

## Heroes of Pymoli Data Analysis

- Over half of players making purchases are ages 15-24.
- Players making in game purchases are most likely male.
- Of purchases, there is not one overwhelming majority purchase, rather they are somewhat evenly distributed among items

```
import pandas as pd
import os
import numpy as np
```

```
json_path1 = os.path.join("purchase_data.json")
json_path2 = os.path.join("purchase_data2.json")
```

```
data1_df = pd.read_json(json_path1)
data2_df = pd.read_json(json_path2)
```

## Player Count

(first dataset will always be shown in the lowest index row, followed by second dataset)

```
ps1 = data1_df["SN"].unique()
ps2 = data2_df["SN"].unique()
```

```
players_df = pd.DataFrame({"Player Count": [len(ps1), len(ps2)]})
players_df
```

Player Count	
0	573
1	74

## Purchasing Analysis (Total)

```
P_analysis_df = pd.DataFrame({"Number of Unique Items": [len(data1_df["Item Name"].unique()),
                                                         len(data2_df["Item Name"].unique())],
                              "Average Purchase Price": ['$ {:.2f}'.format(data1_df["Price"].mean()),
                                                         '$ {:.2f}'.format(data2_df["Price"].mean())],
                              "Total Number of Purchases": [len(data1_df["Price"]),
                                                            len(data2_df["Price"])],
                              "Total Revenue": ['$ {:.2f}'.format(data1_df["Price"].sum()),
                                                 '$ {:.2f}'.format(data2_df["Price"].sum())])
P_analysis_df
```

	Average Purchase Price	Number of Unique Items	Total Number of Purchases	Total Revenue
0	\$2.93	179	780	\$2286.33
1	\$2.92	63	78	\$228.10

## Gender Demographics

```
male = data1_df.loc[data1_df["Gender"] == "Male"]
female = data1_df.loc[data1_df["Gender"] == "Female"]
other = data1_df.loc[data1_df["Gender"] == "Other / Non-Disclosed"]

gender_df = pd.DataFrame({"Gender": ["Male", "Female", "Other / Non-Disclosed"],
                           "Percentage of Players": ['{:0.2f}%'.format(float(100*(len(male["SN"]).unique())/len(data1_df["SN"])),
                                                       '{:0.2f}%'.format(float(100*(len(female["SN"]).unique())/len(data1_df["SN"])),
                                                       '{:0.2f}%'.format(float(100*(len(other["SN"]).unique())/len(data1_df["SN"])),
                           "Total Count": [len(male["SN"].unique()), len(female["SN"].unique()), len(other["SN"].unique())])
gender_df.set_index("Gender", drop=True)
```

Percentage of Players		Total Count
Gender		
Male	81.15%	465
Female	17.45%	100
Other / Non-Disclosed	1.40%	8

Same code, subbing dataset 2:

Percentage of Players		Total Count
Gender		
Male	81.08%	60
Female	17.57%	13
Other / Non-Disclosed	1.35%	1

## Purchasing Analysis (Gender)

```
G_analysis = pd.DataFrame({"Gender": ["Male", "Female", "Other / Non-Disclosed"],
                           "Purchase Count": [len(male["Item Name"]), len(female["Item Name"]), len(other["Item Name"])],
                           "Average Purchase Price": ['${:0.2f}'.format(male["Price"].mean()),
                                                       '${:0.2f}'.format(female["Price"].mean()),
                                                       '${:0.2f}'.format(other["Price"].mean())],
                           "Total Purchase Value": [male["Price"].sum(), female["Price"].sum(), other["Price"].sum()],
                           "Normalized Totals": ['${:0.2f}'.format(male["Price"].sum()/len(male["SN"].unique()),
                                                       '${:0.2f}'.format(female["Price"].sum()/len(female["SN"].unique()),
                                                       '${:0.2f}'.format(other["Price"].sum()/len(other["SN"].unique()))])
G_analysis.set_index("Gender", drop=True)
```

Average Purchase Price		Normalized Totals	Purchase Count	Total Purchase Value
Gender				
Male	\$2.95	\$4.02	633	\$1,867.68
Female	\$2.82	\$3.83	136	\$382.91
Other / Non-Disclosed	\$3.25	\$4.47	11	\$35.74

Same code, subbing dataset 2:

Average Purchase Price		Normalized Totals	Purchase Count	Total Purchase Value
Gender				
Male	\$2.88	\$3.08	64	\$184.60
Female	\$3.18	\$3.18	13	\$41.38
Other / Non-Disclosed	\$2.12	\$2.12	1	\$2.12

## Age Demographics

```
bins = [0,10,15,20,25,30,35,40,100]
age_groups = [<"<10", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39", "40+">]
ages = data1_df
ages["Age Group"] = pd.cut(ages["Age"],bins,labels=age_groups)
```

```
age_group = ages.groupby("Age Group")
age_demo = pd.DataFrame({"Purchase Count":age_group["Item Name"].count(),
                        "Average Purchase Price":age_group["Price"].mean(),
                        "Total Purchase Value":age_group["Price"].sum()})
pd.options.display.float_format = '{:,.2f}'.format
age_demo
```

	Average Purchase Price	Purchase Count	Total Purchase Value
Age Group			
10-14	\$2.87	78	\$224.15
15-19	\$2.87	184	\$528.74
20-24	\$2.96	305	\$902.61
25-29	\$2.89	76	\$219.82
30-34	\$3.07	58	\$178.26
35-39	\$2.90	44	\$127.49
40+	\$2.88	3	\$8.64
<10	\$3.02	32	\$96.62

Same code, subbing dataset 2:

	Average Purchase Price	Purchase Count	Total Purchase Value
Age Group			
10-14	\$3.05	4	\$12.21
15-19	\$2.73	20	\$54.69
20-24	\$3.04	33	\$100.42
25-29	\$2.69	4	\$10.77
30-34	\$2.35	7	\$16.47
35-39	\$3.94	5	\$19.72
40+	\$nan	0	\$nan
<10	\$2.76	5	\$13.82

## Top Spenders

```
sn_totals = {"SN":[], "Purchase Count":[], "Average Purchase Price":[], "Total Purchase Value":[]}
sn_list = datal_df["SN"].unique()
for name in sn_list:
    newobject = datal_df.loc[datal_df["SN"] == name]
    sn_totals["SN"].append(name)
    sn_totals["Purchase Count"].append(len(newobject["Item Name"]))
    sn_totals["Average Purchase Price"].append(newobject["Price"].mean())
    sn_totals["Total Purchase Value"].append(newobject["Price"].sum())
sn_totals_df = pd.DataFrame(sn_totals)
sn_totals_df = sn_totals_df.sort_values("Total Purchase Value", ascending=False)
sn_totals_df = sn_totals_df.reset_index(drop=True)
pd.options.display.float_format = '${:,.2f}'.format
sn_totals_df.head()
```

	Average Purchase Price	Purchase Count	SN	Total Purchase Value
0	\$3.41	5	Undirrala66	\$17.06
1	\$3.39	4	Saedue76	\$13.56
2	\$3.18	4	Mindimnya67	\$12.74
3	\$4.24	3	Haellysu29	\$12.73
4	\$3.86	3	Eoda93	\$11.58

Same code, subbing dataset 2:

	Average Purchase Price	Purchase Count	SN	Total Purchase Value
0	\$2.56	2	Aidaira26	\$5.13
1	\$4.71	1	Alarap40	\$4.71
2	\$4.65	1	Aerithllora36	\$4.65
3	\$4.28	1	Heosurnuru52	\$4.28
4	\$4.12	1	Arithllorin55	\$4.12

## Most Popular Items

```
item_totals = {"Item Name":[], "Purchase Count":[], "Item ID":[], "Item Price":[], "Total Purchase Value":[]}
item_list = datal_df["Item Name"].unique()
for item in item_list:
    newobject = datal_df.loc[datall_df["Item Name"] == item]
    item_totals["Item Name"].append(item)
    item_totals["Purchase Count"].append(len(newobject["Item Name"]))
    item_totals["Item ID"].append(newobject["Item ID"].iloc[0])
    item_totals["Item Price"].append(newobject["Price"].mean())
    item_totals["Total Purchase Value"].append(newobject["Price"].sum())
item_totals_df = pd.DataFrame(item_totals)
item_totals_df = item_totals_df.sort_values("Purchase Count", ascending=False)
item_totals_df = item_totals_df.reset_index(drop=True)
pd.options.display.float_format = '{:,.2f}'.format
item_totals_df.head()
```

	Item ID	Item Name	Item Price	Purchase Count	Total Purchase Value
0	92	Final Critic	\$2.76	14	\$38.60
1	84	Arcane Gem	\$2.23	11	\$24.53
2	39	Betrayal, Whisper of Grieving Widows	\$2.35	11	\$25.85
3	30	Stormcaller	\$3.46	10	\$34.65
4	175	Woeful Adamantite Claymore	\$1.24	9	\$11.16

Same code, subbing dataset 2:

	Item ID	Item Name	Item Price	Purchase Count	Total Purchase Value
0	94	Mourning Blade	\$3.64	3	\$10.92
1	93	Apocalyptic Battlescythe	\$4.49	2	\$8.98
2	90	Betrayer	\$4.12	2	\$8.24
3	60	Wolf	\$2.70	2	\$5.40
4	64	Fusion Pummel	\$2.42	2	\$4.84

## Most Profitable Items

```
item_totals_df = item_totals_df.sort_values("Total Purchase Value", ascending=False)
item_totals_df = item_totals_df.reset_index(drop=True)
pd.options.display.float_format = '{:,.2f}'.format
item_totals_df.head()
```

	Item ID	Item Name	Item Price	Purchase Count	Total Purchase Value
0	92	Final Critic	\$2.76	14	\$38.60
1	34	Retribution Axe	\$4.14	9	\$37.26
2	30	Stormcaller	\$3.46	10	\$34.65
3	115	Spectral Diamond Doomblade	\$4.25	7	\$29.75
4	32	Orenmir	\$4.95	6	\$29.70

Same code, subbing dataset 2:

Item ID		Item Name	Item Price	Purchase Count	Total Purchase Value
0	94	Mourning Blade	\$3.64	3	\$10.92
1	93	Apocalyptic Battlescythe	\$4.49	2	\$8.98
2	90	Betrayer	\$4.12	2	\$8.24
3	60	Wolf	\$2.70	2	\$5.40
4	64	Fusion Pummel	\$2.42	2	\$4.84