In [13]:

```
# Heroes of Pymoli
# By: Jack Cohen

# Import Dependencies and Setup
import pandas as pd

# File to Load
file = 'Resources/purchase_data.csv'

# Read Purchasing File and store into Pandas data frame
purchase_data = pd.read_csv(file)
```

In [14]:

```
# Player Count
player_list = purchase_data.SN.unique()
total_num_players = len(player_list)
print(f"There are {total_num_players} total players.")
player = pd.DataFrame([{'Total Players':total_num_players}])
player
```

```
There are 576 total players.
```

Out[14]:

	Total Players
0	576

In [3]:

```
# Purchasing Analysis
num_unique_items = purchase_data["Item ID"].value_counts().count()
genders = purchase_data.Gender.unique()
avg_price = round(purchase_data["Price"].mean(),2)
num_purchases = purchase_data["SN"].count()
total_rev = purchase_data["Price"].sum()
data_summary = pd.DataFrame([{" ":"Value","Number of Unique Items":num_unique_items,
                "Average Price ($)":avg_price,
                "Number of Purchases":num_purchases,
                "Total Revenue ($)":total_rev}])
data_summary.style.format({"Average Price ($)":'${0:,.2f}',"Total Revenue ($)":'${0:,.2f}'})
.hide_index()
```

Out[3]:

Number of Unique Items	Average Price ($)	Number of Purchases	Total Revenue ($)
Value	179	$3.05	780	$2,379.77

In [4]:

```
# Gender Demographics
no_dups = purchase_data.drop_duplicates(subset=["SN"])
SN_Gender = no_dups[["SN","Gender"]]
SN_Gender = SN_Gender.groupby("Gender").count()
percent_gender = [100*SN_Gender.iloc[:,0][0]/total_num_players,
                  100*SN_Gender.iloc[:,0][1]/total_num_players,
                  100*SN_Gender.iloc[:,0][2]/total_num_players]
SN_Gender.insert(1,"Percentage of Players",percent_gender)
SN_Gender.rename(columns={"SN":"Total Count"},inplace=True)
```

In [5]:

```
SN_Gender.style.format({"Percentage of Players":"{0:,.2f}%"})
```

Out[5]:

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

In [6]:

```
# Purchasing Analysis (Gender)
all_genders = purchase_data.Gender.unique()
gender_df = purchase_data.groupby("Gender")
purchase_count = gender_df['Gender'].count()
average_price = gender_df.mean()["Price"]
total_purchase_val = purchase_count*average_price
avg_tot_purchase_per_person = total_purchase_val/SN_Gender["Total Count"]
```

In [7]:

```
gender_summary = pd.DataFrame({
                                "Purchase Count":purchase_count,
                                "Average Purchase Price":list(average_price),
                                "Total Purchase Value":list(total_purchase_val),
                                "Avg Total Purchase per Person":list(avg_tot_purchase_per_per
son)
                               })
gender_summary = gender_summary.style.format({"Average Purchase Price":'${0:,.2f}',
                             "Total Purchase Value":'${0:,.2f}',
                             "Avg Total Purchase per Person":'${0:,.2f}'})
gender_summary
```

Out[7]:

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

In [8]:

```
# Age Demographics
age_bins = [0,9,14,19,24,29,34,39,200]
age_labels = ['<10','10-14','15-19','20-24','25-29','30-34','35-39','40+']
age_sorted = pd.cut(no_dups['Age'],age_bins, labels=age_labels).value_counts()
percent_age = 100*age_sorted[age_labels]/total_num_players
age_summary = pd.DataFrame({"":age_labels,
                            "Total Count":list(age_sorted[age_labels]),
                            "Percentage of Players":list(percent_age)
                             })
age_summary = age_summary.style.format({"Percentage of Players":'{0:,.2f}%'}).hide_index()
age_summary
```

Out[8]:

	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 [9]:

```
# Purchasing Analysis (Age)
age_analysis = purchase_data
age_analysis = age_analysis.drop(columns=["SN","Gender","Item ID","Item Name","Purchase ID"
])
age_analysis["Age Range"] = pd.cut(purchase_data['Age'],age_bins, labels=age_labels)
age_sorts = age_analysis.groupby("Age Range")
purchase_count_age = age_sorts["Age"].count()
avg_purch_price_age = age_sorts["Price"].mean()
total_purch_val_age = age_sorts["Price"].sum()
avg_tot_purch_per_person_age = total_purch_val_age / age_sorted[age_labels]
age_purchasing_df = pd.DataFrame({'Purchase Count':purchase_count_age,
                                  'Average Purchase Price':avg_purch_price_age,
                                  'Total Purchase Value':total_purch_val_age,
                                  'Avg Total Purchase per Person':avg_tot_purch_per_person_a
ge})
age_purchasing_df = age_purchasing_df.style.format({"Average Purchase Price":'${0:,.2f}',
                            "Total Purchase Value":'${0:,.2f}',
                            "Avg Total Purchase per Person":'${0:,.2f}'})

age_purchasing_df
```

Out[9]:

Age Range	Purchase Count	Average Purchase Price	Total Purchase Value	Avg Total Purchase per Person
<10	23	$3.35	$77.13	$4.54
10-14	28	$2.96	$82.78	$3.76
15-19	136	$3.04	$412.89	$3.86
20-24	365	$3.05	$1,114.06	$4.32
25-29	101	$2.90	$293.00	$3.81
30-34	73	$2.93	$214.00	$4.12
35-39	41	$3.60	$147.67	$4.76
40+	13	$2.94	$38.24	$3.19

In [10]:

```
# Top Spenders
SN_sorted = purchase_data.groupby("SN")
SN_purch_count = SN_sorted["SN"].count()
SN_avg_pp = SN_sorted['Price'].mean()
SN_tot_purch_vol = SN_sorted['Price'].sum()
SN_summary_df = pd.DataFrame({'Purchase Count':SN_purch_count,
                              'Average Purchase Price':SN_avg_pp,
                              'Total Purchase Value':SN_tot_purch_vol})
SN_summary_df = SN_summary_df.sort_values(by='Total Purchase Value',axis=0,ascending=False)
SN_head = SN_summary_df.head()
SN_head.style.format({'Average Purchase Price':'${0:,.2f}',
                      'Total Purchase Value':'${0:,.2f}'})
```

Out[10]:

SN	Purchase Count	Average Purchase Price	Total Purchase Value
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 [11]:

```
# Most Popular Items
popular_items_df = purchase_data[['Item ID','Item Name','Price']]
pop_main = popular_items_df.groupby(['Item ID','Item Name'])
item_prices = popular_items_df.drop_duplicates(subset='Item Name')['Price']
pop_purch_count = pop_main.count()['Price']
pop_item_price = pop_main.mean()['Price']
pop_total_val = pop_main.sum()['Price']
pop_summary = pd.DataFrame({'Purchase Count':pop_purch_count,
                            'Item Price':pop_item_price,
                            'Total Purchase Value':pop_total_val})
pop_summary_count_sorted = pop_summary.sort_values(by='Purchase Count',axis=0,ascending=Fals
e)
pop_head_count = pop_summary_count_sorted.head()
pop_head_count.style.format({'Item Price':'${0:,.2f}',
                       'Total Purchase Value':'${0:,.2f}'})
```

Out[11]:

		Purchase Count	Item Price	Total Purchase Value
Item ID	**Item Name**			
92	**Final Critic**	13	$4.61	$59.99
178	**Oathbreaker, Last Hope of the Breaking Storm**	12	$4.23	$50.76
145	**Fiery Glass Crusader**	9	$4.58	$41.22
132	**Persuasion**	9	$3.22	$28.99
108	**Extraction, Quickblade Of Trembling Hands**	9	$3.53	$31.77

In [12]:

```
# Most Profitable Items
pop_summary_value_sorted = pop_summary.sort_values(by='Total Purchase Value',axis=0,ascendin
g=False)
pop_head_value = pop_summary_value_sorted.head()
pop_head_value.style.format({'Item Price':'${0:,.2f}',
                       'Total Purchase Value':'${0:,.2f}'})
```

Out[12]:

		Purchase Count	Item Price	Total Purchase Value
Item ID	**Item Name**			
92	**Final Critic**	13	$4.61	$59.99
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
103	**Singed Scalpel**	8	$4.35	$34.80

In [ ]: