

part1

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: df=pd.read_csv("shopping_trends.csv")
```

part2

```
In [3]: df.shape
```

```
Out[3]: (3900, 19)
```

```
In [4]: df.head(10)
```

```
Out[4]:
```

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise
5	6	46	Male	Sneakers	Footwear	20	Wyoming	M	White
6	7	63	Male	Shirt	Clothing	85	Montana	M	Gray
7	8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal
8	9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver
9	10	57	Male	Handbag	Accessories	31	Missouri	M	Pink

```
In [5]: df.dtypes
```

```
Out[5]: Customer ID          int64
Age              int64
Gender           object
Item Purchased   object
Category         object
Purchase Amount (USD)  int64
Location         object
Size            object
Color           object
Season          object
Review Rating    float64
Subscription Status object
Payment Method   object
Shipping Type    object
Discount Applied object
Promo Code Used  object
Previous Purchases int64
Preferred Payment Method object
Frequency of Purchases object
dtype: object
```

```
In [6]: 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
```

```
In [7]: print(df.columns.tolist())
```

```
['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']
```

```
In [8]: df.isnull().sum()
```

```
Out[8]: 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 [9]: print(df.astype)
```

<bound method NDFrame.astype of tegory \	Customer ID	Age	Gender	Item Purchased	Ca
0	1	55	Male	Blouse	Clothing
1	2	19	Male	Sweater	Clothing
2	3	50	Male	Jeans	Clothing
3	4	21	Male	Sandals	Footwear
4	5	45	Male	Blouse	Clothing
...
3895	3896	40	Female	Hoodie	Clothing
3896	3897	52	Female	Backpack	Accessories
3897	3898	46	Female	Belt	Accessories
3898	3899	44	Female	Shoes	Footwear
3899	3900	52	Female	Handbag	Accessories

	Purchase Amount (USD)	Location	Size	Color	Season \
0	53	Kentucky	L	Gray	Winter
1	64	Maine	L	Maroon	Winter
2	73	Massachusetts	S	Maroon	Spring
3	90	Rhode Island	M	Maroon	Spring
4	49	Oregon	M	Turquoise	Spring
...
3895	28	Virginia	L	Turquoise	Summer
3896	49	Iowa	L	White	Spring
3897	33	New Jersey	L	Green	Spring
3898	77	Minnesota	S	Brown	Summer
3899	81	California	M	Beige	Spring

Review Rating	Subscription Status	Payment Method	Shipping Type \
0	3.1	Yes Credit Card	Express
1	3.1	Yes Bank Transfer	Express
2	3.1	Yes Cash	Free Shipping
3	3.5	Yes PayPal	Next Day Air
4	2.7	Yes Cash	Free Shipping
...
3895	4.2	No Cash	2-Day Shipping
3896	4.5	No PayPal	Store Pickup
3897	2.9	No Credit Card	Standard
3898	3.8	No PayPal	Express
3899	3.1	No Bank Transfer	Store Pickup

Discount Applied	Promo Code Used	Previous Purchases \
0	Yes	Yes 14
1	Yes	Yes 2
2	Yes	Yes 23
3	Yes	Yes 49
4	Yes	Yes 31
...
3895	No	No 32
3896	No	No 41
3897	No	No 24
3898	No	No 24
3899	No	No 33

Preferred Payment Method	Frequency of Purchases
0	Venmo Fortnightly
1	Cash Fortnightly

2	Credit Card	Weekly
3	PayPal	Weekly
4	PayPal	Annually
...
3895	Venmo	Weekly
3896	Bank Transfer	Bi-Weekly
3897	Venmo	Quarterly
3898	Venmo	Weekly
3899	Venmo	Quarterly

[3900 rows x 19 columns]>

```
In [10]: df.duplicated().sum()
```

```
Out[10]: np.int64(0)
```

par3

Data Cleaning

```
In [11]: df=df[(df['Age']>10)&(df['Age']<90)]
```

```
In [12]: text_cols = ['Item Purchased', 'Location', 'Size', 'Color', 'Season',
                     'Subscription Status', 'Shipping Type']
for col in text_cols:
    df[col]=df[col].str.strip().str.title()
```

```
In [13]: df['Gender']=df['Gender'].astype(str).str.strip().str.title()
df['Gender'] = df['Gender'].replace({
    'Femail': 'Female',
    'Feemale': 'Female',
    'Malee': 'Male'
})
```

```
In [14]: df['Payment Method']=df['Payment Method'].replace({'Creadit Card': 'Credit Card',
                    'Cred Card': 'Credit Card'})
```

```
In [15]: df['Subscription Status']=df['Subscription Status'].replace({'Y': 'Yes',
                                'N': 'No'})
```

```
In [16]: df = df[(df['Review Rating'] >= 1) & (df['Review Rating'] <= 5)]
```

```
In [17]: df = df[df['Purchase Amount (USD)'] > 0]
```

```
In [18]: df.isnull().sum()
```

```
Out[18]: 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
```

part4

```
In [19]: # Age analysis
print("Average age",df['Age'].mean())
print('Age range',df['Age'].max()-df['Age'].min())
```

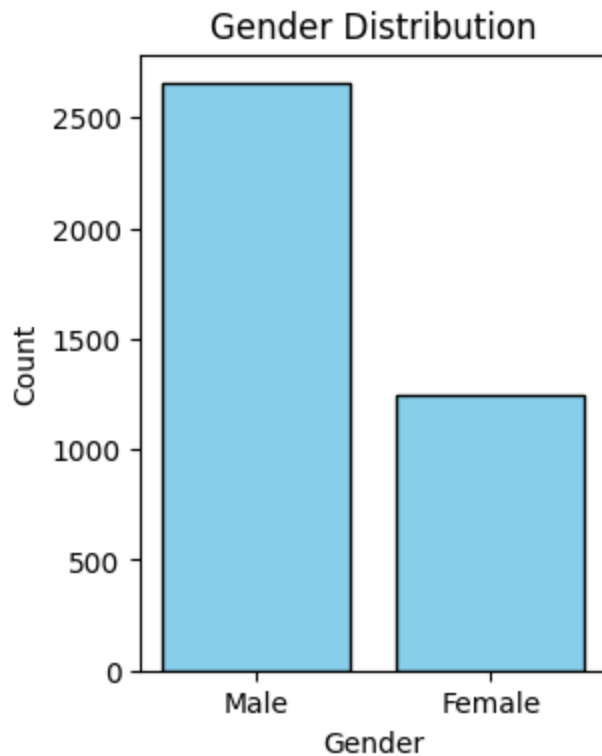
Average age 44.06846153846154
Age range 52

```
In [20]: # gender_Analysis
print(f"\nGender Distribution:")
gender_counts = df['Gender'].value_counts()
print(gender_counts)
```

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

```
In [21]: plt.figure(figsize=(3, 4))
plt.bar(gender_counts.index, gender_counts.values,color='skyblue', edgecolor='black')

plt.xlabel('Gender')
plt.ylabel('Count')
plt.title('Gender Distribution')
plt.show()
```



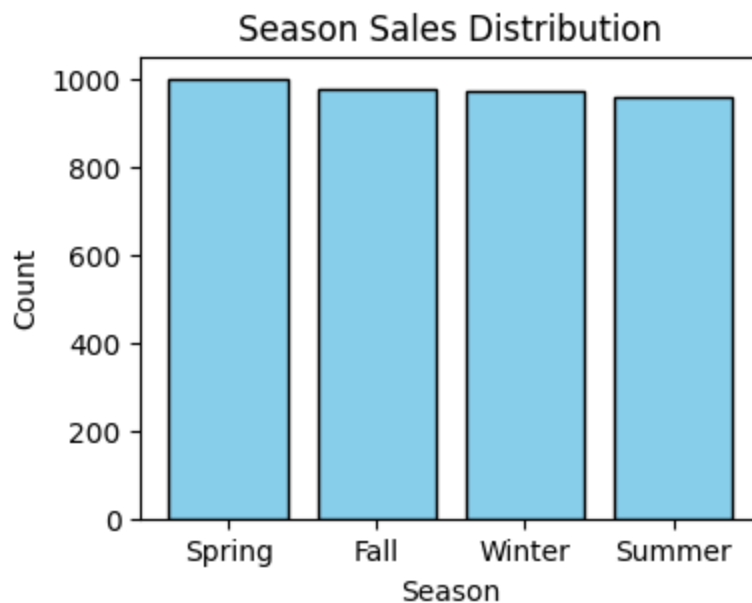
```
In [22]: import matplotlib.pyplot as plt

Season_Sales = df['Season'].value_counts()

plt.figure(figsize=(4, 3))
plt.bar(Season_Sales.index, Season_Sales.values, color='skyblue', edgecolor='black')

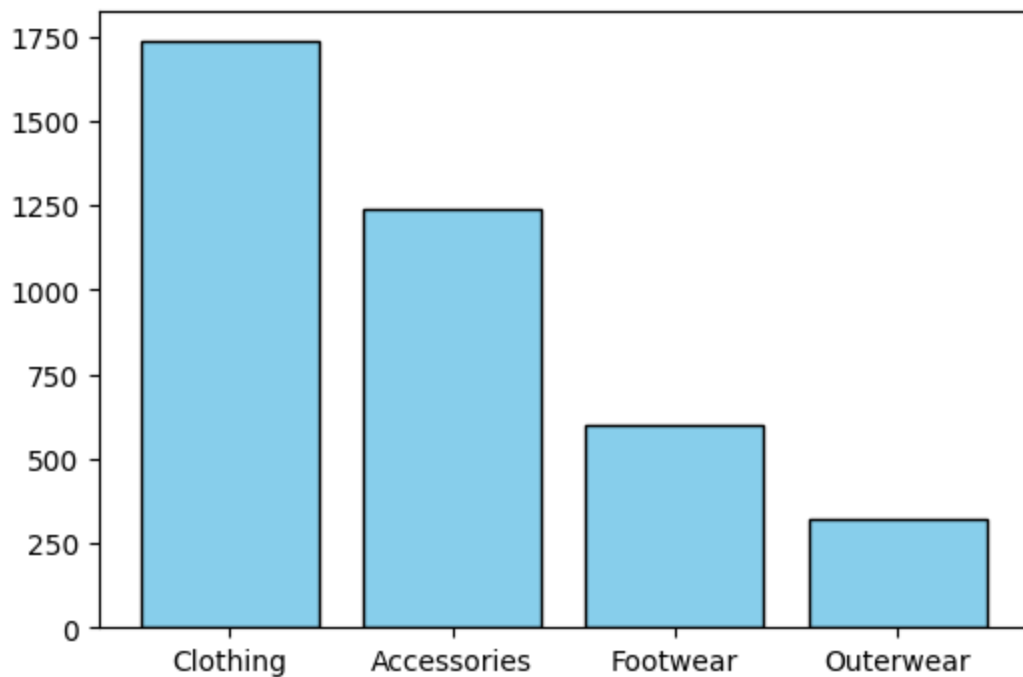
plt.xlabel('Season')
plt.ylabel('Count')
plt.title('Season Sales Distribution')

plt.show()
```



```
In [23]: Category=df['Category'].value_counts()
plt.figure(figsize=(6, 4))
plt.bar(Category.index, Category.values, color='skyblue', edgecolor='black')
```

Out[23]: <BarContainer object of 4 artists>



```
In [24]: PaymentMethod = df['Payment Method'].value_counts()
print("Payment Method:\n", PaymentMethod)

plt.figure(figsize=(6, 4))

colors = ['#6BA292', '#88B4A1', '#A7C5AF', '#C4D7C4', '#E2E8CE']

plt.bar(PaymentMethod.index, PaymentMethod.values, color=colors[:len(PaymentMethod.index)])

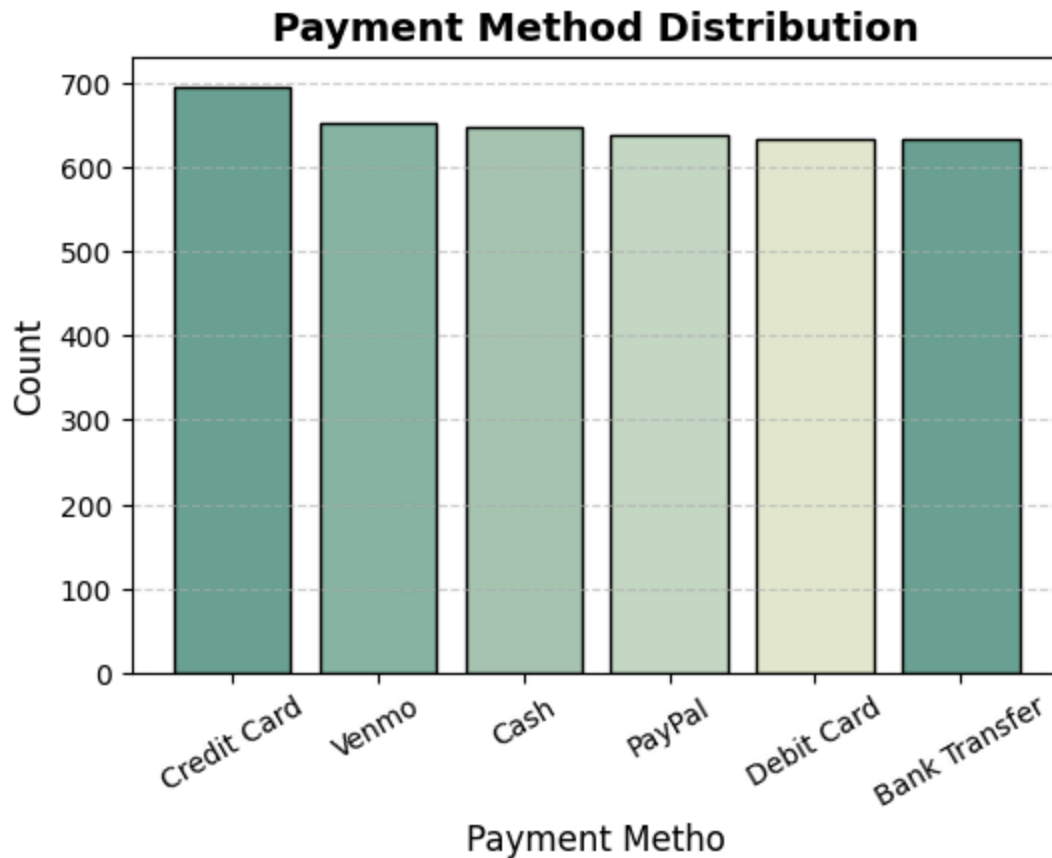
plt.xlabel('Payment Metho', fontsize=12)
plt.ylabel('Count', fontsize=12)
plt.title('Payment Method Distribution', fontsize=14, fontweight='bold')
plt.xticks(rotation=30)
plt.grid(axis='y', linestyle='--', alpha=0.6)

plt.show()
```

Payment Method:

Payment Method	Count
Credit Card	696
Venmo	653
Cash	648
PayPal	638
Debit Card	633
Bank Transfer	632

Name: count, dtype: int64



In [25]: `print(df.columns)`

```
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 [26]: `ItemPurchased=df['Item Purchased'].value_counts(ascending=True).head()`
 ItemPurchased

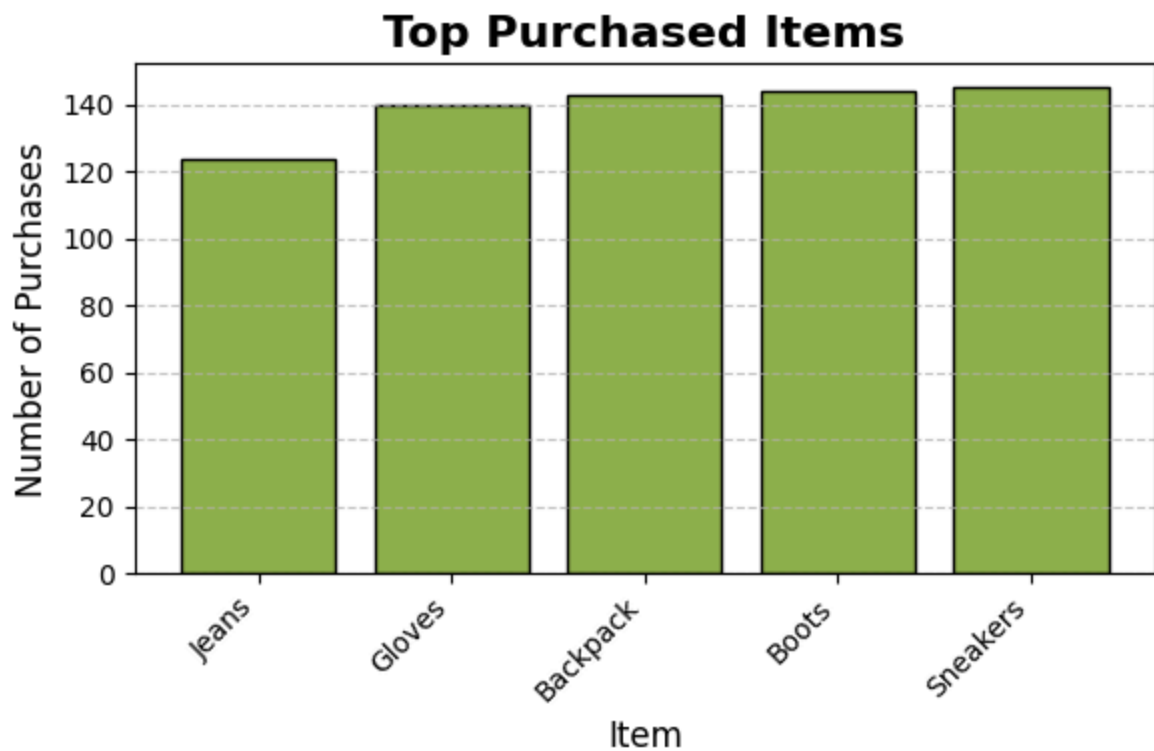
```
Out[26]: Item Purchased
Jeans      124
Gloves     140
Backpack   143
Boots      144
Sneakers   145
Name: count, dtype: int64
```

```
In [27]: plt.figure(figsize=(6,4))
plt.bar(ItemPurchased.index, ItemPurchased.values, color="#90AF4C", edgecolor='black')

plt.title('Top Purchased Items', fontsize=16, fontweight='bold')
plt.xlabel('Item', fontsize=12)
plt.ylabel('Number of Purchases', fontsize=12)

plt.xticks(rotation=45, ha='right')
```

```
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



```
In [28]: TopLocation = df['Location'].value_counts().head(10)

plt.figure(figsize=(10,5))
plt.bar(TopLocation.index, TopLocation.values, color="#5B865F", edgecolor='black')

plt.title('Top Customer Locations', fontsize=16, fontweight='bold')
plt.xlabel('Location', fontsize=12)
plt.ylabel('Number of Purchases', fontsize=12)

plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



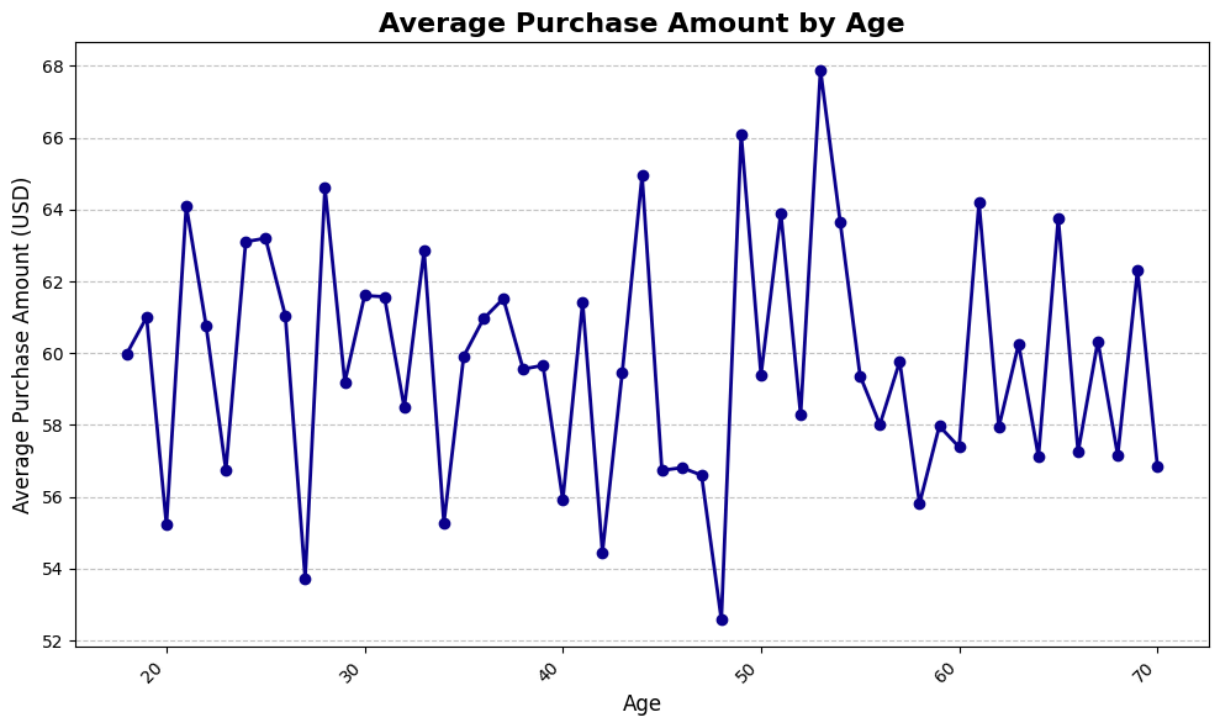
part5

```
In [29]: age_mean_sales = df.groupby('Age')['Purchase Amount (USD)'].mean()

plt.figure(figsize=(10,6))
plt.plot(age_mean_sales.index, age_mean_sales.values, color="#0D048E", marker='o',

plt.title('Average Purchase Amount by Age', fontsize=16, fontweight='bold')
plt.xlabel('Age', fontsize=12)
plt.ylabel('Average Purchase Amount (USD)', fontsize=12)

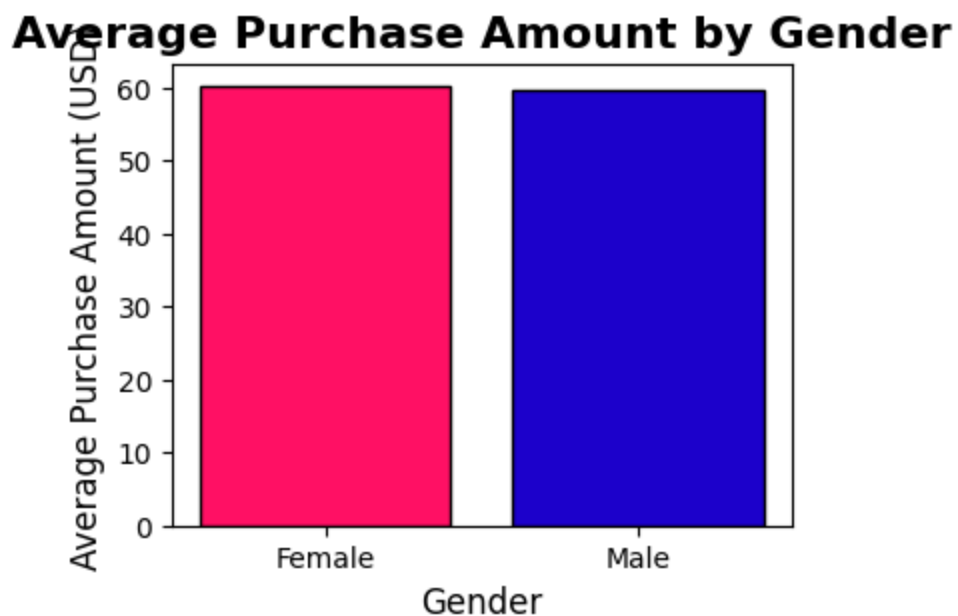
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```



```
In [30]: Gender_Sales = df.groupby('Gender')['Purchase Amount (USD)'].mean()

plt.figure(figsize=(4,3))
bars = plt.bar(Gender_Sales.index, Gender_Sales.values, color=["#FF1064", "#1C01CB"])

plt.title('Average Purchase Amount by Gender', fontsize=16, fontweight='bold')
plt.xlabel('Gender', fontsize=12)
plt.ylabel('Average Purchase Amount (USD)', fontsize=12)
plt.show()
```

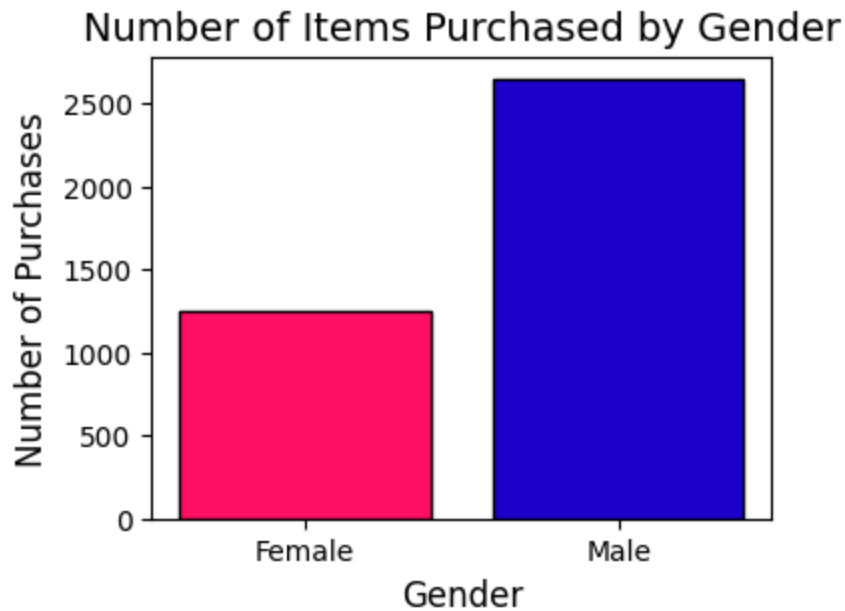


```
In [31]: ItemPurchased_Gender = df.groupby('Gender')['Item Purchased'].count()

plt.figure(figsize=(4,3))
```

```
bars = plt.bar(ItemPurchased_Gender.index, ItemPurchased_Gender.values,
               color=["#FF1064", "#1C01CB"], edgecolor='black')

plt.title('Number of Items Purchased by Gender', fontsize=14)
plt.xlabel('Gender', fontsize=12)
plt.ylabel('Number of Purchases', fontsize=12)
plt.show()
```



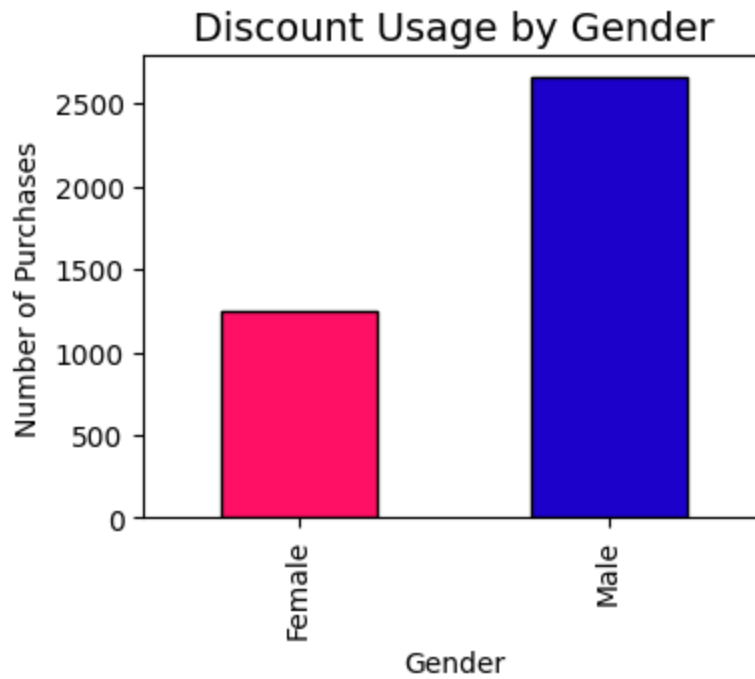
```
In [32]: import pandas as pd
import matplotlib.pyplot as plt

discount_Gender = df.groupby('Gender')['Discount Applied'].count()
print(discount_Gender)

discount_Gender.plot(kind='bar', color=['#FF1064', '#1C01CB'], edgecolor='black', f
plt.title('Discount Usage by Gender', fontsize=14)
plt.xlabel('Gender')
plt.ylabel('Number of Purchases')

plt.show()
```

```
Gender
Female    1248
Male      2652
Name: Discount Applied, dtype: int64
```



```
In [33]: discount_effect = df.groupby('Discount Applied')['Purchase Amount (USD)'].mean()
print(discount_effect)
```

```
plt.figure(figsize=(4,3))
bars = plt.bar(discount_effect.index, discount_effect.values,
               color=["#F28E14", "#0FC60F"], edgecolor='black')

plt.title('Effect of Discount on Purchase Amount', fontsize=14)
plt.xlabel('Discount Applied')
plt.ylabel('Average Purchase Amount (USD)')
plt.show()
```

```
Discount Applied
No      60.130454
Yes     59.279070
Name: Purchase Amount (USD), dtype: float64
```



part6

Average Sales amount by gender and discount

```
In [34]: df.groupby(['Gender', 'Discount Applied'])['Purchase Amount (USD)'].mean().round(0)
```

```
Out[34]: Gender  Discount Applied
Female  No                60.0
Male    No                60.0
         Yes               59.0
Name: Purchase Amount (USD), dtype: float64
```

```
In [35]: # Number of purchases based on Location and product category
df.groupby(['Location', 'Category'])['Item Purchased'].count()
```

```
Out[35]: Location  Category
Alabama  Accessories  25
         Clothing     41
         Footwear     15
         Outerwear     8
Alaska    Accessories  26
         ..
Wisconsin  Outerwear     3
Wyoming   Accessories  23
         Clothing     31
         Footwear     11
         Outerwear     6
Name: Item Purchased, Length: 200, dtype: int64
```

part7

🌱 📊 پروژه تحلیل داده (Shopping Trends Analysis)

در این پروژه، داده‌های مربوط به خرید مشتریان بررسی شد تا الگوهای رفتاری، ترجیحات پرداخت و عوامل تأثیرگذار بر میزان خرید شناسایی شوند. هدف اصلی، استخراج بینش‌های قابل اجرا (Insights)

برای تصمیم‌گیری بهتر در حوزه‌ی بازاریابی و فروش بود.

🔪 مراحل انجام پروژه

1 Data Cleaning:

حذف مقادیر نادرست و پرت

استانداردسازی متون (مثل اصلاح روش پرداخت و جنسیت)

تبدیل نوع داده‌ها برای تحلیل عددی و آماری

2 Univariate Analysis:

بررسی توزیع متغیرهایی مثل سن، جنسیت، فصل خرید و مبلغ خرید

3 Bivariate Analysis:

تحلیل رابطه بین دو متغیر (مثلاً تأثیر سن بر مبلغ خرید)

بررسی الگوهای خرید بر اساس جنسیت و روش پرداخت

🔍 (Key Insights) یافته‌های کلیدی

بازهی سنی ۲۵ تا ۳۵ سال بیشترین میزان خرید را داشته است. 🔪 مشتریان زن و دارای اشتراک Credit 🔪 فعال میانگین خرید بالاتری دارند. 🔪 فصل پاییز بیشترین حجم فروش را نشان می‌دهد. محبوب‌ترین و پربازده‌ترین روش پرداخت بوده است Card.

🧠 (Conclusions) نتیجه‌گیری

♦ مشتریان جوان و دارای اشتراک فعال بیشترین ارزش خرید را ایجاد می‌کنند. ♦ تمرکز بر کمپین‌های بازاریابی ویژه‌ی این گروه می‌تواند سودآوری را افزایش دهد. ♦ پیشنهاد می‌شود تبلیغات فصلی در ماه‌های قبل از پاییز آغاز شوند. ♦ روش پرداخت با کارت اعتباری باید بهینه و تقویت شود چون بیشترین استفاده را دارد.

📦 (Tools & Libraries) ابزارها و کتابخانه‌ها

📁 Python: pandas, matplotlib 📊 Jupyter Notebook / VSCode برای تحلیل گام‌به‌گام 📊 بررسی اولیه داده‌ها

💬 نکته پایانی

است. تمرکز بر کشف (Data Analysis) این پروژه بخشی از مسیر من در یادگیری تحلیل داده‌ها الگوهای واقعی در داده و تبدیل آن‌ها به بینش‌های کاربردی است. خوشحال می‌شم نظرتون رو دوباره‌ی این تحلیل بدونم 🙏

**DataAnalysis #Python #DataCleaning
#EDA #ShoppingTrends #Matplotlib**

#Pandas