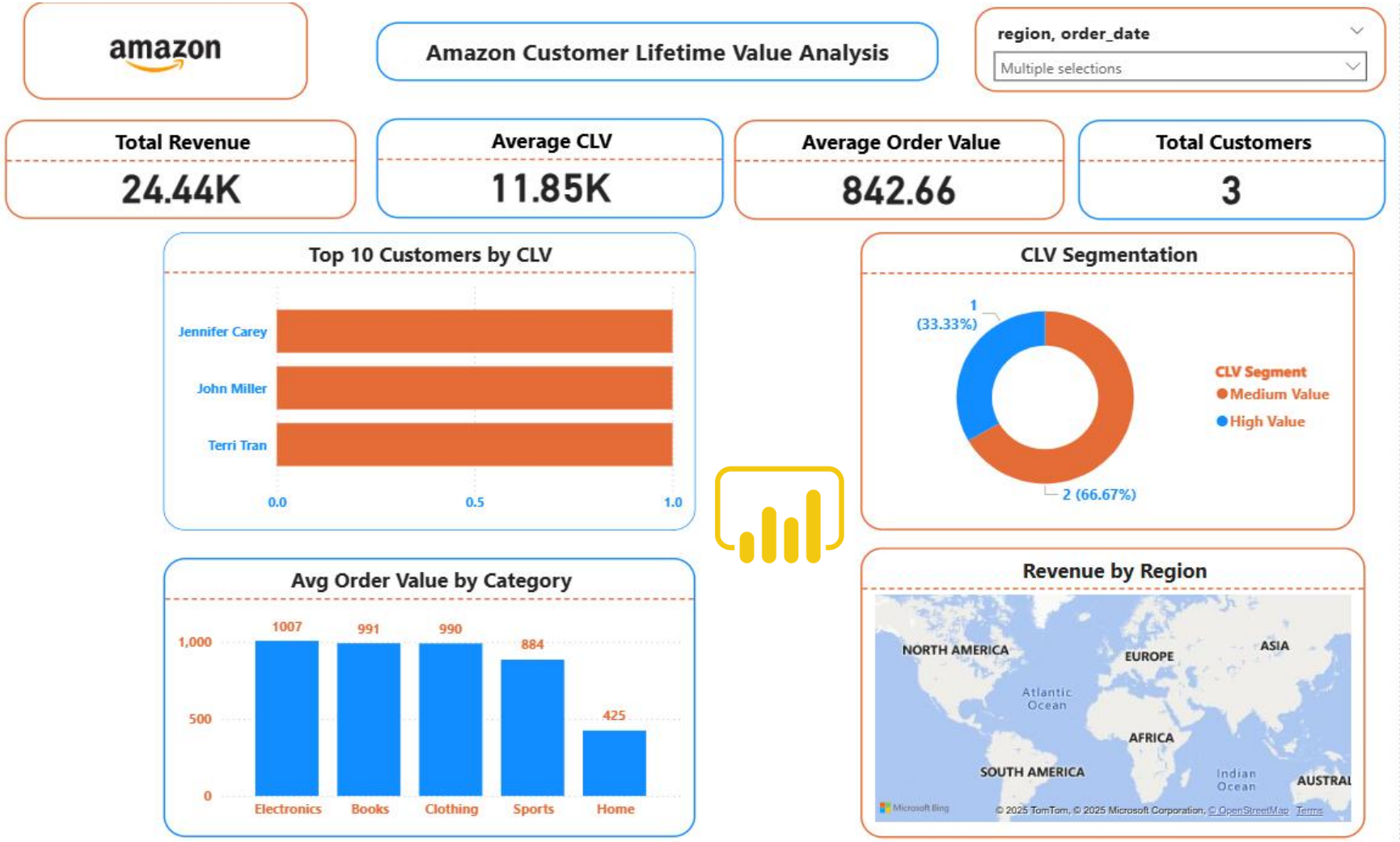
Dashboard Overview

•Title: *Power BI Dashboard*

•Subtitle: *Amazon CLV Analysis*

•Add a short caption:

"Dynamic dashboard showing KPIs and insights across customers, regions, and product categories."

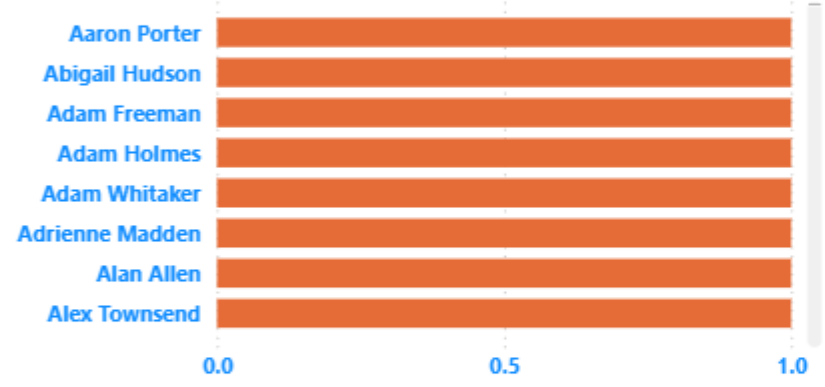


Amazon Customer Lifetime Value Analysis

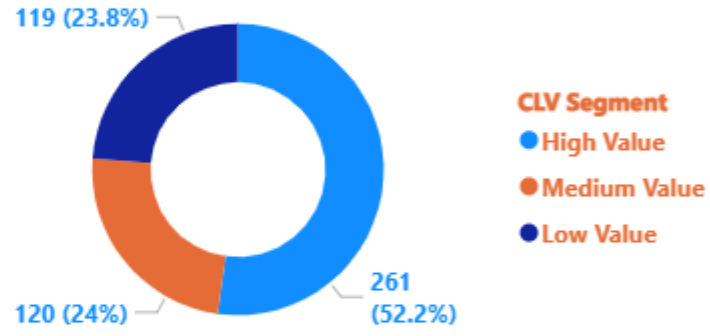
region, order_date: All

Total Revenue	Average CLV	Average Order Value	Total Customers
3.97M	12.18K	793.83	500

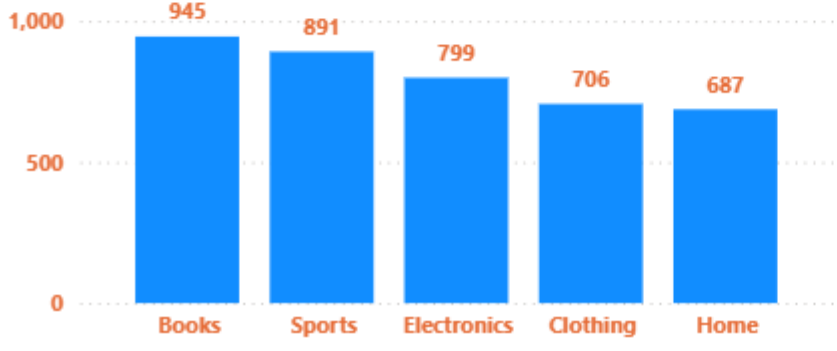
Top 10 Customers by CLV



CLV Segmentation



Avg Order Value by Category

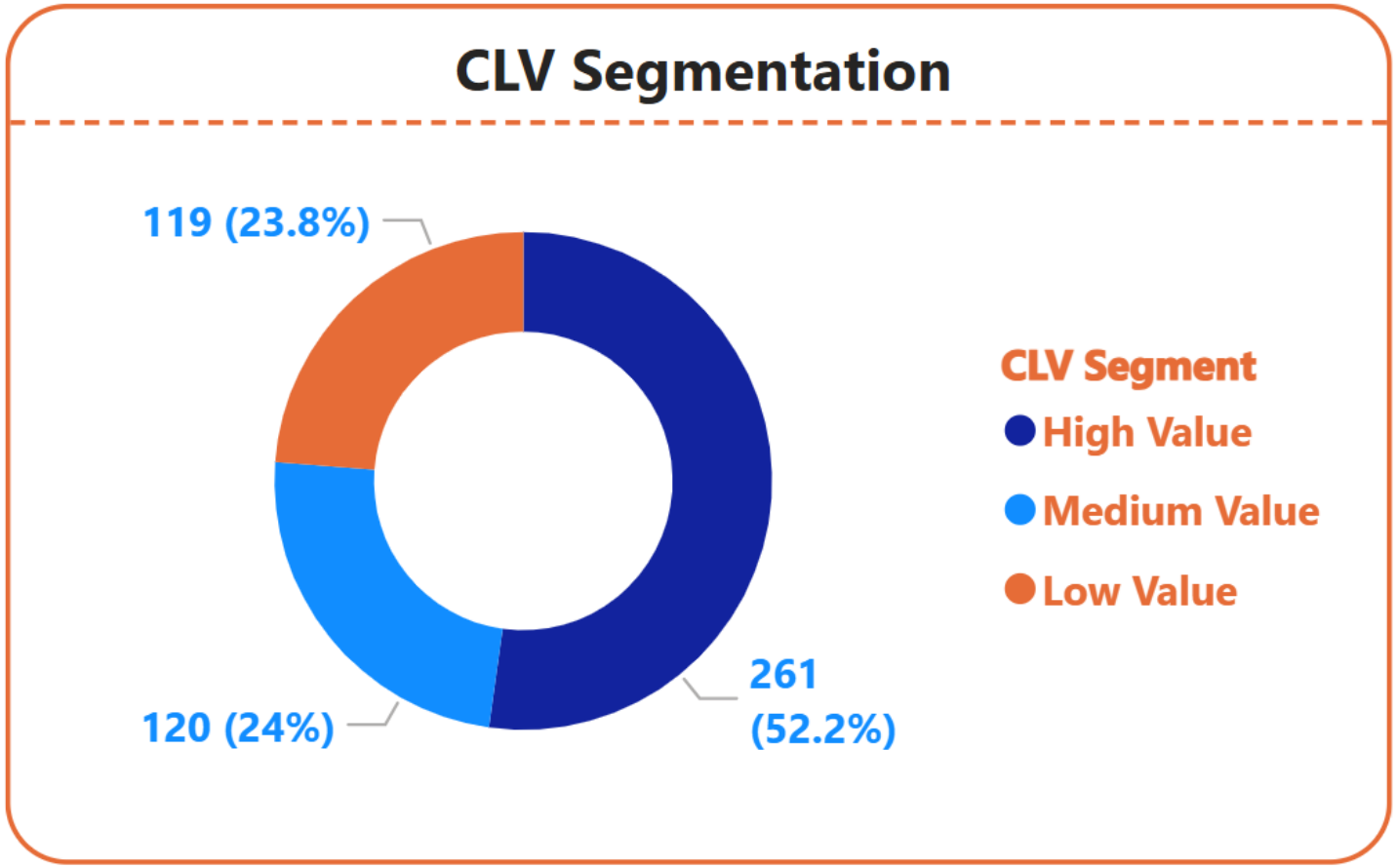


Revenue by Region

Key Insights

•Bullet Points:

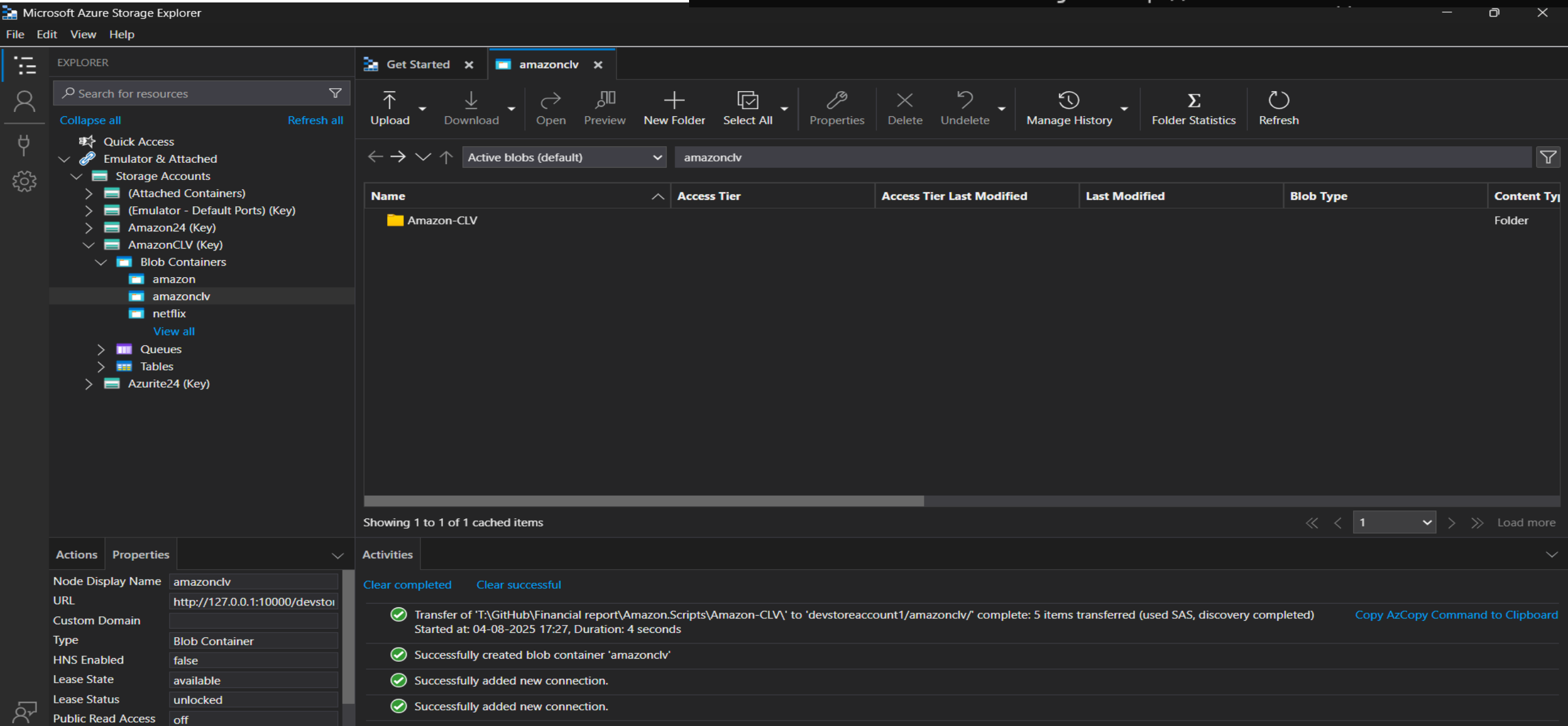
- **Top 10 CLV Customers** → Retention priority.
- **High CLV Segment** → 52% of total revenue.
- **Top Categories:** Books & Sports.
- **Top Region:** North America.



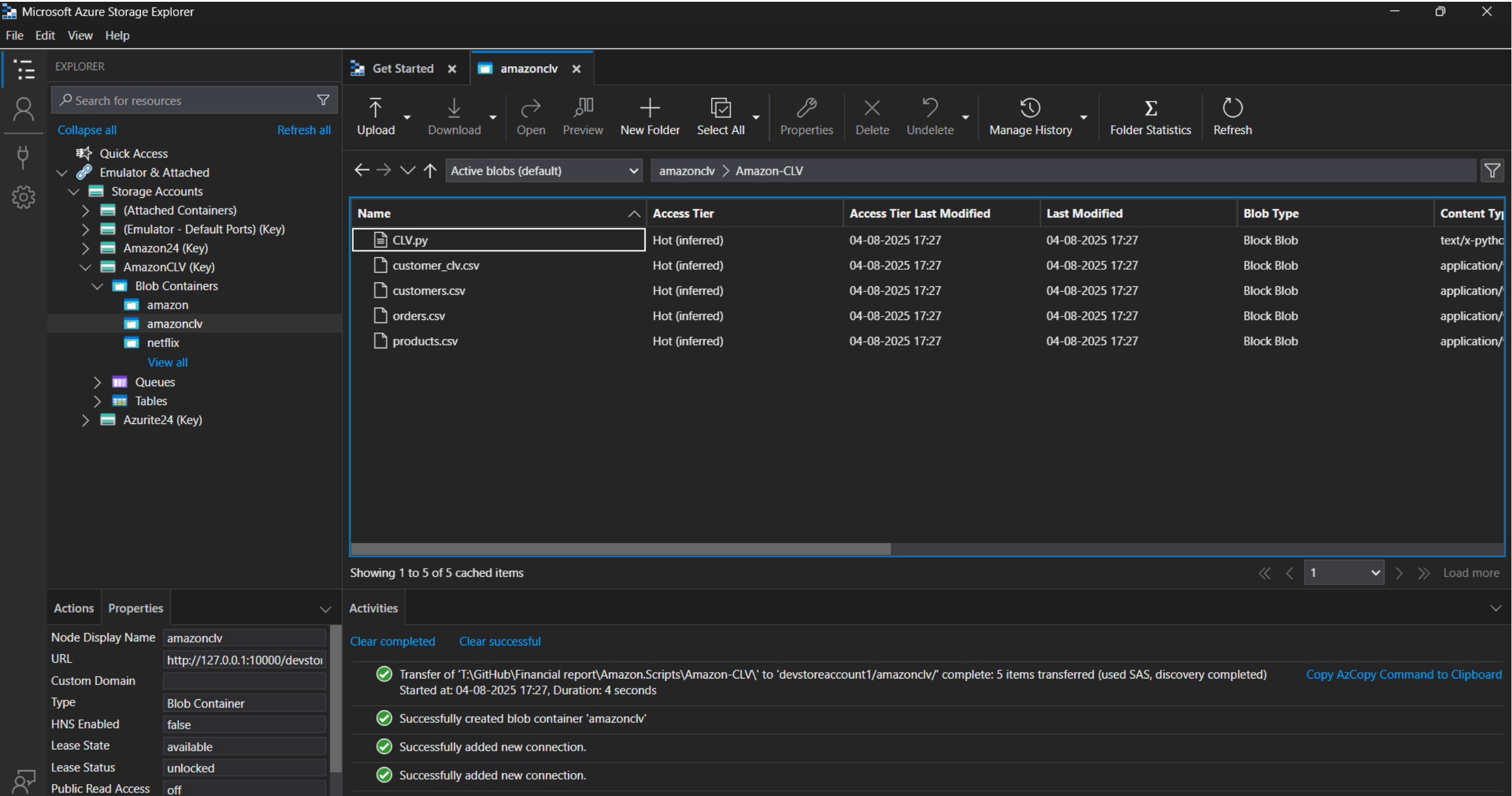
Advanced Features

- Predictive CLV Model in Python.
- Automated MySQL queries for segmentation.
- Azure containerization for cloud readiness.

```
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
```



All CSV Files in Azure Blob Storage:



Conclusion & Future Scope

•Conclusion:

- CLV-based segmentation helps optimize marketing and retention strategies.

•Future Scope:

- Real-time CLV monitoring (Azure Data Factory).
- Churn prediction using ML pipelines.

LinkedIn: http://www.linkedin.com/in/

Github Link :
https://github.com/Tanu272004/Amazon_CLV_Analytics_Project#amazon_clv_analytics_project