E-Commerce Customer Segmentation
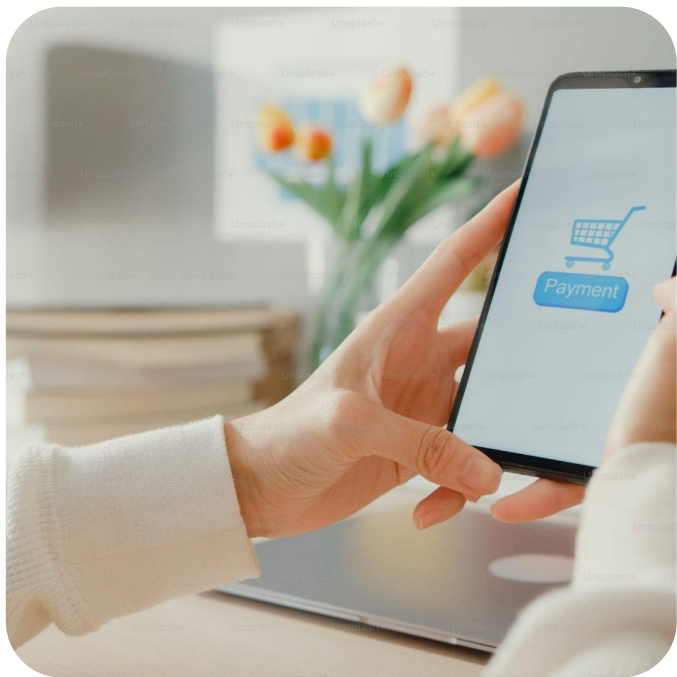

Presentation Contributors
Juan Segovia, Julia McKinnon, Grant Ritzwoller, Daphine Nyangena, Ernesto Garcia

E-Commerce Customer Data

- Dataset obtained from Kaggle
- Synthetic Data with over 3 million rows
- Transaction-level
- Based on one fictional ecommerce store-ShopSpectra
- Data from early 2023- mid 2024
- United States customer locations

Data Preparation

- Aggregated transaction-level data down to customer-level
- Feature groups:
  - Transaction
  - Time-Based
  - Behavioral
  - Location-Based
  - Demographic
  - Merchant Category
  - Transaction Status
- 3 million rows to approx. 38K rows
- 18 original columns to 30 aggregated columns

Problem Statement, Objectives & Project Overview

- Identify distinct customer segments by analyzing data through clustering.
- Segment customers based on purchasing behavior, transaction patterns, and demographics.
- Visualize customer segments and key behavioral trends interactively.
- Utilize unsupervised ML techniques to cluster customers.

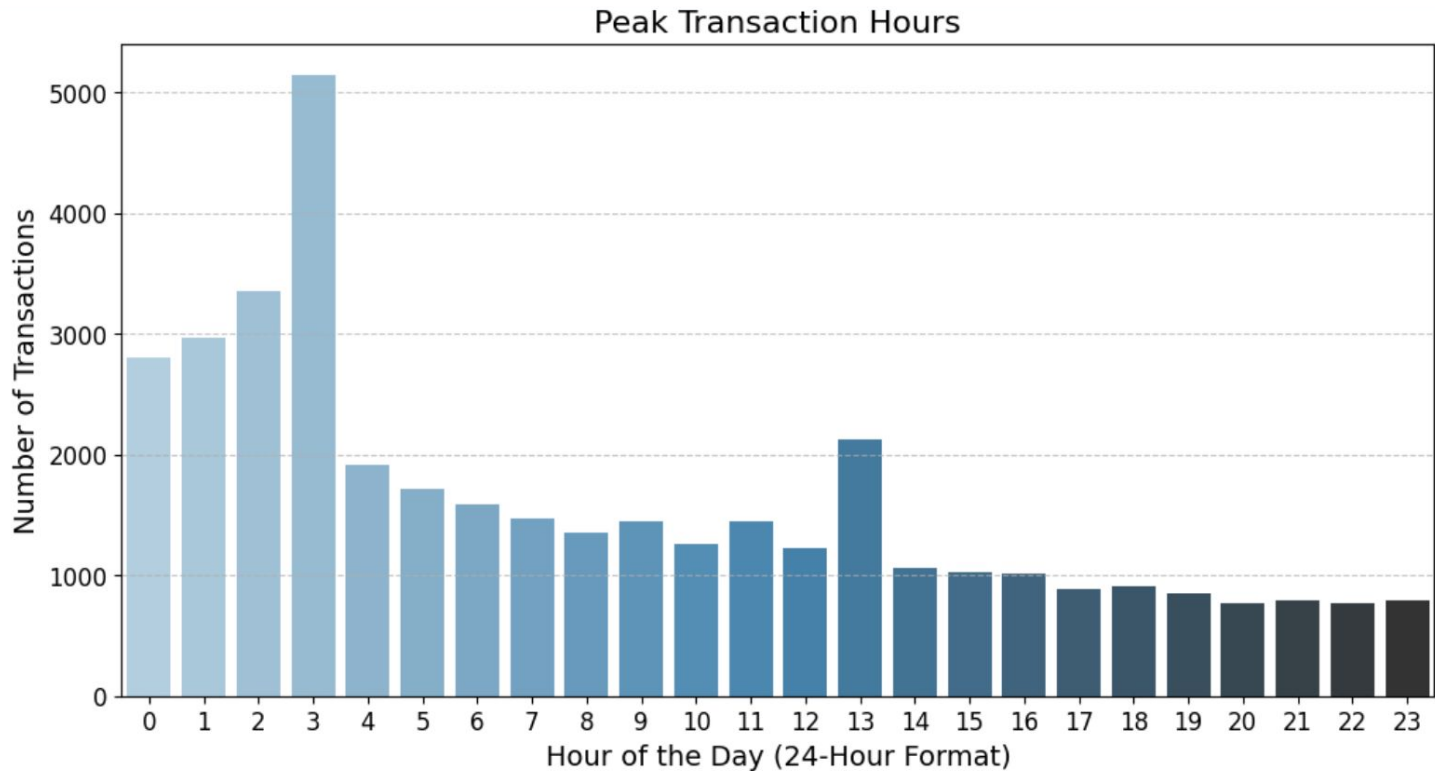# Team Roles

- <u>Data Engineer</u>: Juan Segovia-Collects and preprocesses data.
- <u>Machine Learning Engineer</u>: Julia McKinnon-Develops and optimizes clustering models.
- <u>Visualization Specialists</u>: Grant R./Daphine N./Ernesto G-Visualizes distributions and customer behavior patterns.
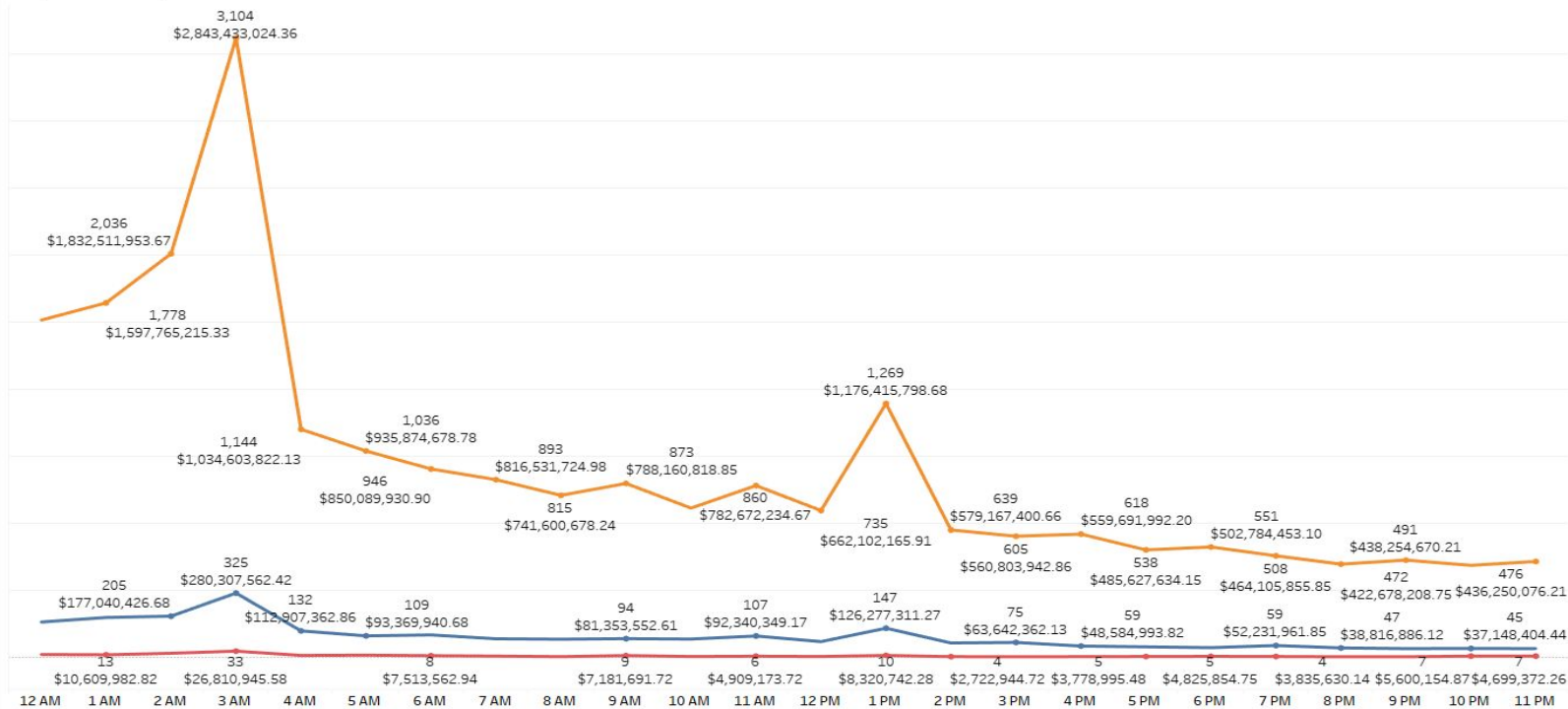- <u>Github Repo/Documentation</u>-Ernesto Garcia
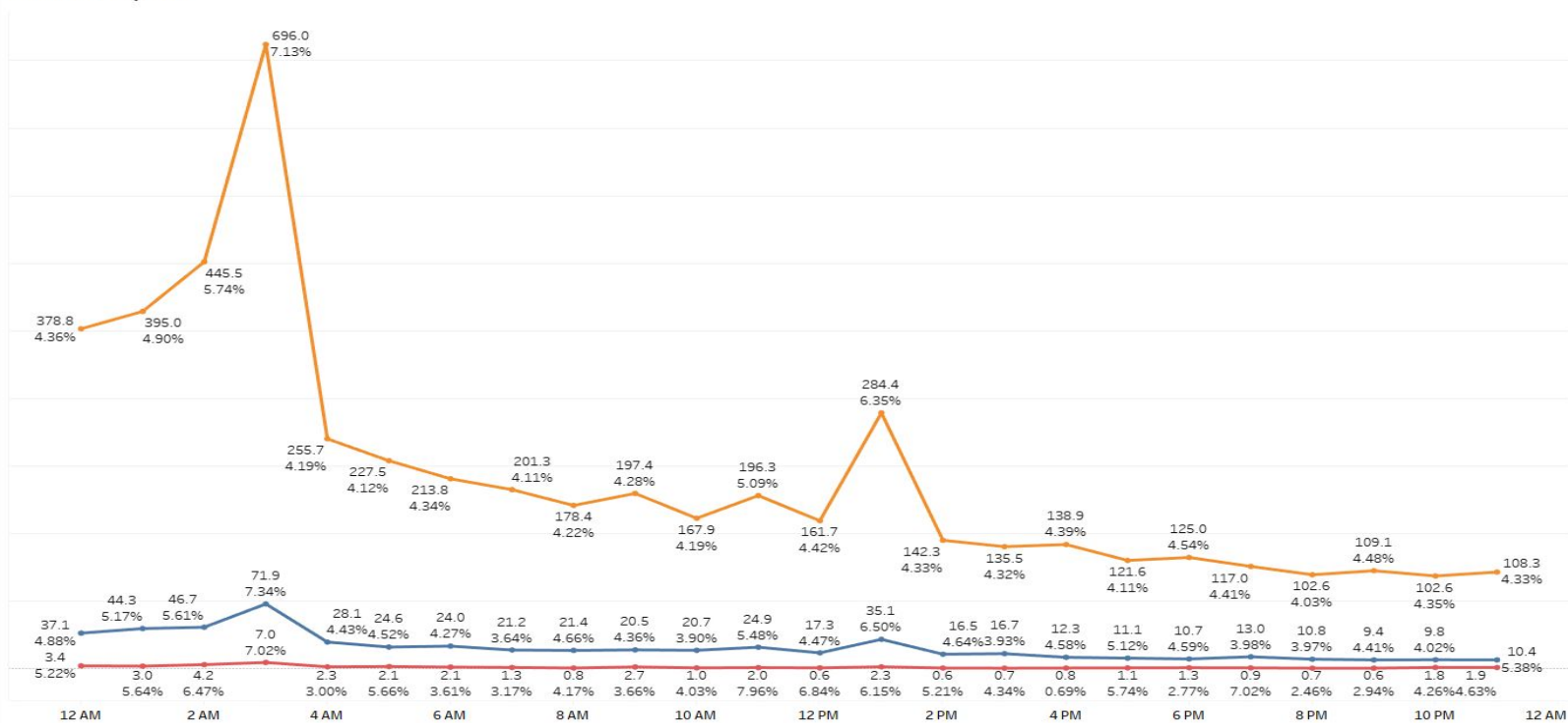
Customer Behavior

Peak Transaction Hours

Completed Rates by Hours

Hour	Mobile (count)	Mobile ($)	Desktop (count)	Desktop ($)	Tablet (count)	Tablet ($)
12 AM	2,036	$1,832,511,953.67	205	$177,040,426.68	13	$10,609,982.82
1 AM	1,778	$1,597,765,215.33				
2 AM						
3 AM	3,104	$2,843,433,024.36	325	$280,307,562.42	33	$26,810,945.58
4 AM	1,144	$1,034,603,822.13	132	$112,907,362.86		
5 AM	946	$850,089,930.90	109	$93,369,940.68		
6 AM	1,036	$935,874,678.78			8	$7,513,562.94
8 AM	893	$816,531,724.98				
9 AM	815	$741,600,678.24	94	$81,353,552.61	9	$7,181,691.72
10 AM	873	$788,160,818.85				
11 AM	860	$782,672,234.67	107	$92,340,349.17	6	$4,909,173.72
12 PM		$662,102,165.91				
1 PM	1,269	$1,176,415,798.68	147	$126,277,311.27	10	$8,320,742.28
2 PM	735	$579,167,400.66			4	$2,722,944.72
3 PM	639	$560,803,942.86	75	$63,642,362.13	5	$3,778,995.48
4 PM	605	$559,691,992.20			6	$4,825,854.75
5 PM	618	$485,627,634.15	59	$48,584,993.82		
6 PM	538	$502,784,453.10			5	$3,835,630.14
7 PM	551	$464,105,855.85	59	$52,231,961.85		
8 PM	508	$438,254,670.21	47	$38,816,886.12	4	$5,600,154.87
9 PM	491	$422,678,208.75				
10 PM	472	$436,250,076.21	45	$37,148,404.44	7	$4,699,372.26
11 PM	476					

12 AM 1 AM 2 AM 3 AM 4 AM 5 AM 6 AM 7 AM 8 AM 9 AM 10 AM 11 AM 12 PM 1 PM 2 PM 3 PM 4 PM 5 PM 6 PM 7 PM 8 PM 9 PM 10 PM 11 PM

Decline Rates by Hours

Time	Orange	Blue	Red
12 AM	378.8 / 4.36%	37.1 / 4.88%	3.4 / 5.22%
	395.0 / 4.90%	44.3 / 5.17%	3.0 / 5.64%
2 AM	445.5 / 5.74%	46.7 / 5.61%	4.2 / 6.47%
	696.0 / 7.13%	71.9 / 7.34%	7.0 / 7.02%
4 AM	255.7 / 4.19%	28.1 / 4.43%	2.3 / 3.00%
	227.5 / 4.12%	24.6 / 4.52%	2.1 / 5.66%
6 AM	213.8 / 4.34%	24.0 / 4.27%	2.1 / 3.61%
	201.3 / 4.11%	21.2 / 3.64%	1.3 / 3.17%
8 AM	178.4 / 4.22%	21.4 / 4.66%	0.8 / 4.17%
	197.4 / 4.28%	20.5 / 4.36%	2.7 / 3.66%
10 AM	167.9 / 4.19%	20.7 / 3.90%	1.0 / 4.03%
	196.3 / 5.09%	24.9 / 5.48%	2.0 / 7.96%
12 PM	161.7 / 4.42%	17.3 / 4.47%	0.6 / 6.84%
	284.4 / 6.35%	35.1 / 6.50%	2.3 / 6.15%
2 PM	142.3 / 4.33%	16.5 / 4.64%	0.6 / 5.21%
	135.5 / 4.32%	16.7 / 3.93%	0.7 / 4.34%
4 PM	138.9 / 4.39%	12.3 / 4.58%	0.8 / 0.69%
	121.6 / 4.11%	11.1 / 5.12%	1.1 / 5.74%
6 PM	125.0 / 4.54%	10.7 / 4.59%	1.3 / 2.77%
	117.0 / 4.41%	13.0 / 3.98%	0.9 / 7.02%
8 PM	102.6 / 4.03%	10.8 / 3.97%	0.7 / 2.46%
	109.1 / 4.48%	9.4 / 4.41%	0.6 / 2.94%
10 PM	102.6 / 4.35%	9.8 / 4.02%	1.8 / 4.26%
	108.3 / 4.33%	10.4 / 5.38%	1.9 / 4.63%

Customer Behavior

Most Commonly Used Devices

- Top 3 Categories
Entertainment, Groceries, Clothing

Transaction Counts by Merchant Category

- Mobile orders are the highest
- Due to easy accessibility

Customer Demographics

Age Distribution of Customers
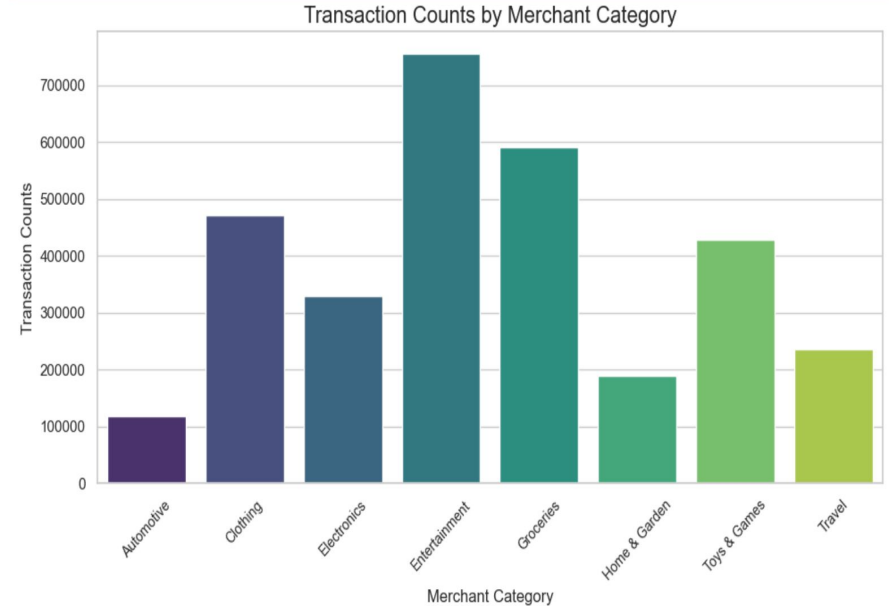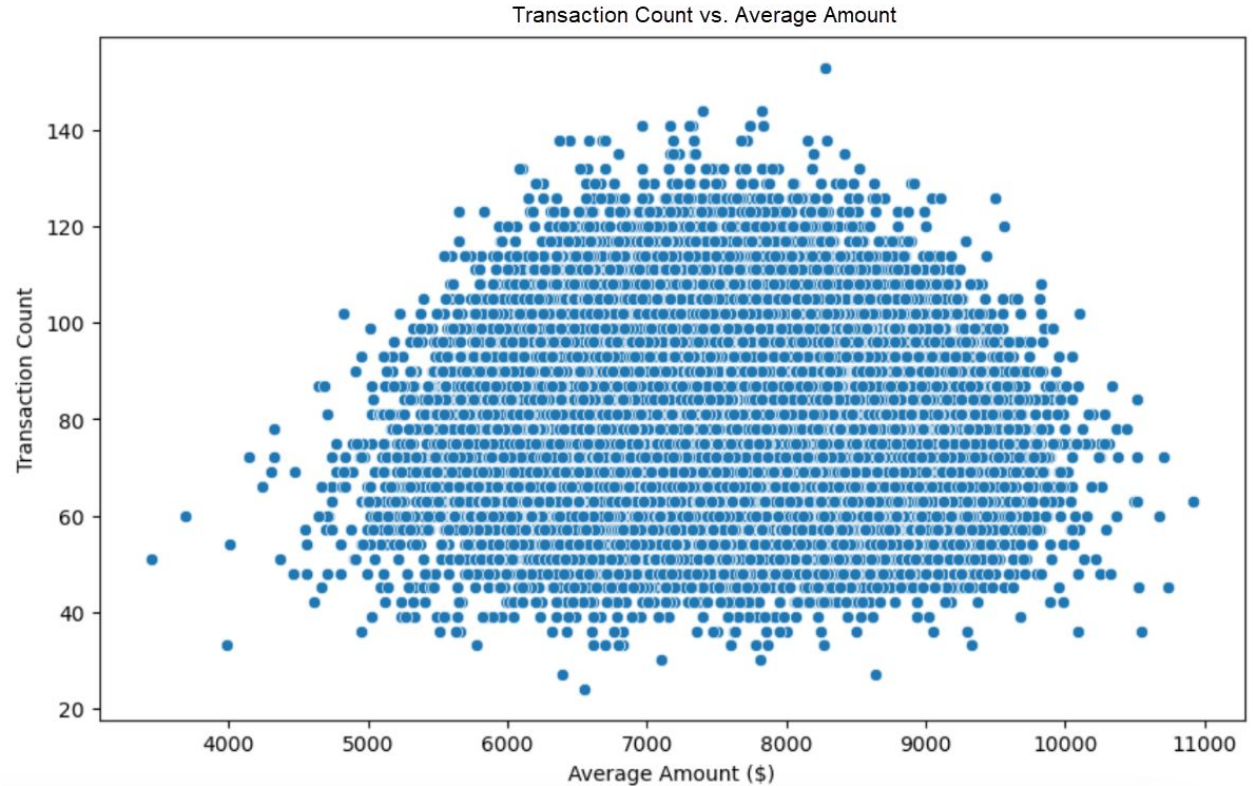
Transaction Status
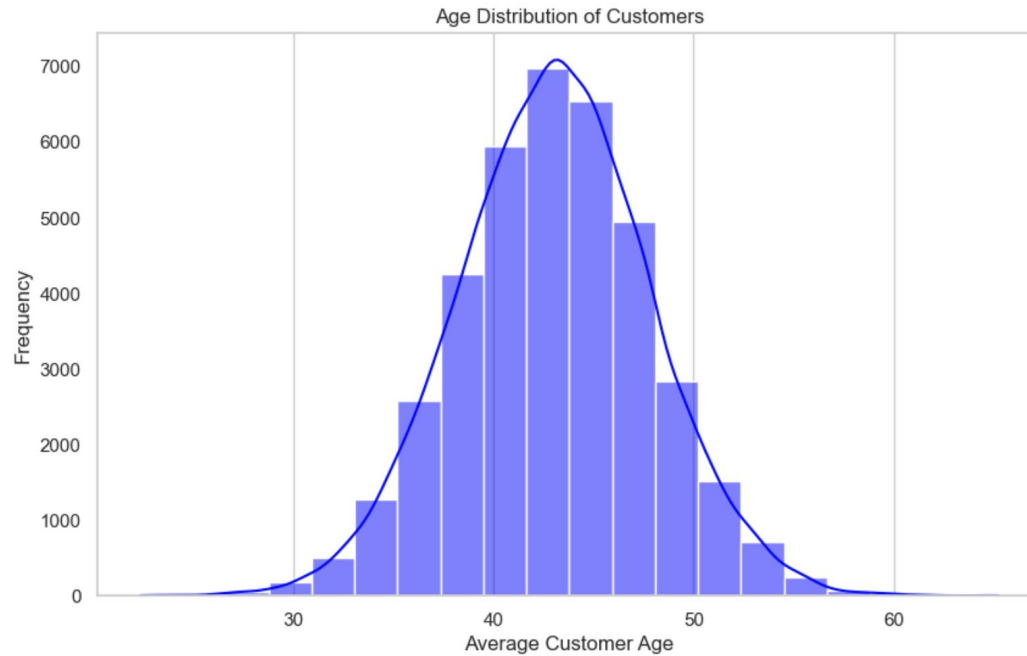
Transaction Status per User (Top 10 Users)
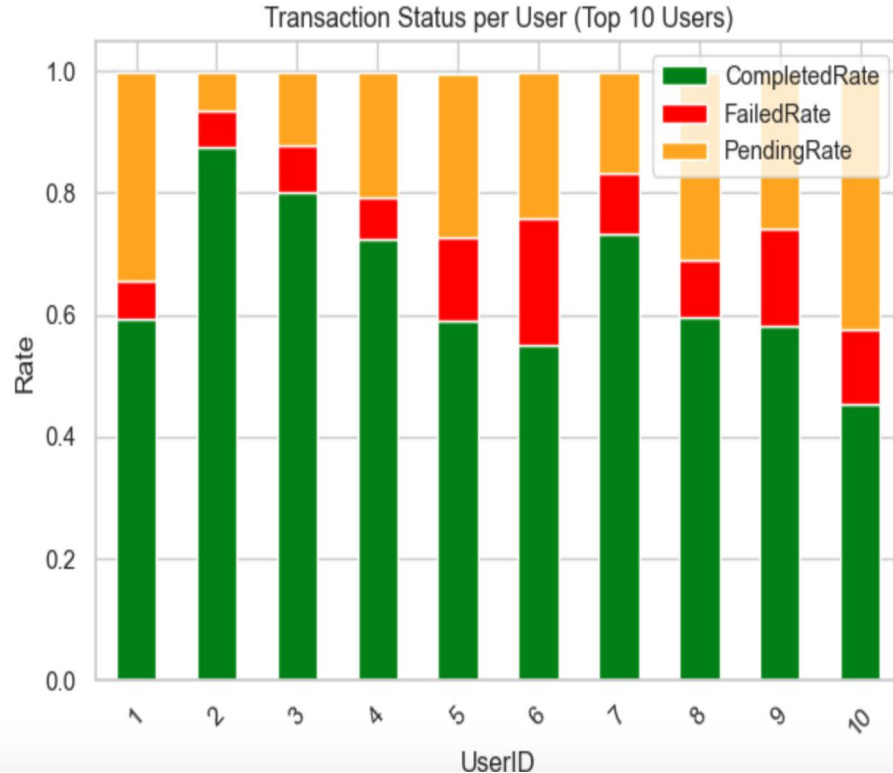
Category Popularity

- Scatter plot demonstrates synthetic data identical in transactions vs avg. amount spent

- Thus very clustered

Transaction Count vs. Average Amount

Conclusion

- Created a segmentation model

- Labeled all the customers

- Synthetic Data not as diverse as real data

- Difficult to identify clear segments

Thank You