**BAIS:3200**

**Final Report**

Michael Cunningham, Broden Hall, Nikko Higgins, Ethan Siegel, Blaine Stockman

URL: <https://apex.oracle.com/pls/apex/r/consumer_behavior_dbm/e-commerce-consumer-behavior/home?session=106342834518042>

**Introduction**

Understanding why people buy things is important for businesses. It helps companies figure out what customers like, how they shop, and what makes them keep coming back. In this project, we analyzed a dataset on consumer behavior to find patterns in shopping habits, customer preferences, and what influences people’s buying decisions. This dataset includes information on things like age, income, shopping frequency, and the category of purchase. By looking at this data, we learned more about different types of customers and how businesses can improve their marketing and sales strategies. This analysis could be useful for business students, marketers, and anyone interested in how data can help companies better understand their customers.

**Data**

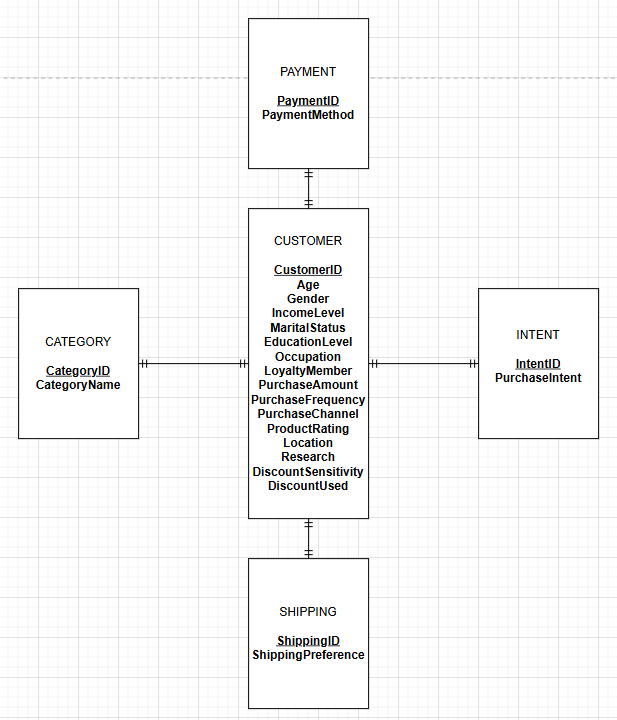
This project uses data from a Kaggle dataset[[1]](#footnote-2) on e-commerce consumer behavior. The dataset contains information about customer interactions with an e-commerce platform, capturing key aspects such as purchase history, browsing behavior, and demographic details. The original dataset includes multiple features that provide insights into consumer preferences, shopping frequency, and factors influencing purchasing decisions. For our analysis, we will focus on key factors like how consumer demographics affect shopping frequency and spending, how education level impacts repeat purchases, and which product categories generate the most revenue across customer segments. We refined the dataset by selecting variables (columns) that were most aligned with our analysis including demographics, purchase statistics, product information, and other customer preferences. Table 1 provides a description of the dataset.

Table 1 Data Dictionary

|  |  |  |
| --- | --- | --- |
| **Field** | **Type** | **Description** |
| CustomerID | Numeric | Unique identifier for each customer |
| Age | Numeric | Age of customer |
| Gender | Text | Gender of customer |
| IncomeLevel | Text | Income level of customer (low, middle, high) |
| MaritalStatus | Text | Marital statues of customer (single, married, divorced, widowed) |
| EducationLevel | Text | Customer’s highest education level (High School, Bachelor’s, Master’s) |
| Occupation | Text | Occupation of customer (High or Middle) |
| Location | Text | Location of customer |
| Research | Numeric | Time spent on research |
| DiscountSensitivity | Text | Sensity to discounts (very sensitive, somewhat sensitive, not sensitive) |
| CategoryID | Text | Unique ID for each category |
| CategoryName | Text | Category of purchased products (electronics, clothing, groceries) |
| ProductRating | Numeric | Rating of product (1-5) |
| PurchaseAmount | Numeric | Amount spent during the purchase |
| PurchaseFreq | Numeric | Number of purchases made per month |
| PurchaseChannel | Text | Purchase method (online, in-store, mixed) |
| PaymentID | Text | Unique ID for each payment type |
| PaymentMethod | Text | Method of payment used for purchase (credit card, debit card, paypal, cash, other) |
| DiscountUsed | Binary | Whether the customer used a discount (1 is true, 0 is false) |
| LoyaltyMem | Binary | Whether the customer is part of a loyalty program (1 is true, 0 is false) |
| IntentID | Text | Unique ID for each purchase intent |
| PurchaseIntent | Text | Intent behind the purchase (impulsive, planned, need-based, want-based) |
| ShippingID | Text | Unique ID for each shipping type |
| ShippingPreference | Text | Preference of shipping (standard, express, no prefernce) |

The primary entity in the database is Customer and each customer has a unique ID which was contained in the original dataset. None of the attributes are optional as the data was pulled from a server.

CUSTOMER was shown as a strong entity with a unique identifier as no customers are repeated in the dataset. INTENTION, CATEGORY, PAYMENT, and SHIPPING are all connected to the CUSTOMER entity with a 1:1 relationship that is mandatory on both sides as each CUSTOMER must have an intention, product category, payment type, and shipping preference. Each of these entities must have a customer as well. Figure 1 displays the entity relationship diagram (ERD for this dataset.



*Figure 1 Entity relationship diagram*

Upon the completion of the ERD, we normalized the tables and made the relational schema containing 5 tables. Since customerID was unique and could identify every row, transitive dependencies were present. They were resolved by creating surrogate IDs for INTENTION, CATEGORY, PAYMENT, and SHIPPING and storing them as foreign keys in the CUSTOMER table. The CUSTOMER table is the child and all the other tables are considered parent tables. Figure 2 displays the relational schema for the database.

A diagram of a company

AI-generated content may be incorrect.

*Figure 2 Graphical relational Schema*

**Database Implementation**

CREATE TABLE commands were written for each table in the relational schema. Any one of the 4 parent tables can be created first, but CUSTOMER must be created last.

CATEGORY

CREATE TABLE CATEGORY (

CategoryID CHAR(3) NOT NULL ,

Category\_Name VARCHAR2(100) NOT NULL,

Constraint Category\_PK Primary Key (CategoryID)

);

PAYMENT

CREATE TABLE PAYMENT (

PaymentID char(3) Not null,

Payment\_Method VARCHAR2(50) not null,

constraint Payment\_PK primary key (PaymentID)

);

INTENTION

CREATE TABLE INTENTION (

IntentID char(3) not null,

Purchase\_Intent VARCHAR(50) not null,

constraint INTENTION\_PK primary key (IntentID)

);

SHIPPING

CREATE TABLE SHIPPING (

ShippingID char(3) Not null,

Shipping\_Preference Varchar2(50) not null,

constraint shipping\_PK primary key (ShippingID)

);

CUSTOMER

CREATE TABLE CUSTOMER (

CustomerID Varchar2(11) Not null,

Age Number(3) Not null,

Gender varchar2(25) Not null,

Income\_Level varchar2(10) check (Income\_Level IN ('High', 'Middle')),

Marital\_Status varchar2(25) check (Marital\_Status IN ('Divorced', 'Married', 'Single', 'Widowed')),

Education\_Level varchar2(25) check (Education\_Level IN ('Bachelor''s','High School', 'Master''s')),

Occupation varchar2(10) check (Occupation IN ('Middle', 'High')),

LoyaltyMem number(1) check (LoyaltyMem IN (1,0)),

Purchase\_Amount number(\*,2) Not Null,

PurchaseFreq number (\*) not null,

Purchase\_Channel varchar2(15) check (Purchase\_Channel IN ('Mixed','In-Store', 'Online')),

Product\_Rating number(1) check (Product\_Rating between 1 and 5),

CategoryID CHAR(3) NOT NULL ,

Location varchar2(100) Not null,

PaymentID char(3) Not null,

Discount\_used number(1) check (Discount\_used IN (1,0)),

Discount\_Sensitivity varchar2(50) check (Discount\_Sensitivity in ('Somewhat Sensitive','Not Sensitive','Very Sensitive')),

IntentID char(3) not null,

ShippingID char(3) Not null,

Research number(\*,1) Not Null,

Constraint CUSTOMER\_PK primary key (CustomerID),

constraint Customer\_FK foreign key (CategoryID) references Category(CategoryID),

constraint Customer\_FK1 foreign key (PaymentID) references Payment(PaymentID),

constraint Customer\_FK2 foreign key (IntentID) references Intention(IntentID),

constraint Customer\_FK3 foreign key (ShippingID) references Shipping(ShippingID)

);

Appropriate data types and field sizes were applied to ensure data integrity. An example of a field level constraint was using CHECK to limit the field LoyaltyMem to 1 or 0 as no other values exist and we do not want to accept any values besides 0 or 1. Many other field level constraints exist in the dataset as seen above. We also ensured the field size of the IDs for each of the parent tables by using CHAR(3). This ensures the value must be 3 characters preventing any erroneous IDs of different lengths being added.

Once data was cleaned and normalized it was inserted into these tables using the “Upload a File” tool in the Object Browser of APEX. The final CATEGORY table contained 24 rows, INTENTION contained 4 rows, PAYMENT contained 5 rows, SHIPPING contained 3 rows, and CUSTOMER had 1,000 records. Example INSERT commands for each table are below:

INSERT INTO CATEGORY (CategoryID, Category\_Name) VALUES (‘C10’, ‘Gardening & Outdoors’);

INSERT INTO PAYMENT (PaymentID, Payment\_Method) VALUES (‘P02’,’ ‘Credit Card’);

INSERT INTO INTENTION (IntentID, Purchase\_Intent) VALUES (‘I02’, ‘Need-based’);

INSERT INTO SHIPPING (ShippingID, Shipping\_preference) VALUES (‘S03’, ‘No Preference’);

INSERT INTO CUSTOMER VALUES (‘37-611-6911’, 22, ‘Female’, ‘Middle’, ‘Married’, ‘Bachelor'’s’, ‘Middle’, 0, 333.8, 4, ‘Mixed’, 5, ‘C10’, ‘Ã‰vry’, ‘P02’, 1, ‘Somewhat Sensitive’, ‘I02’, ‘S03’, 2);

**Analysis**

This analysis is intended to be used by marketers and company officials to understand their target market better. By summarizing key components of their customers, companies are able to further enhance customer loyalty by aligning their strategies with their customers’ consuming patterns in mind.

Question 1: Demographics

Which consumer demographics (age) are most associated with higher purchase frequency and spending? We wrote multiple queries that would analyze several of the demographics of the consumer. The demographics we chose to investigate were age, gender, and education level.

Our strategy was to use a CASE query to analyze the age demographics. We wanted to develop buckets to classify each respondent’s age into a specific group. The buckets we used were the age ranges between 18 and 24, between 25 and 34, between 35 and 44, between 45 and 54, and 55 and over.

SELECT

CASE

WHEN Age BETWEEN 18 AND 24 THEN '18-24'

WHEN Age BETWEEN 25 AND 34 THEN '25-34'

WHEN Age BETWEEN 35 AND 44 THEN '35-44'

WHEN Age BETWEEN 45 AND 54 THEN '45-54'

ELSE '55 and over'

END AS Age\_Class,

COUNT(\*) AS Purchase\_Frequency,

TO\_CHAR(SUM(Purchase\_Amount), '$999,999,999.00') AS Total\_Amount,

TO\_CHAR(AVG(Purchase\_Amount), '$999,999,999.00') AS Avg\_Amount\_Per\_Purchase

FROM CUSTOMER

GROUP BY

CASE

WHEN Age BETWEEN 18 AND 24 THEN '18-24'

WHEN Age BETWEEN 25 AND 34 THEN '25-34'

WHEN Age BETWEEN 35 AND 44 THEN '35-44'

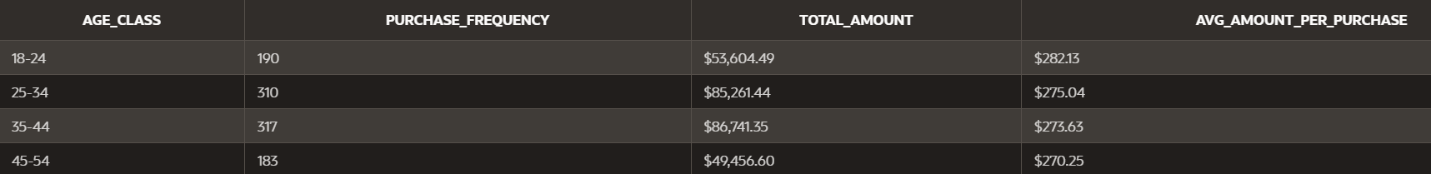
WHEN Age BETWEEN 45 AND 54 THEN '45-54'

ELSE '55 and over'

END

ORDER BY Age\_Class ASC;

The results in Figure 3 show that the consumers in the age range of 35-44 have the most purchase frequencies at 317, a total amount of $86,741.35, and an average amount per purchase at $273.63. We can see that 45-54 is the lowest with a frequency of 183, a total amount of $49,456.60, and an average amount per purchase of $270.25. Our belief is that the age classes of 45-54 and 18-24 had a lower frequency than the other classes because the current stage of life they are in. 45–54-year-olds typically are settled into their lifestyle, and individuals 18-24 years old may not be quite to that stage where they are making planned and researched purchases for themselves or their families.



*Figure 3 Age composition*

Question 2: Customer Satisfaction Score

How does customer satisfaction score impact repeat purchase behavior and customer lifetime value?

Although this question seems to have an intuitive answer, the results told a different story. We had determined the effect any given product score had on the customer’s repetition of purchase and lifetime value. The query made, provided the relevant information to display that the customer’s review of a product, did not leave them less likely to make another purchase, nor does it decrease their lifetime value.

SELECT Product\_Rating,

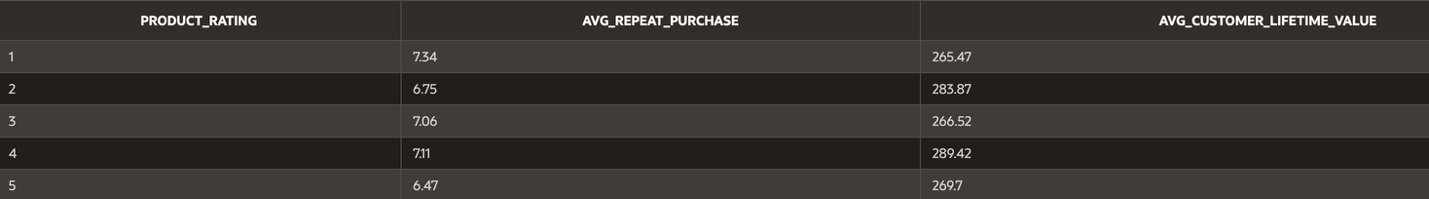
ROUND(AVG(PurchaseFreq), 2) AS Avg\_Repeat\_Purchase, ROUND(AVG(Purchase\_Amount), 2) AS Avg\_Customer\_Lifetime\_Value

FROM customer

GROUP BY Product\_Rating

ORDER BY Product\_Rating;

The following figure is the result of this query. As modeled, a poor product rating had little effect on if a customer would make a repeat purchase, as a product rating of 1 correlated to around 7.5 repeated purchases, while a product rating of 5, was at 6.5 repeated purchases. While it could be said that the customer was less likely to make a repeat purchase because they were satisfied with their original one. There were roughly equivalent lifetime value scores for customers who rated a product 1 vs 5, alluding to the idea a customer who ranked a product usable (2-4), but not excellent (5), would often repeat their purchase.

*Figure 4 Product rating and customer lifetime*

Question 3: Highest Revenue

Which product categories generate the highest revenue, and does this vary across different customer segments?

To answer this question, we first examined total purchase revenue by product category. This helped us determine which categories brought in the most revenue and were more likely to be profitable. This query calculates the total revenue for each product category across all consumers, giving a ranked list in descending order. Regardless of who is buying, this shows us revenue overall.

SELECT

c.CategoryID,

cat.Category\_Name,

SUM(c.Purchase\_Amount) AS Total\_Revenue

FROM

CUSTOMER c

JOIN

CATEGORY cat ON c.CategoryID = cat.CategoryID

GROUP BY

c.CategoryID, cat.Category\_Name

ORDER BY

Total\_Revenue DESC;

The results shown below in Figure 5 are the results of the query. As we can see, Jewelry & Accessories brought in the most revenue at $15,139 followed by categories Sports & Outdoors at $14,610, and Electronics at $13,842. As this displays the products that bring in the most revenue, it does not factor in the average cost of items in the categories, making it difficult to determine profitability. With product knowledge or data on cost, it would lead us to certainty about which category is the most profitable. It might be safe to assume that the cost on Sports and Outdoors could be cheaper than Jewelry & Accessories, leading the second in revenue to be most profitable but based on our question this report will satisfy.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 5 Revenue per category*

We presented the values of the query in a bar chart visualizing total revenue per category in the APEX web application (Figure 5)

After analyzing revenue by category, we wanted to see the revenue breakdown additionally by gender and age group to reflect lifestyle choices.

This query breaks down the total revenue by both customer gender and age group, using a CASE statement to create age brackets. This lets us see which segments (e.g., males aged 25-34) are spending the most in each category.

SELECT Gender,

CASE

WHEN Age < 25 THEN 'Under 25'

WHEN Age BETWEEN 25 AND 34 THEN '25-34'

WHEN Age BETWEEN 35 AND 44 THEN '35-44'

WHEN Age BETWEEN 45 AND 54 THEN '45-54'

ELSE '55 and over'

END AS Age\_Class,

Category\_Name,

SUM(Purchase\_Amount) AS Total\_Revenue

FROM CUSTOMER c

JOIN CATEGORY cat ON c.CategoryID = cat.CategoryID

WHERE Purchase\_Amount > (SELECT AVG(Purchase\_Amount) FROM CUSTOMER)

GROUP BY Gender,

CASE

WHEN Age < 25 THEN 'Under 25'

WHEN Age BETWEEN 25 AND 34 THEN '25-34'

WHEN Age BETWEEN 35 AND 44 THEN '35-44'

WHEN Age BETWEEN 45 AND 54 THEN '45-54'

ELSE '55 and over'

END,

Category\_Name

ORDER BY Gender, Age\_Class, Total\_Revenue DESC;

Looking at our results we can determine the age and gender distribution and association with online purchase categories. We clearly see Packages for females aged 25-34 spending the most at $3,672 followed by Sports & Outdoors male under 25 at $3,161. Insights gained from this report tell us that young men 25 and under like to purchase products online related to outdoor or sporting activities and females aged 35-44 spent $2,948.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 6 Category and age distribution*

To discontinue stating obvious results, what an online retailer would do with these query results is determine which target demographic to advertise a specific purchase category too. With modern advertising techniques it would be costly and inefficient to advertise Baby Products or Jewelry & Accessories to males aged 25 and younger per most circumstances. The wisest option is to match targeted ads of a category’s items to its most profitable demographics.

Question 4: Preferred Shopping Channel

What is the preferred shopping channel (online vs. in-store) for different customer groups, and how does this affect their average order?

To determine which shopping channel is preferred, we first grouped customers by Purchase\_Channel and counted how many customers used each one. This revealed the most common choice between online, in-store, or mixed channels.

SELECT

Purchase\_Channel,

COUNT(CustomerID) AS Total\_Customers

FROM

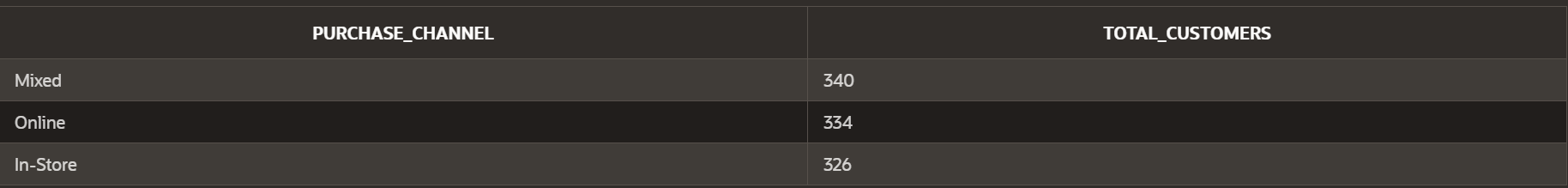
CUSTOMER

GROUP BY

Purchase\_Channel

ORDER BY

Total\_Customers DESC;



*Figure 7 Purchase channels*

Next, we analyzed how average order size varied across Gender and Income\_Level for each channel. We used a GROUP BY in both queries and calculated AVG(Purchase\_Amount) to evaluate spending behavior.

SELECT

Purchase\_Channel,

Gender,

Income\_Level,

ROUND(AVG(Purchase\_Amount), 2) AS Avg\_Order\_Amount

FROM

CUSTOMER

GROUP BY

Purchase\_Channel, Gender, Income\_Level

ORDER BY

Purchase\_Channel, Avg\_Order\_Amount DESC;

A black and white screen with white text

AI-generated content may be incorrect.

*Figure 8 Gender, income level, and average order amount*

These findings suggest that income alone does not fully explain purchase behavior — certain gender identities correlated with significantly higher spending in the in-store channel, particularly in the high- and middle-income tiers. This may reflect unique shopping preferences, loyalty behaviors, or targeted marketing effectiveness among these groups.

Notably, binary genders (male/female) had lower average in-store order amounts compared to several non-binary identities. This insight indicates that diverse gender groups may represent high-value customer segments in the in-store channel — a useful takeaway for brands considering inclusive customer experiences or marketing efforts.

Question 5: Education

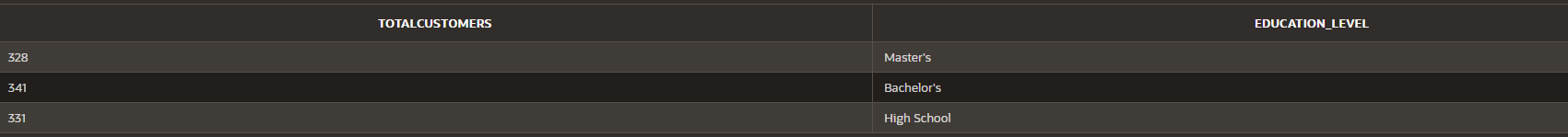
What percentage of customers have a higher degree? Is there a correlation between having a degree and being a loyalty member? What effect might this have on the purchase frequency?

SELECT count(customerID) as TotalCustomers, Education\_level

from customer

group by education\_level;

The results of the query are shown in Figure 9. The data set appears to consist of approximately 1/3 of each of the education levels. This was a surprising result as in general, fewer people tend to be college graduates and even fewer hold a graduate degree.



*Figure 9 Customer education levels*

Higher education levels generally correspond to higher earning potential. To investigate the effect education level had on being a loyalty member we wrote another query seen below.

SELECT

ROUND(

(SELECT COUNT(CustomerID)

FROM customer

WHERE LoyaltyMem = 1 AND Education\_Level = 'High School')

/ (SELECT COUNT(CustomerID)

FROM customer

WHERE Education\_Level = 'High School') \* 100, 2

) || ' %' AS LoyaltyWithHighSchool,

ROUND(

(SELECT COUNT(CustomerID)

FROM customer

WHERE LoyaltyMem = 1 AND Education\_Level = 'Bachelor''s')

/ (SELECT COUNT(CustomerID)

FROM customer

WHERE Education\_Level = 'Bachelor''s') \* 100, 2

) || ' %' AS LoyaltyWithBachelors,

ROUND(

(SELECT COUNT(CustomerID)

FROM customer

WHERE LoyaltyMem = 1 AND Education\_Level = 'Master''s')

/ (SELECT COUNT(CustomerID)

FROM customer

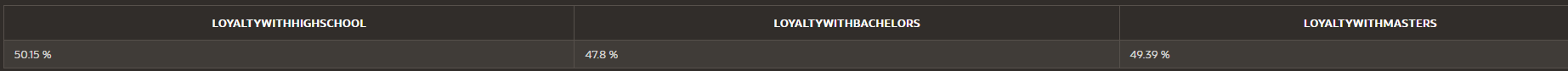
WHERE Education\_Level = 'Master''s') \* 100, 2

) || ' %' AS LoyaltyWithMasters

FROM customer

group by 'All';

The results of the query can be seen in Figure 10. A higher percentage of high school graduates (50.15%) are loyalty members followed by those with a graduate level degree (49.39%). 47.8% of those with a bachelor’s degree are loyalty members. This is indicative that the marketing department should target those with a high school diploma when organizing loyalty member campaigns.



*Figure 10 Loyalty member percentages based on education levels*

Further analysis into monthly purchase frequency would also prove beneficial to the marketing department as repeat customers are more likely to generate revenue for the company. To better understand the target market a query was developed to see which loyalty members (grouped by education) had the highest average purchase frequency.

SELECT

Education\_Level,

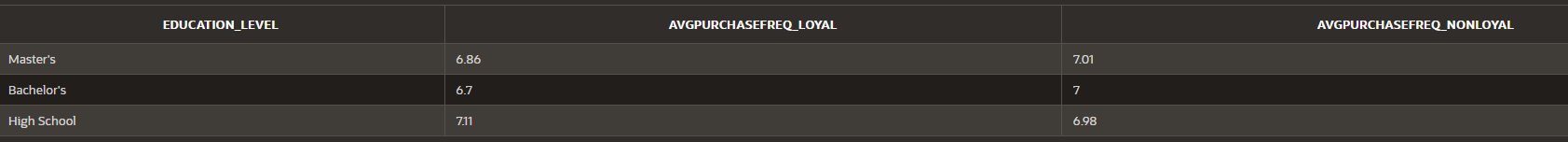
ROUND(AVG(CASE WHEN LoyaltyMem = 1 THEN PurchaseFreq END), 2) AS AvgPurchaseFreq\_Loyal,

ROUND(AVG(CASE WHEN LoyaltyMem = 0 THEN PurchaseFreq END), 2) AS AvgPurchaseFreq\_NonLoyal

FROM customer

GROUP BY Education\_Level;

The results of the query can be seen in Figure 11. The high school category of loyalty members only had a higher purchase frequency (7.11 times a month) compared to non-loyalty high school graduates (6.98). For all other education levels, non-loyalty members had a higher purchase frequency in comparison with their loyalty member counterparts. Loyalty members with a bachelor’s degree had the lowest average monthly purchase frequency out of any category at 6.7 purchases/month.



*Figure 11 Average purchase frequency by education level and loyalty status*

The findings suggest that the marketing team should target high school graduates when developing marketing campaigns for their loyalty program. If they wish to gain a bigger market share, they have plenty of ground to gain when it comes to those with a bachelor’s degree as it appears those consumers do not take advantage of the loyalty program as much compared to other educational groups.

**Web Design**

Homepage

The home page of the web application includes a summary of the project with a hyperlink to the original data source, a main navigation menu featuring nested elements and custom icons, and an image of someone online shopping created by The Financial[[2]](#footnote-3). To maintain visual consistency and highlight key content, a unified accent color is applied throughout the application. The navigation icons are tailored to reflect the type of analysis on each page, and a dropdown menu has been implemented to organize all data tables under a single section. Figure 12 shows a screenshot of the homepage and its navigation menu.

A person using a computer

AI-generated content may be incorrect.

*Figure 12 Home page*

Tables

We developed an interactive report for each database table, enabling users to search, filter, and group the data (Figures 13-17). The column headings and number formats have been customized to make sure the data is easily interpretable. Each page includes a brief description of the data contained in the tables.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 13 CUSTOMER*

*A screenshot of a computer

AI-generated content may be incorrect.*

*Figure 14 CATEGORY*

*A screenshot of a computer

AI-generated content may be incorrect.*

*Figure 15 PAYMENT*

*A screenshot of a computer

AI-generated content may be incorrect.*

*Figure 16 INTENTION*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 17 SHIPPING*

Queries

We presented the results of our analysis on consumer behavior in e-commerce on a single page. The first research question regarding the demographics of consumers and their association with higher purchase frequency and spending was addressed on a single page (Figure 18). This page includes a bar chart visualizing the purchase frequency by age class, and a table showing the total amount spent and average amount per purchase by age class. For clarity, we added a text box explaining the motivating research questions and results. The analysis indicates that consumers in the age class of 35-44 are associated with the highest purchase frequencies and spending.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 18 Demographics of consumers*

The second research question regarding the impact of customer satisfaction scores on repeat purchase behavior and customer lifetime value was addressed on a single dashboard-style page. This page includes a bar chart showing the effect of product ratings on customer lifetime value, and another bar chart presenting the relationship between product ratings and repeated purchases.

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 19 Customer satisfaction score*

Similarly, the third research question analyzes highest revenue generated by product categories was addressed on a single dashboard-style page. This page includes a bar chart showing the total revenue by product category, and another bar chart presenting the revenue by product category and age group.

A screenshot of a graph

AI-generated content may be incorrect.

*Figure 20 Highest revenue*

The fourth research question about the preferred shopping channel (online vs. in-store) for different customer groups and its impact on their average order was addressed on a single dashboard-style page. This page includes a bar chart showing the customer count by shopping channel, and a table presenting the average order amount by channel, gender identity, and income level.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 21 Preferred shopping channel*

The final research question examined the impact of education levels on customer behavior was addressed. This page includes a pie chart showing the distribution of education levels among customers, a table presenting the loyalty member percentage across different education levels, and a bar chart displaying the average purchase frequency by education level and loyalty status.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure 22 Education*

1. <https://www.kaggle.com/datasets/salahuddinahmedshuvo/ecommerce-consumer-behavior-analysis-data?resource=download> [↑](#footnote-ref-2)
2. <https://finchannel.com/e-commerce-continues-to-gain-ground-in-technical-consumer-goods-markets/75413/business-2/2018/09/> [↑](#footnote-ref-3)