

```
In [ ]: # import pandas as pd
```

```
In [157]: json_path = "purchase_data.json"
players_df = pd.read_json(json_path)
players_df.head()
```

Out[157]:

	Age	Gender	Item ID	Item Name	Price	SN
0	38	Male	165	Bone Crushing Silver Skewer	3.37	Aelalis34
1	21	Male	119	Stormbringer, Dark Blade of Ending Misery	2.32	Eolo46
2	34	Male	174	Primitive Blade	2.46	Assastnya25
3	21	Male	92	Final Critic	1.36	Pheusrical25
4	23	Male	63	Stormfury Mace	1.27	Aela59

```
In [158]: #Players Count
#total number of players
Number_of_Players = players_df["SN"].nunique()
num_players = pd.DataFrame({"Total Players":[Number_of_Players]})
num_players
```

Out[158]:

	Total Players
0	573

```
In [159]: #Purchasing Analysis (Total)
#number of uniques items, average price, number of purchases, total rev
count = players_df["Item ID"].nunique()
avg = players_df["Price"].mean()
purchase = players_df["Item ID"].count()
revenue = players_df["Price"].sum()

purch_analysis = pd.DataFrame({"Number of Unique Items":[count],
                              "Average Price":[avg],
                              "Number of Purchases":[purchase],
                              "Total Revenue":[revenue]})
purch_analysis["Average Price"] = purch_analysis["Average Price"].map("${:,.2f}"
".format)
purch_analysis["Total Revenue"] = purch_analysis["Total Revenue"].map("${:,.2
f}".format)
purch_analysis
```

Out[159]:

	Number of Unique Items	Average Price	Number of Purchases	Total Revenue
0	183	\$2.93	780	\$2,286.33

```
In [160]: #Gender Demographics
#group by gender, columns: Percentages of Players, Total Count
gender_total = players_df["Gender"].count()

#male_total = (players_df.loc[(players_df["Gender"]=="Male"), "Gender"]).mean()

grouped_gender = players_df.groupby(['Gender'])

gender_data = pd.DataFrame({"Percentage of Players":
    ((grouped_gender['Gender'].count()/gender_total)*100).round(2),
    "Total Count":grouped_gender["Gender"].count()})

gender_data["Percentage of Players"]= gender_data["Percentage of Players"].map
("{0:,.2f}%".format)
gender_data
```

Out[160]:

	Percentage of Players	Total Count
Gender		
Female	17.44%	136
Male	81.15%	633
Other / Non-Disclosed	1.41%	11

```

In [161]: #Purchase Analysis (Gender)
#purchase count, avg purchase price, total purchase value, normalized totals
players_grouped = players_df.groupby(["Gender"])
players_grouped.count()

gender_total = players_df["Gender"].count()
revenue = players_df["Price"].sum()
normalized_total = (revenue/gender_total)

aggregation={'Price':{'Purchase Count':'count',
                      'Average Purchase Price':'mean',
                      'Total Purchase Value':'sum'}}

players_data = players_df.groupby('Gender').agg(aggregation)

#players_data['Average Purchase Price']= players_data['Average Purchase Price'].map("${:,.2f}".format)
#players_data["Total Purchase Value"]= players_data["Total Purchase Value"].map("${:,.2f}".format)
players_data

```

C:\Users\jpaz5\Anaconda3\lib\site-packages\pandas\core\groupby\groupby.py:4658: FutureWarning: using a dict with renaming is deprecated and will be removed in a future version

```

return super(DataFrameGroupBy, self).aggregate(arg, *args, **kwargs)

```

Out[161]:

	Price		
	Purchase Count	Average Purchase Price	Total Purchase Value
Gender			
Female	136	2.815515	382.91
Male	633	2.950521	1867.68
Other / Non-Disclosed	11	3.249091	35.74

In [162]: *#Age Demographics*

```
bins = [0, 10, 14.90, 19.90, 24.90, 29.90, 34.90, 39.90, 99999]
group_labels = ["<10", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39", "40+"]
players_df["Age Group"] = pd.cut(players_df["Age"], bins, labels=group_labels)

age_group = players_df.groupby(["Age Group"])
age_count = age_group["SN"].count()
player_count = players_df["SN"].nunique()

player_percentage = ((age_count/player_count)*100).round(2)

final_df = pd.DataFrame({"Percentage of Players": player_percentage, "Total Count": age_count})
final_df["Percentage of Players"] = final_df["Percentage of Players"].map("{:0,.2f}%".format)
final_df
```

Out[162]:

	Percentage of Players	Total Count
Age Group		
<10	5.58%	32
10-14	5.41%	31
15-19	23.21%	133
20-24	58.64%	336
25-29	21.82%	125
30-34	11.17%	64
35-39	7.33%	42
40+	2.97%	17

```

In [163]: #Purchasing Analysis (Age)
#Purchase Count,Average Purchase Price, Total Purchase Value, Normalized Total
S
bins = [0, 10, 14.90, 19.90, 24.90, 29.90, 34.90, 39.90, 99999]
group_labels = [<10","10-14","15-19","20-24","25-29","30-34","35-39","40+"]
players_df["Age Group"]=pd.cut(players_df["Age"], bins,labels=group_labels)

avg = players_df.groupby(["Age Group"]).mean()["Price"].rename("Average Purchase Price")
purchase = players_df.groupby(["Age Group"]).count()["Price"].rename("Purchase Count")
revenue = players_df.groupby(["Age Group"]).sum()["Price"].rename("Total Purchase Value")
gender_total = players_df["Gender"].count()
normalized_total = (revenue/gender_total)

purch_df = pd.DataFrame({"Purchase Count": purchase, "Average Purchase Price": avg,
                        "Total Purchase Value":revenue,"Normalized Totals":normalized_total})

purch_df["Average Purchase Price"]= purch_df["Average Purchase Price"].map("{:,.2f}".format)
purch_df["Total Purchase Value"] = purch_df["Total Purchase Value"].map("{:,.2f}".format)
purch_df["Normalized Totals"]= purch_df["Normalized Totals"].map("{:,.2f}".format)

purch_df

```

Out[163]:

	Purchase Count	Average Purchase Price	Total Purchase Value	Normalized Totals
Age Group				
<10	32	\$3.02	\$96.62	\$0.12
10-14	31	\$2.70	\$83.79	\$0.11
15-19	133	\$2.91	\$386.42	\$0.50
20-24	336	\$2.91	\$978.77	\$1.25
25-29	125	\$2.96	\$370.33	\$0.47
30-34	64	\$3.08	\$197.25	\$0.25
35-39	42	\$2.84	\$119.40	\$0.15
40+	17	\$3.16	\$53.75	\$0.07

```

In [165]: #Top Spenders: top 5 spenders by total purchase value (SN,purchase count, average purchase price, total purchase value)
avg = players_df.groupby(["SN"]).mean()["Price"].rename("Average Purchase Price")
purchase = players_df.groupby(["SN"]).count()["Price"].rename("Purchase Count")
revenue = players_df.groupby(["SN"]).sum()["Price"].rename("Total Purchase Value")

top_df= pd.DataFrame({"Purchase Count":purchase,
                      "Average Purchase Price":avg,
                      "Total Purchase Value":revenue})

top_ordered = top_df.sort_values("Total Purchase Value",ascending=False)

top_ordered["Average Purchase Price"]= top_ordered["Average Purchase Price"].map(
"${:,.2f}".format)
top_ordered["Total Purchase Value"] = top_ordered["Total Purchase Value"].map(
"${:,.2f}".format)
top_ordered.head(5)

```

Out[165]:

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

```

In [166]: #Most Popular Items: Item ID, Item Name, Purchase Count, Item Price, Total Purchase Value
popular_df = players_df.loc[:,["Item ID", "Item Name", "Price"]]
#popular_df

purchase = players_df.groupby(["Item ID","Item Name"]).count()["Price"].rename("Purchase Count")
avg = players_df.groupby(["Item ID","Item Name"]).mean()["Price"].rename("Item Price")
revenue = players_df.groupby(["Item ID","Item Name"]).sum()["Price"].rename("Total Purchase Value")

popular_item= pd.DataFrame({"Purchase Count":purchase,
                           "Item Price": avg,
                           "Total Purchase Value":revenue})

popular_item = popular_item.sort_values("Purchase Count",ascending=False)

popular_item["Item Price"]= popular_item["Item Price"].map("${:,.2f}".format)
popular_item["Total Purchase Value"] = popular_item["Total Purchase Value"].map("${:,.2f}".format)
popular_item.head(5)

```

Out[166]:

		Purchase Count	Item Price	Total Purchase Value
Item ID	Item Name			
39	Betrayal, Whisper of Grieving Widows	11	\$2.35	\$25.85
84	Arcane Gem	11	\$2.23	\$24.53
31	Trickster	9	\$2.07	\$18.63
175	Woeful Adamantite Claymore	9	\$1.24	\$11.16
13	Serenity	9	\$1.49	\$13.41

```

In [167]: #Most Popular Items: Item ID, Item Name, Purchase Count, Item Price, Total Purchase Value
popular_df = players_df.loc[:,["Item ID", "Item Name", "Price"]]
#popular_df

purchase = players_df.groupby(["Item ID","Item Name"]).count()["Price"].rename("Purchase Count")
avg = players_df.groupby(["Item ID","Item Name"]).mean()["Price"].rename("Item Price")
revenue = players_df.groupby(["Item ID","Item Name"]).sum()["Price"].rename("Total Purchase Value")

popular_item= pd.DataFrame({"Purchase Count":purchase,
                           "Item Price": avg,
                           "Total Purchase Value":revenue})

popular_item = popular_item.sort_values("Total Purchase Value",ascending=False)

popular_item["Item Price"]= popular_item["Item Price"].map("${:,.2f}".format)
popular_item["Total Purchase Value"] = popular_item["Total Purchase Value"].map("${:,.2f}".format)
popular_item.head(5)

```

Out[167]:

		Purchase Count	Item Price	Total Purchase Value
Item ID	Item Name			
34	Retribution Axe	9	\$4.14	\$37.26
115	Spectral Diamond Doomblade	7	\$4.25	\$29.75
32	Orenmir	6	\$4.95	\$29.70
103	Singed Scalpel	6	\$4.87	\$29.22
107	Splitter, Foe Of Subtlety	8	\$3.61	\$28.88