Milestone Report

How Many Unique Products were sold?
3,631

How Much Volume was generated during this time frame?
$5,095,812,742

What information can we find on customer trips/purchases?

count	5891.000000
mean	93.374300
std	107.190049
min	6.000000
25%	26.000000
50%	54.000000
75%	117.000000
max	1026.000000

- There were 5,891 unique customers
- The customer purchased on average 93 products during this time period
- The top customer purchased 1,026 products
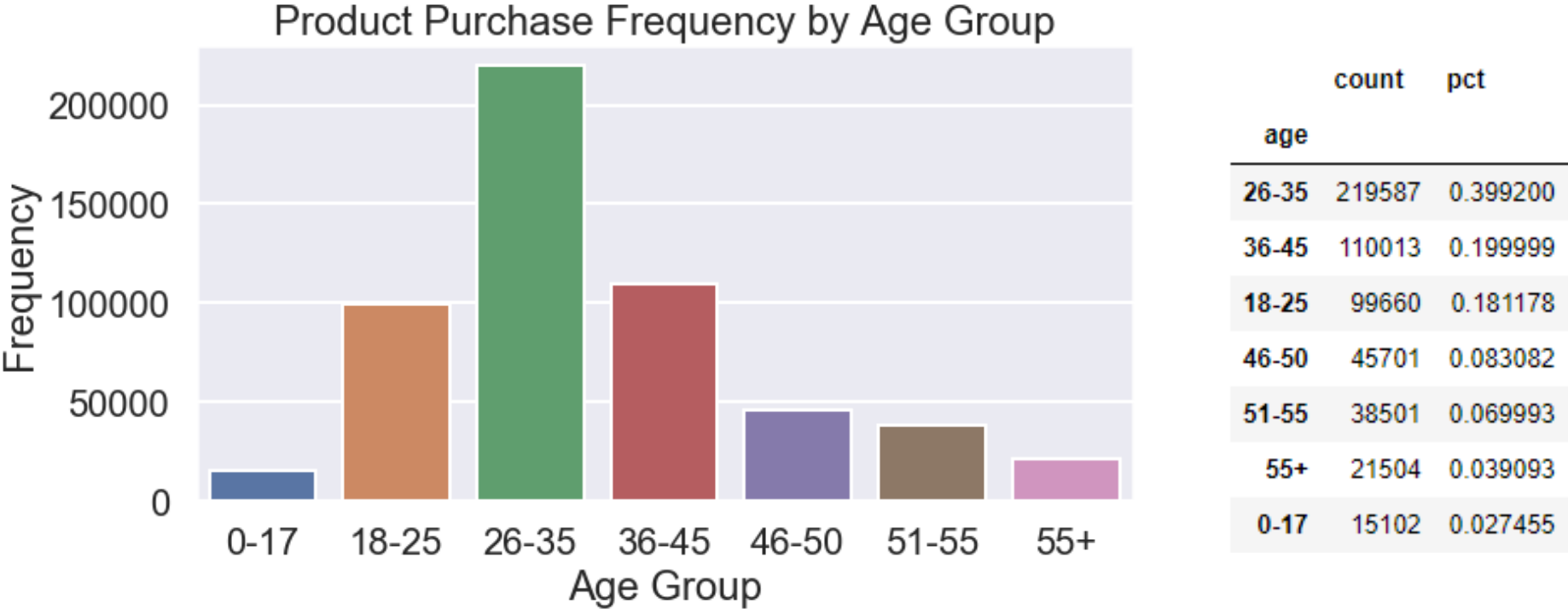- There was a wide range of product purchases from 6 to 1,026 items

Are there insights on Purchase Amounts?

- Customers purchased 550,068 products were sold
- The average purchase amount of products was $9,264
- The top purchase amount was $23,961
- There was a wide range of product purchase amounts from $12 to $23,961 items

```
count      550068.000000
mean         9263.968713
std          5023.065394
min            12.000000
25%          5823.000000
50%          8047.000000
75%         12054.000000
max         23961.000000
```

What is the predominate age group that shopped?

Product Purchase Frequency by Age Group

	count	pct
age		
26-35	219587	0.399200
36-45	110013	0.199999
18-25	99660	0.181178
46-50	45701	0.083082
51-55	38501	0.069993
55+	21504	0.039093
0-17	15102	0.027455

The age group 26-35 shop the most at this retailer, making up 40% of purchases.

Is there a difference in purchases or spend between Genders?

	count	pct
gender		
M	414259	0.753105
F	135809	0.246895

	Purchase
Gender	
F	8734.565765
M	9437.526040

Males make up 75% of the customers for this retailer and spend $703 more than Female customers.

Are there any shopping differences among occupations?

Purchase Frequency by Occupation

occupation	count	pct
4	72308	0.131453
0	69638	0.126599
7	59133	0.107501
1	47426	0.086218
17	40043	0.072796
20	33562	0.061014
12	31179	0.056682
14	27309	0.049647
2	26588	0.048336
16	25371	0.046123

Volume Totals by Occupation

purchase	count	pct
4	666244484	0.130744
0	635406958	0.124692
7	557371587	0.109378
1	424614144	0.083326
17	393281453	0.077177
12	305449446	0.059941
20	296570442	0.058199
14	259454692	0.050915
16	238346955	0.046773
2	238028583	0.046711

Occupations 4, 0 and 7 account for 37% of this retailers customers. Most of the occupations' spend match the fre3quency of the observed occupations, which means that the average spend across occupations is rather similar.

Is there a difference in purchases or spend by Marital Status?

0 = Single
1 = Married

	count	pct
marital_status		
0	324731	0.590347
1	225337	0.409653

	Purchase
Marital_Status	
0	9265.907619
1	9261.174574

60% of shoppers are single, and both married and single customers spend similar amounts on average.
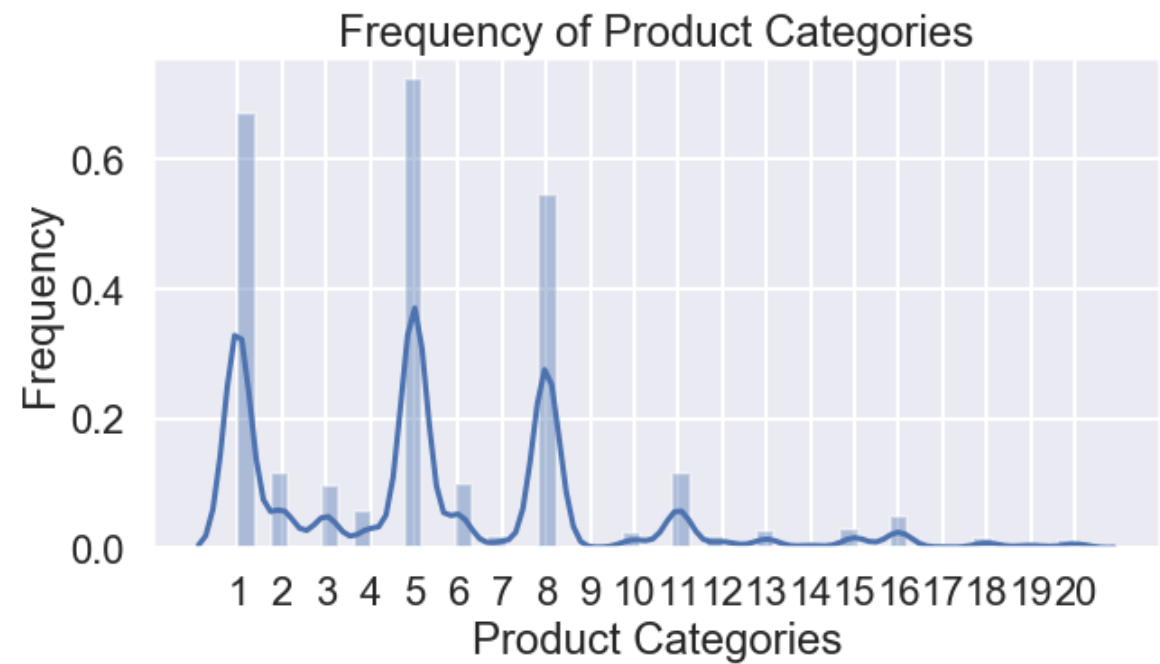
Where Do Our Customers Live?
Does Spend Vary by City Location?

	count	pct
city_category		
B	231173	0.420263
C	171175	0.311189
A	147720	0.268549

	Purchase
City_Category	
A	8911.939216
B	9151.300563
C	9719.920993

City Category B has the most amount of customers at 42%, however, City Category C spends the most with the retailer.

Which Product Category Sold The Most?

Frequency of Product Categories

Product_Category_1	Times Purchased
5	150933
1	140378
8	113925
11	24287
2	23864
6	20466
3	20213
4	11753

Top 8 categories by # of times purchased

Categories 5, 1 and 8 account for the majority of product purchases

What are the differences category prices?

Purhcase Totals by Product Category 1 Categories
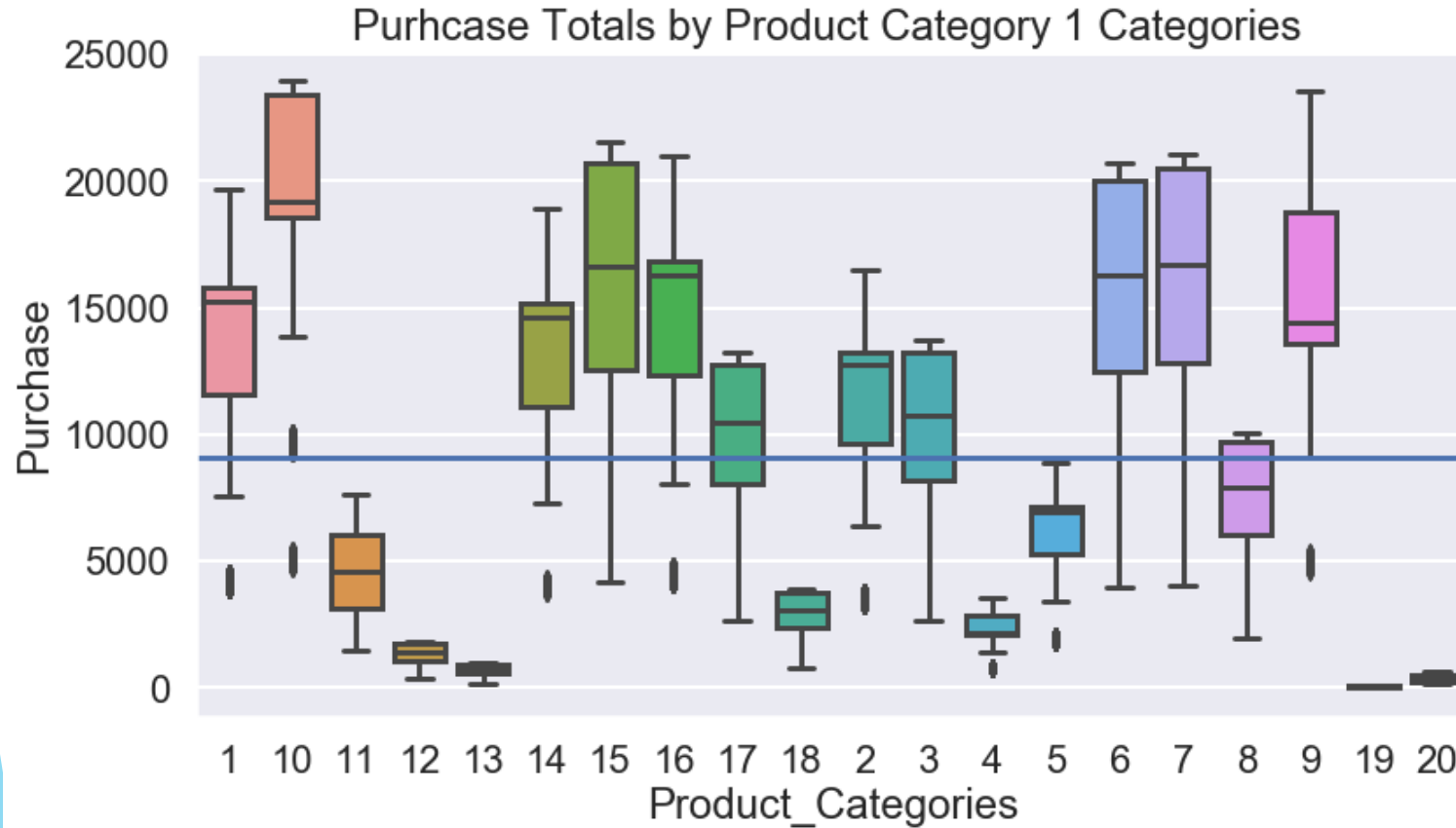
Top 5 categories
by total volume

Product_Categories	Purchase
1	1910013754
5	941835229
8	854318799
6	324150302
2	268516186

There is large variability among the prices of different product categories.

There is little variability among purchase amounts for customer demographic features. There is variability among various product categories and price. Is there a correlation between purchase categories and customer demographic features? Taking a look at the data, (the top 8 categories by count) Product Category 5 and 1 decrease while product category 8 increases, there are similar trends when grouped by Occupation and City_Category.

Frequency Of The Top 8 Categories Purchased by Age Group

There is little variability among purchase amounts for customer demographic features. There is variability among various product categories and price. Is there a correlation between purchase categories and customer demographic features? Taking a look at the data, (the top 8 categories by count) Product Category 5 and 1 decrease while product category 8 increases, there are similar trends when grouped by Occupation and City_Category.

There is little variability among purchase amounts for customer demographic features. There is variability among various product categories and price. Is there a correlation between purchase categories and customer demographic features? Taking a look at the data, (the top 8 categories by count) Product Category 5 and 1 decrease while product category 8 increases, there are similar trends when grouped by Occupation and City_Category.

Frequency Of The Top 8 Categories Purchased by Gender
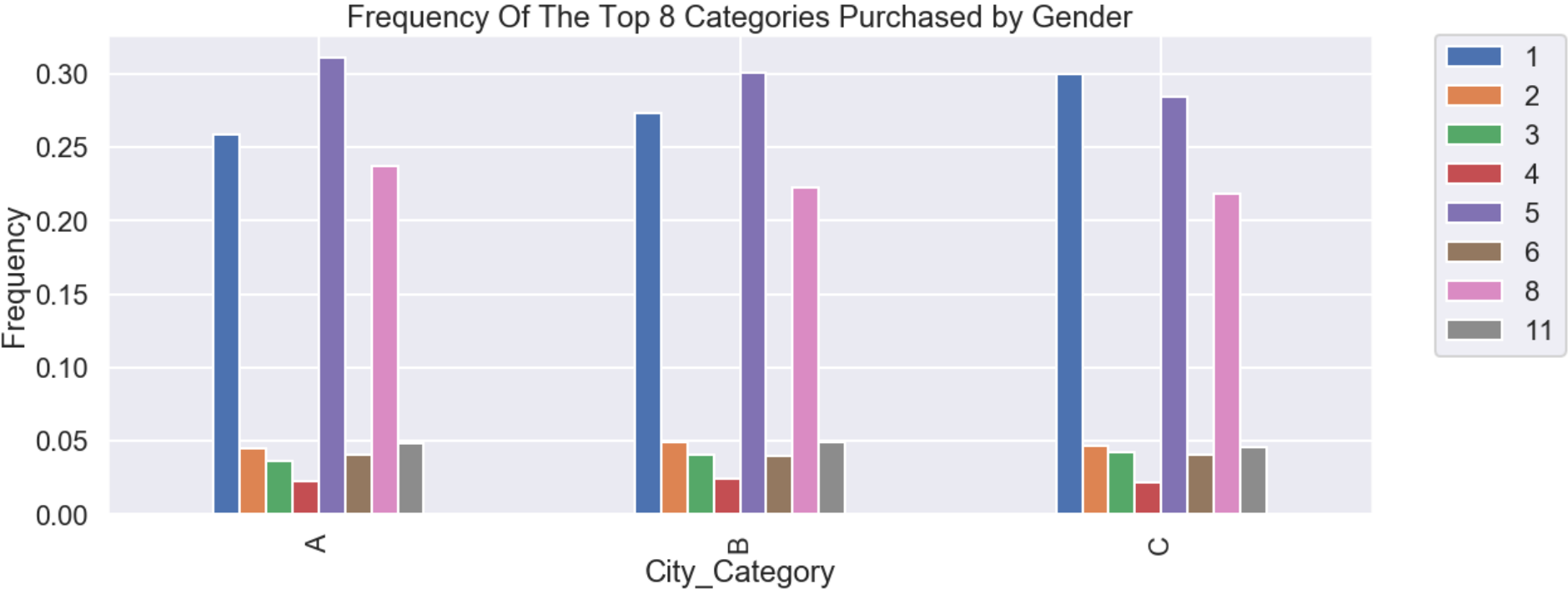
Purchase Price Correlation
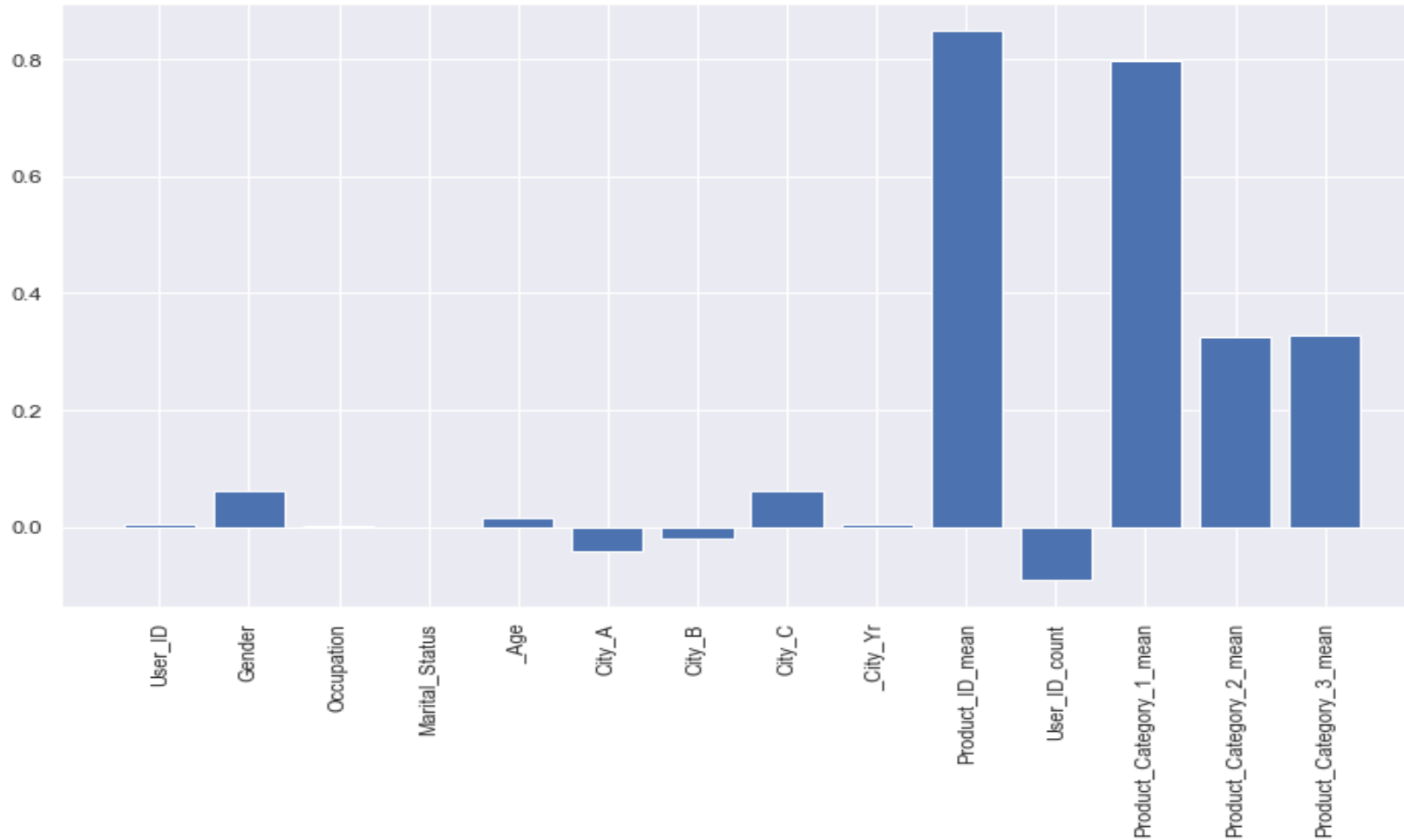
These correlation values verify that there is a high correlation of product and the purchase amount of that product

Hypothesis

I hypothesize that due to the lack of variability or correlation between the customer demographic features and the purchase amount, in order to predict purchase amounts, I will need to focus on either building new features and/or focus on using customer demographic data to predict the category a customer will purchase. The categories have a high correlation with the purchase price.