# Dependencies and Setup

import pandas as pd

import numpy as np

# File to Load (Load file as per path)

file\_to\_load = "Resources/purchase\_data.csv"

# Read Purchasing File and store into Pandas data frame

purchase\_data = pd.read\_csv(file\_to\_load)

purchase\_data.head()

numberofplayers=purchase\_data['SN'].value\_counts()

players\_df = pd.DataFrame({

"Total Players": [len(numberofplayers)]

})

players\_df

#Take Price Column and Sum

Total\_Revenue=purchase\_data["Price"].sum()

#print(f"Total Revenue :",Total\_Revenue)

#Take Price Column and find Mean

Avg\_Purchase\_Price=purchase\_data["Price"].mean()

#print(f"Average Purchase Price :",round(Avg\_Purchase\_Price,2))

#Take Price Column and find total count

Total\_Number\_Purchase=purchase\_data["Price"].count()

#print(f"Total Number of Purchase :", Total\_Number\_Purchase)

# FInd number of Unique Value

Number\_Unique\_Item=purchase\_data["Item Name"].nunique()

#print(f"Number of Unique Item :", Number\_Unique\_Item)

# Create DataFrame Using above varibales

purchasing\_df = pd.DataFrame({

"Number of Unique Items": [Number\_Unique\_Item],

"Average Price $" : [round(Avg\_Purchase\_Price,2)],

"Number of Purchases" : [Total\_Number\_Purchase],

"Total Revenue $" : [Total\_Revenue]

})

purchasing\_df

#Group by Gender & SN, find unique values

gender\_count = purchase\_data.groupby('Gender')['SN'].nunique()

Total\_Gender\_Count = purchase\_data.groupby('Gender')['SN'].nunique().sum()

percentage\_gender = (gender\_count/Total\_Gender\_Count)\*100

#print(percentage\_gender)

gender\_demo = pd.DataFrame({

"Total Counts": gender\_count,

"Percentage of Players": round(percentage\_gender,2),

})

# Index Added

# gender\_demo.index = (["Female", "Male", "Other / Non-Disclosed"])

gender\_demo.sort\_values('Total Counts',ascending = False)

# Group by Gender

gender\_group\_purchase\_data = purchase\_data.groupby(["Gender"])

gender\_group\_purchase\_data.count()

purchase\_count=gender\_group\_purchase\_data["SN"].count()

#print(purchase\_count)

avg\_purchase\_price=gender\_group\_purchase\_data["Price"].mean()

#print(avg\_purchase\_price)

total\_purchase\_value=gender\_group\_purchase\_data["Price"].sum()

#print(total\_purchase\_value)

avg\_purchase\_price\_perperson = round((gender\_group\_purchase\_data["Price"].sum() / gender\_count),2)

#print(avg\_purchase\_price\_perperson)

# Create new DataFrame

purchase\_analysis\_gender = pd.DataFrame({ "Purchase Count" : purchase\_count,

"Average Purchase Price" : round(avg\_purchase\_price,2),

"Total Purchase Value" : total\_purchase\_value,

"Avg Total Purchase per Person": avg\_purchase\_price\_perperson})

purchase\_analysis\_gender

age\_bins = [0, 9.99, 14.99, 19.99, 24.99, 29.99, 34.99, 39.99, 100]

group\_names = ["<10", "10-14", "15-19", "20-24", "25-29", "30-34", "35-39", "40+"]

purchase\_data["Age\_Summary"]= pd.cut(purchase\_data["Age"], age\_bins, labels=group\_names)

purchase\_data

purchase\_data.head()

group\_data = purchase\_data.groupby(['SN', 'Age\_Summary'])['Age'].count()

#print(group\_data)

total\_counts = group\_data.groupby('Age\_Summary').count()

total\_counts

# %Calcualtions

percentage =round((total\_counts/len(group\_data))\*100,2)

percentage

age\_demographics = pd.DataFrame ({

"Total Counts" : total\_counts,

"Percentage of Players" : percentage,

})

age\_demographics

# Count purchase data using SN by age bin

purchase\_data\_item = purchase\_data.groupby('Age\_Summary')['SN'].count()

#print(purchase\_data\_item)

average\_purchase\_price = purchase\_data.groupby('Age\_Summary')['Price'].mean()

#print(round(average\_purchase\_price,2))

total\_purchase\_value = purchase\_data.groupby('Age\_Summary')['Price'].sum()

#print(total\_purchase\_value)

avg\_total\_purchase\_per\_person = total\_purchase\_value/total\_counts

#print(round(avg\_total\_purchase\_per\_person,2))

purchase\_analysis\_df = pd.DataFrame ({"Purchase Count": purchase\_data\_item,

"Average Purchase Price ": round(average\_purchase\_price,2),

"Total Purchase Value": total\_purchase\_value,

"Avg Total Purchase per Person": round(avg\_total\_purchase\_per\_person,2),

})

purchase\_analysis\_df

top\_spenders= purchase\_data['SN'].value\_counts()

#print(top\_spenders)

average\_purchase\_price\_spenders = purchase\_data.groupby('SN')['Price'].mean()

#print(round(average\_purchase\_price\_spenders,2))

total\_purchase\_value\_spenders = purchase\_data.groupby('SN')['Price'].sum()

#print(round(total\_purchase\_value\_spenders,2))

# Cretaing DataFrame

spender\_analysis\_df = pd.DataFrame ({ "Purchase Count" : top\_spenders,

"Average Purchase Price" : round(average\_purchase\_price\_spenders,2),

"Total Purchase Value" : total\_purchase\_value\_spenders,

})

spender\_analysis\_df.sort\_values('Purchase Count', ascending = False).head()

#Group by Item ID & Item Name than perform calculations

most\_popular\_items= purchase\_data.groupby('Item ID')['Item Name'].value\_counts()

#print(most\_popular\_items)

most\_popular\_items\_price= purchase\_data.groupby(['Item ID','Item Name'])['Price'].mean()

#print(most\_popular\_items\_price)

most\_popular\_items\_price\_total= purchase\_data.groupby(['Item ID','Item Name'])['Price'].sum()

#print(most\_popular\_items\_price\_total)

# Cretaing DataFrame

most\_popular\_items\_df = pd.DataFrame ({ "Purchase Count" : most\_popular\_items,

"Item Price" : most\_popular\_items\_price,

"Total Purchase Value" : most\_popular\_items\_price\_total,

})

most\_popular\_items\_df.sort\_values('Purchase Count', ascending = False).head()

# Above DataFrame in Ascending Order

most\_popular\_items\_df.sort\_values('Purchase Count', ascending = True).head()