

```
In [48]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib as plt
```

```
In [49]: df=pd.read_csv("Original_data_with_more_rows.csv")
```

```
In [50]: df
```

	0	0	female	group B	bachelor's degree	standard	none	72	72	74
	1	1	female	group C	some college	standard	completed	69	90	88
	2	2	female	group B	master's degree	standard	none	90	95	93
	3	3	male	group A	associate's degree	free/reduced	none	47	57	44
	4	4	male	group C	some college	standard	none	76	78	75
	--	--	--	--	--	--	--	--	--	--
30636	995	male	group C	some high school	standard	none	56	47	51	
30637	996	male	group E	associate's degree	free/reduced	none	74	75	72	
30638	997	male	group C	some college	standard	none	36	29	27	
30639	998	male	group A	some high school	free/reduced	completed	43	34	39	
30640	999	female	group D	associate's degree	standard	none	52	68	66	
30641 rows x 11 columns										

30641 rows x 9 columns

```
In [51]: print(df.head())

0      Unnamed: 0  Gender  EthnicGroup  ParentEduc  LunchType  \
0      0      female  group B  bachelor's degree  standard
1      1      female  group C      some college  standard
2      2      female  group B  master's degree  standard
3      3      male   group A  associate's degree  free/reduced
4      4      male   group C      some college  standard

      TestPrep  MathScore  ReadingScore  WritingScore
0      none       72         72         74
1  completed      69         90         88
2      none       90         95         93
3      none       47         57         44
4      none       76         78         75
```

```
In [52]: df.describe()
```

	Unnamed: 0	MathScore	ReadingScore	WritingScore
count	30641.000000	30641.000000	30641.000000	30641.000000
mean	499.556607	66.749355	69.624980	68.468327
std	288.747894	15.206049	14.671572	15.307814
min	0.000000	0.000000	10.000000	5.000000
25%	249.000000	56.000000	60.000000	58.000000
50%	500.000000	67.000000	70.000000	69.000000
75%	750.000000	78.000000	80.000000	79.000000
max	999.000000	100.000000	100.000000	100.000000

```
In [53]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30641 entries, 0 to 30640
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  --
0  Unnamed: 0    30641 non-null    int64
1  Gender        30641 non-null    object
2  EthnicGroup   30641 non-null    object
3  ParentEduc    30641 non-null    object
4  LunchType     30641 non-null    object
5  TestPrep      30641 non-null    object
6  MathScore     30641 non-null    int64
7  ReadingScore  30641 non-null    int64
8  WritingScore  30641 non-null    int64
dtypes: int64(4), object(5)
memory usage: 2.1+ MB
```

```
In [54]: df.isnull().sum()
```

```
Out[54]: Unnamed: 0    0
Gender          0
EthnicGroup     0
ParentEduc      0
LunchType       0
TestPrep        0
MathScore       0
ReadingScore    0
WritingScore    0
dtype: int64
```

DROP UNNAMED COLUMN

```
In [55]: df.columns = df.columns.str.strip()
```

```
In [56]: # Check if 'Unnamed: 0' is in the list of columns
if 'Unnamed: 0' in df.columns:
    df = df.drop('Unnamed: 0', axis=1)
```

```
In [57]: print(df.head())

      Gender  EthnicGroup  ParentEduc  LunchType  TestPrep  MathScore  \
0  female   group B      bachelor's degree  standard    none         72
1  female   group C      some college      standard  completed         69
2  female   group B      master's degree      standard    none         90
3  male     group A  associate's degree  free/reduced  none         47
4  male     group C      some college      standard    none         76

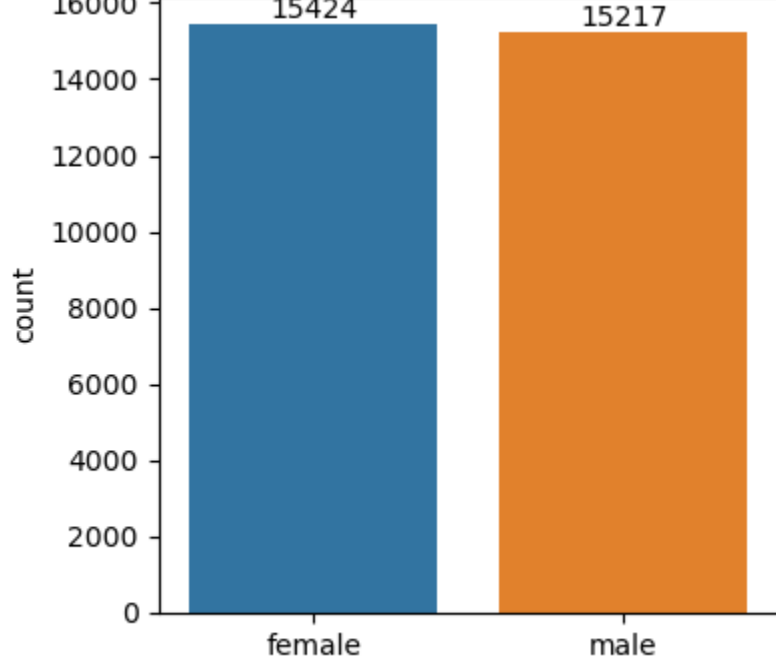
      ReadingScore  WritingScore
0                72              74
1                69              88
2                95              93
3                57              44
4                78              75
```

change weekly study hours column

```
In [58]: print(df.columns)

Index(['Gender', 'EthnicGroup', 'ParentEduc', 'LunchType', 'TestPrep',
      'MathScore', 'ReadingScore', 'WritingScore'],
      dtype='object')
```

```
In [64]: plt.figure(figsize=(4,4))
sns.countplot(data=df,x="Gender")
ax.bar_label(ax.containers[0])
plt.show()
```



```
In [65]: gb = df.groupby("ParentEduc").agg(
    "MathScore": "mean",
    "ReadingScore": "mean",
    "WritingScore": "mean" # Correct the column name here
)
```

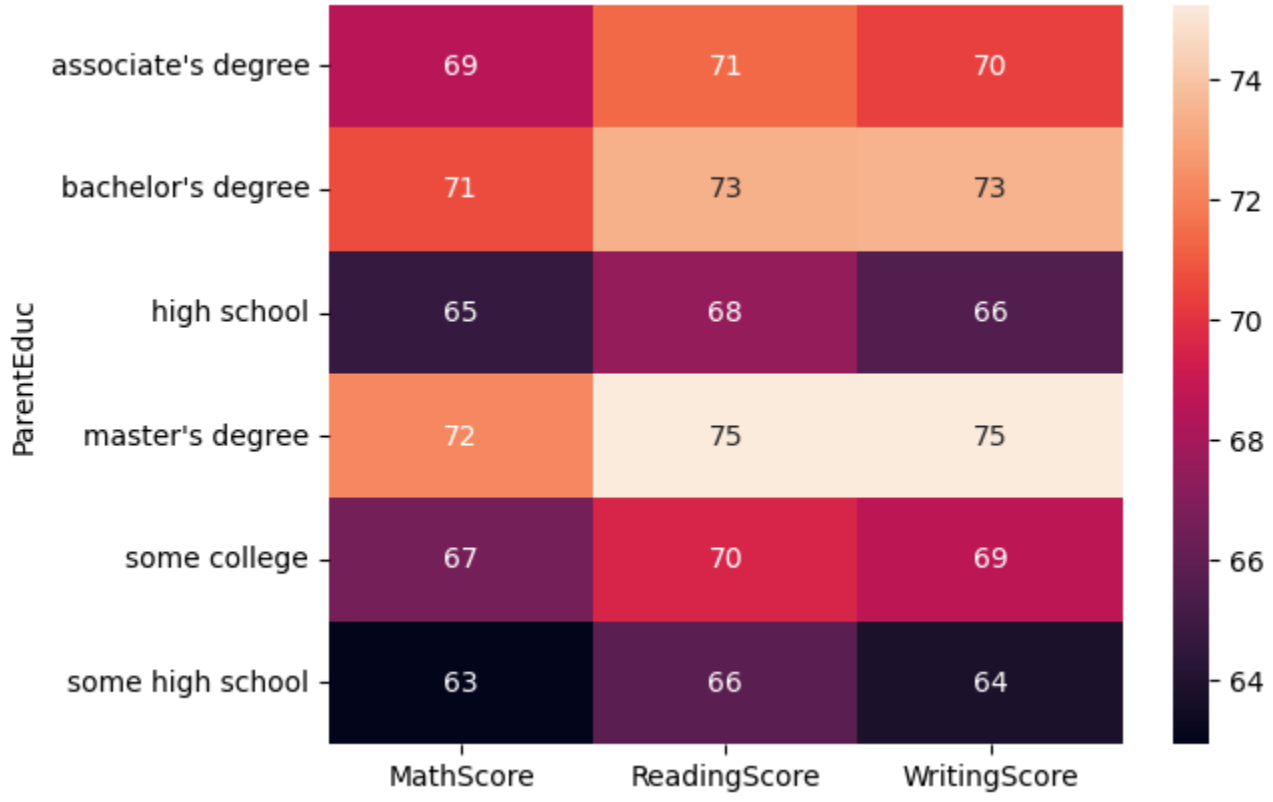
```
print(gb)

      ParentEduc  MathScore  ReadingScore  WritingScore
associate's degree    68.527787      71.368350      70.374619
bachelor's degree     70.635000      73.375556      73.423611
high school           64.670038      61.522467      65.365909
master's degree       72.168137      75.202044      75.246168
some college          66.603434      69.528093      68.653661
some high school      62.323359      65.826816      63.415473
```

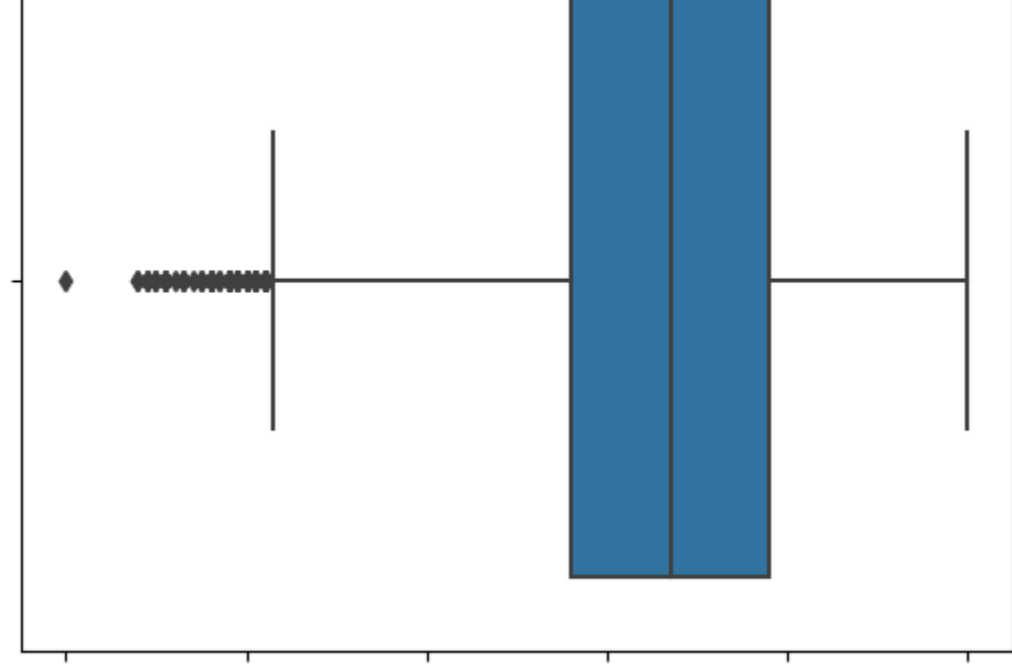
```
In [69]: import matplotlib.pyplot as plt # Import the necessary module

# Your code for creating the heatmap goes here
sns.heatmap(gb,annot = True)
```

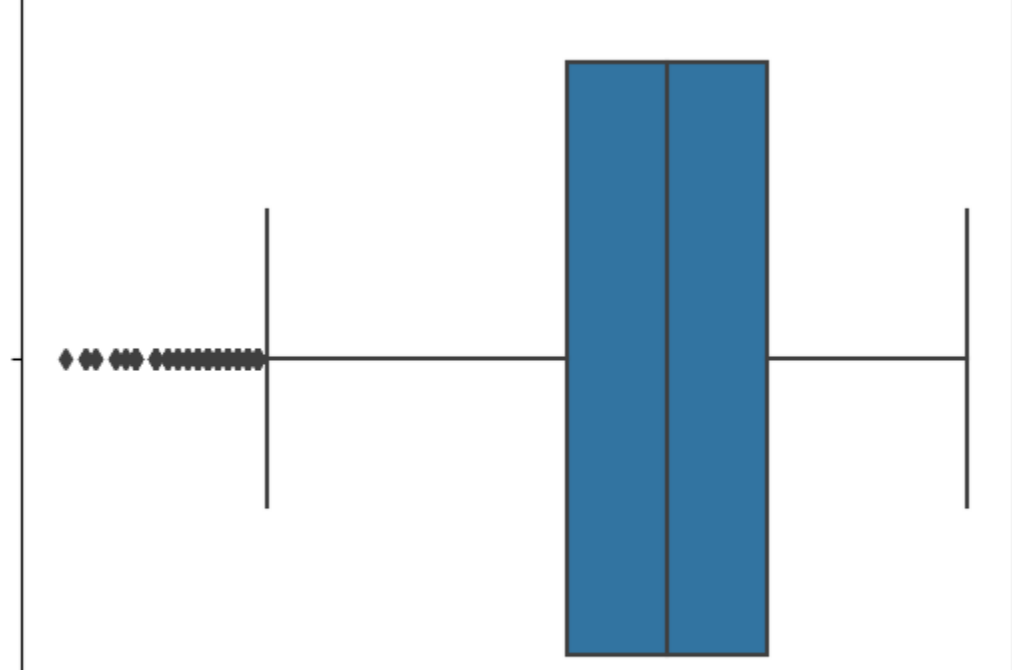
```
plt.show() # Use plt.show() to display the heatmap
```



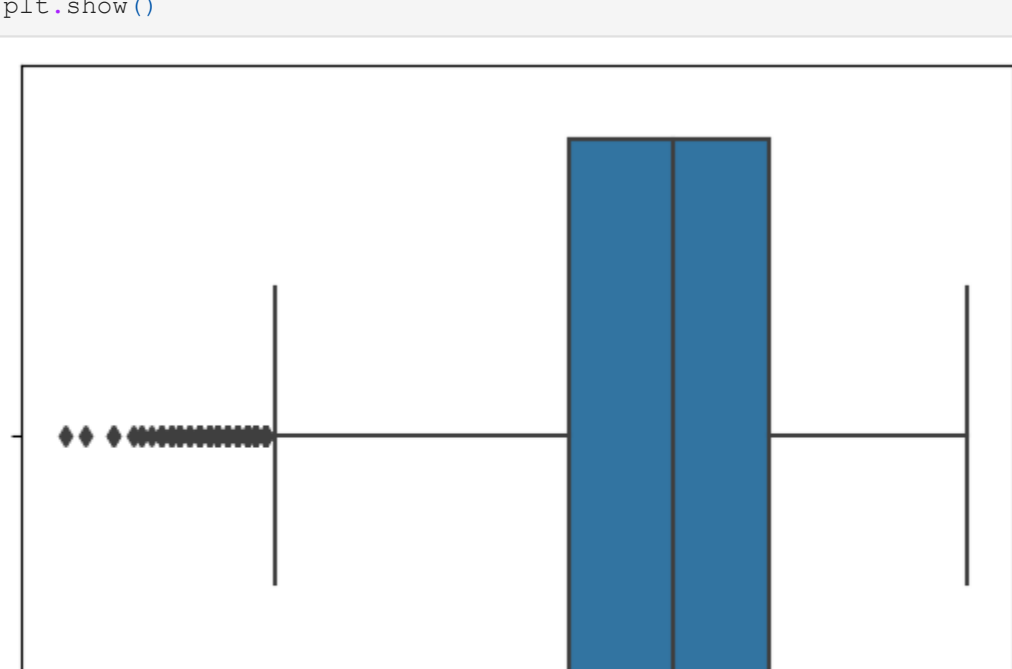
```
In [71]: sns.boxplot(data=df,x="MathScore")
plt.show()
```



```
In [73]: sns.boxplot(data=df,x="ReadingScore")
plt.show()
```



```
In [75]: sns.boxplot(data=df,x="WritingScore")
plt.show()
```



```
In [78]: print(df["EthnicGroup"].unique())

['group B' 'group C' 'group A' 'group D' 'group E']
```

Distribution ethic group

```
In [81]: groupA = df.loc[(df["EthnicGroup"] == "group A").count()]
print(groupA)

Gender      2359
EthnicGroup  2359
ParentEduc   2359
LunchType    2359
TestPrep     2359
MathScore    2359
ReadingScore  2359
WritingScore  2359
dtype: int64
```

```
In [85]: groupA = df.loc[(df["EthnicGroup"] == "group A").count()]
groupB = df.loc[(df["EthnicGroup"] == "group B").count()]
groupC = df.loc[(df["EthnicGroup"] == "group C").count()]
groupD = df.loc[(df["EthnicGroup"] == "group D").count()]
groupE = df.loc[(df["EthnicGroup"] == "group E").count()]

In [91]: mlist=[groupA["EthnicGroup"],groupB["EthnicGroup"],groupC["EthnicGroup"],groupD["EthnicGroup"],groupE["EthnicGroup"]]
plt.pie(mlist)

Cell In[91], line 1
mlist=[groupA["EthnicGroup"],groupB["EthnicGroup"],groupC["EthnicGroup"],groupD["EthnicGroup"],groupE["EthnicGroup"]]
SyntaxError: closing parenthesis ')' does not match opening parenthesis '['
```

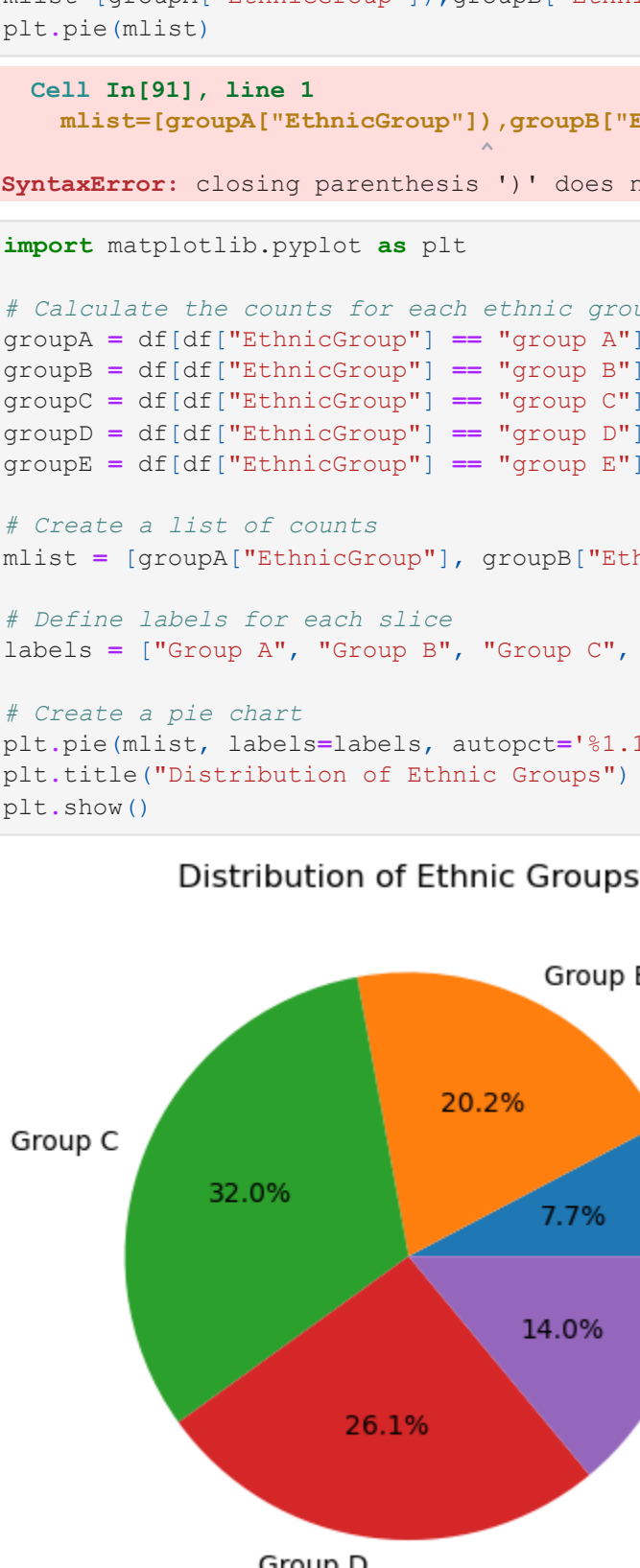
```
In [92]: import matplotlib.pyplot as plt

# Calculate the counts for each ethnic group
groupA = df[df["EthnicGroup"] == "group A"].count()
groupB = df[df["EthnicGroup"] == "group B"].count()
groupC = df[df["EthnicGroup"] == "group C"].count()
groupD = df[df["EthnicGroup"] == "group D"].count()
groupE = df[df["EthnicGroup"] == "group E"].count()

# Create a list of counts
mlist = [groupA["EthnicGroup"], groupB["EthnicGroup"], groupC["EthnicGroup"], groupD["EthnicGroup"], groupE["EthnicGroup"]]

# Define labels for each slice
labels = ["Group A", "Group B", "Group C", "Group D", "Group E"]

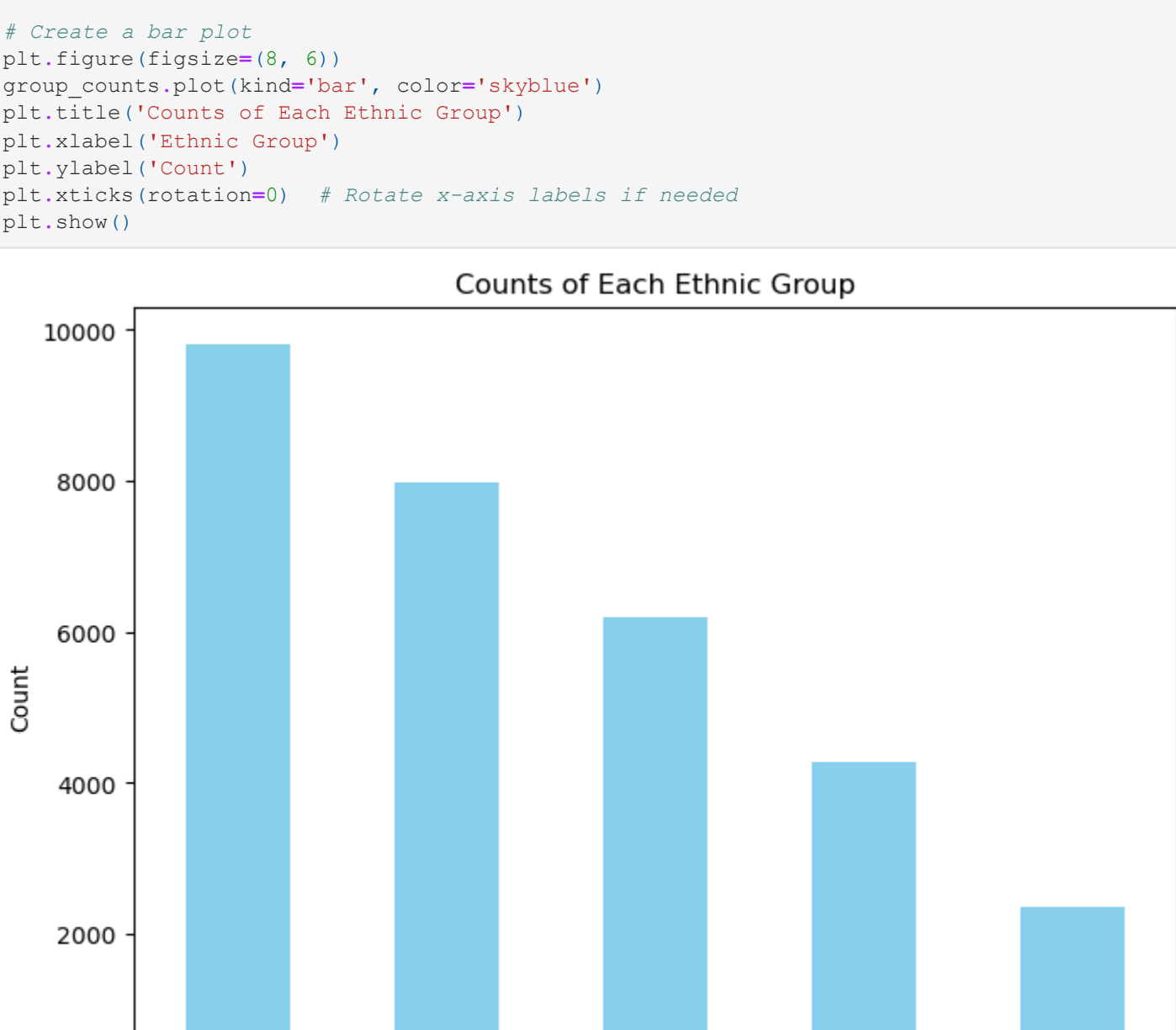
# Create a pie chart
plt.pie(mlist, labels=labels, autopct='%1.1f%%')
plt.title("Distribution of Ethnic Groups")
plt.show()
```



```
In [93]: import matplotlib.pyplot as plt

# Calculate the counts for each ethnic group
group_counts = df["EthnicGroup"].value_counts()

# Create a bar plot
plt.figure(figsize=(8, 6))
group_counts.plot(kind='bar', color='skyblue')
plt.title('Counts of Each Ethnic Group')
plt.xlabel('Ethnic Group')
plt.ylabel('Count')
plt.xticks(rotation=0) # Rotate x-axis labels if needed
plt.show()
```



```
In [94]: import matplotlib.pyplot as plt

# Calculate the counts for each ethnic group
group_counts = df["EthnicGroup"].value_counts()

# Create a bar plot
plt.figure(figsize=(8, 6))
ax = group_counts.plot(kind='bar', color='skyblue')
plt.title('Counts of Each Ethnic Group')
plt.xlabel('Ethnic Group')
plt.ylabel('Count')
plt.xticks(rotation=0) # Rotate x-axis labels if needed

# Add values on top of the bars
for i, v in enumerate(group_counts):
    ax.text(i, v + 10, str(v), color='black', ha='center')

plt.show()
```



```
In [ ] :
```