

In [1]:

```
import pandas as pd
```

In [2]:

```
file_to_import = "../purchase_data.csv"
purchase_data = pd.read_csv('purchase_data.csv')
purchase_data.head(1)
```

Out[2]:

	Purchase ID	SN	Age	Gender	Item ID	Item Name	Price
0	0	Lisim78	20	Male	108	Extraction, Quickblade Of Trembling Hands	3.53

In [3]:

```
# Total Players
total_players = len(purchase_data["SN"].unique())
total_players
```

Out[3]:

576

In [4]:

```
# Number of Unique Item
number_unique = len(purchase_data["Item ID"].unique())
number_unique
```

Out[4]:

183

In [5]:

```
# Average Price
average_item_price = purchase_data["Price"].mean()
pd.options.display.float_format = '${:,.2f}'.format
average_item_price.round(2)
```

Out[5]:

3.05

In [6]:

```
# Number of Purchases
number_purchases = len(purchase_data["Purchase ID"].value_counts())
number_purchases
```

Out[6]:

780

In [7]:

```
# Total Revenue
total_revenue = purchase_data["Price"].sum()
pd.options.display.float_format = '${:,.2f}'.format
total_revenue
```

Out[7]:

2379.77

In [8]:

```
#Purchase Analysis Total
summary_table_one = pd.DataFrame({"Total Number of Players": total_players,
                                   "Unique Items": number_unique,
                                   "Average Price per Item": average_item_price.round(2),
                                   "Number of Purchases": number_purchases,
                                   "Total Revenue": total_revenue}, index=[0])

summary_table_one
```

Out[8]:

	Total Number of Players	Unique Items	Average Price per Item	Number of Purchases	Total Revenue
0	576	183	\$3.05	780	\$2,379.77

In [9]:

```
#Gender demo Male
group_male = purchase_data.groupby(["Gender"]).get_group('Male')
count_male = len(group_male['SN'].unique())
print(count_male)

percent_male = (count_male/total_players)
print(percent_male)
```

484  
0.8402777777777778

In [10]:

```
#Gender demo Female
group_female = purchase_data.groupby(["Gender"]).get_group('Female')
count_female = len(group_female['SN'].unique())
print(count_female)

percent_female = (count_female/total_players)
print(percent_female)
```

81  
0.140625

In [11]:

```
#Gender demo Other
group_other = purchase_data.groupby(["Gender"]).get_group('Other / Non-Disclosed')
count_other = len(group_other['SN'].unique())
print(count_other)

percent_other = (count_other/total_players)
print(percent_other)
```

11  
0.019097222222222224

In [12]:

```
gender_dataframe = {'Player Count by Gender': [count_male, count_female, count_other],
                    'Gender': ['Male', 'Female', 'Other/Non-Disclosed'],
                    'Percentage of Players by Gender': [percent_male, percent_female, percent_other]}

gender_final = pd.DataFrame(gender_dataframe)
gender_final = gender_final.set_index('Gender')
gender_final[['Percentage of Players by Gender', 'Player Count by Gender']]
```

```
gender_final
```

```
format_dict = {'Percentage of Players by Gender': '{:.2%}'}
gender_final.style.format(format_dict)
```

Out[12]:

Player Count by Gender		Percentage of Players by Gender
Gender		
Male	484	84.03%
Female	81	14.06%
Other/Non-Disclosed	11	1.91%

In [13]:

```
# Purchasing Analysis (Gender) - Male
group_male = purchase_data.groupby(["Gender"]).get_group('Male')
purchase_count_male = len(group_male['Price'])
print(purchase_count_male)

purchase_total_male = group_male['Price'].sum()
print(purchase_total_male)

avg_purchase_price_male = purchase_total_male / purchase_count_male
print(avg_purchase_price_male)

count_male = len(group_male['SN'].unique())
print(count_male)

avg_per_person_male = purchase_total_male / count_male
print(avg_per_person_male)
```

```
652
1967.64
3.0178527607361967
484
4.065371900826446
```

In [14]:

```
# Purchasing Analysis (Gender) - Female
group_female = purchase_data.groupby(["Gender"]).get_group('Female')
purchase_count_female = len(group_female['Price'])
print(purchase_count_female)

purchase_total_female = group_female['Price'].sum()
print(purchase_total_female)

avg_purchase_price_female = purchase_total_female / purchase_count_female
print(avg_purchase_price_female)

count_female = len(group_female['SN'].unique())
print(count_female)

avg_per_person_female = purchase_total_female / count_female
print(avg_per_person_female)
```

```
113
361.94
3.203008849557522
81
4.468395061728395
```

In [15]:

```
# Purchasing Analysis (Gender) - Other
group_other = purchase_data.groupby(["Gender"]).get_group('Other / Non-Disclosed')
purchase_count_other = len(group_other['Price'])
print(purchase_count_other)
```

```

purchase_total_other = group_other['Price'].sum()
print(purchase_total_other)

avg_purchase_price_other = purchase_total_other / purchase_count_other
print(avg_purchase_price_other)

count_other = len(group_other['SN'].unique())
print(count_other)

avg_per_person_other = purchase_total_other / count_other
print(avg_per_person_other)

```

```

15
50.19
3.3459999999999996
11
4.5627272727272725

```

In [16]:

```

#Data Frame for Purchasing Analysis (Gender)
gender_purchase_analysis_dataframe = {'Purchase Count': [purchase_count_male,
                                                         ,purchase_count_other],
                                     'Gender': ['Male', 'Female', 'Other/Non-Disclosed'],
                                     'Avg Purchase Price': [avg_purchase_price_male, avg_purchase_price_female, avg_purchase_price_other],
                                     'Total Purchase Value': [purchase_total_male, purchase_total_female, purchase_total_other],
                                     'Avg Total Purchase Per Person': [avg_per_person_male, avg_per_person_female, avg_per_person_other]}

gender_purchase_analysis_final = pd.DataFrame(gender_purchase_analysis_dataframe)
gender_purchase_analysis_final = gender_purchase_analysis_final.set_index('Gender')
gender_purchase_analysis_final[['Purchase Count', 'Avg Purchase Price', 'Total Purchase Value',
                               'Avg Total Purchase Per Person']]

gender_purchase_analysis_final

```

Out[16]:

	Purchase Count	Avg Purchase Price	Total Purchase Value	Avg Total Purchase Per Person
Gender				
Male	652	\$3.02	\$1,967.64	\$4.07
Female	113	\$3.20	\$361.94	\$4.47
Other/Non-Disclosed	15	\$3.35	\$50.19	\$4.56

In [17]:

```

# Create bins & labels
bins = [0, 9, 14, 19, 24, 29, 34, 39, 45]
bin_labels = ["<10", "10 - 14", "15 - 19", "20 - 24", "25 - 29", "30 - 34", "35 - 39", "40+"]

total_players = len(purchase_data["SN"].unique())

# Cut to bin data, adds age group column to dataframe
purchase_data["Age Group"] = pd.cut(purchase_data["Age"], bins=bins, labels=bin_labels)
purchase_data

purchase_data_group = purchase_data.groupby("Age Group")

total_count_age = purchase_data_group["SN"].nunique()

percent_age = (total_count_age/total_players)*100
percent_age

age_demographics = pd.DataFrame({"Total Count": total_count_age, "Percentage of Players": percent_age})
age_demographics.index.name = None
age_demographics.style.format({"Percentage of Players": "{:.2f}%"})

```

```
age_demographics.style.format({"Percentage of Players": "{:.2f}%",
```

Out[17]:

Total Count		Percentage of Players
<10	17	2.95%
10 - 14	22	3.82%
15 - 19	107	18.58%
20 - 24	258	44.79%
25 - 29	77	13.37%
30 - 34	52	9.03%
35 - 39	31	5.38%
40+	12	2.08%

In [18]:

```
#Purchase Analysis by Age
bins = [0, 9, 14, 19, 24, 29, 34, 39, 45]
bin_labels = ["<10", "10 - 14", "15 - 19", "20 - 24", "25 - 29", "30 - 34", "35 - 39", "40+"]

purchase_data["Age Group"] = pd.cut(purchase_data["Age"], bins=bins, labels=bin_labels)
purchase_data

purchase_data_group = purchase_data.groupby("Age Group")
purchase_data_group

purchase_count_age = purchase_data_group["Purchase ID"].count()
avg_purchase_price = purchase_data_group["Price"].mean()
total_purchase_value = purchase_data_group["Price"].sum()
avg_purchase_person = total_purchase_value/purchase_count_age

age_purchase_analysis = pd.DataFrame({"Purchase Count": purchase_count_age,
                                     "Avg Purchase Price": avg_purchase_price,
                                     "Total Purchase Value": total_purchase_value,
                                     "Avg Total Purchase per Person": avg_purchase_person})

age_purchase_analysis.index.name = None

# Format with currency style
age_purchase_analysis.style.format({"Avg Purchase Price": "${:,.2f}",
                                   "Total Purchase Value": "${:,.2f}",
                                   "Avg Total Purchase per Person": "${:,.2f}"})
```

Out[18]:

Purchase Count		Avg Purchase Price	Total Purchase Value	Avg Total Purchase per Person
<10	23	\$3.35	\$77.13	\$3.35
10 - 14	28	\$2.96	\$82.78	\$2.96
15 - 19	136	\$3.04	\$412.89	\$3.04
20 - 24	365	\$3.05	\$1,114.06	\$3.05
25 - 29	101	\$2.90	\$293.00	\$2.90
30 - 34	73	\$2.93	\$214.00	\$2.93
35 - 39	41	\$3.60	\$147.67	\$3.60
40+	13	\$2.94	\$38.24	\$2.94

In [19]:

```
#top spenders
top_spenders = purchase_data.groupby("SN")
count_top_spenders = top_spenders["Purchase ID"].count()
avg_pur_price_top_spenders = top_spenders["Price"].mean()
total_pur_value_top_spenders = top_spenders["Price"].sum()

top_spenders_df = pd.DataFrame({"Purchase Count": count_top_spenders,
                                "Average Purchase Price": avg_pur_price_top_spenders,
                                "Total Purchase Value per Person":
                                    total_pur_value_top_spenders})
top_spenders_df_formatted = top_spenders_df.sort_values(["Total Purchase Value per Person"],
                                                         ascending=False).head()

top_spenders_df_formatted.style.format({"Average Purchase Price": "${:,.2f}",
                                         "Total Purchase Value per Person": "${:,.2f}"})
```

Out[19]:

	Purchase Count	Average Purchase Price	Total Purchase Value per Person
SN			
Lisosia93	5	\$3.79	\$18.96
Idastidru52	4	\$3.86	\$15.45
Chamjask73	3	\$4.61	\$13.83
Iral74	4	\$3.40	\$13.62
Iskadarya95	3	\$4.37	\$13.10

In [20]:

```
#popular items
item_df = purchase_data[["Item ID", "Item Name", "Price"]]
item_df_group = item_df.groupby(["Item ID", "Item Name"])

item_count = item_df_group["Price"].count()
total_item_value = (item_df_group["Price"].sum())

price_per_item = total_item_value/item_count

pop_item = pd.DataFrame({"Purchase Count": item_count,
                          "Item Price": price_per_item,
                          "Total Purchase Value": total_item_value})
pop_item_sorted = pop_item.sort_values(["Purchase Count"],
                                         ascending=False).head()

pop_item_sorted.style.format({"Item Price": "${:,.2f}",
                              "Item Price": "${:,.2f}",
                              "Total Purchase Value": "${:,.2f}"})
```

Out[20]:

		Purchase Count	Item Price	Total Purchase Value
Item ID	Item Name			
178	Oathbreaker, Last Hope of the Breaking Storm	12	\$4.23	\$50.76
145	Fiery Glass Crusader	9	\$4.58	\$41.22
108	Extraction, Quickblade Of Trembling Hands	9	\$3.53	\$31.77
82	Nirvana	9	\$4.90	\$44.10
19	Pursuit, Cudgel of Necromancy	8	\$1.02	\$8.16

In [21]:

```
#profitable items
```

```
pop_item_sorted = pop_item.sort_values(["Total Purchase Value"],
                                       ascending=False).head()

# Format with currency style
pop_item_sorted.style.format({"Item Price": "${:,.2f}",
                             "Total Purchase Value": "${:,.2f}"})
```

Out[21]:

		Purchase Count	Item Price	Total Purchase Value
Item ID	Item Name			
178	Oathbreaker, Last Hope of the Breaking Storm	12	\$4.23	\$50.76
82	Nirvana	9	\$4.90	\$44.10
145	Fiery Glass Crusader	9	\$4.58	\$41.22
92	Final Critic	8	\$4.88	\$39.04
103	Singed Scalpel	8	\$4.35	\$34.80