Consumer Behavior and Shopping Habits

Data Analytics with Python

Team 2: Dhvanil, Hardik, Harini, Sharath

Introduction

Overview of the Dataset: An extensive compilation of data offering insights into consumer preferences, shopping patterns, and decision-making processes.

Key Components:

- **Demographics :** Age and gender.
- **Purchase History :** Past purchases, Highlighting Trends and Customer Loyalty.
- **Product Preferences:** Data on preferred products, understanding market demand.
- **Shopping frequency:** How often consumer shop (annual , Bi-weekly, quarterly).
- **Seasonal trends :** Influences the buying pattern (Fall, winter, spring & summer).
- **Discounts**: Impact of discounts on sales volume,various promotional strategies.

Data Preparation and Cleaning

Few steps which we took to clean our dataset:

1. Renaming original columns:

Original:

```
df.columns
```

```
Index(['Age', 'Gender', 'Item Purchased', 'Category', 'Purchase Amount (USD)',
       'Location', 'Size', 'Color', 'Season', 'Review Rating',
       'Subscription Status', 'Shipping Type', 'Discount Applied',
       'Promo Code Used', 'Previous Purchases', 'Payment Method',
       'Frequency of Purchases'],
      dtype='object')
```

New:

```
df.columns = ['Age',
              'Gender',
              'Item_Purchased',
              'Category',
              'Purchase_Amount_(USD)',
              'Location',
              'Size',
              'Color',
              'Season',
              'Review_Rating',
              'Subscription_Status',
              'Shipping_Type',
              'Discount_Applied',
              'PromoCode_Applied',
              'Previous_Purchases',
              'Payment_Method',
              'Purchase_Frequency']
```

2. Replacing value of ‘Yes’ and ‘No’ with 1 and 0 respectively:

```
df.Subscription_Status.replace({'Yes': 1, 'No': 0}, inplace=True)

df.Discount_Applied.replace({'Yes': 1, 'No': 0}, inplace=True)

df.PromoCode_Applied.replace({'Yes': 1, 'No': 0}, inplace=True)

df
```

3. Dropping “PromoCode_Applied” column, as Discount will be applied if and only if promocode is applied.

4. Introducing new column (“Age_Category”) on the basis of age:

a. 12 <= age <= 27 : Gen-Z
b. 28 <= age <= 43 : Millennials
c. 44 <= age <= 59 : Gen-X
d. age > 59 : Boomers

5. Introducing new column (“Amount_Category”) on the basis of Purchase_Amount:

a. 0 <= amount <= 30 : Cheap
b. 31 <= amount <= 60 : Average
c. amount > 60 : Expensive

6. Introducing new column (“Rating_Category”) on the basis of Review_Rating:

a. 0 <= rating <= 2.5 : Poor
b. 2.6 <= rating <= 3.5 : Fair
c. rating > 3.5 : Excellent

Findings

Finding #1: Impact of Discounts and Promotions on Sales

Finding #2: Product Preferences and Seasonality Effects

Finding #3: Customer Loyalty and Subscription Services

- Predictive Analytics for Customer Subscription Behavior: A Decision Tree Approach
- Regression

Finding #4: Demographic Influence on Purchasing Pattern

Finding #1 Impact of Discounts and Promotions on Sales

```
count    1677.000000
mean       59.279070
std        23.610697
min        20.000000
25%        38.000000
50%        60.000000
75%        80.000000
max       100.000000
Name: Purchase_Amount_(USD), dtype: float64
```

```
num_transactions_discount
```

```
average_amount_with_discount      59.27906976744186

average_amount_without_discount   60.13045434098656
```

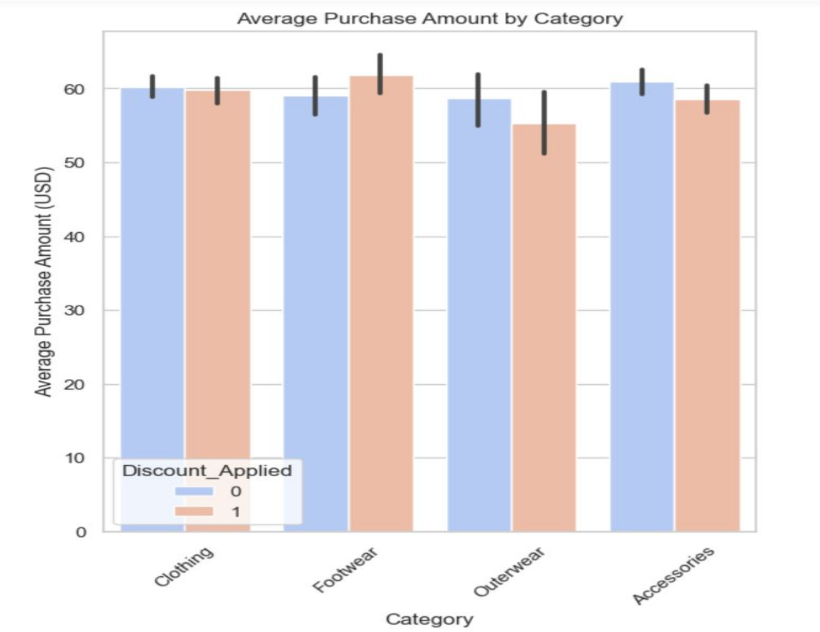
```
count    2223.000000
mean       60.130454
std        23.740327
min        20.000000
25%        39.000000
50%        60.000000
75%        81.000000
max       100.000000
Name: Purchase_Amount_(USD), dtype: float64
```

```
num_transactions_no_discount
```

Transactions with Discounts: There were 1,677 transactions with discounts applied. The average purchase amount for these transactions was approximately 59.28 ,with a standard deviation of about 23.61 . The minimum and maximum purchase amounts were 20 and 100, respectively.

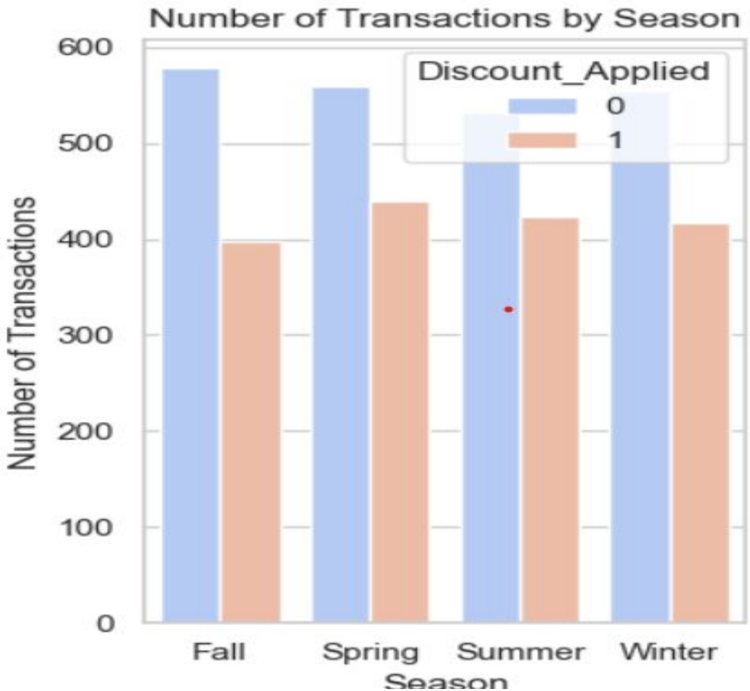
Transactions without Discounts: There were 2,223 transactions without any discounts applied. The average purchase amount for these transactions was slightly higher, at approximately 60.13, with a standard deviation of about 23.74. The minimum and maximum purchase amounts were 20 and 100, respectively.

1. Average Purchase Amount by Category



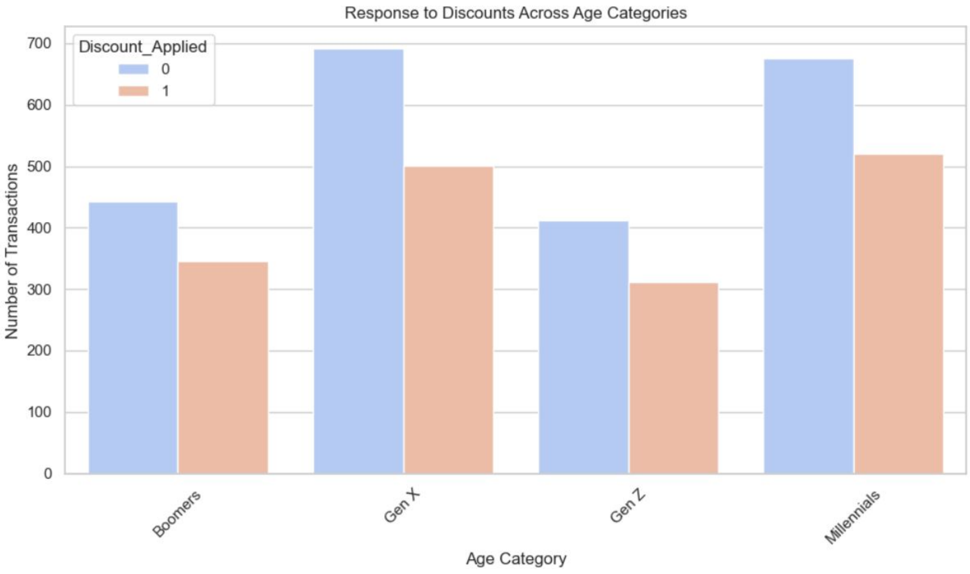
The chart compares the average purchase amount across categories with and without discounts, highlighting product sensitivity to price changes and identifying potential areas for promotional strategy optimization.

2. Number of Transactions by Season



The second chart illustrates the number of transactions in each season, comparing those with and without discounts. Seasonal trends in discount responsiveness can inform targeted promotional campaigns, helping to optimize sales during specific periods.

3. Response to Discounts Across Age Categories



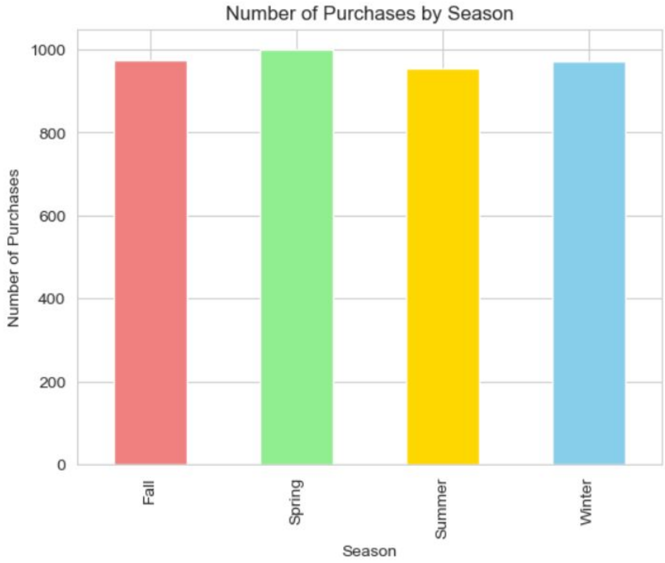
The third chart shows the response to discounts across different age categories. It highlights which age groups are more likely to engage with discounts, suggesting targeted marketing strategies that could resonate well with specific demographic segments.

Managerial Insights

- **Discount Impact:** The focus could be on strategically offering discounts to either clear inventory or attract new customers rather than increasing the transaction value.
- **Customer Segmentation:** Analyzing customer responsiveness to discounts based on segments such as age, gender, or previous purchase history could help tailor more effective promotional strategies.
- **Optimizing Discount Strategy:** It could be beneficial to explore different types of discounts to see which ones have a greater impact on both the volume of purchases and the average purchase amount.

Finding #2 Product Preferences and Seasonality Effects

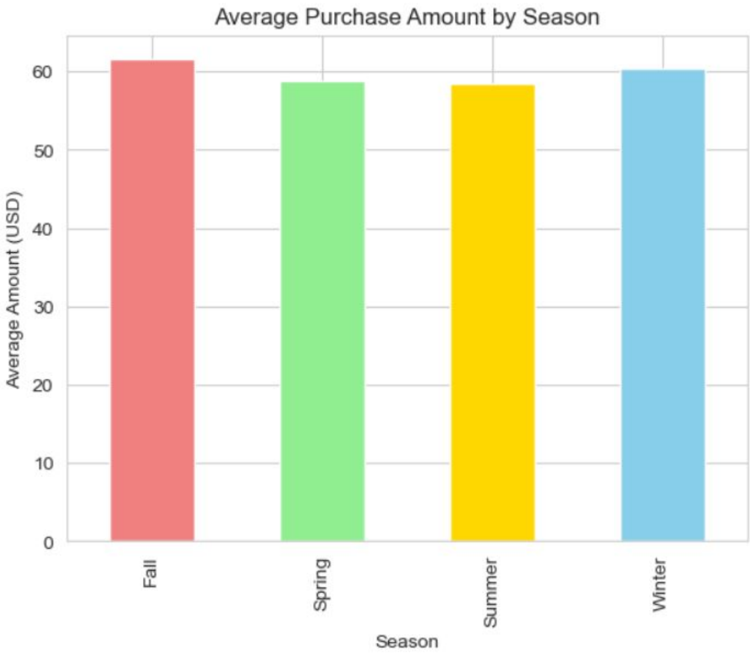
1. Purchases by Season




```
Season
Spring    999
Fall      975
Winter    971
Summer    955
Name: count, dtype: int64
```

This indicates a slightly higher number of purchases in Spring, suggesting a potential increase in consumer buying behavior during this season.

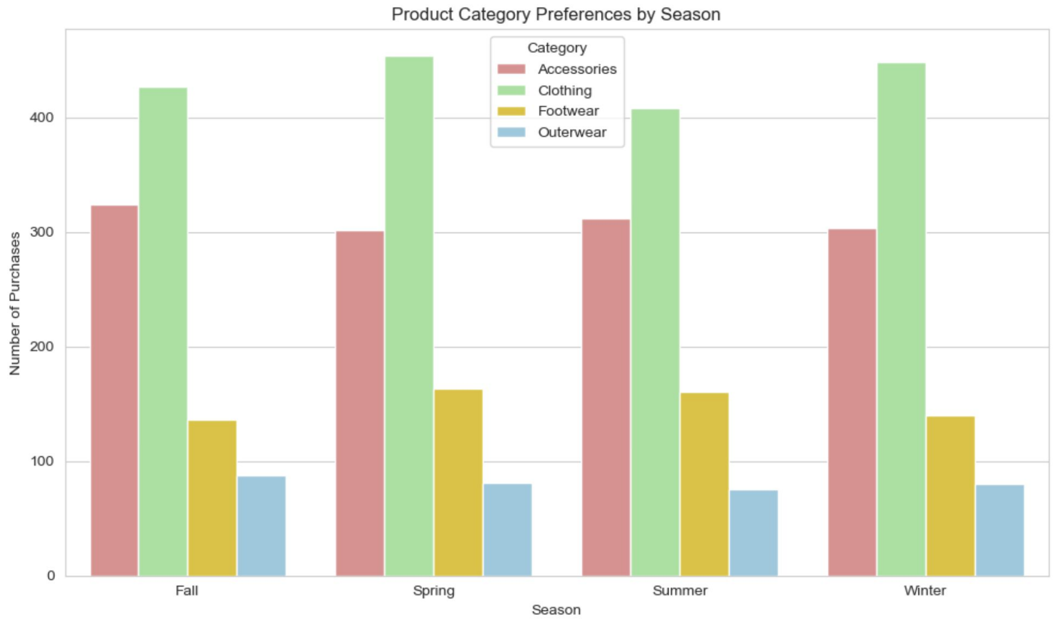
2. Average Purchase Amount by Season (USD):




```
Season
Fall      61.556923
Spring    58.737738
Summer    58.405236
Winter    60.357364
Name: Purchase_Amount_(USD), dtype: float64
```

The average purchase amount is slightly higher in Fall and Winter, indicating that consumers may spend more during these seasons.

3. Product Category Preferences by Season:



	Season	Category	Number_of_Purchases
0	Fall	Accessories	324
1	Fall	Clothing	427
2	Fall	Footwear	136
3	Fall	Outerwear	88
4	Spring	Accessories	301
5	Spring	Clothing	454
6	Spring	Footwear	163
7	Spring	Outerwear	81
8	Summer	Accessories	312
9	Summer	Clothing	408
10	Summer	Footwear	160
11	Summer	Outerwear	75
12	Winter	Accessories	303
13	Winter	Clothing	448
14	Winter	Footwear	140
15	Winter	Outerwear	80

Clothing is consistently the most purchased category across all seasons, with Spring (454 purchases) and Winter (448 purchases) being the top seasons.

Accessories and Footwear also show seasonal variations, with Accessories being slightly more popular in Fall (324 purchases)

Outerwear purchases are relatively lower compared to other categories, with the highest number of purchases in Fall (88 purchases).

Managerial insights

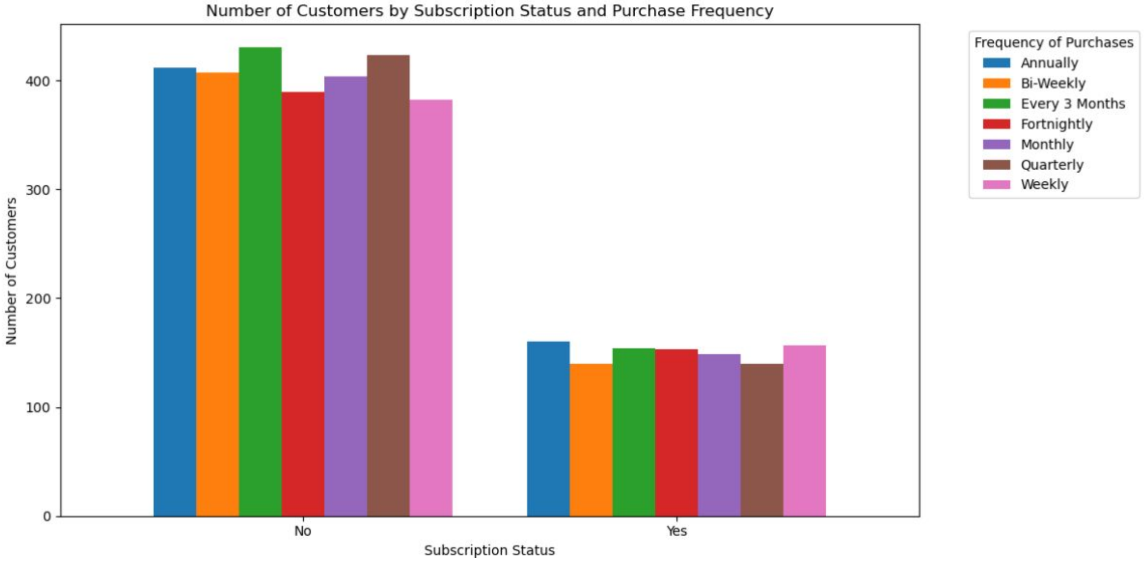
- **Leverage Seasonal Peaks in Consumer Purchasing**
  Insight: There is a noticeable increase in purchases during Spring.
  Action: Launch marketing campaigns and promotions specifically tailored for the Spring season. Consider introducing new product lines or seasonal collections to capitalize on this peak buying period.

- **Optimize Pricing and Promotions Based on Seasonal Spending Patterns**
  Insight: There is higher spending in Fall and Winter.
  Action: Implement dynamic pricing strategies to adjust prices during high spending seasons. Plan high-value promotions or bundled offers during Fall and Winter to increase average purchase amounts.

- **Tailor Product Offerings and Stock to Seasonal Preferences**
  Insight: Preferences for product categories vary by season, with Clothing being the dominant category. Accessories and Footwear also show significant seasonal preferences.
  Action: Adjust inventory levels and product displays to match seasonal preferences. Increase the stock of Clothing, especially during Spring and Winter when purchases peak. Develop targeted advertising for Accessories and Footwear during seasons they are most popular, encouraging consumers to explore these categories.

Finding #3 Customer Loyalty and Subscription Services

1. Purchase frequency vs Subscription Status

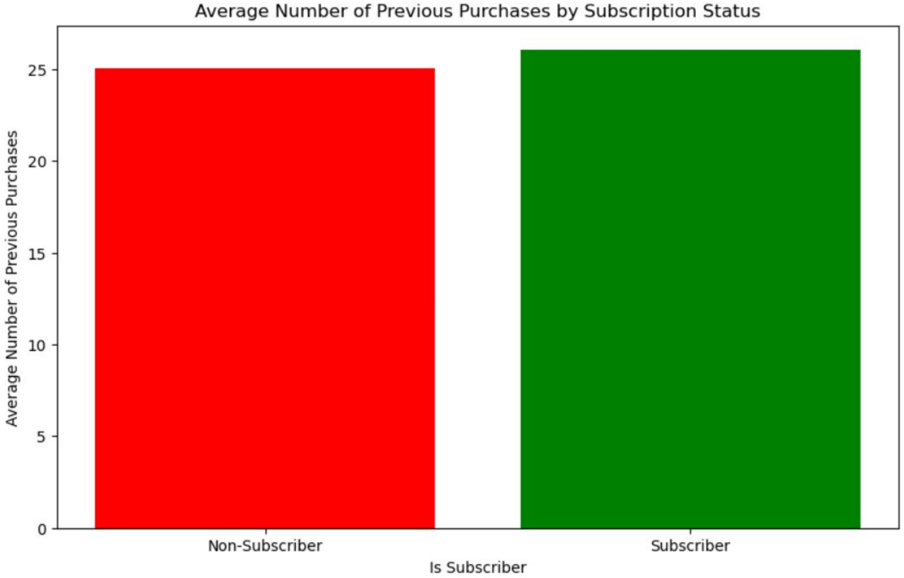


Subscription Status	
No	2847
Yes	1053

	Purchase Frequency	Percentage
0	Every 3 Months	14.974359
1	Annually	14.666667
2	Quarterly	14.435897
3	Monthly	14.179487
4	Bi-Weekly	14.025641
5	Fortnightly	13.897436
6	Weekly	13.820513

Non-subscribers have a higher count of customers making frequent purchases every 3 months than subscribers.

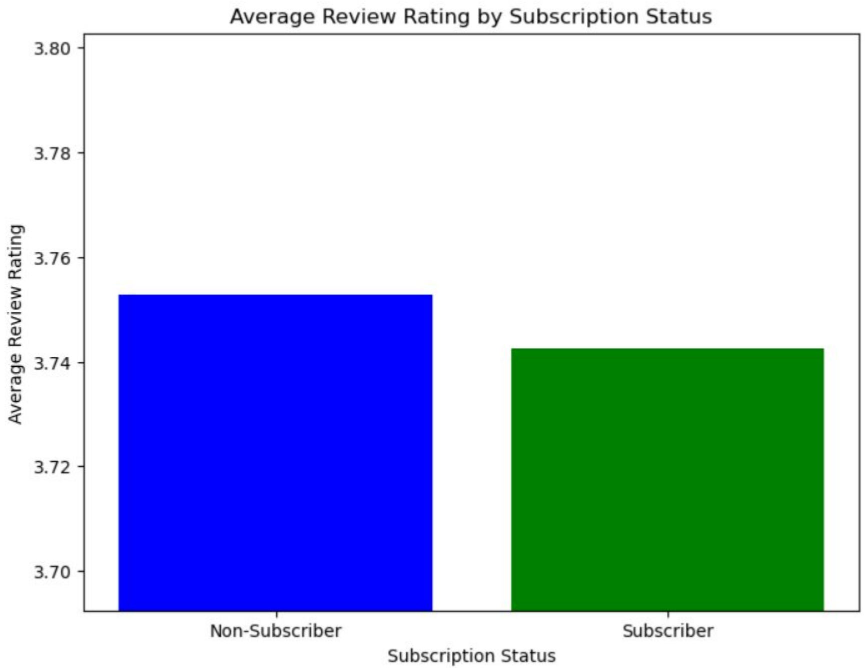
2. Average number of previous purchase by Subscription Status




```
Average Number of Previous Purchases:
Is_Subscriber
0    25.080436
1    26.084520
```

Subscribers have a higher average number of previous purchases compared to Non-Subscribers, as indicated by the taller green bar in the graph.

3. Average review rating by Subscription Status




```
Average Review Rating:
Is_Subscriber
0    3.752722
1    3.742450
```

Non-Subscribers have a marginally higher average rating compared to Subscribers, based on the height of the blue bar being slightly taller than the green bar.

Managerial Insights

- Introduce special offers for monthly subscription plans.
- Create a loyalty scheme that recognizes and rewards the consistency and frequency of purchases.
- Engage with your most loyal non-subscribed customers by offering promotional deals.

Machine Learning

- Predictive Analytics for Customer Subscription Behavior: A Decision Tree Approach

```
▾            DecisionTreeClassifier
DecisionTreeClassifier(random_state=42)
```

```
Accuracy: 0.7782051282051282

Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.84      0.84       554
           1       0.62      0.62      0.62       226

    accuracy                           0.78       780
   macro avg       0.73      0.73      0.73       780
weighted avg       0.78      0.78      0.78       780


Confusion Matrix:
[[468  86]
 [ 87 139]]
```

Accuracy (0.768 or 76.8%): This is the overall percentage of correct predictions made by the model out of all predictions. An accuracy of 76.8% means that, on average, the model correctly predicts whether a customer will subscribe or not about 77 times out of 100.

Managerial Insights

1. Targeting and Efficiency: The model is more precise in predicting non-subscribers than subscribers. While it's effective at identifying those unlikely to subscribe, there's room to improve in accurately identifying potential subscribers. Marketing efforts could be adjusted to better target the latter group.

2. Opportunities for Improvement: The relatively lower recall for subscribers suggests that the model (and potentially marketing strategies) might be missing out on a significant portion of potential subscribers. Initiatives to better understand and engage this segment could lead to higher conversion rates.

3. Balancing Precision and Recall: Depending on business objectives, it might be desirable to adjust the model to prioritize recall (capturing as many actual subscribers as possible) even at the expense of precision (accepting more false positives), especially if the cost of false positives is low compared to the benefit of acquiring a new subscriber.

4. Strategic Adjustments: The insights from the confusion matrix can guide strategic adjustments. For instance, reducing false negatives (missed subscribers) might be a priority if the lifetime value of a subscriber is high. This could involve reevaluating the features used, the model itself, or the threshold for classifying a prediction as a subscriber.

5. Resource Allocation: Understanding which areas the model performs well in and where it struggles can help in allocating resources more efficiently. Efforts could be focused on improving model performance where it has the most significant impact on business outcomes.

Managerial Insights

6. Customized Communication: Insights from model performance can help tailor communication strategies to different customer segments based on their predicted subscription likelihood, optimizing marketing spend and messaging for maximum impact.

7. Feature Importance: Understanding which features are most influential in predicting subscription status can help focus marketing and product development efforts. For instance, if Review_Rating and Previous_Purchases are significant predictors, efforts to improve customer satisfaction and encourage repeat purchases could be beneficial.

8. Customer Segmentation: The model can help identify characteristics of customers who are more likely to subscribe. These insights can be used to tailor marketing strategies to target similar customer profiles more effectively.

9. Subscription Drivers: Analyzing the decision tree can provide insights into the main drivers behind subscription decisions. For example, if customers with a high number of Previous_Purchases and a positive Review_Rating are more likely to subscribe, these factors could be leveraged in promotional strategies.

10. Improvement Areas: The classification report and confusion matrix can highlight areas where the model (and potentially the business) struggles to predict accurately. For example, if there's a high rate of false negatives (customers predicted not to subscribe but who do), it might indicate missed opportunities that can be addressed through targeted campaigns or offers.

Managerial Insights

11. Data-Driven Decision Making: The overall accuracy and detailed metrics provide a baseline for data-driven decision-making. By understanding the model's strengths and limitations, management can make informed decisions about resource allocation, marketing strategies, and product development priorities.

12. Personalization: Insights from the model can inform personalized engagement strategies, targeting users with specific characteristics or behaviors with tailored messages, offers, and products likely to drive subscription.

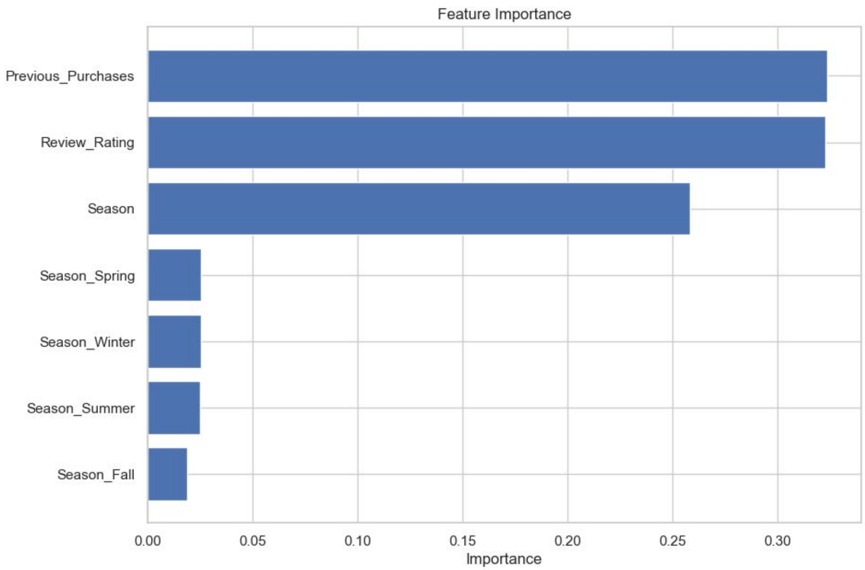
By leveraging the insights gained from the model, management can make more informed decisions that align with the goal of increasing subscription rates and improving customer satisfaction.

By closely examining both the overall and detailed performance of the model, management can derive actionable insights to refine marketing strategies, enhance customer targeting, and ultimately drive better business outcomes.

Machine Learning

- Regression

```
▼ RandomForestRegressor
RandomForestRegressor(random_state=42)
```



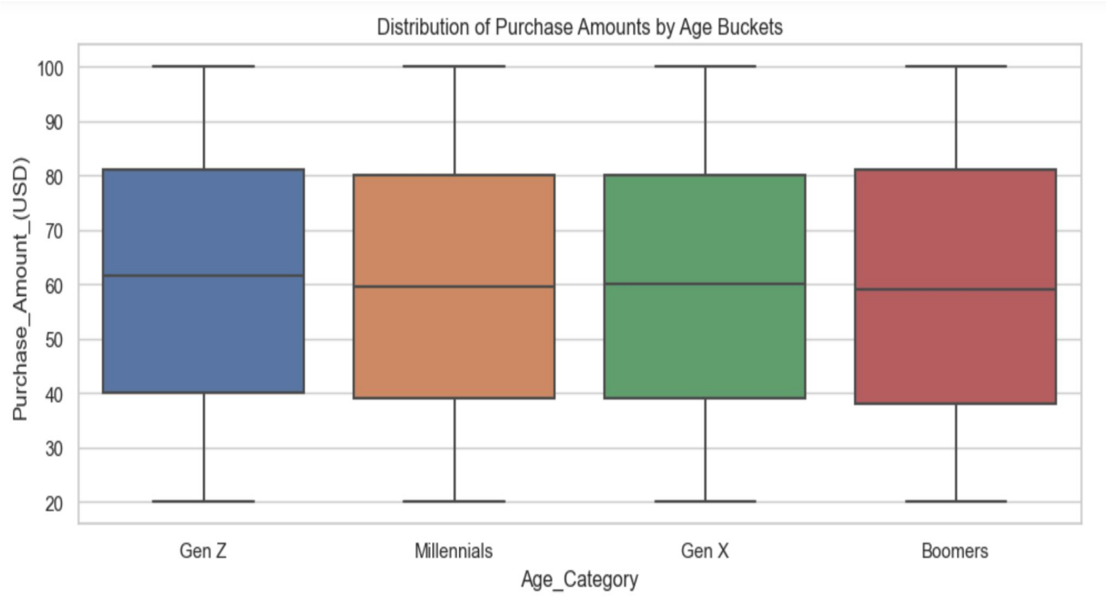
Machine Learning

- The Mean Squared Error (MSE) is 631.0077, which is the average of the squares of the errors between what the model predicted and the actual values. Since Purchase_Amount_(USD) is in dollars, this value can be quite high, depending on the scale of purchase amounts in our dataset.
- The R^2 Score is -0.1276, which is a measure of the proportion of variance for the dependent variable that's explained by the independent variables in the model. An R^2 score of 1 means the model explains all the variability of the response data around its mean. In our case, a negative R^2 score suggests that the model is performing worse than a simple model that would predict the mean Purchase_Amount_(USD) for all observations. This could happen if the model is overfitting on the training data or if important predictive features are missing.

Managerial Insights

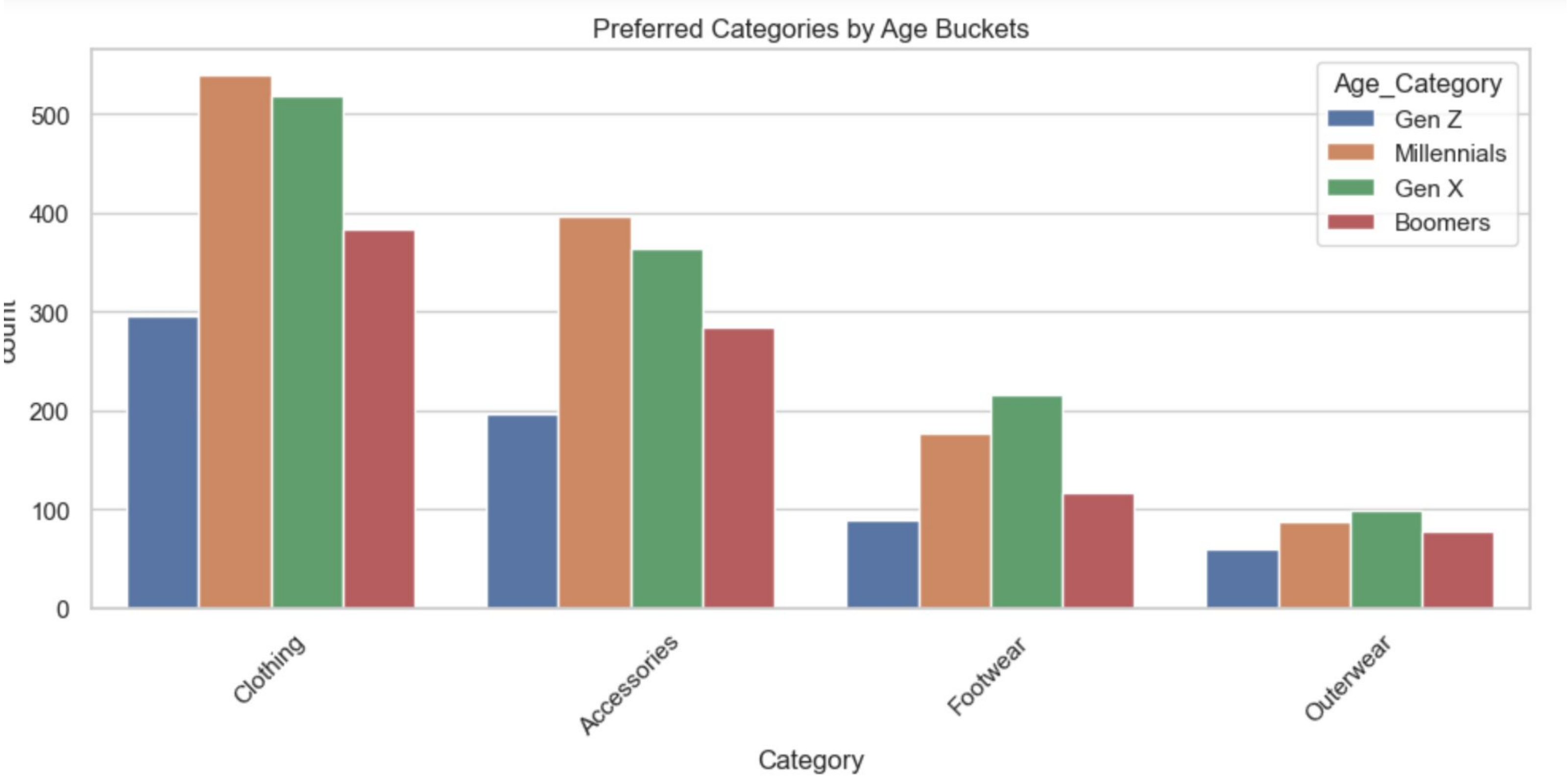
- Given that Review_Rating is a strong predictor, it implies customer satisfaction is highly linked to spending behavior. Improving customer service and product quality to boost satisfaction could increase spending.
- Since Previous_Purchases is also a significant predictor, creating targeted marketing strategies for repeat customers, such as loyalty programs or personalized discounts, could encourage more spending.
- As the seasonal features are less impactful, focusing marketing efforts based on seasonality might not yield as significant a return compared to efforts made to improve overall customer satisfaction and retention.

Finding #4 Demographic Influence on Purchasing Pattern

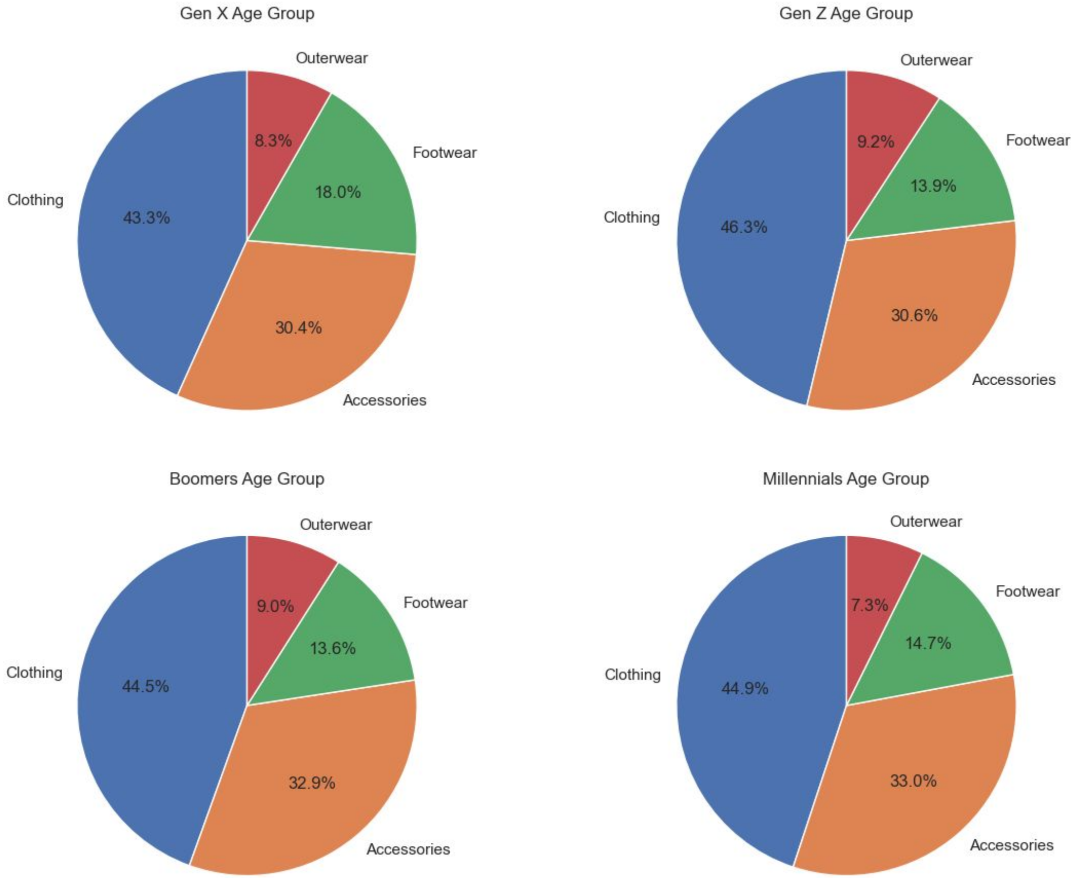


	Age_Category	max	min	mean	median
0	Gen Z	100	20	60.69	61.5
1	Millennials	100	20	59.39	59.5
2	Gen X	100	20	59.92	60.0
3	Boomers	100	20	59.38	59.0

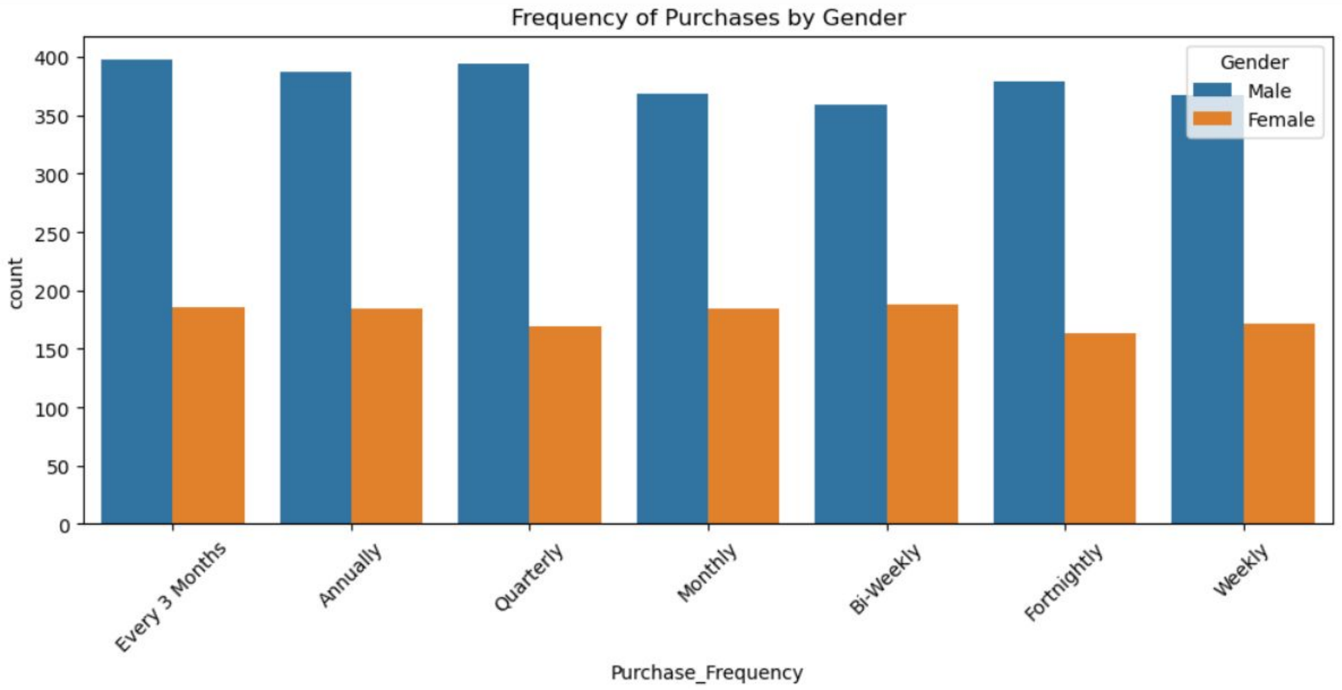
2.Preferred categories by Age Group

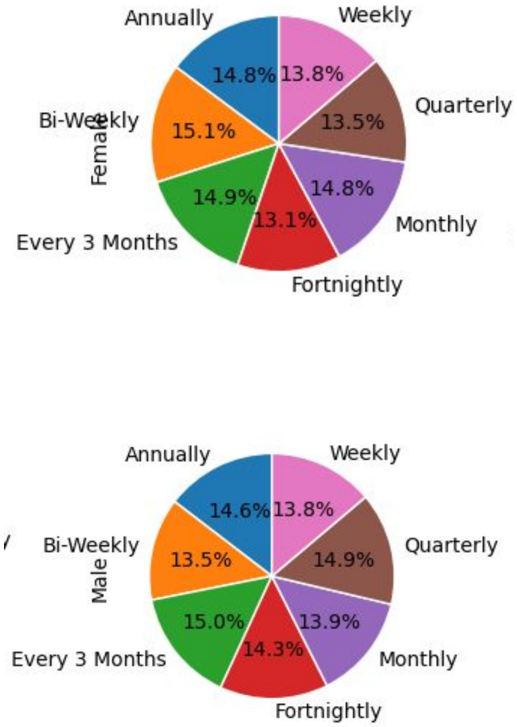


2.Preferred categories by Age Group

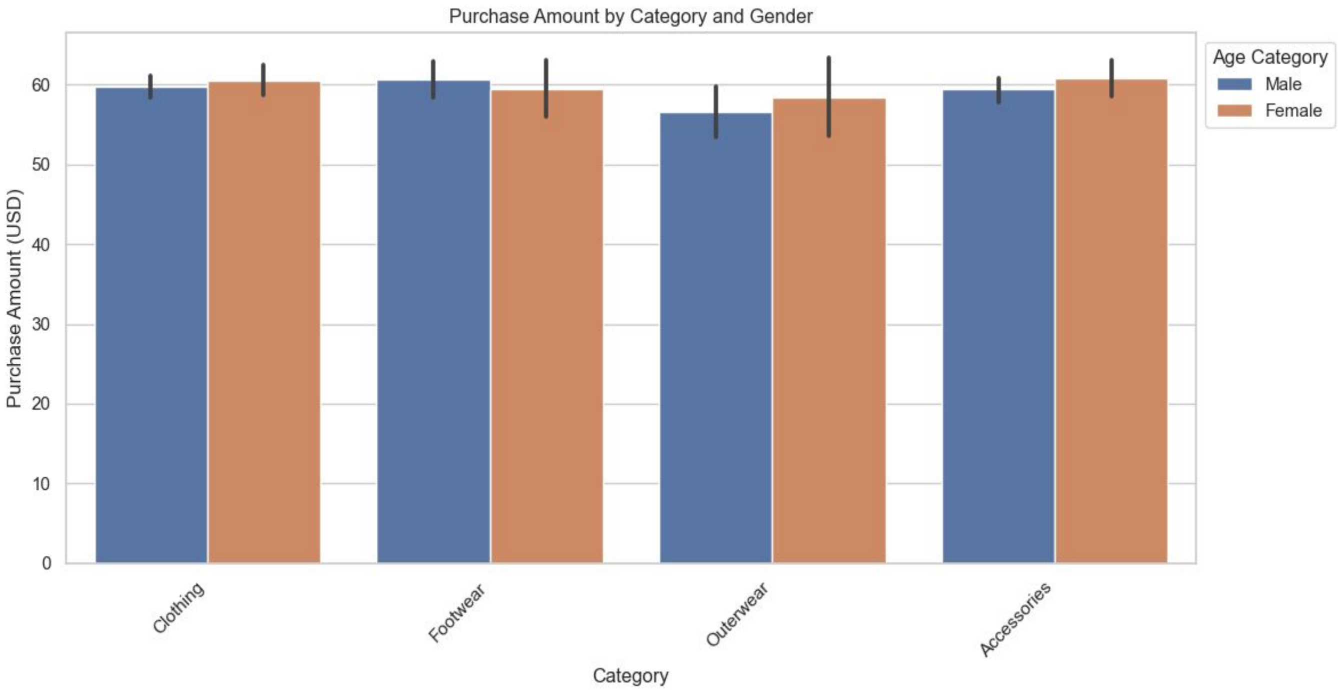


3. Frequency of purchase by Gender

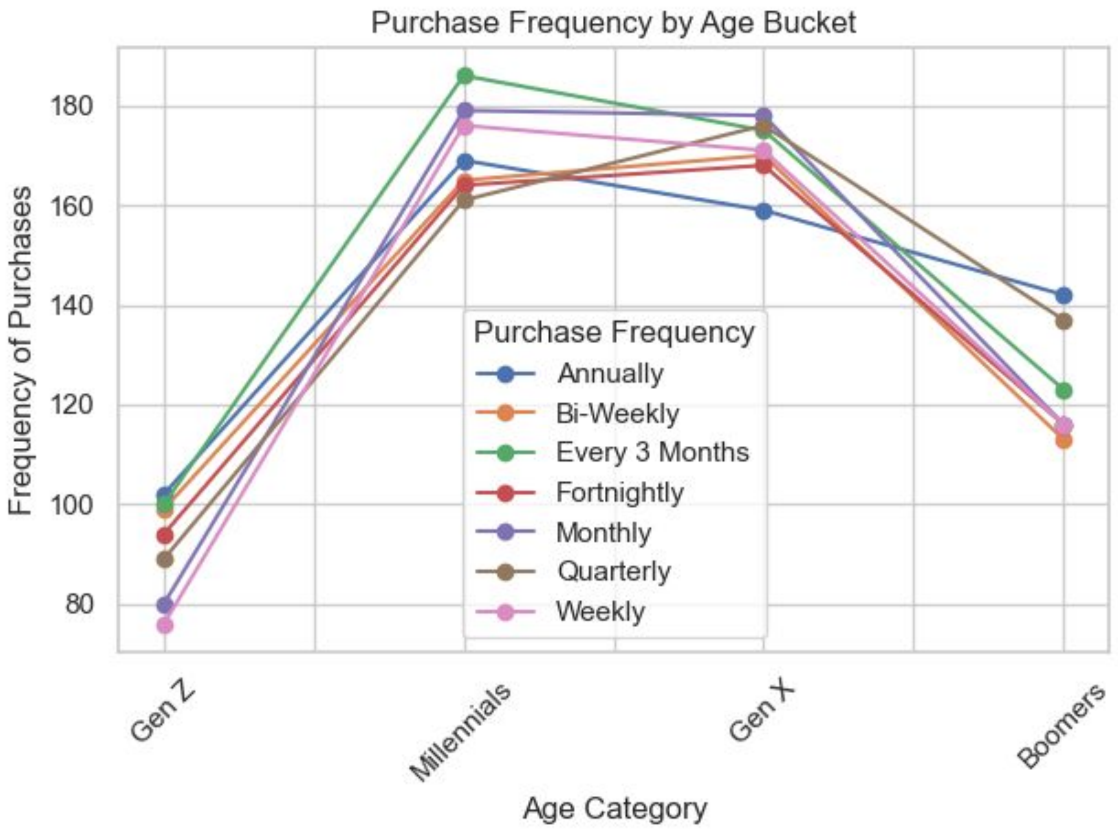




4. Purchase Amount by Category and Gender



5. Purchase Frequency by Age Bucket



Managerial Insights

- **Tailoring Marketing Campaigns:** These insights can guide tailored marketing campaigns. For example, older demographics (Boomers and Gen X) showing a willingness to spend more could be targeted with premium product lines.
- **Product Offerings and Promotions:** Understanding preferred categories by age can help in curating product lines and promotions that appeal to each demographic segment.
- **Engagement Strategies:** Frequency of purchase data suggests the need for different engagement strategies by gender.

Q/A Session

Any Questions!?

If not, we are getting back to our pizzas!

Thank you!