Group 10 Assignments Names of the Members are: 1. Oduselu-Hassan, Oladayo Emmanuel 2. Adebayo Stephen Opeyemi 3. Fayomi Oluwafemi 4. Ibeneme Christogonus 5. Onyekwere GraceA dataset of Electronic sales from Sept 2023 to Sept 2024 data set was giving by Mr. Dada Dayo and we generate the following objectives to guide the analysis for the set of data sets OBJECTIVES 1. Basic Analysis: • Which product has the highest unit price across all. • Which customer ID has the highest rating • Which gender bought the most product • Which product has the highest rating • What payment method and what shipping type is most used • Which customer bought the highest quantity and which product • What is the average unit price for each product? 2. Profit and Pricing Analysis: • What is the total profit generated for each product type. • What payment method has the highest total price 3. Date-Based Analysis: yearly trend • What is the total unit price in the year 2024 • What is the highest total unit price in year 2023 • Identify the month with the highest use of standard, express, or overnight shipping 4. Data Visualization: • o Create a line chart to visualize total price over time • o Can you spot any seasonality in the total unit price? • Showcase the top 5 product types purchased using bar charts. • Create a stacked or detached bar chart to visualize which product is the most purchase to the most shipping type

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

In [2]:
```
df = pd.read_csv(r"C:\Users\PC\Downloads\Electronic_sales_Sep2023-Sep2024(Group 10).xlsx.csv")
df
```

Out[2]:

	Customer ID	Age	Gender	Loyalty Member	Product Type	SKU	Rating	Order Status	Payment Method	Total Price	Unit Price	Quantity	Purchase Date	Shipping Type	Add-ons Purchased	Add-on Total
0	1000	53	Male	No	Smartphone	SKU1004	2	Cancelled	Credit Card	5538.33	791.19	7	3/20/2024	Standard	Accessory,Accessory,Accessory	40.21
1	1000	53	Male	No	Tablet	SKU1002	3	Completed	Paypal	741.09	247.03	3	4/20/2024	Overnight	Impulse Item	26.09
2	1002	41	Male	No	Laptop	SKU1005	3	Completed	Credit Card	1855.84	463.96	4	10/17/2023	Express	NaN	0.00
3	1002	41	Male	Yes	Smartphone	SKU1004	2	Completed	Cash	3164.76	791.19	4	8/9/2024	Overnight	Impulse Item,Impulse Item	60.16
4	1003	75	Male	Yes	Smartphone	SKU1001	5	Completed	Cash	41.50	20.75	2	5/21/2024	Express	Accessory	35.56
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
19995	19996	27	Female	No	Smartphone	SMP234	4	Completed	Bank Transfer	6838.08	1139.68	6	6/15/2024	Expedited	NaN	0.00
19996	19996	27	Female	Yes	Laptop	LTP123	4	Cancelled	Credit Card	2697.28	674.32	4	7/18/2024	Standard	NaN	0.00
19997	19996	27	Female	No	Headphones	HDP456	4	Completed	Bank Transfer	1805.90	361.18	5	8/26/2024	Standard	Impulse Item, Extended Warranty, Accessory	198.98
19998	19997	27	Male	No	Headphones	HDP456	1	Cancelled	Bank Transfer	2528.26	361.18	7	1/6/2024	Expedited	Extended Warranty, Accessory	101.34
19999	19998	27	NaN	Yes	Laptop	LTP123	4	Completed	Bank Transfer	674.32	674.32	1	1/29/2024	Expedited	NaN	0.00

20000 rows × 16 columns

In [3]:
```
df.describe()
```

Out[3]:

	Customer ID	Age	Rating	Total Price	Unit Price	Quantity	Add-on Total
count	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000
mean	10483.526550	48.994100	3.093950	3180.133419	578.631867	5.485550	62.244848
std	5631.732525	18.038745	1.223764	2544.978675	312.274076	2.870854	58.058431
min	1000.000000	18.000000	1.000000	20.750000	20.750000	1.000000	0.000000
25%	5478.000000	33.000000	2.000000	1139.680000	361.180000	3.000000	7.615000
50%	10499.500000	49.000000	3.000000	2534.490000	463.960000	5.000000	51.700000
75%	15504.000000	65.000000	4.000000	4639.600000	791.190000	8.000000	93.842500
max	19998.000000	80.000000	5.000000	11396.800000	1139.680000	10.000000	292.770000

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

Out[4]:

	Customer ID	Age	Gender	Loyalty Member	Product Type	SKU	Rating	Order Status	Payment Method	Total Price	Unit Price	Quantity	Purchase Date	Shipping Type	Add-ons Purchased	Add-on Total
0	1000	53	Male	No	Smartphone	SKU1004	2	Cancelled	Credit Card	5538.33	791.19	7	3/20/2024	Standard	Accessory,Accessory,Accessory	40.21
1	1000	53	Male	No	Tablet	SKU1002	3	Completed	Paypal	741.09	247.03	3	4/20/2024	Overnight	Impulse Item	26.09
2	1002	41	Male	No	Laptop	SKU1005	3	Completed	Credit Card	1855.84	463.96	4	10/17/2023	Express	NaN	0.00
3	1002	41	Male	Yes	Smartphone	SKU1004	2	Completed	Cash	3164.76	791.19	4	8/9/2024	Overnight	Impulse Item,Impulse Item	60.16
4	1003	75	Male	Yes	Smartphone	SKU1001	5	Completed	Cash	41.50	20.75	2	5/21/2024	Express	Accessory	35.56

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

Out[5]:

	Customer ID	Age	Gender	Loyalty Member	Product Type	SKU	Rating	Order Status	Payment Method	Total Price	Unit Price	Quantity	Purchase Date	Shipping Type	Add-ons Purchased	Add-on Total
0	1000	53	Male	No	Smartphone	SKU1004	2	Cancelled	Credit Card	5538.33	791.19	7	3/20/2024	Standard	Accessory,Accessory,Accessory	40.21
1	1000	53	Male	No	Tablet	SKU1002	3	Completed	Paypal	741.09	247.03	3	4/20/2024	Overnight	Impulse Item	26.09
2	1002	41	Male	No	Laptop	SKU1005	3	Completed	Credit Card	1855.84	463.96	4	10/17/2023	Express	NaN	0.00
3	1002	41	Male	Yes	Smartphone	SKU1004	2	Completed	Cash	3164.76	791.19	4	8/9/2024	Overnight	Impulse Item,Impulse Item	60.16
4	1003	75	Male	Yes	Smartphone	SKU1001	5	Completed	Cash	41.50	20.75	2	5/21/2024	Express	Accessory	35.56
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
19995	19996	27	Female	No	Smartphone	SMP234	4	Completed	Bank Transfer	6838.08	1139.68	6	6/15/2024	Expedited	NaN	0.00
19996	19996	27	Female	Yes	Laptop	LTP123	4	Cancelled	Credit Card	2697.28	674.32	4	7/18/2024	Standard	NaN	0.00
19997	19996	27	Female	No	Headphones	HDP456	4	Completed	Bank Transfer	1805.90	361.18	5	8/26/2024	Standard	Impulse Item, Extended Warranty, Accessory	198.98
19998	19997	27	Male	No	Headphones	HDP456	1	Cancelled	Bank Transfer	2528.26	361.18	7	1/6/2024	Expedited	Extended Warranty, Accessory	101.34
19999	19998	27	NaN	Yes	Laptop	LTP123	4	Completed	Bank Transfer	674.32	674.32	1	1/29/2024	Expedited	NaN	0.00

20000 rows × 16 columns

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
Customer ID             0
Age                     0
Gender                  1
Loyalty Member          0
Product Type            0
SKU                     0
Rating                  0
Order Status            0
Payment Method          0
Total Price             0
Unit Price              0
Quantity                0
Purchase Date           0
Shipping Type           0
Add-ons Purchased    4868
Add-on Total            0
dtype: int64
```

In [7]:
```
#replacing the empty column with 'unknown'
df.fillna('unknown')
```

Out[7]:

	Customer ID	Age	Gender	Loyalty Member	Product Type	SKU	Rating	Order Status	Payment Method	Total Price	Unit Price	Quantity	Purchase Date	Shipping Type	Add-ons Purchased	Add-on Total
0	1000	53	Male	No	Smartphone	SKU1004	2	Cancelled	Credit Card	5538.33	791.19	7	3/20/2024	Standard	Accessory,Accessory,Accessory	40.21
1	1000	53	Male	No	Tablet	SKU1002	3	Completed	Paypal	741.09	247.03	3	4/20/2024	Overnight	Impulse Item	26.09
2	1002	41	Male	No	Laptop	SKU1005	3	Completed	Credit Card	1855.84	463.96	4	10/17/2023	Express	unknown	0.00
3	1002	41	Male	Yes	Smartphone	SKU1004	2	Completed	Cash	3164.76	791.19	4	8/9/2024	Overnight	Impulse Item,Impulse Item	60.16
4	1003	75	Male	Yes	Smartphone	SKU1001	5	Completed	Cash	41.50	20.75	2	5/21/2024	Express	Accessory	35.56
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
19995	19996	27	Female	No	Smartphone	SMP234	4	Completed	Bank Transfer	6838.08	1139.68	6	6/15/2024	Expedited	unknown	0.00
19996	19996	27	Female	Yes	Laptop	LTP123	4	Cancelled	Credit Card	2697.28	674.32	4	7/18/2024	Standard	unknown	0.00
19997	19996	27	Female	No	Headphones	HDP456	4	Completed	Bank Transfer	1805.90	361.18	5	8/26/2024	Standard	Impulse Item, Extended Warranty, Accessory	198.98
19998	19997	27	Male	No	Headphones	HDP456	1	Cancelled	Bank Transfer	2528.26	361.18	7	1/6/2024	Expedited	Extended Warranty, Accessory	101.34
19999	19998	27	unknown	Yes	Laptop	LTP123	4	Completed	Bank Transfer	674.32	674.32	1	1/29/2024	Expedited	unknown	0.00

20000 rows × 16 columns

In [8]:
```
#•      Which product has the highest unit price across all
#we first pass an argument
all_unit_price_product = df.groupby('Product Type')['Unit Price'].sum()

#we then use the argument
highest_unit_price_product = all_unit_price_product.idxmax()
print('the product that has the highest unit price:',highest_unit_price_product)
```

```
the product that has the highest unit price: Smartphone
```

In [9]:
```
#•      Which customer ID has the highest rating

#we first pass an argument
all_customerID_ratings = df.groupby('Customer ID')['Rating'].sum()

highest_customerID_rating = all_customerID_ratings.idxmax()
print('HIGHEST CUSTOMER ID RATING:', highest_customerID_rating)
```

```
HIGHEST CUSTOMER ID RATING: 18304
```

In [18]:
```
#•      Which gender bought the most product
all_products_purchased_by_both_genders = df.groupby('Gender')['Product Type'].sum()
gender_with_most_purchased_product = all_products_purchased_by_both_genders.idxmax()
print('the most purchased product gender:',gender_with_most_purchased_product)
```

```
the most purchased product gender: Male
```

In [21]:
```
#•      Which product has the highest rating
#we pass an argument
all_product_rating = df.groupby('Product Type')['Rating'].sum()

highest_rated_product = all_product_rating.idxmax()
print('HIGHEST RATED PRODUCT:',highest_rated_product)
```

```
HIGHEST RATED PRODUCT: Smartphone
```

In [23]:
```
#product rating
print('Rating by Each Product:\n', all_product_rating)
```

```
Rating by Each Product:
 Product Type
Headphones     6020
Laptop        11859
Smartphone    19841
Smartwatch    11780
Tablet        12379
Name: Rating, dtype: int64
```

In [25]:
```
#•      What payment method and what shipping type is most used to purchase product
all_payments_method = df.groupby('Payment Method')['Product Type'].sum()

most_used_payment_method = all_payments_method.idxmax()
print('MOST USED PAYMENT METHOD:', most_used_payment_method)
```

```
MOST USED PAYMENT METHOD: Paypal
```

In [27]:
```
#•      Which customer bought the highest quantity
all_customers_quantity = df.groupby('Customer ID')['Product Type'].sum()
highest_quantity_purchased_customer = all_customers_quantity.idxmax()
highest_quantity_purchased_customer
```

Out[27]: 11309

In [29]:
```
#•      What is the average unit price for each product?
average_unit_price = df.groupby('Product Type')['Unit Price'].mean()
print('Average unit purchased per product:\n',average_unit_price)
```

```
Average unit purchased per product:
 Product Type
Headphones    361.411159
Laptop        567.907654
Smartphone    652.880878
Smartwatch    650.891668
Tablet        518.034279
Name: Unit Price, dtype: float64
```

In [31]:
```
#•      What is the total price generated for each product type.
total_price_per_product = df.groupby('Product Type')['Total Price'].sum()
total_price_per_product
```

Out[31]:
```
Product Type
Headphones     4041400.24
Laptop        12296239.97
Smartphone    21516754.69
Smartwatch    14036273.06
Tablet        11712000.41
Name: Total Price, dtype: float64
```

In [33]:
```
#•      What payment method has the highest total price
all_payment_method_total_price = df.groupby('Payment Method')['Total Price'].sum()
highest_payment_method_price = all_payment_method_total_price.idxmax()
print('HIGHEST PAYMENT METHOD PRICE:',highest_payment_method_price)
```

```
HIGHEST PAYMENT METHOD PRICE: Credit Card
```

In [35]:
```
#Which customer bought the highest quantity
customer_purchase_quantity = df.groupby("Customer ID")["Quantity"].sum()
highest_quantity_purchase_by_customer = customer_purchase_quantity.idxmax()
print('Customer With Highest Purchase:\n', highest_quantity_purchase_by_customer)
```

```
Customer With Highest Purchase:
 2447
```

In [37]:
```
#•      What is the total unit price in the year 2023 and 2024

#converting date column to datetime format
df['Purchase Date']=pd.to_datetime(df['Purchase Date'])
#extract the year
df['year'] = df['Purchase Date'].dt.year

#•      What is the total unit price in the year
total_unit_price = df.groupby('year')['Total Price'].sum()
total_unit_price
```

Out[37]:
```
year
2023     6849562.61
2024    56753105.76
Name: Total Price, dtype: float64
```

In [39]:
```
#•      Identify the most used type of shipping type in all months
#extract the month
df['month'] = df['Purchase Date'].dt.month
all_shiping_type = df.groupby('Shipping Type')['month'].sum()
most_used_shipping_fee=all_shiping_type.idxmax
most_used_shipping_fee
```

Out[39]:
```
<bound method Series.idxmax of Shipping Type
Expedited    16236
Express      21400
Overnight    21952
Same Day     16197
Standard     38415
Name: month, dtype: int32>
```

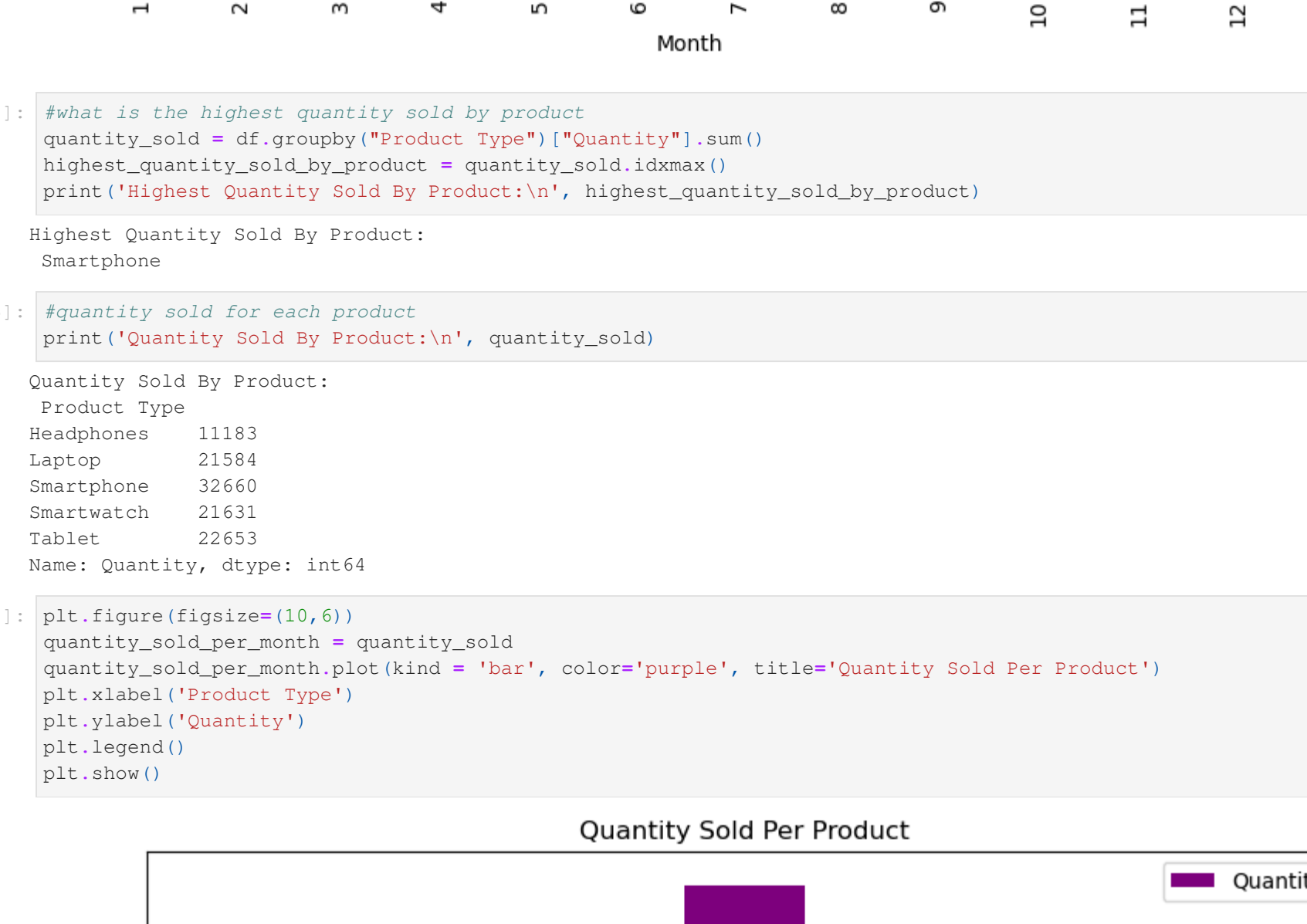
In [41]:
```
#What payment method is most use
payment_method = df.groupby("Payment Method")["Product Type"].sum()
most_used_payment_method = payment_method.idxmax()
print('Most Used Payment Method:\n', most_used_payment_method)
# what shipping type is most used
shipping_type = df.groupby("Shipping Type")["Product Type"].sum()
most_used_Shipping_type = shipping_type.idxmax()
print('Most Used Shipping Type:\n', most_used_Shipping_type)
```

```
Most Used Payment Method:
 Paypal
Most Used Shipping Type:
 Overnight
```

In [47]:
```
#converting date column to datetime format(month)
df['Month'] = df['Purchase Date'].dt.month

 #what is the total profit generated by month
profit_generated_by_month = df.groupby("Month")["Total Price"].sum()
print('Total Profit Generated for Each Month:\n', profit_generated_by_month)
```

```
Total Profit Generated for Each Month:
 Month
1     6620172.47
2     5733696.06
3     6324367.84
4     6418253.62
5     6709042.93
6     6668633.59
7     6535129.52
8     6706118.61
9     5519652.91
10    2318466.35
11    2068434.14
12    1980700.33
Name: Total Price, dtype: float64
```

In [49]:
```
plt.figure(figsize=(10,6))
total_profit_per_month = profit_generated_by_month
total_profit_per_month.plot(kind = 'bar', color='blue', title='Total Profit Per Month')
plt.xlabel('Month')
plt.ylabel('Total Price')
plt.legend()
plt.show()
```



In [51]:
```
#what is the highest quantity sold by product
quantity_sold = df.groupby("Product Type")["Quantity"].sum()
highest_quantity_sold_by_product = quantity_sold.idxmax()
print('Highest Quantity Sold By Product:\n', highest_quantity_sold_by_product)
```

```
Highest Quantity Sold By Product:
 Smartphone
```

In [53]:
```
#quantity sold for each product
print('Quantity Sold By Product:\n', quantity_sold)
```

```
Quantity Sold By Product:
 Product Type
Headphones    11183
Laptop        21584
Smartphone    32660
Smartwatch    21631
Tablet        22653
Name: Quantity, dtype: int64
```

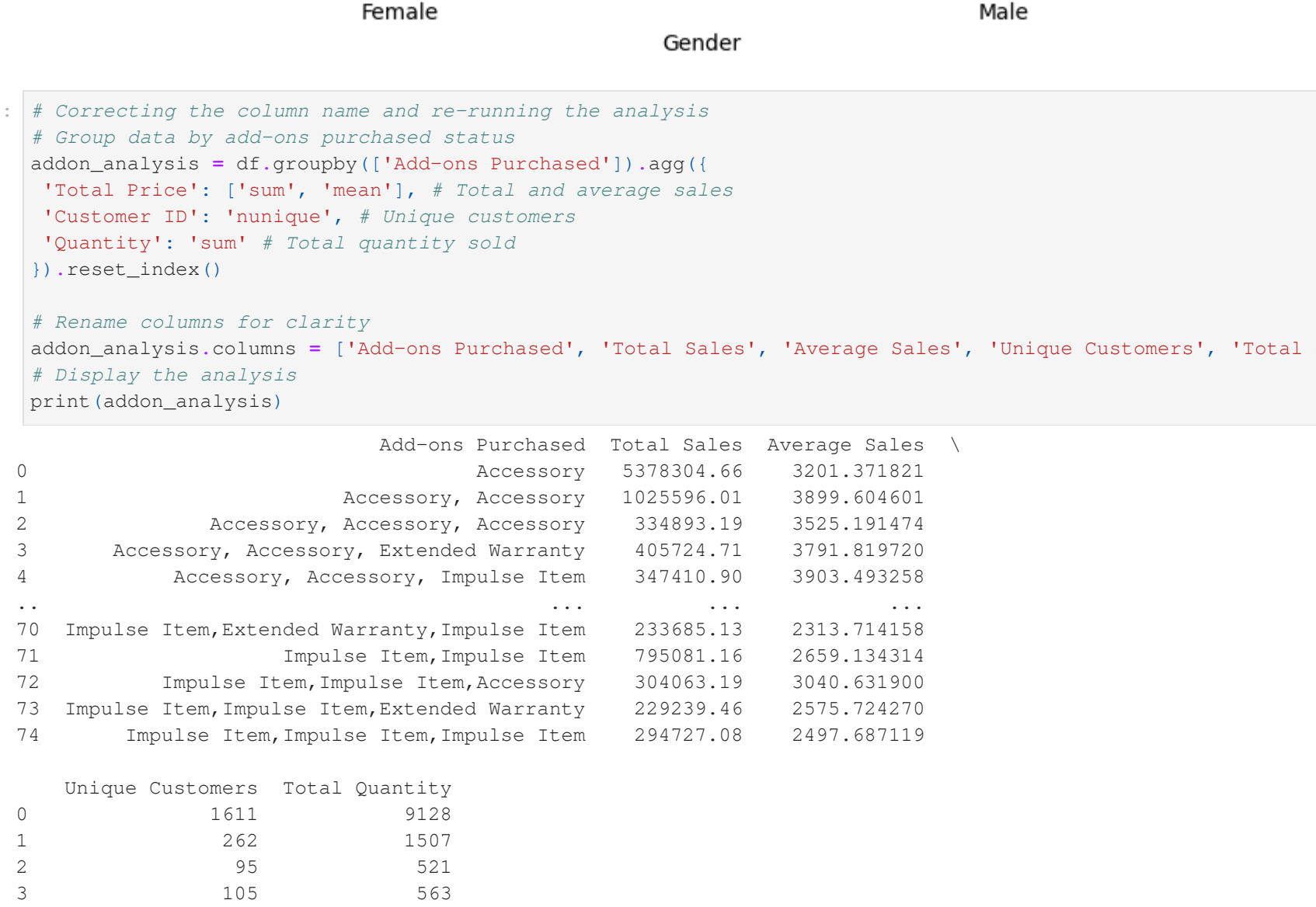
In [61]:
```
plt.figure(figsize=(10,6))
quantity_sold_per_month = quantity_sold
quantity_sold_per_month.plot(kind = 'bar', color='purple', title='Quantity Sold Per Product')
plt.xlabel('Product Type')
plt.ylabel('Quantity')
plt.legend()
plt.show()
```

In [73]:
```
# Group data by demographics
customer_segments = df.groupby(['Gender', 'Loyalty Member']).agg({
 'Customer ID': 'nunique', # Count unique customers
 'Age': 'mean', # Average age
 'Total Price': 'sum' # Total spending
}).reset_index()
```

In [75]:
```
# Rename columns for clarity
customer_segments.rename(columns={
 'Customer ID': 'Unique Customers',
 'Age': 'Average Age',
 'Total Price': 'Total Spending'
}, inplace=True)
```

In [81]:
```
# Display the segmentation table
print(customer_segments)
# Visualize the segmentation
plt.figure(figsize=(10, 6))
sns.barplot(data=customer_segments, x='Gender', y='Total Spending', hue='Loyalty Member')
plt.title('Total Spending by Gender and Loyalty Membership')
plt.xlabel('Gender')
plt.ylabel('Total Spending')
plt.legend(title='Loyalty Member')
plt.show()
```

```
   Gender Loyalty Member  Unique Customers  Average Age  Total Spending
0  Female             No              5166    48.718295     24781797.11
1  Female            Yes              1889    49.246377      6631380.87
2    Male             No              5406    49.259892     25194954.13
3    Male            Yes              1944    48.762143      6993861.94
```

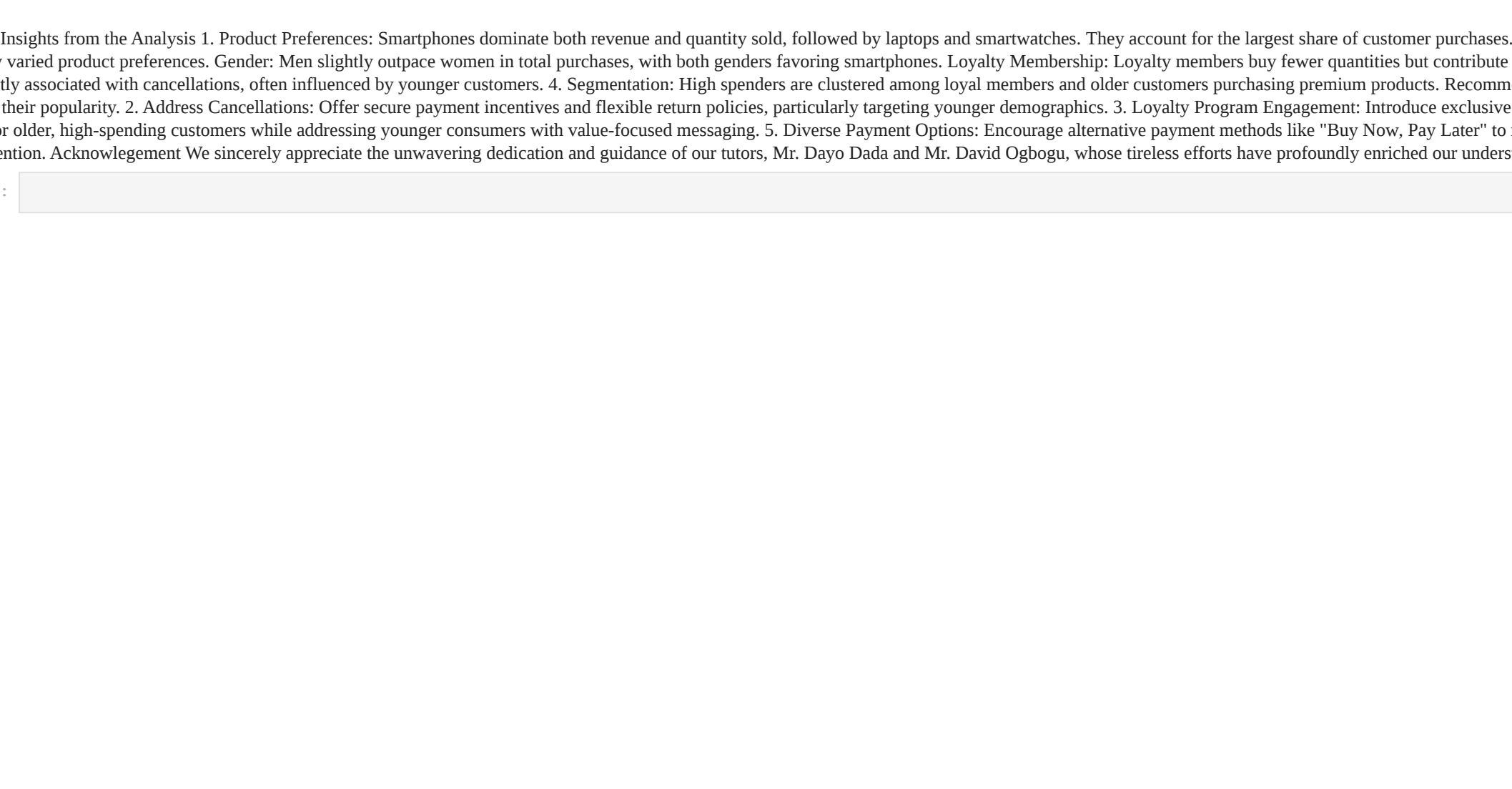


In [97]:
```
# Correcting the column name and re-running the analysis
# Group data by add-ons purchased status
addon_analysis = df.groupby(['Add-ons Purchased']).agg({
 'Total Price': ['sum', 'mean'], # Total and average sales
 'Customer ID': 'nunique', # Unique customers
 'Quantity': 'sum' # Total quantity sold
}).reset_index()

# Rename columns for clarity
addon_analysis.columns = ['Add-ons Purchased', 'Total Sales', 'Average Sales', 'Unique Customers', 'Total Quantity']
# Display the analysis
print(addon_analysis)
```

```
                              Add-ons Purchased  Total Sales  Average Sales  \
0                                     Accessory   5378304.66    3201.371821
1                          Accessory, Accessory   1025596.01    3899.604601
2               Accessory, Accessory, Accessory    334893.19    3525.191474
3       Accessory, Accessory, Extended Warranty    405724.71    3791.819720
4            Accessory, Accessory, Impulse Item    347410.90    3903.493258
..                                          ...          ...            ...
70  Impulse Item,Extended Warranty,Impulse Item    233685.13    2313.714158
71                    Impulse Item,Impulse Item    795081.16    2659.134314
72          Impulse Item,Impulse Item,Accessory    304063.19    3040.631900
73  Impulse Item,Impulse Item,Extended Warranty    229239.46    2575.724270
74       Impulse Item,Impulse Item,Impulse Item    294727.08    2497.687119

    Unique Customers  Total Quantity
0               1611            9128
1                262            1507
2                 95             521
3                105             563
4                 88             523
..               ...             ...
70               101             480
71               297            1585
72                99             583
73                89             450
74               117             637

[75 rows x 5 columns]
```

In [106…
```
# Visualize the impact of add-ons on total sales
plt.figure(figsize=(10, 6))
sns.barplot(data=addon_analysis, x='Add-ons Purchased', y='Total Sales',
palette='viridis')
plt.title('Impact of Add-ons on Total Sales')
plt.xlabel('Add-ons Purchased')
plt.ylabel('Total Sales')
plt.show()
```

```
C:\Users\PC\AppData\Local\Temp\ipykernel_10536\2134790395.py:3: 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=addon_analysis, x='Add-ons Purchased', y='Total Sales',
```



Summarised Insights from the Analysis 1. Product Preferences: Smartphones dominate both revenue and quantity sold, followed by laptops and smartwatches. They account for the largest share of customer purchases. 2. Demographic Trends: Age: Older age groups (51-75) contribute the most to sales, while younger groups (18-25) show varied product preferences. Gender: Men slightly outpace women in total purchases, with both genders favoring smartphones. Loyalty Membership: Loyalty members buy fewer quantities but contribute significantly to high-value products. 3. Cancellation Trends: Smartphones and credit card payments are most frequently associated with cancellations, often influenced by younger customers. 4. Segmentation: High spenders are clustered among loyal members and older customers purchasing premium products. Recommendations 1. Enhance Smartphone Offerings: Invest in promotions and add-ons for smartphones to capitalize on their popularity. 2. Address Cancellations: Offer secure payment incentives and flexible return policies, particularly targeting younger demographics. 3. Loyalty Program Engagement: Introduce exclusive benefits for loyalty members, such as discounts on high-value items. 4. Customized Marketing: Tailor campaigns for older, high-spending customers while addressing younger consumers with value-focused messaging. 5. Diverse Payment Options: Encourage alternative payment methods like "Buy Now, Pay Later" to reduce credit card-linked cancellations. These strategies will boost sales, reduce cancellations, and improve customer retention. Acknowlegement We sincerely appreciate the unwavering dedication and guidance of our tutors, Mr. Dayo Dada and Mr. David Ogbogu, whose tireless efforts have profoundly enriched our understanding and proficiency in data analysis.

In [ ]: