In [4]:
```
#Load all packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Set a global seaborn theme with magma palette
sns.set_theme(style="darkgrid", palette="magma")
```

In [5]:
```
# Load the dataset and display the first few rows
shopdata = pd.read_csv(r"C:\Users\ADMIN\Desktop\shopping_trends.csv")
shopdata.head()
```

Out[5]:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size
0	1	55	Male	Blouse	Clothing	53	Kentucky	L
1	2	19	Male	Sweater	Clothing	64	Maine	L
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M
4	5	45	Male	Blouse	Clothing	49	Oregon	M

In [6]:
```
# Check number of rows and columns in dataset
shopdata.shape
```

Out[6]:
```
(3900, 19)
```

In [7]:
```
# Summary statistics of the dataset
shopdata.describe()
```

Out[7]:

	Customer ID	Age	Purchase Amount (USD)	Review Rating	Previous Purchases
count	3900.000000	3900.000000	3900.000000	3900.000000	3900.000000
mean	1950.500000	44.068462	59.764359	3.749949	25.351538
std	1125.977353	15.207589	23.685392	0.716223	14.447125
min	1.000000	18.000000	20.000000	2.500000	1.000000
25%	975.750000	31.000000	39.000000	3.100000	13.000000
50%	1950.500000	44.000000	60.000000	3.700000	25.000000
75%	2925.250000	57.000000	81.000000	4.400000	38.000000
max	3900.000000	70.000000	100.000000	5.000000	50.000000

In [8]:
```
# Get a concise summary of the DataFrame including data types and non-nul
shopdata.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
```

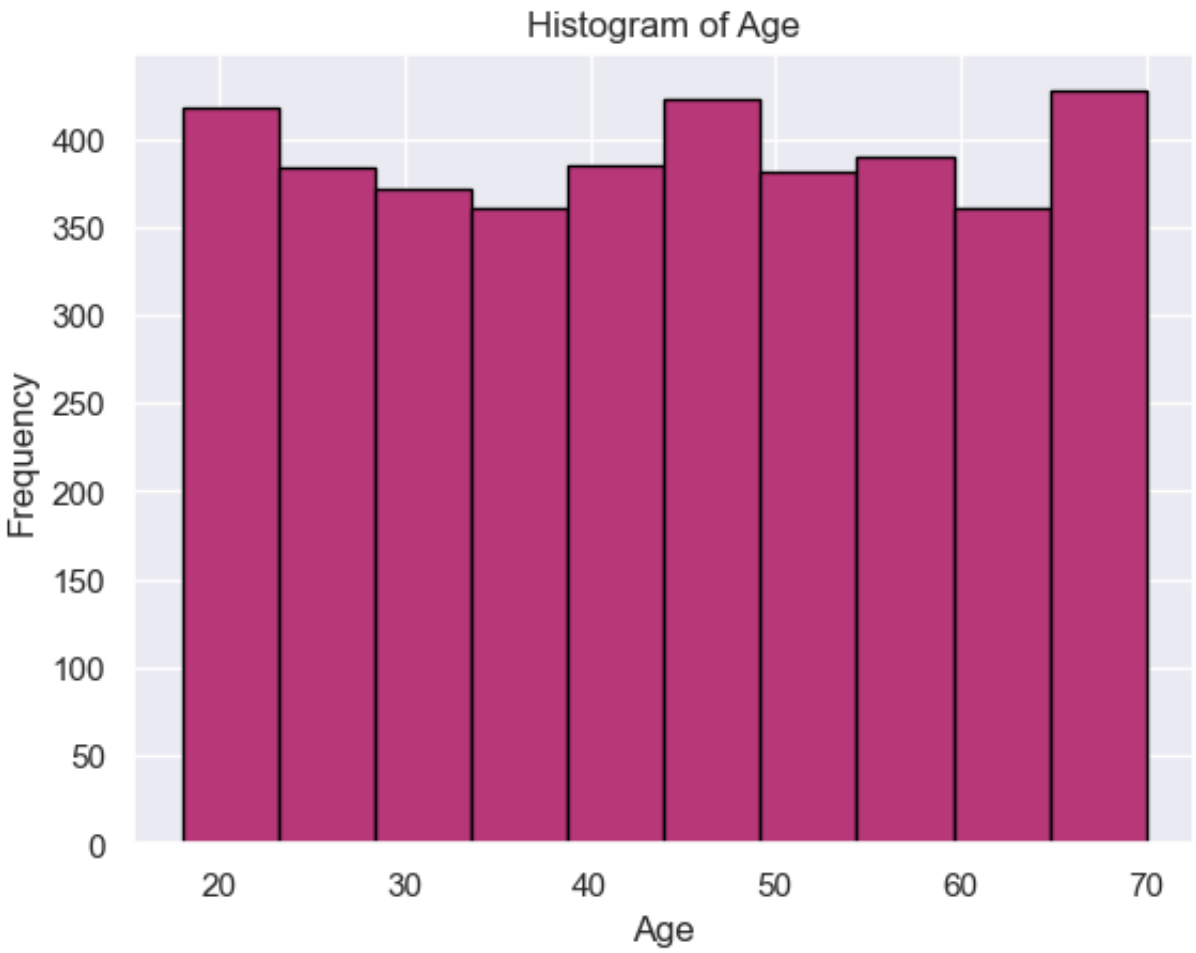
In [9]:
```
# Checking for missing values in each column
shopdata.isna().sum()
```

Out[9]:
```
Customer ID                 0
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
```

In [10]:
```
# Checking for duplicate rows in the dataset
shopdata.duplicated().sum()
```

Out[10]:
```
0
```

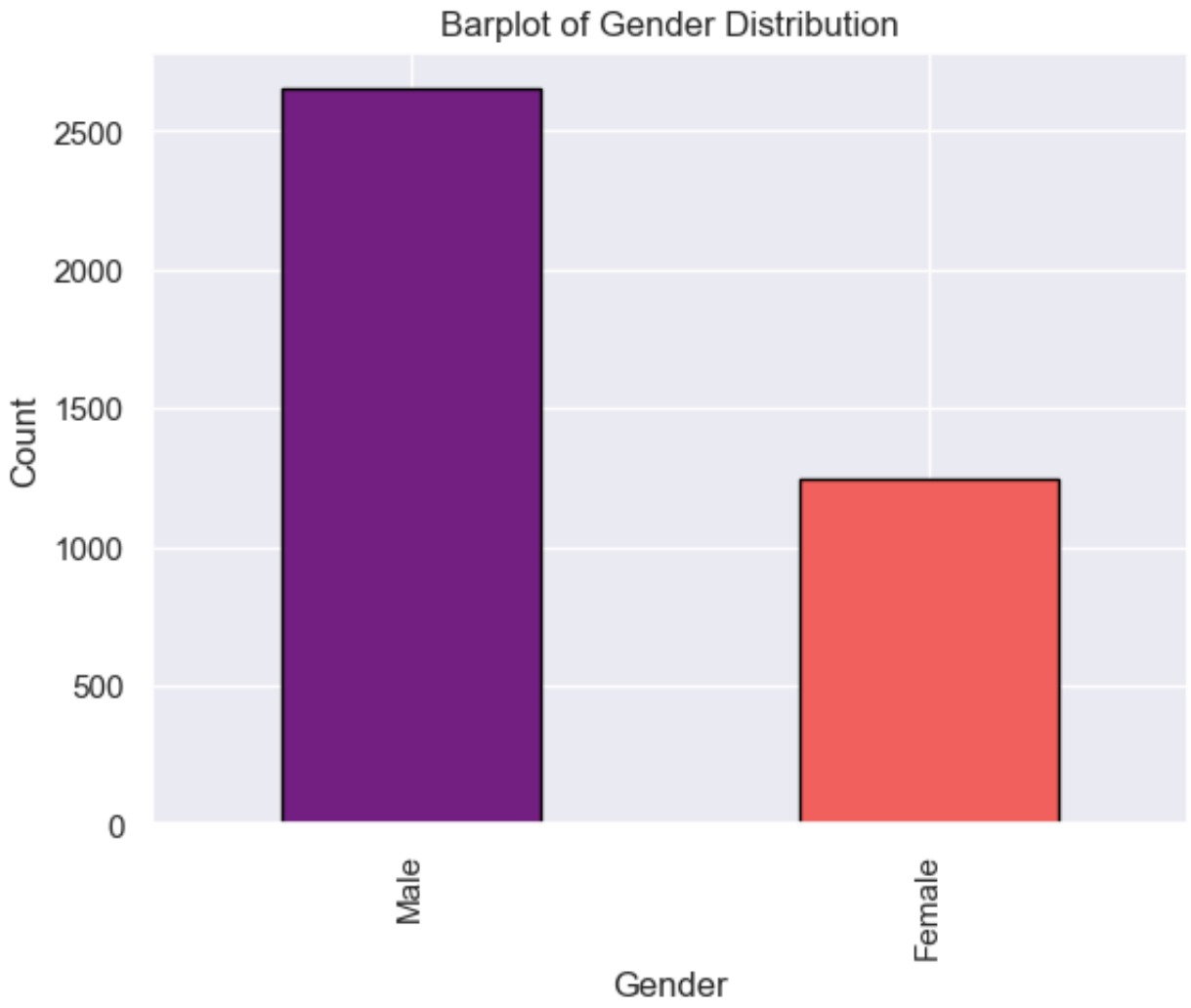
In [11]:
```
# VISUALIZATION
# Histogram showing the frequency distribution of the Age
age_color = sns.color_palette("magma", 1)[0]
shopdata["Age"].plot(kind="hist", edgecolor="black", color=age_color)
plt.title("Histogram of Age")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()
```



In [12]:
```
# List all columns present in the dataset
shopdata.columns
```

Out[12]:
```
Index(['Customer ID', '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')
```

In [18]:
```
# Barplot of Gender Distribution
gender_counts = shopdata["Gender"].value_counts()
gender_colors = sns.color_palette("magma", len(gender_counts))
gender_counts.plot(kind="bar", color=gender_colors, edgecolor="black")
plt.title("Barplot of Gender Distribution")
plt.xlabel("Gender")
plt.ylabel("Count")
plt.show()
```



In [21]:
```
# Calculate and display the top 2 categories with the highest average pur
shopdata.groupby("Category")["Purchase Amount (USD)"].mean().sort_values(
```

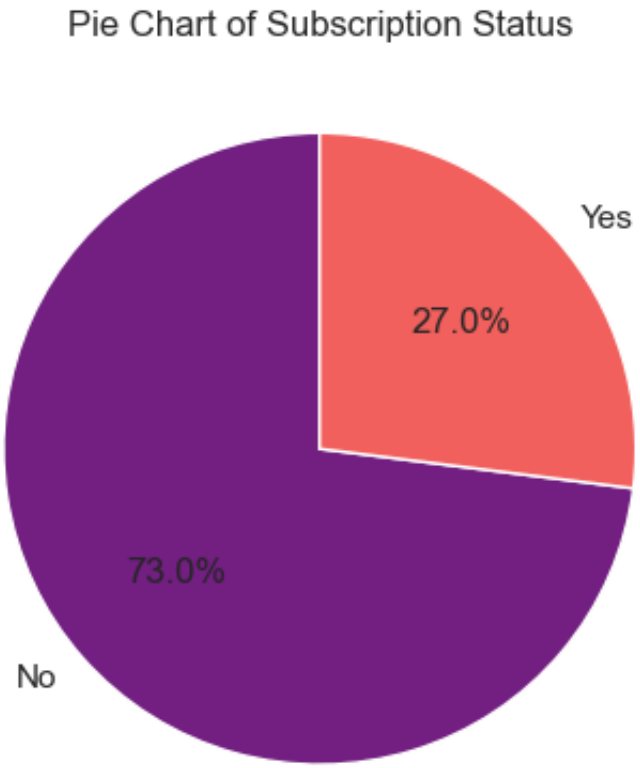
Out[21]:
```
Category
Footwear    60.255426
Clothing    60.025331
Name: Purchase Amount (USD), dtype: float64
```

In [24]:
```
# Sorting dataset based on Review Rating in descending order
shopdata.sort_values("Review Rating", ascending = False, inplace = True)
shopdata.head()
```

Out[24]:

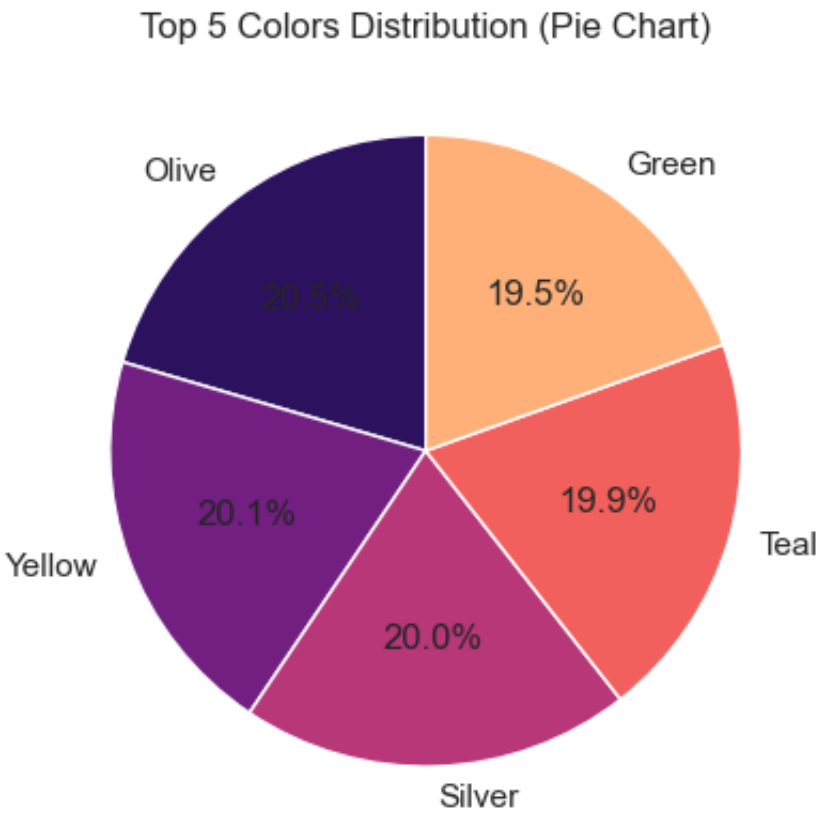
	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size
776	777	49	Male	Shirt	Clothing	60	Alabama	M
1664	1665	19	Male	Handbag	Accessories	53	Minnesota	M
1277	1278	19	Male	Blouse	Clothing	97	Rhode Island	L
2632	2633	24	Male	Scarf	Accessories	27	Alaska	M
965	966	43	Male	Boots	Footwear	55	Delaware	L

In [26]:
```
# Pie chart to display the proportion of each subscription status.
subs_counts = shopdata["Subscription Status"].value_counts()
subs_colors = sns.color_palette("magma", len(subs_counts))
subs_counts.plot(kind="pie", autopct="%1.1f%%", startangle=90, colors=sub
plt.title("Pie Chart of Subscription Status")
plt.ylabel("")  # Hide default y-label
plt.savefig("Pie_Chart_of_Subscription_Status", dpi=300, bbox_inches ="ti
plt.show()
```



In [30]:
```
# Pie chart for top 5 colors by frequency (head of value counts)
```

```
top5_colors = shopdata["Color"].value_counts().head()
colors_palette = sns.color_palette("magma", len(top5_colors))
top5_colors.plot(kind="pie", autopct="%1.1f%%", startangle=90, colors=col
plt.title("Top 5 Colors Distribution (Pie Chart)")
plt.ylabel("")
plt.savefig("Top 5 Colors Distribution (Pie Chart)", dpi=300, bbox_inches
plt.show()
```



In [34]:
```
# Scatter plot showing the relationship between Previous Purchases and Re

scatter_color = sns.color_palette("magma", 1)[0]
shopdata.plot(x="Previous Purchases", y="Review Rating", kind="scatter",
plt.title("Scatter Plot: Previous Purchases vs. Review Rating")
plt.xlabel("Previous Purchases")
plt.ylabel("Review Rating")
plt.savefig("ScatterPlot_PreviousPurchases_vs_Review_Rating", dpi=300, bb
plt.show()
```

In [62]:
```
# Sum of Purchase Amounts grouped by Season
shopdata.groupby("Season")["Purchase Amount (USD)"].sum()
```

Out[62]:
```
Season
Fall      60018
Spring    58679
Summer    55777
Winter    58607
Name: Purchase Amount (USD), dtype: int64
```

In [36]:
```
# Boxplot of Purchase Amount by Frequency of Purchases
# Helps identify the distribution and outliers in purchase amounts across

shopdata.boxplot(column="Purchase Amount (USD)", by="Frequency of Purchas
                 boxprops=dict(facecolor=sns.color_palette("magma", 1)[0]
plt.title("Boxplot of Purchase Amount by Frequency of Purchases")
plt.suptitle("")  # Remove automatic subtitle
plt.xlabel("Frequency of Purchases")
plt.ylabel("Purchase Amount (USD)")
plt.xticks(rotation=45, fontsize=8)
plt.savefig("Boxplot_of_Purchase_Amount_by_Frequency_of_Purchases", dpi=3
plt.show()
```

In [28]:
```
# Compare Preferred Payment Method frequency (lowest vs. highest)
low = shopdata["Preferred Payment Method"].value_counts().sort_values(asc
high = shopdata["Preferred Payment Method"].value_counts().sort_values(as
print("Least frequent Preferred Payment Method:", low)
print("Most frequent Preferred Payment Method:", high)
```

```
Least frequent Preferred Payment Method: Preferred Payment Method
Bank Transfer    612
Name: count, dtype: int64
Most frequent Preferred Payment Method: Preferred Payment Method
PayPal    677
Name: count, dtype: int64
```

In [82]:
```
# Pivot Table: Average Purchase Amount by Location and Item Purchased
pd.pivot_table(shopdata, values = "Purchase Amount (USD)", index = "Locat
```

Out[82]:

Item Purchased	Backpack	Belt	Blouse	Boots	Coat	Dre
Location						
Alabama	74.000000	26.500000	52.000000	73.666667	54.500000	55.2000
Alaska	74.600000	76.750000	67.500000	56.000000	45.500000	74.5000
Arizona	48.600000	84.500000	61.000000	NaN	65.333333	71.0000

Arkansas	90.000000	55.000000	66.000000	50.000000	54.250000	62.8000
California	57.400000	61.666667	39.500000	62.500000	55.000000	58.4285
Colorado	57.000000	68.500000	55.000000	NaN	46.000000	39.7500
Connecticut	60.000000	47.666667	67.400000	64.250000	44.500000	65.0000
Delaware	33.000000	41.500000	42.200000	44.500000	57.500000	46.0000
Florida	59.000000	94.000000	31.000000	51.500000	51.166667	48.5000
Georgia	44.750000	62.000000	64.857143	85.000000	93.000000	69.7500
Hawaii	96.000000	44.000000	50.600000	74.000000	69.666667	42.0000
Idaho	53.250000	61.000000	65.000000	60.333333	65.250000	47.5000
Illinois	93.000000	66.800000	55.000000	60.250000	44.000000	61.3333
Indiana	44.500000	63.400000	33.333333	62.333333	48.666667	Na
Iowa	51.000000	49.666667	59.500000	49.333333	40.666667	92.5000
Kansas	20.000000	74.500000	48.833333	57.400000	62.000000	69.0000
Kentucky	53.500000	42.666667	77.333333	69.000000	58.833333	58.0000
Louisiana	54.000000	52.000000	67.500000	55.000000	44.000000	61.3333
Maine	NaN	34.000000	73.250000	55.600000	59.000000	81.0000
Maryland	42.000000	66.000000	52.500000	57.000000	74.000000	59.5000
Massachusetts	57.500000	64.000000	52.000000	48.000000	55.000000	66.8333
Michigan	35.500000	64.750000	57.000000	97.000000	39.000000	89.0000
Minnesota	45.000000	54.000000	55.500000	38.000000	55.000000	51.0000
Mississippi	85.000000	52.250000	54.833333	73.000000	70.250000	47.6000
Missouri	94.000000	36.333333	22.000000	72.666667	40.000000	44.0000
Montana	45.200000	80.250000	61.500000	52.750000	65.375000	84.6666
Nebraska	59.000000	70.500000	45.000000	62.200000	68.000000	50.3333
Nevada	64.900000	73.428571	64.333333	48.500000	25.000000	60.6666
New Hampshire	68.250000	69.000000	64.142857	62.000000	52.200000	49.6666
New Jersey	45.000000	64.142857	37.333333	82.000000	33.000000	Na
New Mexico	62.666667	41.400000	74.000000	48.000000	64.500000	Na
New York	72.000000	75.666667	59.800000	75.000000	78.500000	53.7500
North Carolina	52.500000	65.000000	82.000000	48.000000	90.000000	67.5000
North Dakota	49.000000	66.200000	68.000000	44.000000	61.166667	74.3333

Ohio	40.000000	45.000000	56.333333	80.400000	27.500000	66.0000
Oklahoma	55.000000	50.000000	51.000000	59.750000	59.500000	99.0000
Oregon	73.000000	39.000000	58.800000	69.333333	40.666667	81.6666
Pennsylvania	94.000000	78.500000	73.666667	81.000000	60.000000	71.4000
Rhode Island	80.500000	40.666667	79.000000	66.000000	90.000000	41.5000
South Carolina	70.500000	75.333333	55.000000	74.400000	43.333333	57.0000
South Dakota	59.600000	78.000000	60.333333	54.666667	26.250000	81.0000
Tennessee	79.000000	77.666667	26.000000	69.750000	67.000000	52.2500
Texas	73.000000	59.500000	85.333333	31.000000	72.333333	45.4000
Utah	NaN	42.000000	56.666667	66.250000	59.000000	84.0000
Vermont	59.500000	50.000000	76.500000	67.000000	60.000000	67.5000
Virginia	71.400000	82.000000	56.000000	48.750000	27.000000	44.0000
Washington	65.333333	60.000000	75.333333	82.000000	73.000000	65.0000
West Virginia	59.500000	82.000000	71.250000	72.750000	69.375000	79.0000
Wisconsin	35.000000	59.750000	79.285714	86.000000	NaN	43.0000
Wyoming	76.250000	46.500000	73.750000	68.000000	62.200000	64.5000

50 rows × 25 columns

In [84]:
```
# Mean Purchase Amount by Size
shopdata.groupby("Size")["Purchase Amount (USD)"].mean()
```

Out[84]:
```
Size
L     58.563153
M     59.924217
S     61.037707
XL    60.090909
Name: Purchase Amount (USD), dtype: float64
```

In [30]:
```
# Sum of Purchase Amount by Payment Method
shopdata.groupby("Payment Method")["Purchase Amount (USD)"].sum()
```

Out[30]:
```
Payment Method
Bank Transfer    37123
Cash             38833
Credit Card      42567
Debit Card       37118
PayPal           37449
Venmo            39991
Name: Purchase Amount (USD), dtype: int64
```

In [90]:
```
# Mean Review Rating by Discount Applied
shopdata.groupby("Discount Applied")["Review Rating"].mean()
```

Out[90]:
```
Discount Applied
No     3.757715
Yes    3.739654
Name: Review Rating, dtype: float64
```

In [92]:
```
# Count of Categories per Color
shopdata.groupby("Color")["Category"].value_counts()
```

Out[92]:
```
Color   Category
Beige   Clothing       65
        Accessories    44
        Footwear       24
        Outerwear      14
Black   Clothing       81
                       ..
White   Outerwear      15
Yellow  Clothing       75
        Accessories    56
        Footwear       30
        Outerwear      13
Name: count, Length: 100, dtype: int64
```

In [94]:
```
# Median of Previous Purchases grouped by Frequency of Purchases
shopdata.groupby("Frequency of Purchases")["Previous Purchases"].median()
```

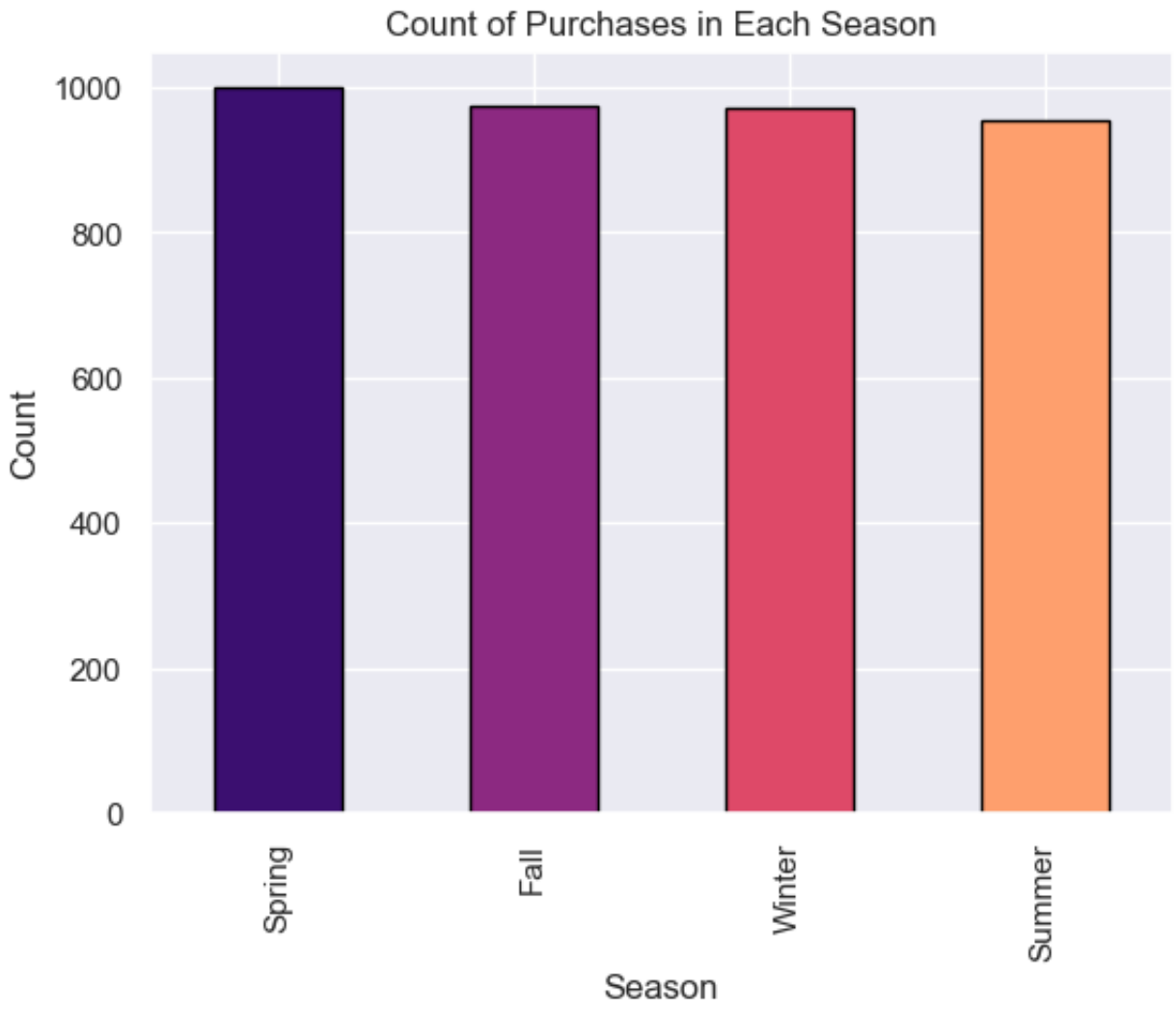
Out[94]:
```
Frequency of Purchases
Annually           24.0
Bi-Weekly          24.0
Every 3 Months     24.0
Fortnightly        26.0
Monthly            25.0
Quarterly          28.0
Weekly             26.0
Name: Previous Purchases, dtype: float64
```

In [96]:
```
# Mean Review Rating by Season
shopdata.groupby("Season")["Review Rating"].mean()
```

Out[96]:
```
Season
Fall      3.729949
Spring    3.790591
Summer    3.725654
Winter    3.752111
Name: Review Rating, dtype: float64
```

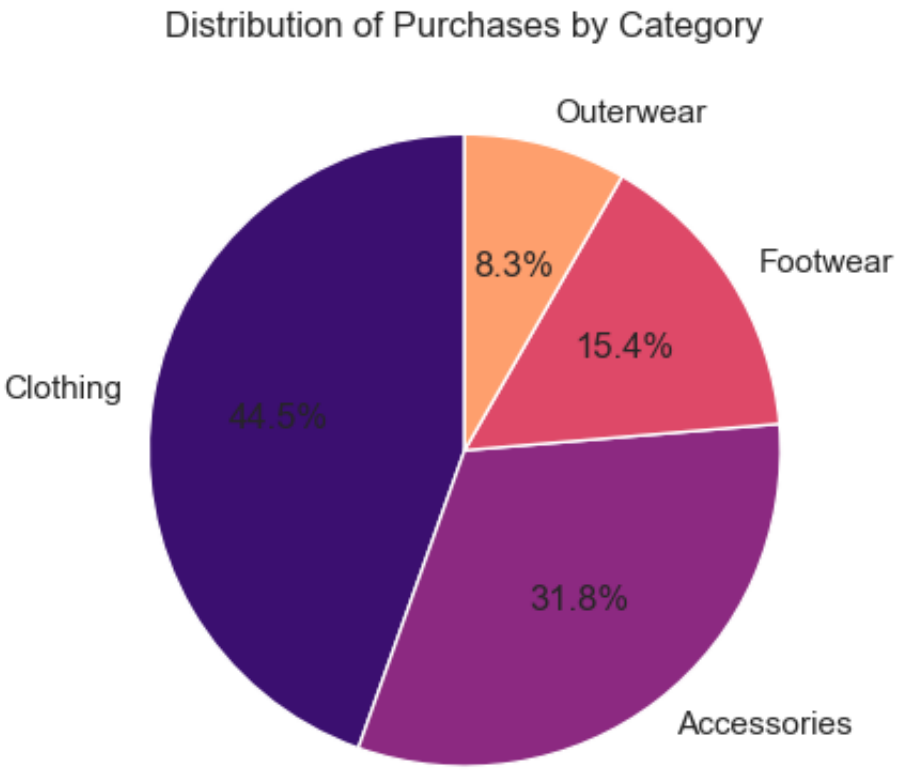
In [38]:
```
# Bar plot counting the number of purchases per Season

season_counts = shopdata["Season"].value_counts()
season_colors = sns.color_palette("magma", len(season_counts))
season_counts.plot(kind="bar", color=season_colors, edgecolor="black")
plt.xlabel("Season")
plt.ylabel("Count")
plt.title("Count of Purchases in Each Season")
plt.savefig("Count_of_Purchases_in_Each_Season", dpi=300, bbox_inches ="t
plt.show()
```

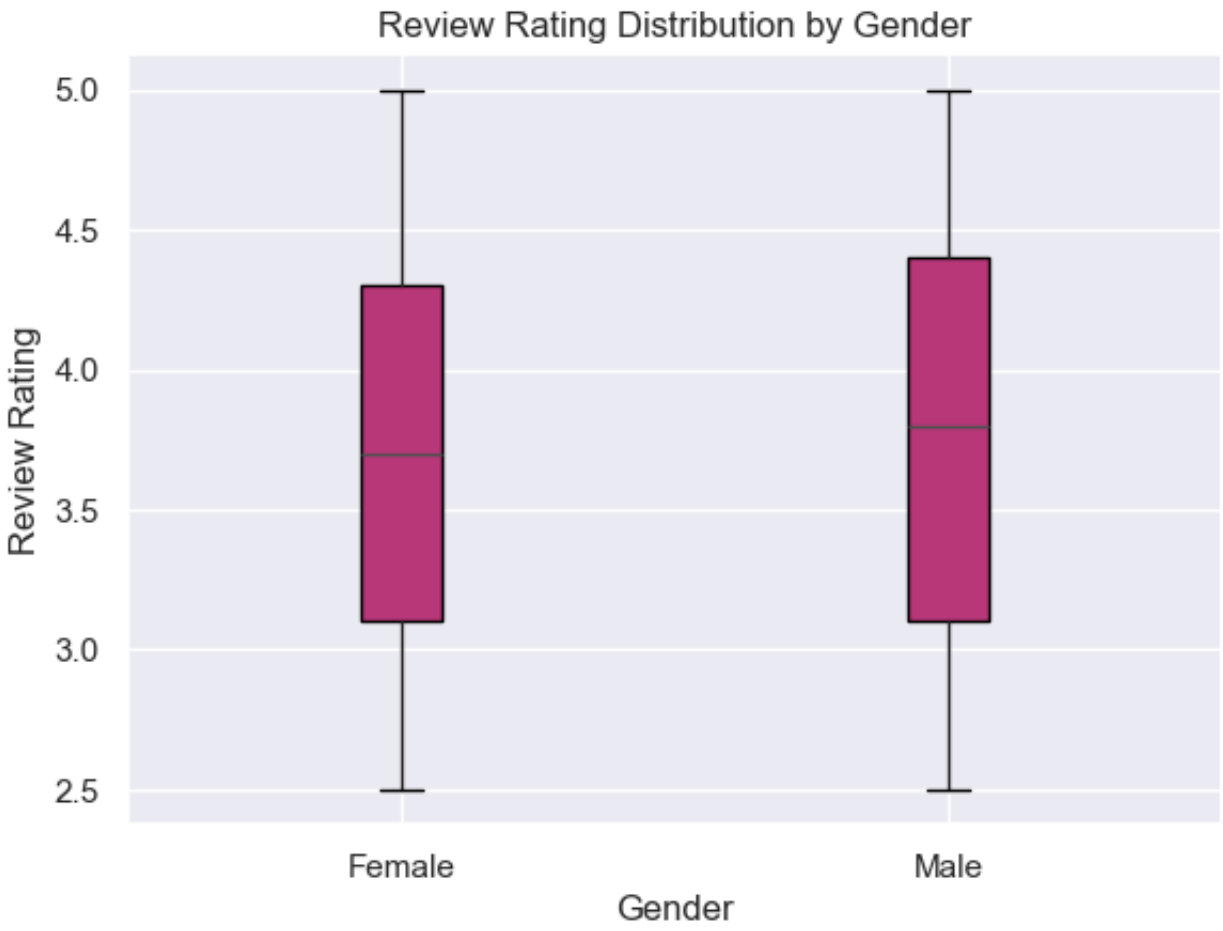


In [40]:
```
# Pie chart for distribution of Purchases by Category

cat_counts = shopdata["Category"].value_counts()
cat_colors = sns.color_palette("magma", len(cat_counts))
cat_counts.plot(kind="pie", autopct="%1.1f%%", startangle=90, colors=cat_
plt.title("Distribution of Purchases by Category")
plt.ylabel("")
plt.savefig("Distribution_of_Purchases_by_Category(Pie_Chart)", dpi=300,
plt.show()
```

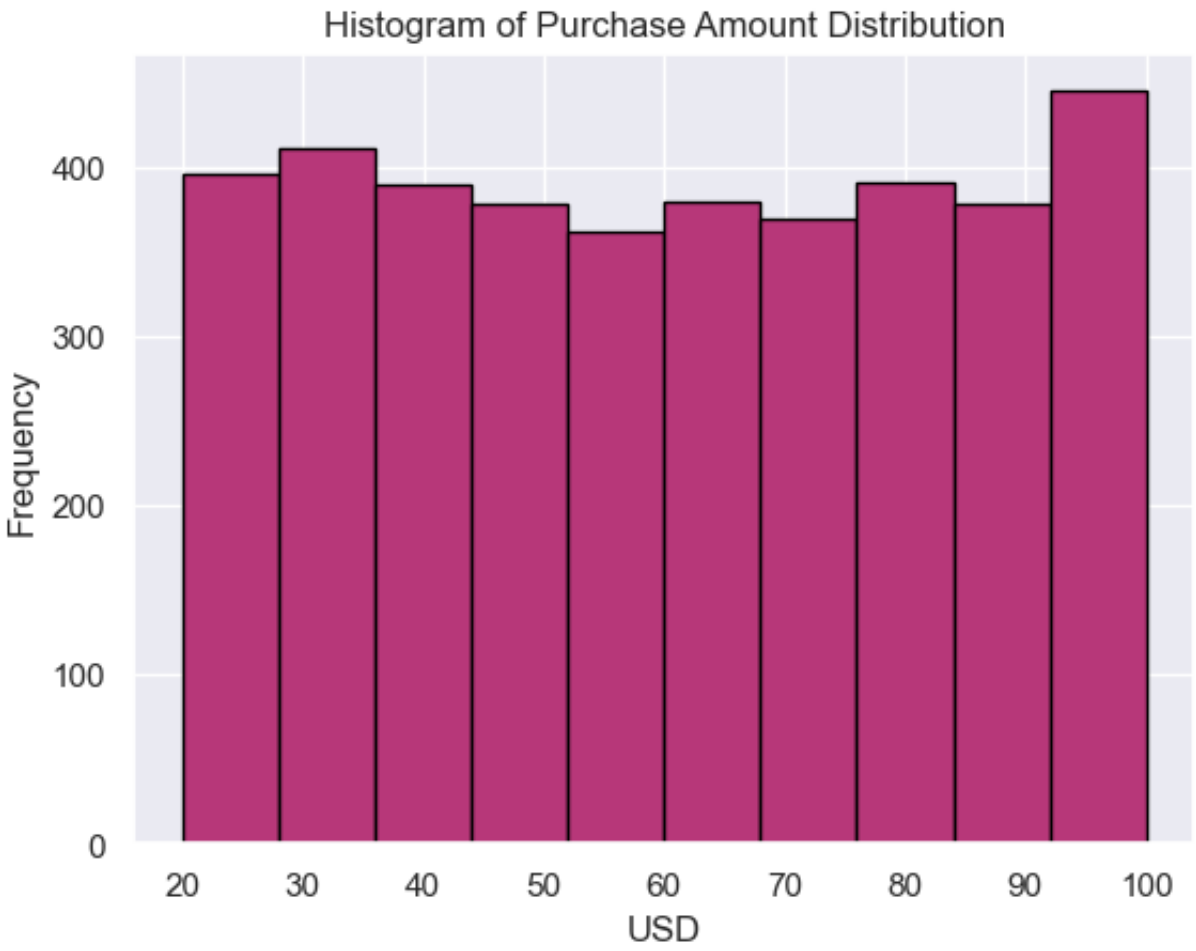


In [42]:
```
# Visualization: Boxplot of Review Rating Distribution by gender to show
# ------------------------------------------------------------------------
shopdata.boxplot(column="Review Rating", by="Gender", patch_artist=True,
                 boxprops=dict(facecolor=sns.color_palette("magma", 1)[0]
plt.xlabel("Gender")
plt.ylabel("Review Rating")
plt.title("Review Rating Distribution by Gender")
plt.suptitle("")
plt.savefig("Review Rating Distribution by Gender", dpi=300, bbox_inches
plt.show()
```



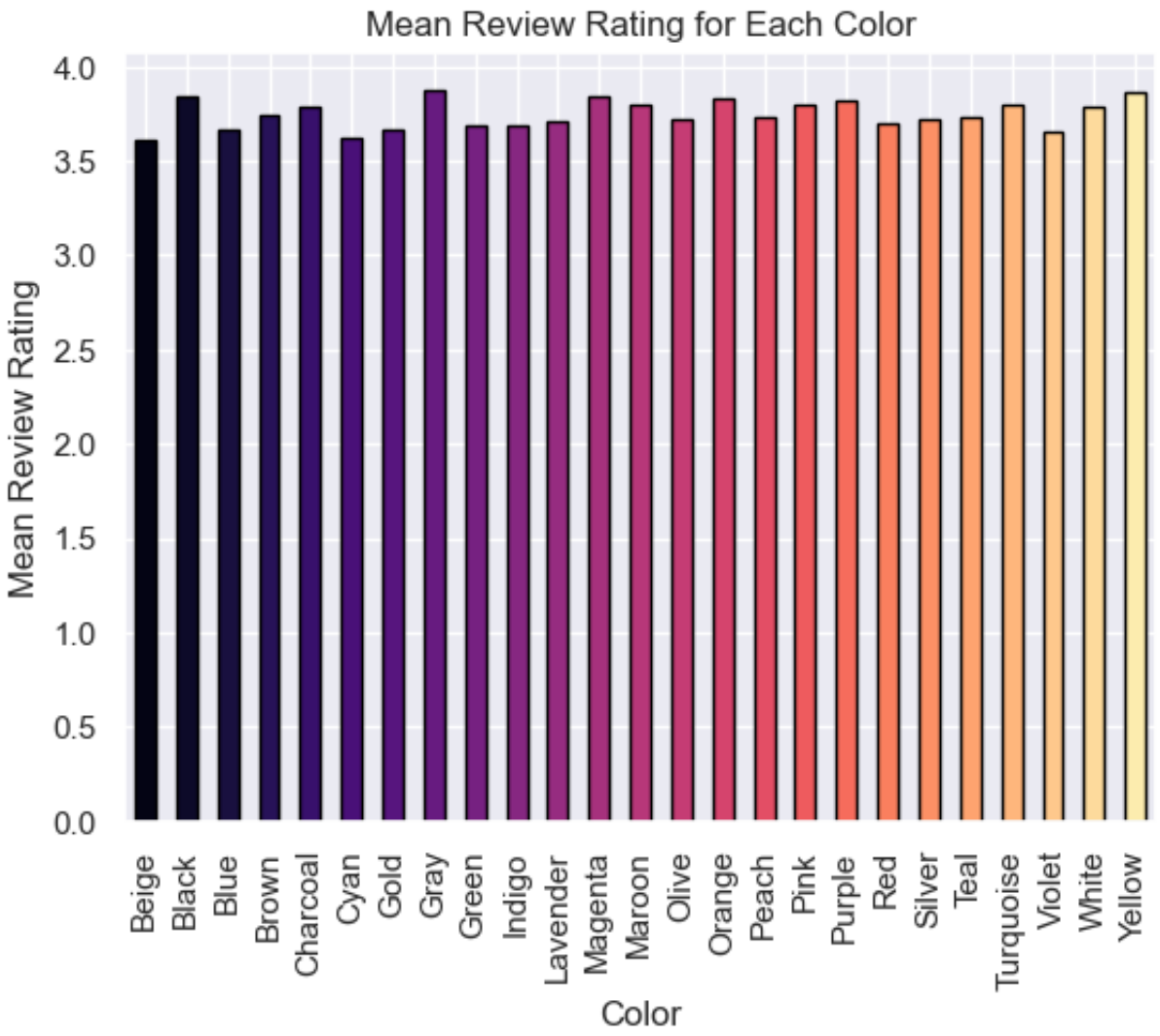
In [44]:
```
# Histogram for Purchase Amount Distribution

hist_color = sns.color_palette("magma", 1)[0]
shopdata["Purchase Amount (USD)"].plot(kind="hist", bins=10, edgecolor="b
plt.xlabel("USD")
plt.ylabel("Frequency")
plt.title("Histogram of Purchase Amount Distribution")
plt.savefig("Histogram of Purchase Amount Distribution", dpi=300, bbox_in
plt.show()
```



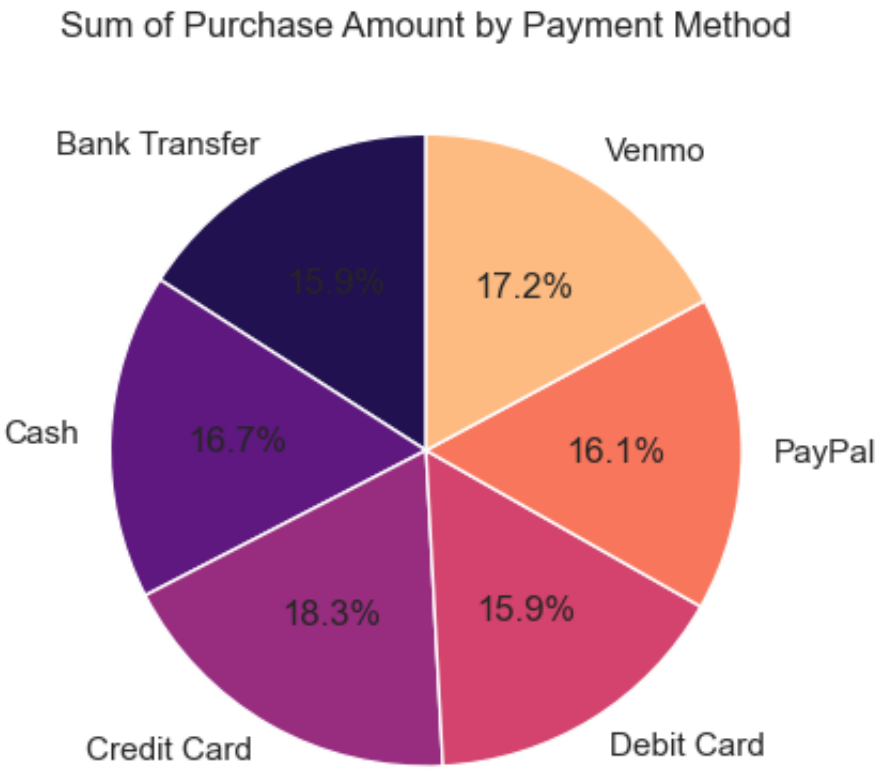
In [46]:
```
# Bar plot of Mean Review Rating for each Color

mean_reviewrating = shopdata.groupby("Color")["Review Rating"].mean()
color_bar = sns.color_palette("magma", len(mean_reviewrating))
mean_reviewrating.plot(kind="bar", color=color_bar, edgecolor="black")
plt.xlabel("Color")
plt.ylabel("Mean Review Rating")
plt.title("Mean Review Rating for Each Color")
plt.savefig("Mean Review Rating for Each Color", dpi=300, bbox_inches ="t
plt.show()
```



In [48]:
```
# Visualization: Pie chart for Sum of Purchase Amount by Payment Method

pay_sum = shopdata.groupby("Payment Method")["Purchase Amount (USD)"].sum
pay_colors = sns.color_palette("magma", len(pay_sum))
pay_sum.plot(kind="pie", autopct="%1.1f%%", startangle=90, colors=pay_col
plt.title("Sum of Purchase Amount by Payment Method")
plt.ylabel("")
plt.savefig("Sum of Purchase Amount by Payment Method", dpi=300, bbox_inc
plt.show()
```

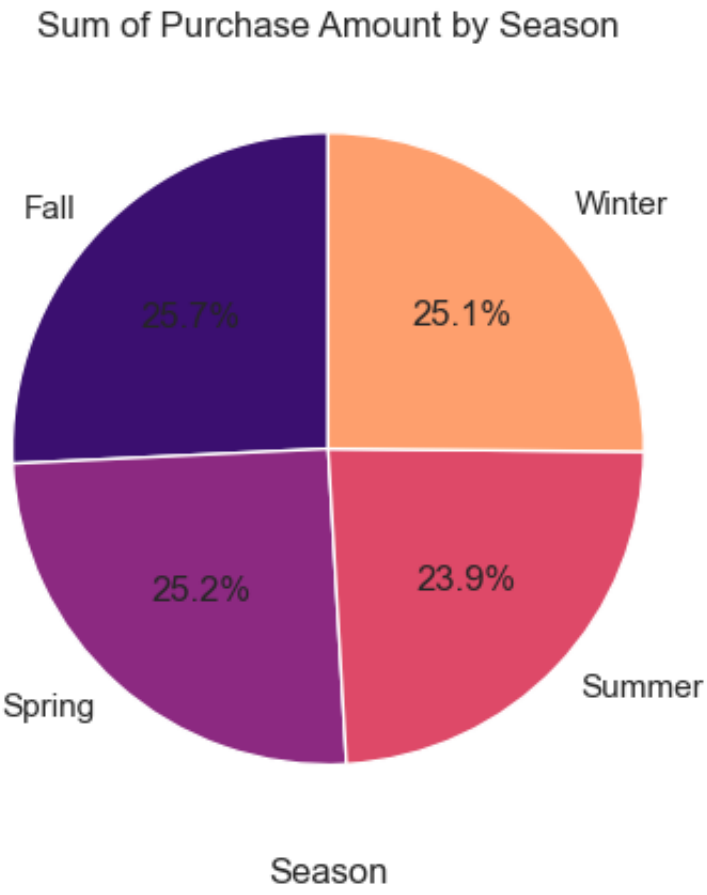


In [50]:
```
# Boxplot of Purchase Amount Distribution by Frequency of Purchases

shopdata.boxplot(column="Purchase Amount (USD)", by="Frequency of Purchas
                 boxprops=dict(facecolor=sns.color_palette("magma", 1)[0]
plt.xlabel("Frequency of Purchases")
plt.ylabel("Purchase Amount (USD)")
plt.title("Purchase Amount Distribution by Frequency of Purchases")
plt.xticks(rotation=45, fontsize=8)
plt.suptitle("")
plt.savefig("Purchase Amount Distribution by Frequency of Purchases", dpi
plt.show()
```

In [54]:
```
# Pie chart for Sum(percentages) of Purchase Amount by Season

season_sum = shopdata.groupby("Season")["Purchase Amount (USD)"].sum()
season_sum_colors = sns.color_palette("magma", len(season_sum))
season_sum.plot(kind="pie", autopct="%1.1f%%", startangle=90, colors=seas
plt.xlabel("Season")
plt.ylabel(" ")
plt.title("Sum of Purchase Amount by Season")
plt.savefig("Sum of Purchase Amount by Season", dpi=300, bbox_inches ="ti
plt.show()
```


Sum of Purchase Amount by Season
Fall
Winter
25.7%
25.1%
25.2%
23.9%
Spring
Summer
Season

In [ ]: