nitinrawat2502@gmail.com
https://www.linkedin.com/in/nitin-rawat-a38536270/
7982893096

mzsfgt3gr

December 12, 2024

0.1 Decoding Customer Shopping Trends: A Modern Analysis

[452]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing, cluster
from pandas.plotting import scatter_matrix

import os
```

[321]:

```
df = pd.read_csv("shopping_trends.csv")
```

[322]:

```
df.head()
```

[322]:

```
   Customer ID  Age Gender Item Purchased  Category  Purchase Amount (USD)  \
0            1   55   Male         Blouse  Clothing                      53
1            2   19   Male        Sweater  Clothing                      64
2            3   50   Male          Jeans  Clothing                      73
3            4   21   Male        Sandals  Footwear                      90
4            5   45   Male         Blouse  Clothing                      49

        Location Size      Color  Season  Review Rating Subscription Status  \
0       Kentucky    L       Gray  Winter            3.1                 Yes
1          Maine    L     Maroon  Winter            3.1                 Yes
2  Massachusetts    S     Maroon  Spring            3.1                 Yes
3   Rhode Island    M     Maroon  Spring            3.5                 Yes
4         Oregon    M  Turquoise  Spring            2.7                 Yes

  Payment Method  Shipping Type Discount Applied Promo Code Used  \
0    Credit Card        Express              Yes             Yes
1  Bank Transfer        Express              Yes             Yes
2           Cash  Free Shipping              Yes             Yes
3         PayPal   Next Day Air              Yes             Yes
4           Cash  Free Shipping              Yes             Yes

   Previous Purchases Preferred Payment Method Frequency of Purchases
0                  14                    Venmo            Fortnightly
```

```
1                    2                      Cash            Fortnightly
2                   23               Credit Card                 Weekly
3                   49                    PayPal                 Weekly
4                   31                    PayPal               Annually
```

[323]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 19 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Customer ID               3900 non-null   int64
 1   Age                       3900 non-null   int64
 2   Gender                    3900 non-null   object
 3   Item Purchased            3900 non-null   object
 4   Category                  3900 non-null   object
 5   Purchase Amount (USD)     3900 non-null   int64
 6   Location                  3900 non-null   object
 7   Size                      3900 non-null   object
 8   Color                     3900 non-null   object
 9   Season                    3900 non-null   object
 10  Review Rating             3900 non-null   float64
 11  Subscription Status       3900 non-null   object
 12  Payment Method            3900 non-null   object
 13  Shipping Type             3900 non-null   object
 14  Discount Applied          3900 non-null   object
 15  Promo Code Used           3900 non-null   object
 16  Previous Purchases        3900 non-null   int64
 17  Preferred Payment Method  3900 non-null   object
 18  Frequency of Purchases    3900 non-null   object
dtypes: float64(1), int64(4), object(14)
memory usage: 579.0+ KB
```

[324]:
```
df.index = df["Customer ID"]
```

[325]:
```
del df["Customer ID"]
```

[326]:
```
df
```

[326]:
```
             Age  Gender Item Purchased     Category   Purchase Amount (USD)  \
Customer ID
1             55    Male         Blouse     Clothing                      53
2             19    Male        Sweater     Clothing                      64
3             50    Male          Jeans     Clothing                      73
4             21    Male        Sandals     Footwear                      90
5             45    Male         Blouse     Clothing                      49
```

```
…          …     …            …          …                    …
3896          40  Female         Hoodie     Clothing                     28
3897          52  Female       Backpack  Accessories                     49
3898          46  Female           Belt  Accessories                     33
3899          44  Female          Shoes     Footwear                     77
3900          52  Female        Handbag  Accessories                     81

                  Location Size      Color  Season  Review Rating  \
Customer ID
1                 Kentucky    L       Gray  Winter            3.1
2                    Maine    L     Maroon  Winter            3.1
3             Massachusetts    S     Maroon  Spring            3.1
4              Rhode Island    M     Maroon  Spring            3.5
5                   Oregon    M  Turquoise  Spring            2.7
…                    …  …        …     …            …
3896              Virginia    L  Turquoise  Summer            4.2
3897                  Iowa    L      White  Spring            4.5
3898            New Jersey    L      Green  Spring            2.9
3899             Minnesota    S      Brown  Summer            3.8
3900            California    M      Beige  Spring            3.1

            Subscription Status Payment Method   Shipping Type  \
Customer ID
1                           Yes    Credit Card         Express
2                           Yes  Bank Transfer         Express
3                           Yes           Cash   Free Shipping
4                           Yes         PayPal    Next Day Air
5                           Yes           Cash   Free Shipping
…                         …            …             …
3896                         No           Cash  2-Day Shipping
3897                         No         PayPal    Store Pickup
3898                         No    Credit Card        Standard
3899                         No         PayPal         Express
3900                         No  Bank Transfer    Store Pickup

            Discount Applied Promo Code Used  Previous Purchases  \
Customer ID
1                        Yes             Yes                  14
2                        Yes             Yes                   2
3                        Yes             Yes                  23
4                        Yes             Yes                  49
5                        Yes             Yes                  31
…                      …             …                 …
3896                      No              No                  32
3897                      No              No                  41
3898                      No              No                  24
3899                      No              No                  24
```

```
3900                      No              No                   33

            Preferred Payment Method Frequency of Purchases
Customer ID
1                                Venmo            Fortnightly
2                                 Cash            Fortnightly
3                          Credit Card                 Weekly
4                               PayPal                 Weekly
5                               PayPal               Annually
…                                    …                      …
3896                             Venmo                 Weekly
3897                     Bank Transfer              Bi-Weekly
3898                             Venmo              Quarterly
3899                             Venmo                 Weekly
3900                             Venmo              Quarterly

[3900 rows x 18 columns]
```

```
[327]: df.shape
```

```
[327]: (3900, 18)
```

```
[328]: df.count()
```

```
[328]: Age                         3900
       Gender                      3900
       Item Purchased              3900
       Category                    3900
       Purchase Amount (USD)       3900
       Location                    3900
       Size                        3900
       Color                       3900
       Season                      3900
       Review Rating               3900
       Subscription Status         3900
       Payment Method              3900
       Shipping Type               3900
       Discount Applied            3900
       Promo Code Used             3900
       Previous Purchases          3900
       Preferred Payment Method    3900
       Frequency of Purchases      3900
       dtype: int64
```

```
[329]: df.isnull().sum()
```

```
[329]: Age                         0
       Gender                      0
       Item Purchased              0
       Category                    0
       Purchase Amount (USD)       0
       Location                    0
       Size                        0
       Color                       0
       Season                      0
       Review Rating               0
       Subscription Status         0
       Payment Method              0
       Shipping Type               0
       Discount Applied            0
       Promo Code Used             0
       Previous Purchases          0
       Preferred Payment Method    0
       Frequency of Purchases      0
       dtype: int64
```

```
[330]: categorical = df.dtypes[df.dtypes == "object"].index
```

```
[331]: categorical
```

```
[331]: Index(['Gender', 'Item Purchased', 'Category', 'Location', 'Size', 'Color',
              'Season', 'Subscription Status', 'Payment Method', 'Shipping Type',
              'Discount Applied', 'Promo Code Used', 'Preferred Payment Method',
              'Frequency of Purchases'],
             dtype='object')
```

```
[332]: for i in categorical:
           res = sorted(df[i].unique())
           print(i, ": ",res,"\n")
```

```
Gender :  ['Female', 'Male']

Item Purchased :  ['Backpack', 'Belt', 'Blouse', 'Boots', 'Coat', 'Dress',
'Gloves', 'Handbag', 'Hat', 'Hoodie', 'Jacket', 'Jeans', 'Jewelry', 'Pants',
'Sandals', 'Scarf', 'Shirt', 'Shoes', 'Shorts', 'Skirt', 'Sneakers', 'Socks',
'Sunglasses', 'Sweater', 'T-shirt']

Category :  ['Accessories', 'Clothing', 'Footwear', 'Outerwear']

Location :  ['Alabama', 'Alaska', 'Arizona', 'Arkansas', 'California',
'Colorado', 'Connecticut', 'Delaware', 'Florida', 'Georgia', 'Hawaii', 'Idaho',
'Illinois', 'Indiana', 'Iowa', 'Kansas', 'Kentucky', 'Louisiana', 'Maine',
'Maryland', 'Massachusetts', 'Michigan', 'Minnesota', 'Mississippi', 'Missouri',
'Montana', 'Nebraska', 'Nevada', 'New Hampshire', 'New Jersey', 'New Mexico',
'New York', 'North Carolina', 'North Dakota', 'Ohio', 'Oklahoma', 'Oregon',
'Pennsylvania', 'Rhode Island', 'South Carolina', 'South Dakota', 'Tennessee',
'Texas', 'Utah', 'Vermont', 'Virginia', 'Washington', 'West Virginia',
'Wisconsin', 'Wyoming']

Size :  ['L', 'M', 'S', 'XL']

Color :  ['Beige', 'Black', 'Blue', 'Brown', 'Charcoal', 'Cyan', 'Gold', 'Gray',
'Green', 'Indigo', 'Lavender', 'Magenta', 'Maroon', 'Olive', 'Orange', 'Peach',
'Pink', 'Purple', 'Red', 'Silver', 'Teal', 'Turquoise', 'Violet', 'White',
'Yellow']

Season :  ['Fall', 'Spring', 'Summer', 'Winter']

Subscription Status :  ['No', 'Yes']

Payment Method :  ['Bank Transfer', 'Cash', 'Credit Card', 'Debit Card',
'PayPal', 'Venmo']

Shipping Type :  ['2-Day Shipping', 'Express', 'Free Shipping', 'Next Day Air',
'Standard', 'Store Pickup']

Discount Applied :  ['No', 'Yes']

Promo Code Used :  ['No', 'Yes']

Preferred Payment Method :  ['Bank Transfer', 'Cash', 'Credit Card', 'Debit
Card', 'PayPal', 'Venmo']

Frequency of Purchases :  ['Annually', 'Bi-Weekly', 'Every 3 Months',
'Fortnightly', 'Monthly', 'Quarterly', 'Weekly']
```

[333]:

```
df.dtypes.index
```

[333]:

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

0.1.1 Age by Groups

[334]:
```
bins = [0,12,19,35,64,100]
labels = ["children" , "teenagers","young adults","adults","seniors"]
new_range = pd.cut(df["Age"],bins = bins ,labels=labels,right=True)
df["Age"] = new_range
```

[335]:
```
df
```

[335]:
```
                      Age  Gender Item Purchased     Category  \
Customer ID
1                  adults    Male         Blouse     Clothing
2               teenagers    Male        Sweater     Clothing
3                  adults    Male          Jeans     Clothing
4            young adults    Male        Sandals     Footwear
5                  adults    Male         Blouse     Clothing
…                   …     …            …          …
3896               adults  Female         Hoodie     Clothing
3897               adults  Female       Backpack  Accessories
3898               adults  Female           Belt  Accessories
3899               adults  Female          Shoes     Footwear
3900               adults  Female        Handbag  Accessories

             Purchase Amount (USD)       Location Size      Color  Season  \
Customer ID
1                               53       Kentucky    L       Gray  Winter
2                               64          Maine    L     Maroon  Winter
3                               73  Massachusetts    S     Maroon  Spring
4                               90   Rhode Island    M     Maroon  Spring
5                               49         Oregon    M  Turquoise  Spring
…                            …            …  …        …     …
3896                            28       Virginia    L  Turquoise  Summer
3897                            49           Iowa    L      White  Spring
3898                            33     New Jersey    L      Green  Spring
3899                            77      Minnesota    S      Brown  Summer
3900                            81     California    M      Beige  Spring

             Review Rating Subscription Status Payment Method   Shipping Type  \
Customer ID
1                      3.1                 Yes    Credit Card         Express
2                      3.1                 Yes  Bank Transfer         Express
3                      3.1                 Yes           Cash   Free Shipping
4                      3.5                 Yes         PayPal    Next Day Air
5                      2.7                 Yes           Cash   Free Shipping
…                    …                  …            …             …
3896                   4.2                  No           Cash  2-Day Shipping
3897                   4.5                  No         PayPal    Store Pickup
```

```
3898                   2.9                  No    Credit Card        Standard
3899                   3.8                  No         PayPal         Express
3900                   3.1                  No  Bank Transfer    Store Pickup

            Discount Applied Promo Code Used  Previous Purchases  \
Customer ID
1                        Yes             Yes                  14
2                        Yes             Yes                   2
3                        Yes             Yes                  23
4                        Yes             Yes                  49
5                        Yes             Yes                  31
…                      …             …                 …
3896                      No              No                  32
3897                      No              No                  41
3898                      No              No                  24
3899                      No              No                  24
3900                      No              No                  33

            Preferred Payment Method Frequency of Purchases
Customer ID
1                              Venmo            Fortnightly
2                               Cash            Fortnightly
3                        Credit Card                 Weekly
4                             PayPal                 Weekly
5                             PayPal               Annually
…                              …                    …
3896                           Venmo                 Weekly
3897                   Bank Transfer              Bi-Weekly
3898                           Venmo              Quarterly
3899                           Venmo                 Weekly
3900                           Venmo              Quarterly

[3900 rows x 18 columns]
```

0.1.2 Defining the order of the season based on the Northern Hemisphere.

```
[336]: df["Season"]
```

```
[336]: Customer ID
1        Winter
2        Winter
3        Spring
4        Spring
5        Spring
          …
3896     Summer
3897     Spring
```

```
3898    Spring
3899    Summer
3900    Spring
Name: Season, Length: 3900, dtype: object
```

[337]:
```
df["Season"]
```

[337]:
```
Customer ID
1       Winter
2       Winter
3       Spring
4       Spring
5       Spring
         …
3896    Summer
3897    Spring
3898    Spring
3899    Summer
3900    Spring
Name: Season, Length: 3900, dtype: object
```

[338]:
```
season_order = ["Spring", "Summer", "Fall", "Winter"]
new_season = pd.Categorical(df["Season"], categories=season_order, ordered=True)
df["Season"] = new_season


df["Season"].unique()
```

[338]:
```
['Winter', 'Spring', 'Summer', 'Fall']
Categories (4, object): ['Spring' < 'Summer' < 'Fall' < 'Winter']
```

[339]:
```
df
```

[339]:
```
                      Age  Gender Item Purchased     Category  \
Customer ID
1                  adults    Male         Blouse     Clothing
2               teenagers    Male        Sweater     Clothing
3                  adults    Male          Jeans     Clothing
4            young adults    Male        Sandals     Footwear
5                  adults    Male         Blouse     Clothing
…                     …     …             …         …
3896               adults  Female         Hoodie     Clothing
3897               adults  Female       Backpack  Accessories
3898               adults  Female           Belt  Accessories
3899               adults  Female          Shoes     Footwear
3900               adults  Female        Handbag  Accessories
```

```
             Purchase Amount (USD)       Location Size      Color  Season  \
Customer ID
1                               53       Kentucky    L       Gray  Winter
2                               64          Maine    L     Maroon  Winter
3                               73  Massachusetts    S     Maroon  Spring
4                               90   Rhode Island    M     Maroon  Spring
5                               49         Oregon    M  Turquoise  Spring
…                            …            …  …        …     …
3896                            28       Virginia    L  Turquoise  Summer
3897                            49           Iowa    L      White  Spring
3898                            33     New Jersey    L      Green  Spring
3899                            77      Minnesota    S      Brown  Summer
3900                            81     California    M      Beige  Spring

             Review Rating Subscription Status Payment Method   Shipping Type  \
Customer ID
1                      3.1                 Yes    Credit Card         Express
2                      3.1                 Yes  Bank Transfer         Express
3                      3.1                 Yes           Cash   Free Shipping
4                      3.5                 Yes         PayPal    Next Day Air
5                      2.7                 Yes           Cash   Free Shipping
…                    …                 …            …             …
3896                   4.2                  No           Cash  2-Day Shipping
3897                   4.5                  No         PayPal    Store Pickup
3898                   2.9                  No    Credit Card        Standard
3899                   3.8                  No         PayPal         Express
3900                   3.1                  No  Bank Transfer    Store Pickup

            Discount Applied Promo Code Used  Previous Purchases  \
Customer ID
1                        Yes             Yes                  14
2                        Yes             Yes                   2
3                        Yes             Yes                  23
4                        Yes             Yes                  49
5                        Yes             Yes                  31
…                      …             …                 …
3896                      No              No                  32
3897                      No              No                  41
3898                      No              No                  24
3899                      No              No                  24
3900                      No              No                  33

            Preferred Payment Method Frequency of Purchases
Customer ID
1                              Venmo            Fortnightly
2                               Cash            Fortnightly
3                        Credit Card                 Weekly
```

```
4                             PayPal                 Weekly
5                             PayPal               Annually
…                              …                    …
3896                           Venmo                 Weekly
3897                   Bank Transfer              Bi-Weekly
3898                           Venmo              Quarterly
3899                           Venmo                 Weekly
3900                           Venmo              Quarterly

[3900 rows x 18 columns]
```

[340]:

```
df["Season"]
```

[340]:

```
Customer ID
1       Winter
2       Winter
3       Spring
4       Spring
5       Spring
         …
3896    Summer
3897    Spring
3898    Spring
3899    Summer
3900    Spring
Name: Season, Length: 3900, dtype: category
Categories (4, object): ['Spring' < 'Summer' < 'Fall' < 'Winter']
```

0.1.3 Rating Conversion

[341]:

```
to_percent_rating = df["Review Rating"].apply(lambda X:X/5)
df["Review Rating"] = to_percent_rating
```

[342]:

```
df.head(2)
```

[342]:

```
                   Age Gender Item Purchased  Category  Purchase Amount (USD)  \
Customer ID
1               adults   Male         Blouse  Clothing                     53
2            teenagers   Male        Sweater  Clothing                     64

             Location Size   Color  Season Review Rating Subscription Status  \
Customer ID
1            Kentucky    L    Gray  Winter          0.62                 Yes
2               Maine    L  Maroon  Winter          0.62                 Yes

            Payment Method Shipping Type Discount Applied Promo Code Used  \
Customer ID
```

```
1               Credit Card       Express              Yes             Yes
2             Bank Transfer       Express              Yes             Yes

             Previous Purchases Preferred Payment Method  \
Customer ID                                                
1                            14                    Venmo   
2                             2                     Cash   

            Frequency of Purchases  
Customer ID                         
1                      Fortnightly  
2                      Fortnightly  
```

[343]:
```
obj  = df.dtypes[df.dtypes == "object"].index
obj
```

[343]:
```
Index(['Gender', 'Item Purchased', 'Category', 'Location', 'Size', 'Color',
       'Subscription Status', 'Payment Method', 'Shipping Type',
       'Discount Applied', 'Promo Code Used', 'Preferred Payment Method',
       'Frequency of Purchases'],
      dtype='object')
```

The analysis of categorical and boolean columns highlights significant customer trends. A majority of the customer base, **2,652 out of 3,900,** consists of males. The most popular item among purchases is **blouses,** categorized under clothing. Seasonal trends indicate that spring accounts for the highest number of transactions. Additionally, most customers have not subscribed to any services. While **PayPal** emerges as the most preferred payment method, **credit cards** account for the majority of completed transactions.

[344]:
```
df[obj].describe()
```

[344]:
```
       Gender Item Purchased  Category Location  Size  Color  \
count    3900           3900      3900     3900  3900   3900   
unique      2             25         4       50     4     25   
top      Male         Blouse  Clothing  Montana     M  Olive   
freq     2652            171      1737       96  1755    177   

       Subscription Status Payment Method  Shipping Type Discount Applied  \
count                 3900           3900           3900             3900   
unique                   2              6              6                2   
top                     No    Credit Card  Free Shipping               No   
freq                  2847            696            675             2223   

       Promo Code Used Preferred Payment Method Frequency of Purchases
count             3900                     3900                   3900
unique               2                        6                      7
top                 No                   PayPal         Every 3 Months
```

```
freq               2223                       677                     584
```

[345]:

```
df.describe()
```

[345]:

```
       Purchase Amount (USD)  Review Rating  Previous Purchases
count            3900.000000    3900.000000         3900.000000
mean               59.764359       0.749990           25.351538
std                23.685392       0.143245           14.447125
min                20.000000       0.500000            1.000000
25%                39.000000       0.620000           13.000000
50%                60.000000       0.740000           25.000000
75%                81.000000       0.880000           38.000000
max               100.000000       1.000000           50.000000
```

Gender Distribution As shown, most customers are **Male** with a total of 2652 (68%). Meanwhile, **Female** garners a total of 1248 (32%).

[346]:

```
df["Gender"].value_counts()
```

[346]:

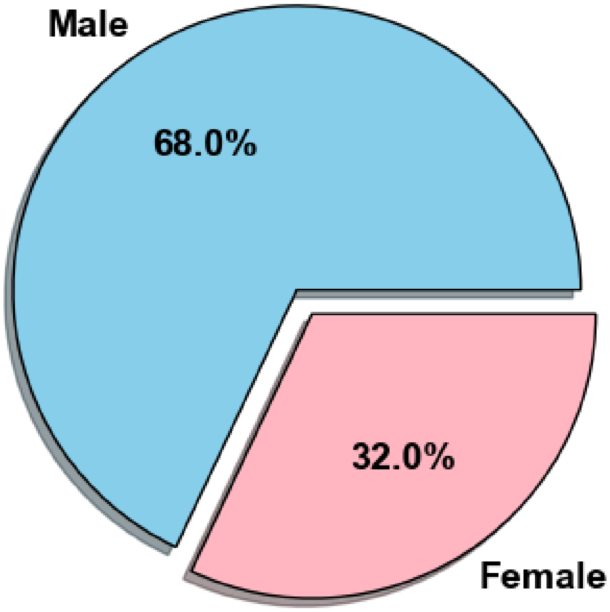
```
Gender
Male      2652
Female    1248
Name: count, dtype: int64
```

[347]:

```
gender = df["Gender"].value_counts(normalize=True)
```

[348]:

```
plt.pie(gender,
        labels=gender.index,
        autopct='%1.1f%%',
        explode=(0.1, 0),
        colors=["skyblue", "lightpink"],
        wedgeprops={"edgecolor":"black", "linewidth":1},
        shadow=True,
        textprops={'fontsize': 14, 'color': 'black', 'weight': 'bold'})
plt.show()
```



Age Distribution Most costumers are **Adults** followed by **Young Adults and Seniors**. Meanwhile, **teenangers** got the lowest number in distribution.

[349]:
```
df.Age.value_counts()
```

[349]:
```
Age
adults          2160
young adults    1163
seniors          427
teenagers        150
children           0
Name: count, dtype: int64
```

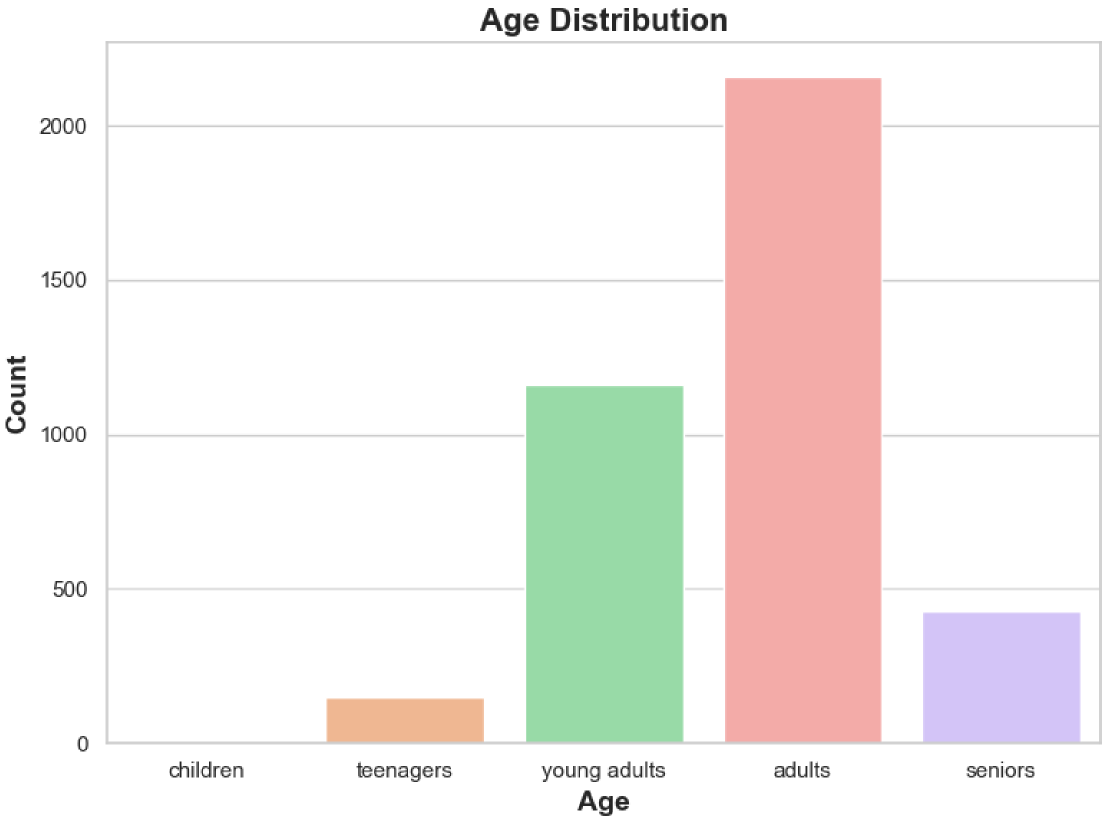
[350]:
```
plt.figure(figsize=(8, 6))
sns.countplot(data=df, x="Age", palette="pastel")
plt.xlabel('Age', fontsize=14, weight='bold')
plt.ylabel('Count', fontsize=14, weight='bold')
plt.title('Age Distribution', fontsize=16, weight='bold')
plt.tight_layout()
plt.show()
```

```
C:\Users\nitin\AppData\Local\Temp\ipykernel_7964\1117022828.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
```

```
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(data=df, x="Age", palette="pastel")
```



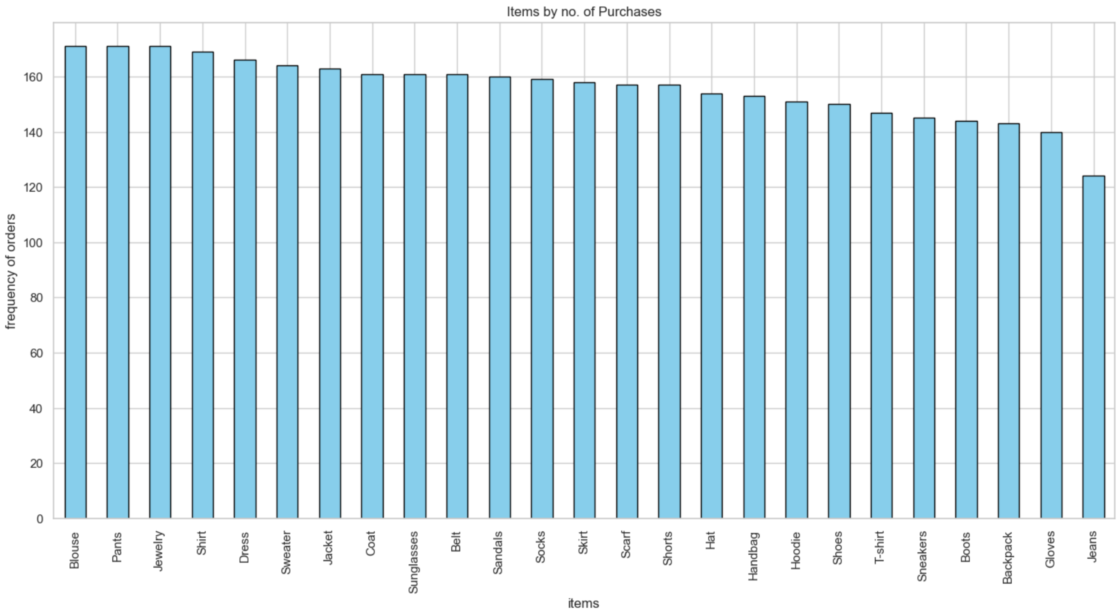
0.1.4 Sales Analysis:

The sales data reveals that **Jewelry**, **Blouse**, and **Pants** are the most popular items, with a combined total of **171 purchases** across these categories. **Shirts** follow closely, totaling **169 purchases**, demonstrating strong demand as well. On the other hand, **Jeans** recorded the lowest sales, with only **124 purchases**, indicating relatively lower consumer interest compared to the other categories.

[351]:
```
item = df["Item Purchased"].value_counts(sort=True)
```

[352]:
```
item.plot(kind="bar", title="Items by no. of Purchases", ylabel="frequency of␣
↪orders", xlabel="items", figsize=(17, 8),
          legend=False, edgecolor="black", grid=True, color="skyblue")
```

[352]:
```
<Axes: title={'center': 'Items by no. of Purchases'}, xlabel='items',
ylabel='frequency of orders'>
```



[353]:
```
df.columns
```

[353]:
```
Index(['Age', 'Gender', 'Item Purchased', 'Category', 'Purchase Amount (USD)',
       'Location', 'Size', 'Color', 'Season', 'Review Rating',
       'Subscription Status', 'Payment Method', 'Shipping Type',
       'Discount Applied', 'Promo Code Used', 'Previous Purchases',
       'Preferred Payment Method', 'Frequency of Purchases'],
      dtype='object')
```

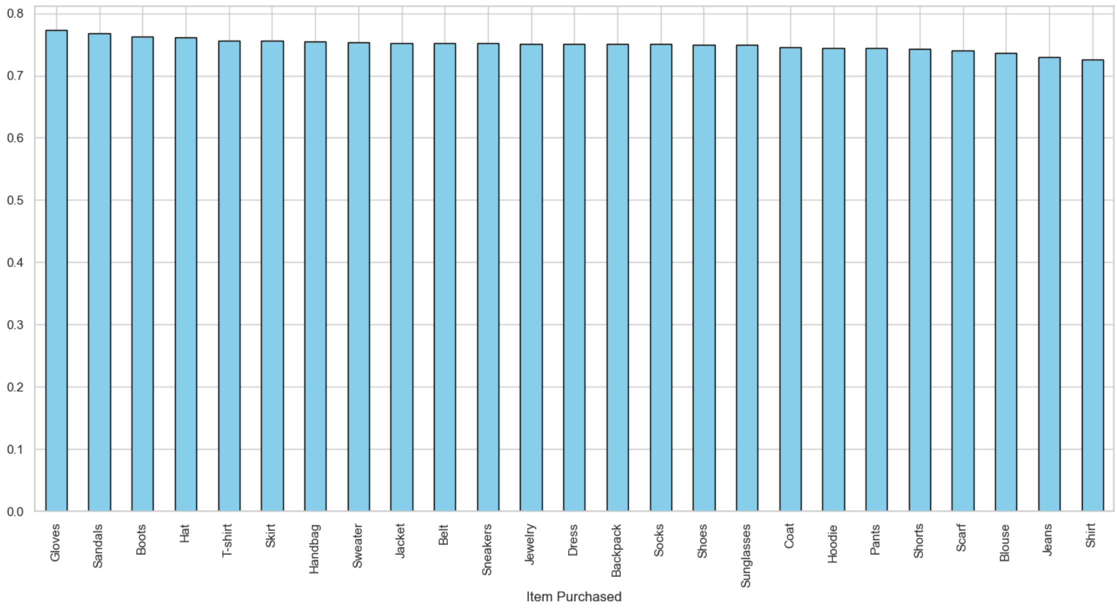
0.1.5 Review Ratings Analysis

The analysis of the review ratings reveals that **Gloves** exhibit the highest mean rating, indicating strong customer satisfaction in this category. Conversely, **Shirts** have the lowest mean rating, suggesting relatively lower customer satisfaction. This insight highlights that customers are generally more satisfied with their purchases of gloves compared to shirts, which may require further investigation into factors affecting customer dissatisfaction in that category.

[354]:
```
item_rating = df.groupby("Item Purchased")["Review Rating"].mean().
 ↪sort_values(ascending=False).reset_index()
```

[355]:
```
item_rating.plot(kind = "bar", x ="Item Purchased",y ="Review␣
 ↪Rating",grid=True, edgecolor="black",␣
 ↪color="skyblue",legend=False,figsize=(17, 8))
```

[355]:
```
<Axes: xlabel='Item Purchased'>
```



0.1.6 Number of Purchases by Location

The analysis of purchase data indicates that **Montana** has the highest number of purchases, totaling **96**. In contrast, **Rhode Island** recorded the lowest number of purchases. This disparity suggests that regional factors may influence purchasing behavior, and further investigation could help identify opportunities to increase engagement and sales in underperforming locations.

[356]:
```
df["Location"].value_counts().reset_index().head()
```

[356]:
```
     Location  count
0     Montana     96
1  California     95
2       Idaho     93
3    Illinois     92
4     Alabama     89
```

[357]:
```
top_mon = df[df["Location"] =="Montana"]["Item Purchased"].value_counts()
```

0.1.7 Analysis of Top Items by Location

The analysis reveals the top items purchased across various locations. For instance, in **Alabama**, the most purchased item is **Jewelry**. Interestingly, even though **Shirt** has the lowest average rating, it remains the top item purchased in **Illinois** and **Wyoming**. This suggests that certain items may have a higher purchase frequency regardless of their ratings, highlighting regional preferences.

[358]:
```
top_mon.head()
```

```
[358]: Item Purchased
       Coat       8
       Sweater    7
       Socks      6
       Handbag    6
       Jeans      6
       Name: count, dtype: int64
```

```
[359]: unique_location = sorted(df["Location"].unique())
```

```
[360]: unique_location
```

```
[360]: ['Alabama',
        'Alaska',
        'Arizona',
        'Arkansas',
        'California',
        'Colorado',
        'Connecticut',
        'Delaware',
        'Florida',
        'Georgia',
        'Hawaii',
        'Idaho',
        'Illinois',
        'Indiana',
        'Iowa',
        'Kansas',
        'Kentucky',
        'Louisiana',
        'Maine',
        'Maryland',
        'Massachusetts',
        'Michigan',
        'Minnesota',
        'Mississippi',
        'Missouri',
        'Montana',
        'Nebraska',
        'Nevada',
        'New Hampshire',
        'New Jersey',
        'New Mexico',
        'New York',
        'North Carolina',
        'North Dakota',
        'Ohio',
        'Oklahoma',
        'Oregon',
        'Pennsylvania',
        'Rhode Island',
        'South Carolina',
        'South Dakota',
        'Tennessee',
        'Texas',
        'Utah',
        'Vermont',
        'Virginia',
        'Washington',
        'West Virginia',
        'Wisconsin',
        'Wyoming']
```

0.1.8 Most Frequent Item Purchased by Location

[361]:
```
result = df.groupby('Location')['Item Purchased'].agg(lambda x: x.mode()[0])
print(result)
```

```
Location
Alabama            Jewelry
Alaska            Backpack
Arizona           Backpack
Arkansas            Gloves
California           Dress
Colorado            Hoodie
Connecticut           Coat
Delaware             Pants
Florida               Coat
Georgia              Dress
Hawaii             Handbag
Idaho                 Belt
Illinois             Shirt
Indiana              Skirt
Iowa               Sweater
Kansas              Blouse
Kentucky           Handbag
Louisiana           Hoodie
Maine                Shoes
Maryland             Skirt
Massachusetts        Boots
Michigan           Sandals
Minnesota            Skirt
Mississippi        Handbag
Missouri          Sneakers
```

```
Montana                Coat
Nebraska           Backpack
Nevada             Backpack
New Hampshire        Blouse
New Jersey             Belt
New Mexico           Hoodie
New York             Gloves
North Carolina       Gloves
North Dakota         Jacket
Ohio                Sandals
Oklahoma             Jacket
Oregon                Shirt
Pennsylvania          Scarf
Rhode Island          Dress
South Carolina          Hat
South Dakota       Backpack
Tennessee             Scarf
Texas                 Shoes
Utah                 Shorts
Vermont               Pants
Virginia             Shorts
Washington            Socks
West Virginia          Coat
Wisconsin            Blouse
Wyoming               Shirt
Name: Item Purchased, dtype: object
```

```
[362]: result.shape
```

```
[362]: (50,)
```

0.1.9 Category Distribution Analysis

Looking at the data, **clothing** is the most popular category among consumers, with **accessories** coming in second. On the other hand, **outerwear** is the least purchased category, meaning fewer people are buying it compared to the others. This might be because people generally prefer clothing all year round, while outerwear purchases tend to happen more in colder months.

```
[363]: cate = df["Category"].value_counts(ascending=False).reset_index()
       cate
```

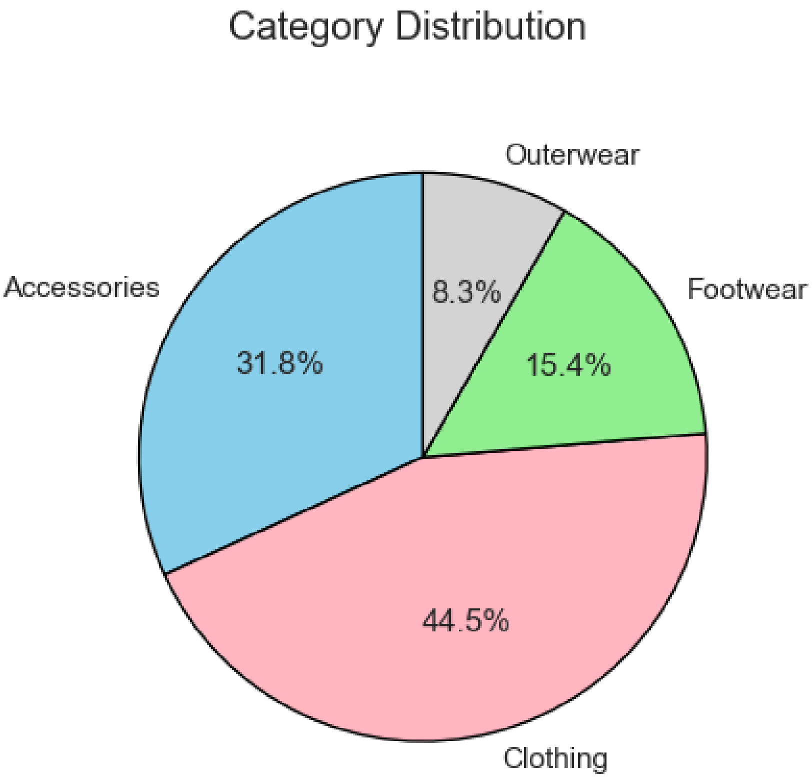
```
[363]:       Category  count
       0     Clothing   1737
       1  Accessories   1240
       2     Footwear    599
       3    Outerwear    324
```

```
[364]: cat = pd.crosstab(df["Category"], columns="counts")
       cat
```

```
[364]: col_0        counts
       Category
       Accessories    1240
       Clothing       1737
       Footwear        599
       Outerwear       324
```

```
[365]: colors = ["skyblue", "lightpink", "lightgreen", "lightgrey"]
       cat.plot(kind="pie", subplots=True, title="Category Distribution", ylabel="",␣
        ↪legend=False,
                 wedgeprops={"edgecolor":"black"},autopct='%1.1f%%', startangle=90,␣
        ↪colors=colors )
```

```
[365]: array([<Axes: >], dtype=object)
```



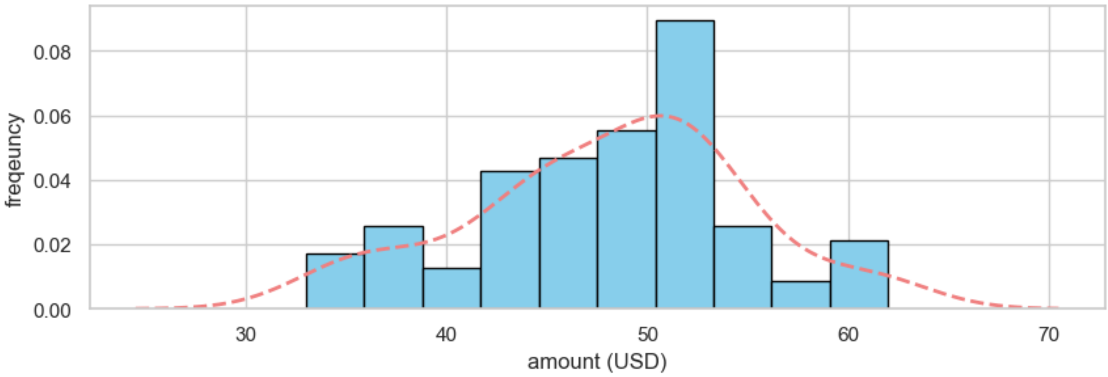
0.1.10 Purchase Amount (USD) Frequency¶

[366]:
```
amount = pd.crosstab(df["Purchase Amount (USD)"], columns="counts")
amount.sort_values(by="counts", ascending=False).head(5)
```

[366]:
```
col_0                  counts
Purchase Amount (USD)
36                         62
32                         62
94                         62
51                         61
90                         60
```

[367]:
```
amount.plot.hist(grid=True, legend=False, xlabel="amount (USD)",␣
 ↪ylabel="freqeuncy", density=True,
                 edgecolor="black", color="skyblue", figsize=(10,3))
sns.kdeplot(amount["counts"], color="lightcoral", lw=2, linestyle="--")
```

[367]:
```
<Axes: xlabel='amount (USD)', ylabel='freqeuncy'>
```



0.1.11 Average Purchase Amount Among Age Groups

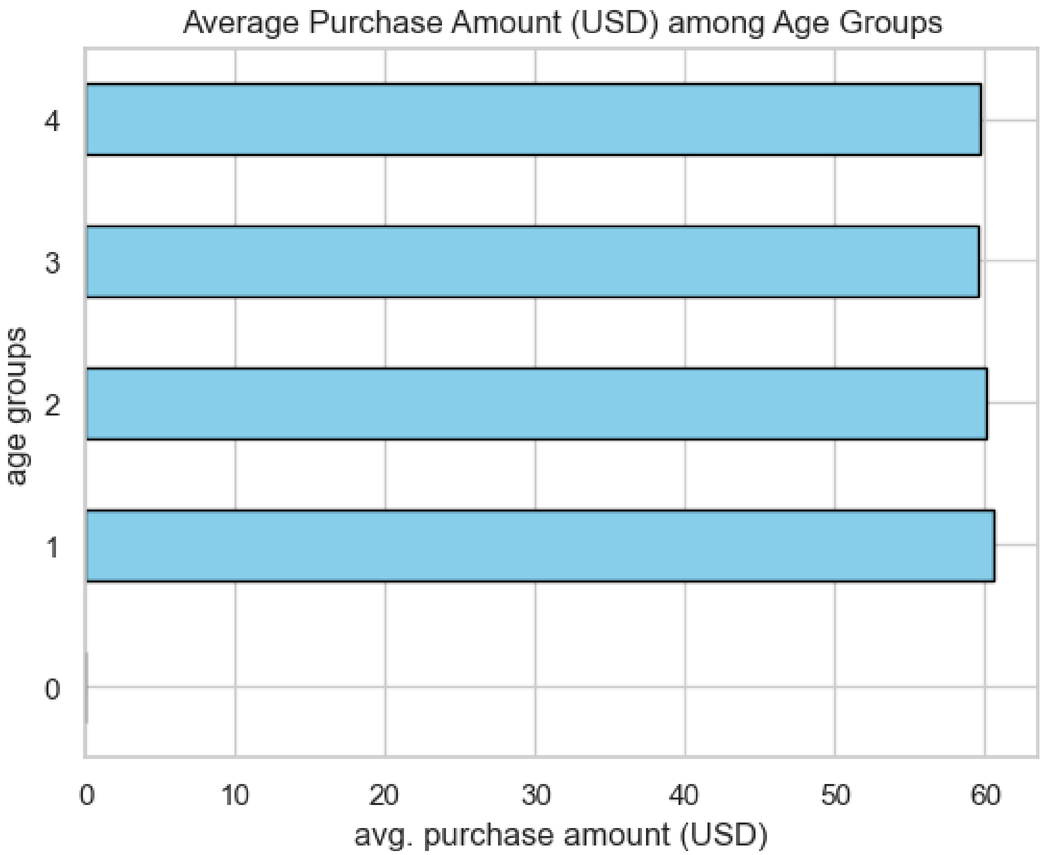
From the data, we can see that **teenagers** spend the most on average compared to other age groups. Right after them, **young adults** come in second place for the highest average purchase amount.

[368]:
```
avg_purchase = df.groupby("Age")["Purchase Amount (USD)"].mean().reset_index()
```

```
C:\Users\nitin\AppData\Local\Temp\ipykernel_7964\4079642796.py:1: FutureWarning:
The default of observed=False is deprecated and will be changed to True in a
future version of pandas. Pass observed=False to retain current behavior or
observed=True to adopt the future default and silence this warning.
  avg_purchase = df.groupby("Age")["Purchase Amount (USD)"].mean().reset_index()
```

[369]:
```
avg_purchase.plot.barh(legend=False, grid=True, xlabel="avg. purchase amount␣
↪(USD)", ylabel="age groups",
                          title="Average Purchase Amount (USD) among Age Groups",␣
↪edgecolor="black", color="skyblue")
```

[369]:
```
<Axes: title={'center': 'Average Purchase Amount (USD) among Age Groups'},
xlabel='avg. purchase amount (USD)', ylabel='age groups'>
```



0.1.12 Size Frequency

Most people buy items in **Medium** size, with more than **1,755 purchases**. On the other hand, **Extra Large** size is the least popular.

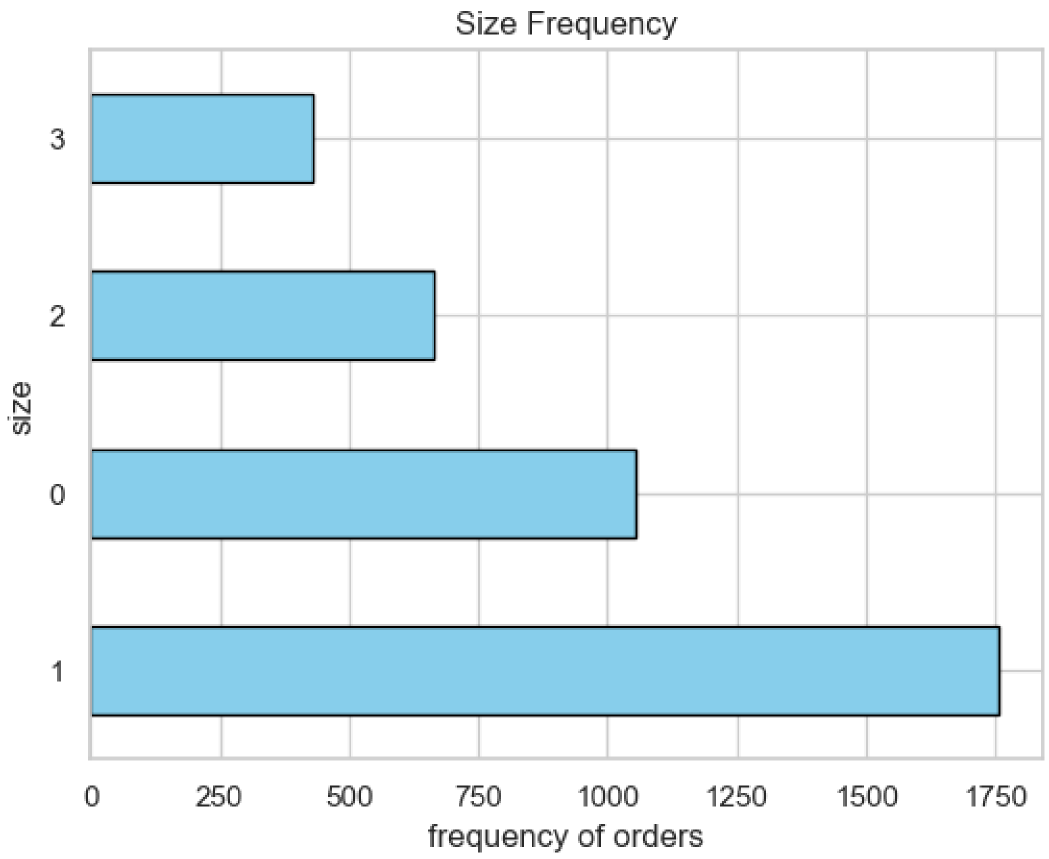
[370]:
```
size = df.groupby("Size").size().reset_index(name="counts").
↪sort_values(by="counts", ascending=False)
print(size)
```

```
  Size  counts
1    M    1755
0    L    1053
```

```
2    S      663
3   XL      429
```

[371]:
```
size.plot.barh(title="Size Frequency", grid=True, legend=False,␣
↪edgecolor="black",
                   xlabel="frequency of orders", ylabel="size", color="skyblue")
```

[371]:
```
<Axes: title={'center': 'Size Frequency'}, xlabel='frequency of orders',
ylabel='size'>
```



0.1.13 Seasonal Purchases Frequency

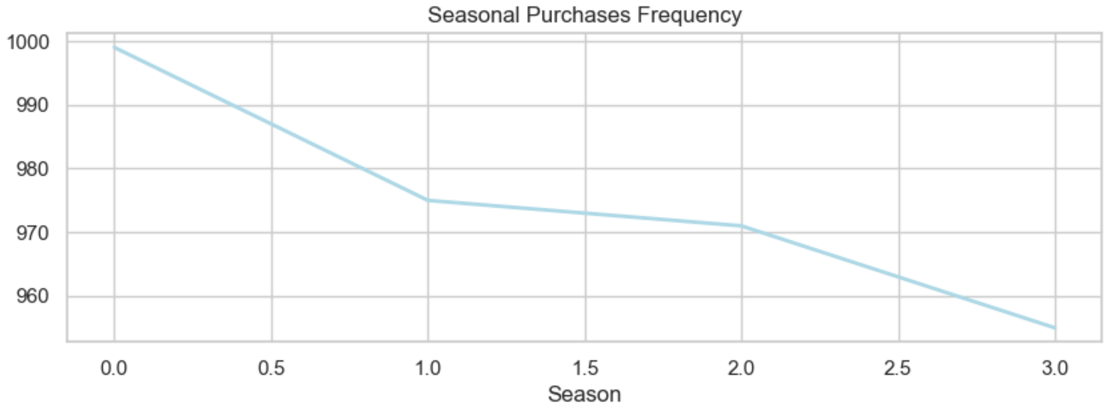
Spring has the highest number of purchases with **999**, making it the most popular season. **Summer** has the lowest number of purchases with **955**.

[372]:
```
seasons = df["Season"].value_counts().reset_index()
```

[373]:
```
seasons
```

[373]:
```
   Season  count
0  Spring    999
1    Fall    975
2  Winter    971
3  Summer    955
```

[374]:
```
seasons.plot(kind="line", title="Seasonal Purchases Frequency",␣
↪xlabel="Season", linewidth=2,
            figsize=(10,3), grid=True, legend=False, color="lightblue")
plt.show()
```



0.1.14 Seasonal Average Expenses

Consumers tend to spend the **most in Fall and Winter**, indicating a higher average expense during these seasons. On the other hand, **Summer and Spring see slightly lower spending**, with consumers spending a bit less during these warmer months.

[375]:
```
df.columns
```

[375]:
```
Index(['Age', 'Gender', 'Item Purchased', 'Category', 'Purchase Amount (USD)',
       'Location', 'Size', 'Color', 'Season', 'Review Rating',
       'Subscription Status', 'Payment Method', 'Shipping Type',
       'Discount Applied', 'Promo Code Used', 'Previous Purchases',
       'Preferred Payment Method', 'Frequency of Purchases'],
      dtype='object')
```

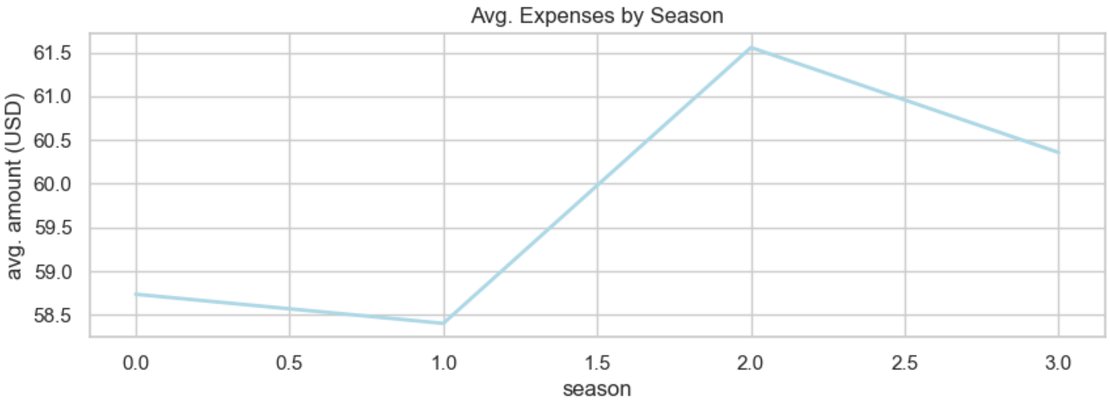
[376]:
```
season_amount = df.groupby("Season")["Purchase Amount (USD)"].mean().
↪reset_index()
```

```
C:\Users\nitin\AppData\Local\Temp\ipykernel_7964\4283095270.py:1: FutureWarning:
The default of observed=False is deprecated and will be changed to True in a
future version of pandas. Pass observed=False to retain current behavior or
observed=True to adopt the future default and silence this warning.
```

```
  season_amount = df.groupby("Season")["Purchase Amount
(USD)"].mean().reset_index()
```

[377]:
```
season_amount.plot(kind="line", legend=False, grid=True, color="lightblue",␣
↪linewidth=2,
                    ylabel="avg. amount (USD)", xlabel="season", figsize=(10,3),␣
↪title="Avg. Expenses by Season")
```

[377]:
```
<Axes: title={'center': 'Avg. Expenses by Season'}, xlabel='season',
ylabel='avg. amount (USD)'>
```



0.1.15 Seasonal Item Purchase

[378]:
```
seasonal_item = pd.crosstab(df["Item Purchased"],df["Season"])
seasonal_item = seasonal_item.T
seasonal_item
```

[378]:
```
Item Purchased  Backpack  Belt  Blouse  Boots  Coat  Dress  Gloves  Handbag  \
Season
Spring                39    41      46     40    46     43      42       36
Summer                45    39      43     38    42     47      29       35
Fall                  34    41      42     35    34     36      37       48
Winter                25    40      40     31    39     40      32       34

Item Purchased  Hat  Hoodie  …  Scarf  Shirt  Shoes  Shorts  Skirt  \
Season                       …
Spring           27      36  …     41     42     40      47     46
Summer           37      31  …     43     38     46      40     28
Fall             50      36  …     40     39     26      35     46
Winter           40      48  …     33     50     38      35     38

Item Purchased  Sneakers  Socks  Sunglasses  Sweater  T-shirt
```

```
Season
Spring                39     40           33       52       38
Summer                36     42           37       28       30
Fall                  31     42           39       42       39
Winter                39     35           52       42       40

[4 rows x 25 columns]
```

[379]:
```
seasonal_item.idxmax(axis=1)
```

[379]:
```
Season
Spring       Sweater
Summer         Pants
Fall          Jacket
Winter    Sunglasses
dtype: object
```

0.1.16 Colors

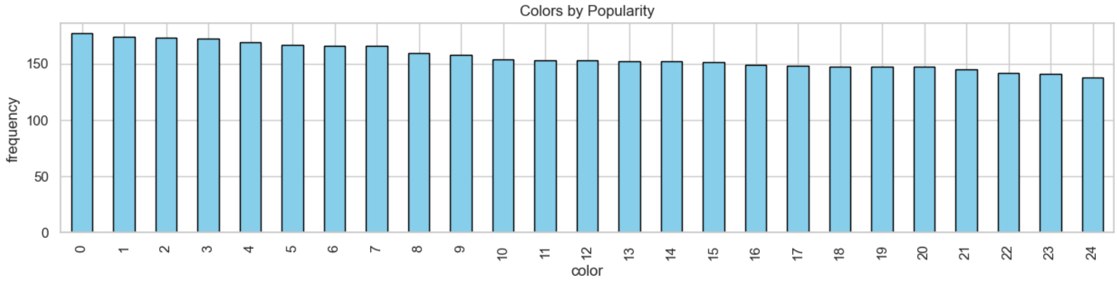
Most consumers prefer the **color Olive**, followed by **Yellow**, **Silver**, and **Teal**. Meanwhile,**Gold** is the least preferred color.

[380]:
```
colors = df["Color"].value_counts().reset_index()
colors.head()
```

[380]:
```
    Color  count
0   Olive    177
1  Yellow    174
2  Silver    173
3    Teal    172
4   Green    169
```

[381]:
```
colors.plot(kind="bar", title="Colors by Popularity", xlabel="color",␣
 ↪ylabel="frequency",
            figsize=(15,3), grid=True, edgecolor="black", legend=False,␣
 ↪color="skyblue")
```

[381]:
```
<Axes: title={'center': 'Colors by Popularity'}, xlabel='color',
ylabel='frequency'>
```



0.1.17 Subscription Status

Most customers **do not have a subscription** with a total **2847**. This is against the total of **1053** who have a **current subscription**.

[382]:
```
subs = pd.crosstab(df["Subscription Status"],columns = "counts").sort_values(by␣
 ↪= "counts",ascending=True)
subs
```

[382]:
```
col_0                counts
Subscription Status
Yes                    1053
No                     2847
```

[383]:
```
subsi = df["Subscription Status"].value_counts(ascending= False).reset_index()
subsi
```

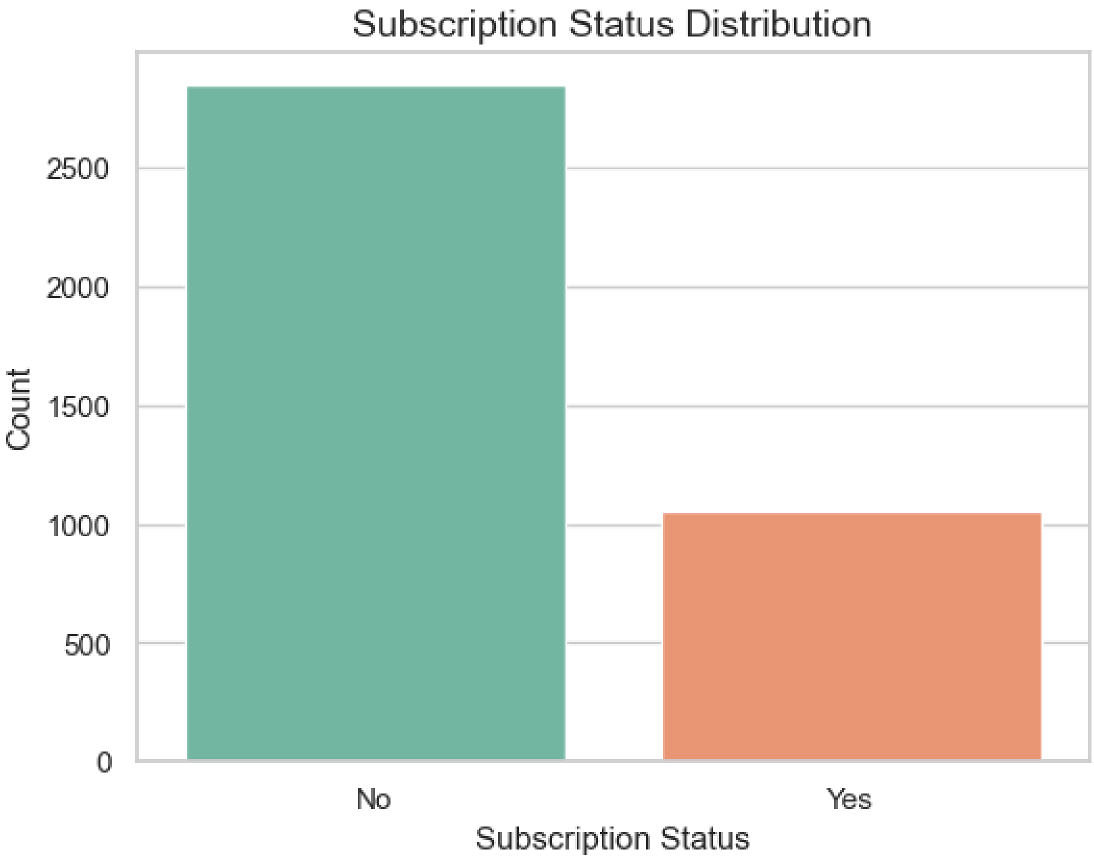
[383]:
```
  Subscription Status  count
0                  No   2847
1                 Yes   1053
```

[384]:
```
sns.barplot(data=subsi,x = "Subscription Status",y = "count",palette='Set2')
plt.title('Subscription Status Distribution', fontsize=14)
plt.xlabel('Subscription Status', fontsize=12)
plt.ylabel('Count', fontsize=12)
plt.show()
```

```
C:\Users\nitin\AppData\Local\Temp\ipykernel_7964\2664632087.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(data=subsi,x = "Subscription Status",y = "count",palette='Set2')
```



0.1.18 Payment Method

As observed, the most commonly preferred payment method is **Credit Card**. Following that, people tend to use **Venmo**, **Cash**, and **Paypal** as their payment options. However, **Debit Card** and **Bank Transfer** are the least preferred choices for payment, with fewer people using these methods compared to others.

```
[385]: payment_type = df["Payment Method"].value_counts().reset_index()
       payment_type
```

```
[385]:   Payment Method  count
       0    Credit Card    696
       1          Venmo    653
       2           Cash    648
       3         PayPal    638
       4     Debit Card    633
       5  Bank Transfer    632
```

```
[386]:
```

```
plt.figure(figsize=(8, 6))
df.set_index('Payment Method')['Count'].plot.pie(autopct='%1.1f%%',␣
 ↪startangle=140, colors=plt.cm.Paired.colors, legend=False)
plt.title('Distribution of Payment Methods')
plt.show()
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
File c:
 ↪\Users\nitin\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\indexes\
 ↪py:3805, in Index.get_loc(self, key)
   3804 try:
-> 3805     return self._engine.get_loc(casted_key)
   3806 except KeyError as err:

File index.pyx:167, in pandas._libs.index.IndexEngine.get_loc()

File index.pyx:196, in pandas._libs.index.IndexEngine.get_loc()

File pandas\\_libs\\hashtable_class_helper.pxi:7081, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()

File pandas\\_libs\\hashtable_class_helper.pxi:7089, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()

KeyError: 'Count'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
Cell In[386], line 2
      1 plt.figure(figsize=(8, 6))
----> 2 df.set_index('Payment Method')['Count'].plot.pie(autopct='%1.1f%%',␣
 ↪startangle=140, colors=plt.cm.Paired.colors, legend=False)
      3 plt.title('Distribution of Payment Methods')
      4 plt.show()

File c:
 ↪\Users\nitin\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\frame.
 ↪py:4102, in DataFrame.__getitem__(self, key)
   4100 if self.columns.nlevels > 1:
   4101     return self._getitem_multilevel(key)
-> 4102 indexer = self.columns.get_loc(key)
   4103 if is_integer(indexer):
   4104     indexer = [indexer]
```

```
File c:
 ↪\Users\nitin\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\indexes\
 ↪py:3812, in Index.get_loc(self, key)
   3807     if isinstance(casted_key, slice) or (
   3808         isinstance(casted_key, abc.Iterable)
   3809         and any(isinstance(x, slice) for x in casted_key)
   3810     ):
   3811         raise InvalidIndexError(key)
-> 3812     raise KeyError(key) from err
   3813 except TypeError:
   3814     # If we have a listlike key, _check_indexing_error will raise
   3815     #  InvalidIndexError. Otherwise we fall through and re-raise
   3816     #  the TypeError.
   3817     self._check_indexing_error(key)

KeyError: 'Count'
```

```
<Figure size 800x600 with 0 Axes>
```

[390]:
```
df.columns
```

[390]:
```
Index(['Age', 'Gender', 'Item Purchased', 'Category', 'Purchase Amount (USD)',
       'Location', 'Size', 'Color', 'Season', 'Review Rating',
       'Subscription Status', 'Payment Method', 'Shipping Type',
       'Discount Applied', 'Promo Code Used', 'Previous Purchases',
       'Preferred Payment Method', 'Frequency of Purchases'],
      dtype='object')
```

0.1.19 Shipping Type Distribution¶

[400]:
```
shipping_type = df["Shipping Type"].value_counts().reset_index()
shipping_type
```

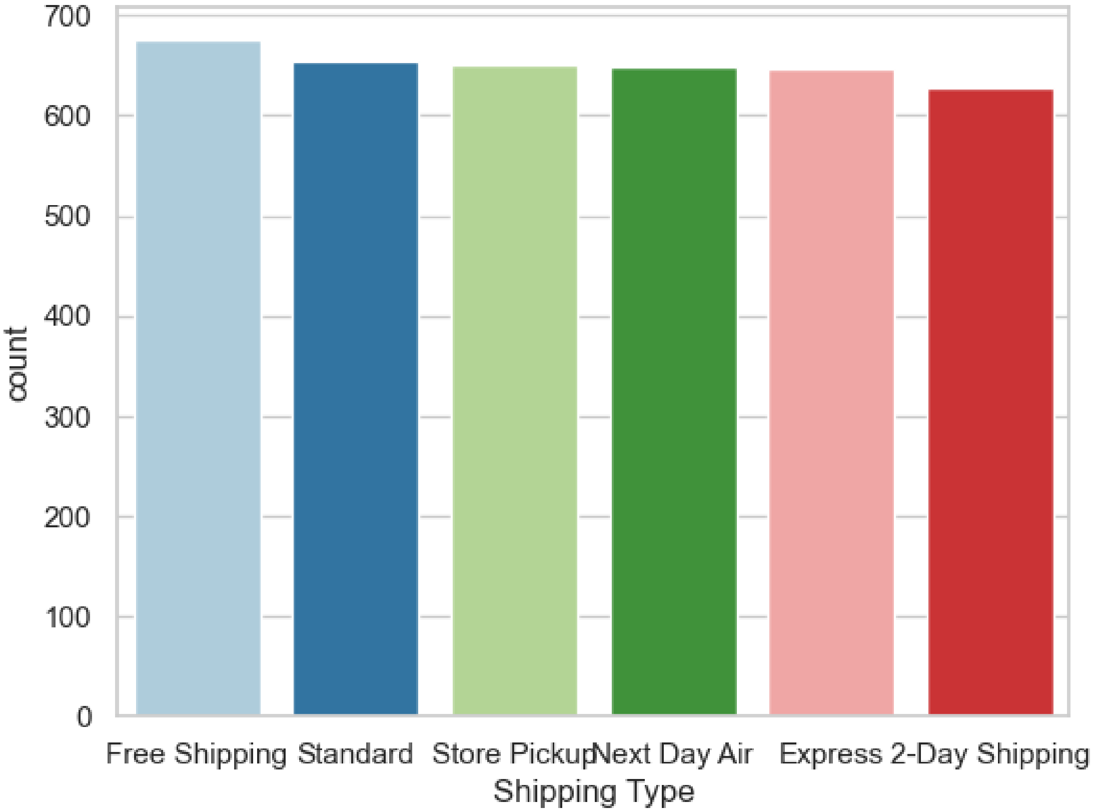
[400]:
```
    Shipping Type  count
0   Free Shipping    675
1        Standard    654
2    Store Pickup    650
3    Next Day Air    648
4         Express    646
5  2-Day Shipping    627
```

[408]:
```
sns.barplot(data = shipping_type ,x ="Shipping Type",y = "count"␣
 ↪,palette="Paired")
plt.show()
```

```
C:\Users\nitin\AppData\Local\Temp\ipykernel_7964\4157131399.py:1: FutureWarning:
```

```
Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.barplot(data = shipping_type ,x ="Shipping Type",y = "count"
,palette="Paired")
```



0.1.20 Most Preferred Shipping Type among Locations

[410]:
```
ship_loc = pd.crosstab(df["Location"], columns=df["Shipping Type"])
ship_loc
```

[410]:
```
Shipping Type   2-Day Shipping  Express  Free Shipping  Next Day Air  \
Location
Alabama                     16       20              9            17
Alaska                      13       13             14             8
Arizona                     10       12             10            11
Arkansas                    12       16             14            17
California                  14       16             16            20
Colorado                    12        4             13            14
Connecticut                 16       11              6            18
```

```
Delaware                    14       14             20            16
Florida                     10       14             17             8
Georgia                     13       13             18             9
Hawaii                       7       13             10            11
Idaho                       15       14             16            16
Illinois                    13       17             14             9
Indiana                     13       14             12            10
Iowa                         7       14             10             9
Kansas                       9       12             12            11
Kentucky                     9       20             12            12
Louisiana                   12       13             14            20
Maine                        9       17             20             8
Maryland                    18       13             15            15
Massachusetts               14        6              8            19
Michigan                    12       10             15             9
Minnesota                   15       15             13            15
Mississippi                 16       10             15            13
Missouri                     8       16             13            13
Montana                     19        7             22            15
Nebraska                    12       16             12            12
Nevada                      15       11             17            18
New Hampshire               12       13             11            11
New Jersey                  11       12             10            11
New Mexico                  12       13             15            16
New York                    14       13             20            15
North Carolina              14        8             15            13
North Dakota                14       19             18             9
Ohio                        18        7             13            11
Oklahoma                     5       17              8            14
Oregon                      19       12             11            13
Pennsylvania                14       12             15            14
Rhode Island                 7       12              7            11
South Carolina              13       14             12            14
South Dakota                15       14              7            11
Tennessee                   17       10             12            12
Texas                        7       12             19            11
Utah                         8       10             16            15
Vermont                     17       11             13            14
Virginia                     9       17             14             9
Washington                  13       12             14            13
West Virginia               15       13             15            14
Wisconsin                   12       11             13            12
Wyoming                      8       13             10            12

Shipping Type   Standard  Store Pickup
Location
Alabama                8            19
```

Alaska	17	7
Arizona	10	12
Arkansas	11	9
California	16	13
Colorado	14	18
Connecticut	12	15
Delaware	9	13
Florida	13	6
Georgia	8	18
Hawaii	10	14
Idaho	14	18
Illinois	15	24
Indiana	10	20
Iowa	12	17
Kansas	10	9
Kentucky	13	13
Louisiana	14	11
Maine	13	10
Maryland	18	7
Massachusetts	14	11
Michigan	16	11
Minnesota	16	14
Mississippi	11	15
Missouri	19	12
Montana	17	16
Nebraska	16	19
Nevada	16	10
New Hampshire	9	15
New Jersey	12	11
New Mexico	12	13
New York	15	10
North Carolina	13	15
North Dakota	12	11
Ohio	11	17
Oklahoma	13	18
Oregon	12	7
Pennsylvania	11	8
Rhode Island	14	12
South Carolina	14	9
South Dakota	12	11
Tennessee	13	13
Texas	12	16
Utah	15	7
Vermont	20	10
Virginia	17	11
Washington	9	12
West Virginia	9	15

```
Wisconsin             16            11
Wyoming               11            17
```

[413]:

```
most_preferred = pd.DataFrame(ship_loc.idxmax(axis=1)).rename(columns={0:"Most␣
 ↪Preferred Shipping Type"})
most_preferred
```

[413]:

```
               Most Preferred Shipping Type
Location
Alabama                             Express
Alaska                             Standard
Arizona                             Express
Arkansas                       Next Day Air
California                     Next Day Air
Colorado                       Store Pickup
Connecticut                    Next Day Air
Delaware                      Free Shipping
Florida                       Free Shipping
Georgia                       Free Shipping
Hawaii                         Store Pickup
Idaho                          Store Pickup
Illinois                       Store Pickup
Indiana                        Store Pickup
Iowa                           Store Pickup
Kansas                              Express
Kentucky                            Express
Louisiana                      Next Day Air
Maine                         Free Shipping
Maryland                     2-Day Shipping
Massachusetts                  Next Day Air
Michigan                           Standard
Minnesota                          Standard
Mississippi                  2-Day Shipping
Missouri                           Standard
Montana                       Free Shipping
Nebraska                       Store Pickup
Nevada                         Next Day Air
New Hampshire                  Store Pickup
New Jersey                          Express
New Mexico                     Next Day Air
New York                      Free Shipping
North Carolina                Free Shipping
North Dakota                        Express
Ohio                         2-Day Shipping
Oklahoma                       Store Pickup
Oregon                       2-Day Shipping
Pennsylvania                  Free Shipping
```

```
Rhode Island                           Standard
South Carolina                          Express
South Dakota                     2-Day Shipping
Tennessee                        2-Day Shipping
Texas                             Free Shipping
Utah                              Free Shipping
Vermont                                Standard
Virginia                                Express
Washington                        Free Shipping
West Virginia                    2-Day Shipping
Wisconsin                              Standard
Wyoming                            Store Pickup
```

0.1.21 Discount Applied

Most orders **do not have any discount applied** with a total of **2223**. Whilst the total no. of orders with **discount applied** is **1677**.

[414]:

```
discount = pd.crosstab(df["Discount Applied"],columns="counts")
discount
```

[414]:

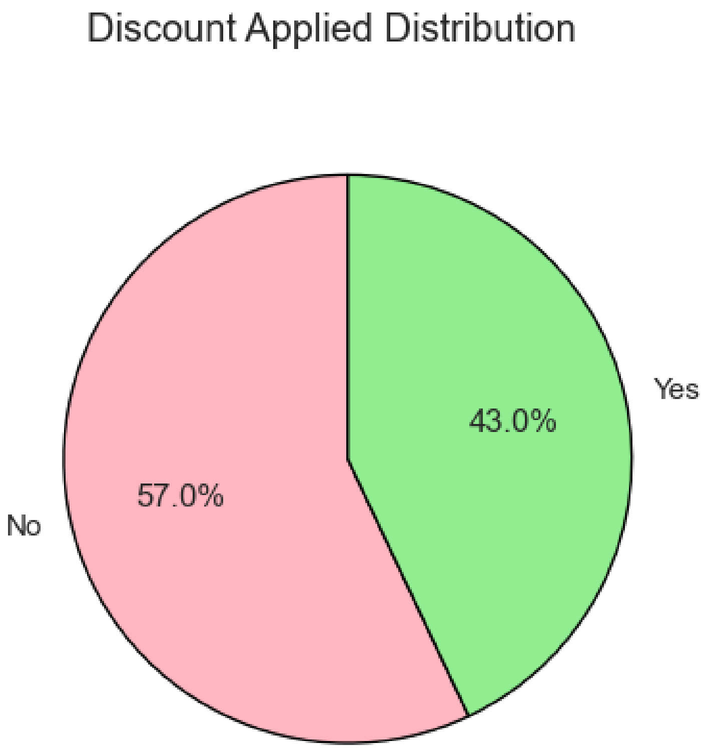
```
col_0             counts
Discount Applied
No                  2223
Yes                 1677
```

[415]:

```
discount.plot(kind="pie", subplots=True, xlabel="", ylabel="", title="Discount␣
 ↪Applied Distribution",
              autopct='%1.1f%%', startangle=90, legend=False,␣
 ↪colors=["lightpink", "lightgreen"],
             wedgeprops={"edgecolor": "black"})
```

[415]:

```
array([<Axes: >], dtype=object)
```

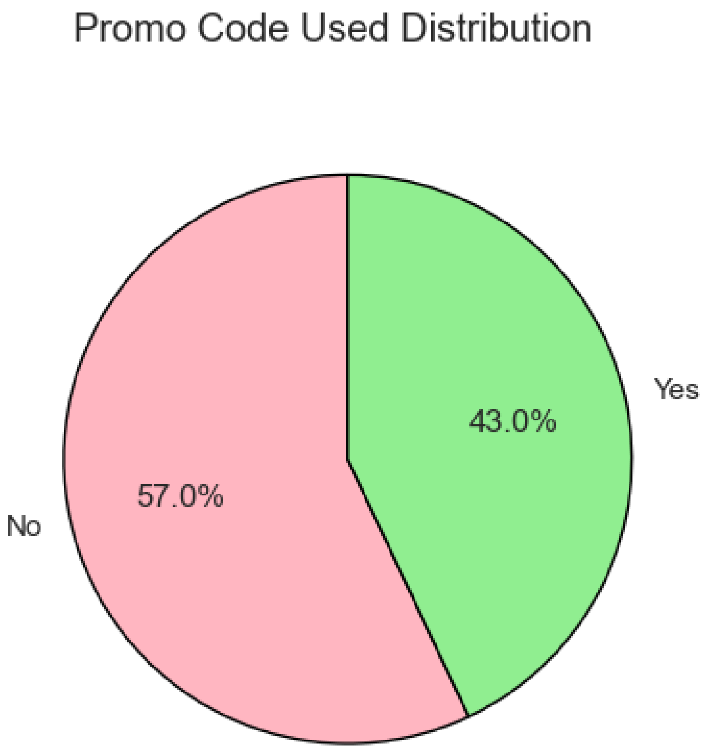


```
[417]: promo = pd.crosstab(df["Promo Code Used"], columns="counts")
       promo
```

```
[417]: col_0            counts
       Promo Code Used
       No                 2223
       Yes                1677
```

```
[421]: promo.plot(kind="pie", subplots=True, xlabel="", ylabel="", title="Promo Code␣
        ↪Used Distribution",
                     autopct='%1.1f%%', startangle=90, legend=False,␣
        ↪colors=["lightpink", "lightgreen"],
                    wedgeprops={"edgecolor": "black"})
```

```
[421]: array([<Axes: >], dtype=object)
```



0.1.22 Frequency of Purchases

Most consumers purchase **every 3 months**, followed by those who purchased **annually, quarterly, and monthly.**

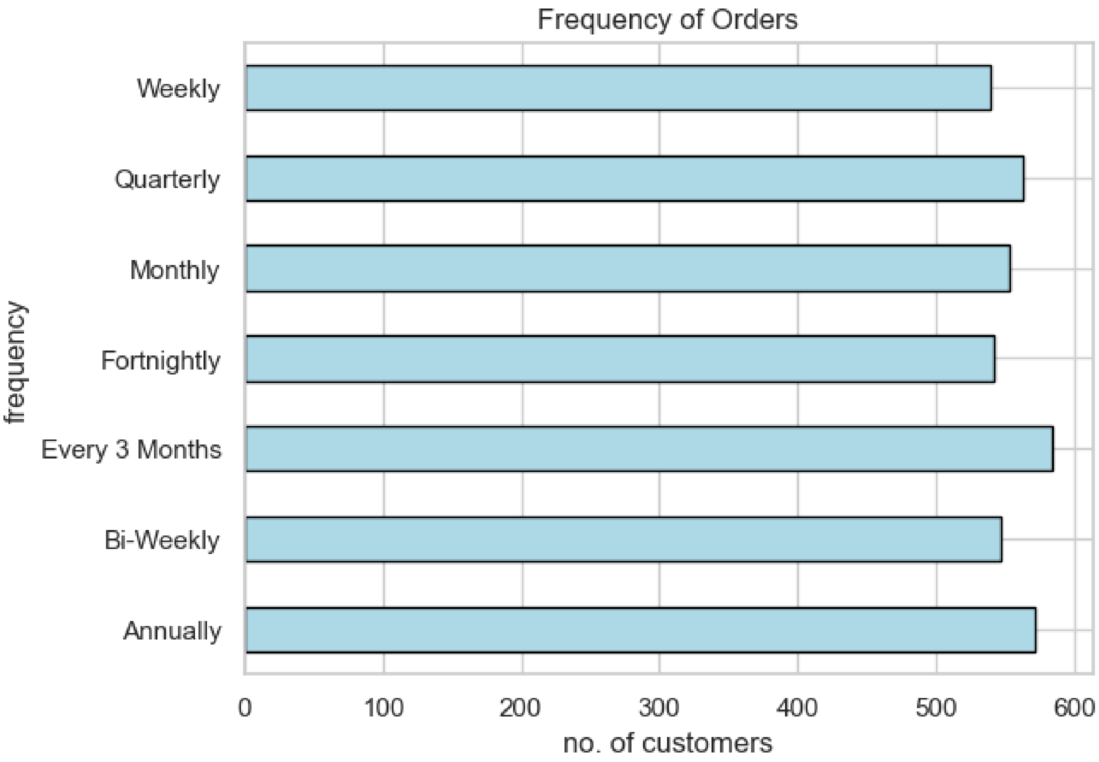
```
[426]: fre_purchase = pd.crosstab(df["Frequency of Purchases"],columns="Counts",)
       fre_purchase.sort_values(by = "Counts",ascending= False)
```

```
[426]: col_0                   Counts
       Frequency of Purchases
       Every 3 Months             584
       Annually                   572
       Quarterly                  563
       Monthly                    553
       Bi-Weekly                  547
       Fortnightly                542
       Weekly                     539
```

```
[428]:
```

```
fre_purchase.plot.barh(title="Frequency of Orders", xlabel="no. of customers",␣
↪ylabel="frequency",
                    grid=True, legend=False, color="lightblue",␣
↪edgecolor="black")
```

```
[428]: <Axes: title={'center': 'Frequency of Orders'}, xlabel='no. of customers',
ylabel='frequency'>
```



0.1.23 Previous Purchases

- The highest number of previous purchases is 50, with a total of 77 customers.
- The most common number of previous purchases among customers is 37.

```
[435]: prev_purchase = df["Previous Purchases"].value_counts().reset_index().head()
prev_purchase
```

```
[435]:    Previous Purchases  count
0                       31     97
1                       21     96
2                       14     92
3                        4     91
4                        3     91
```

[441]:

```
prev_purchase = prev_purchase.sort_values(by ="count")
prev_purchase
```

[441]:

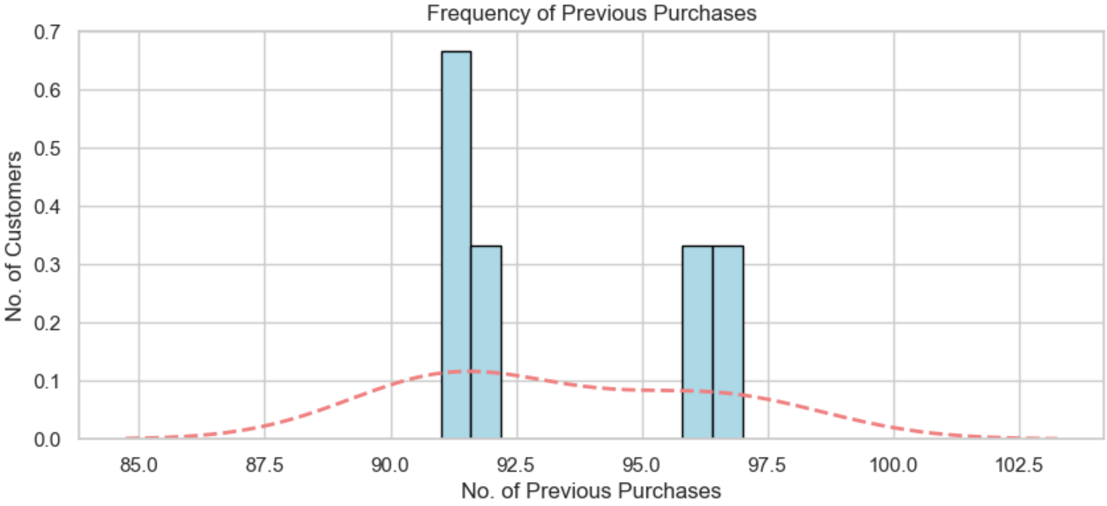
```
   Previous Purchases  count
3                   4     91
4                   3     91
2                  14     92
1                  21     96
0                  31     97
```

[447]:

```
ax = prev_purchase.plot.hist(y="count", color="lightblue", edgecolor="black",
                             title="Frequency of Previous Purchases",
                             xlabel="No. of Previous Purchases", ylabel="No. of␣
 ↪Customers",
                             grid=True, legend=False, figsize=(10, 4),␣
 ↪density=True)
sns.kdeplot(prev_purchase["count"], color="lightcoral", lw=2, linestyle="--")
```

[447]:

```
<Axes: title={'center': 'Frequency of Previous Purchases'}, xlabel='No. of
Previous Purchases', ylabel='No. of Customers'>
```



0.1.24 Previous Purchases Among Age Groups

The **Seniors** tend to have more previous purchases compared to other age groups, which is then followed by **Adults** and **Young Adults**. Meanwhile, **Teenangers** have the lowest mean for previous purchases.

[449]:

```
age_prevpur = df.groupby("Age")["Previous Purchases"].mean()
age_prevpur = pd.DataFrame(age_prevpur).rename_axis("Age Groups")
```

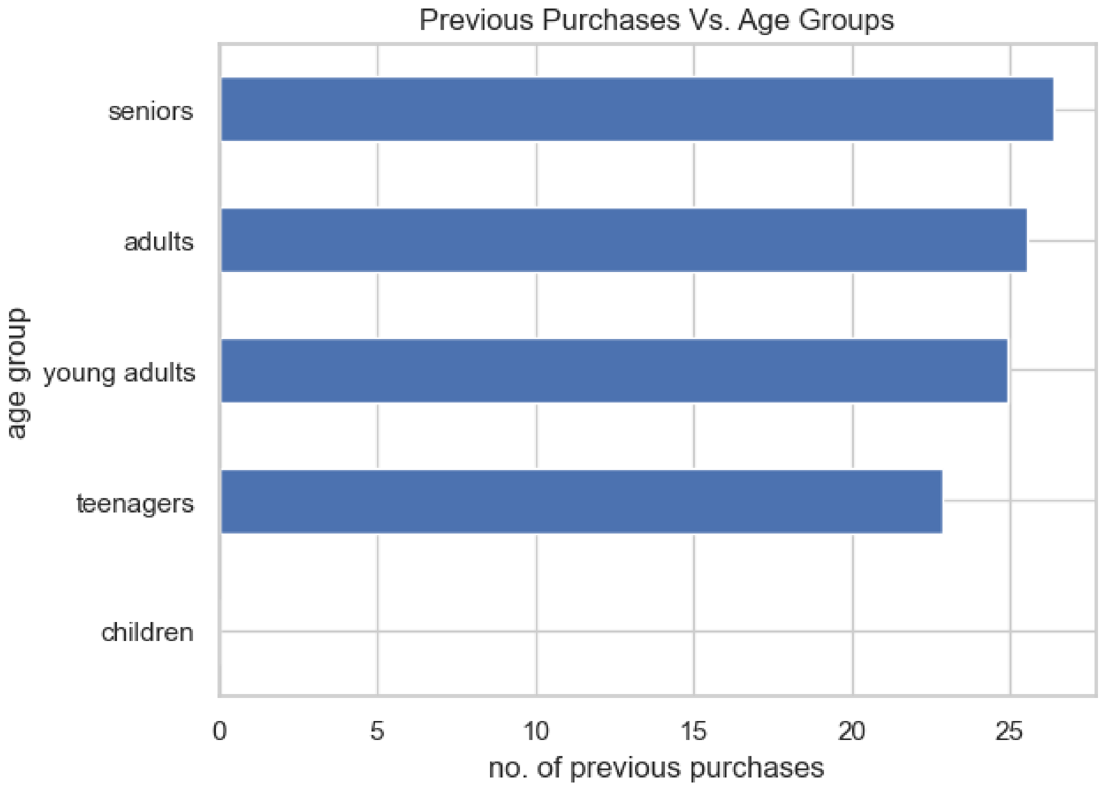
```
age_prevpur
```

```
C:\Users\nitin\AppData\Local\Temp\ipykernel_7964\2746215112.py:1: FutureWarning:
The default of observed=False is deprecated and will be changed to True in a
future version of pandas. Pass observed=False to retain current behavior or
observed=True to adopt the future default and silence this warning.
  age_prevpur = df.groupby("Age")["Previous Purchases"].mean()
```

[449]:
```
              Previous Purchases
Age Groups
children                     NaN
teenagers              22.893333
young adults           24.938951
adults                 25.539352
seniors                26.388759
```

[450]:
```
age_prevpur.plot.barh(grid=True, legend=False, xlabel="no. of previous␣
↪purchases", ylabel="age group",
                      title="Previous Purchases Vs. Age Groups")
```

[450]:
```
<Axes: title={'center': 'Previous Purchases Vs. Age Groups'}, xlabel='no. of
previous purchases', ylabel='age group'>
```



[ ]:

```
```