```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb


from google.colab import files
uploaded = files.upload()
```

```
Choose Files  student_scores.csv
   • student_scores.csv(text/csv) - 2909278 bytes, last modified: 4/14/2023 - 100% done
   Saving student_scores.csv to student_scores.csv
```

```
df = pd.read_csv("student_scores.csv")
```

```
df.columns
```

```
Index(['Gender', 'EthnicGroup', 'ParentEduc', 'LunchType', 'TestPrep',
       'ParentMaritalStatus', 'PracticeSport', 'IsFirstChild', 'NrSiblings',
       'TransportMeans', 'WklyStudyHours', 'MathScore', 'ReadingScore',
       'WritingScore'],
      dtype='object')
```

Data Cleaning

```
df.describe()
```

```
df.isna().sum()
```

```
del[df["Unnamed: 0"]]
```

```
#drop unnamed column
df.shape
```

```
#replacing wrong values in a column to correct values
df["WklyStudyHours"] = df["WklyStudyHours"].str.replace("5-Oct","5-10")
```

Data Analysis

```
we = sb.countplot(x = "Gender", data = df)
sb.set(rc = {"figure.figsize" : (3,3)})
plt.ylabel("No of Students")
plt.xlabel("Gender")
plt.title("Gender wise Distribution")
for bars in we.containers:
  we.bar_label(bars)
```
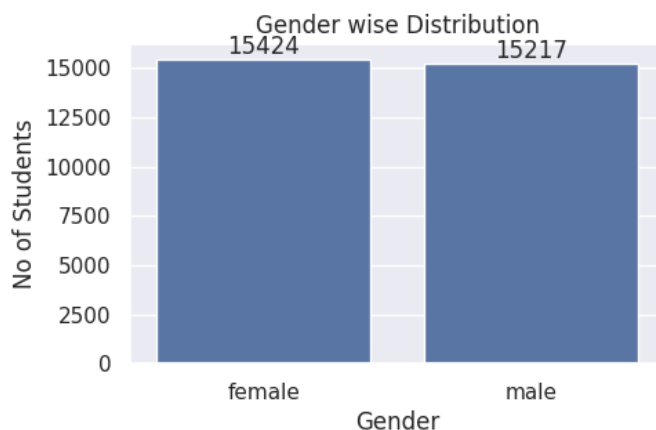


```
# from the above chart , we found out the number of females are slighly more than number of males.
```

```
gp = df.groupby("ParentEduc").agg({"WritingScore":"mean","MathScore":"mean","ReadingScore":"mean"})
sb.heatmap(gp, annot = True)
sb.set(rc = {"figure.figsize": (10,5)})
```
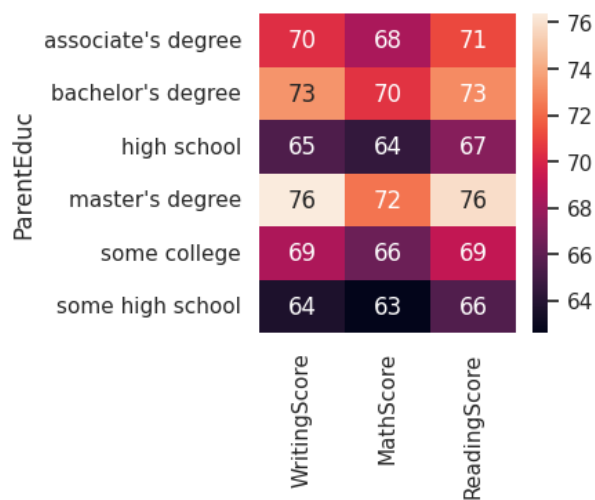
```
# from above heatmap, we concluded that parent's educatiom have good impact on student's score.
# For eg- kids of parents with master's degree have better average score compared to other parents.


gp1 = df.groupby("ParentMaritalStatus").agg({"MathScore":"mean","ReadingScore":"mean","WritingScore":"mean"})
sb.heatmap(gp1, annot = True)
```
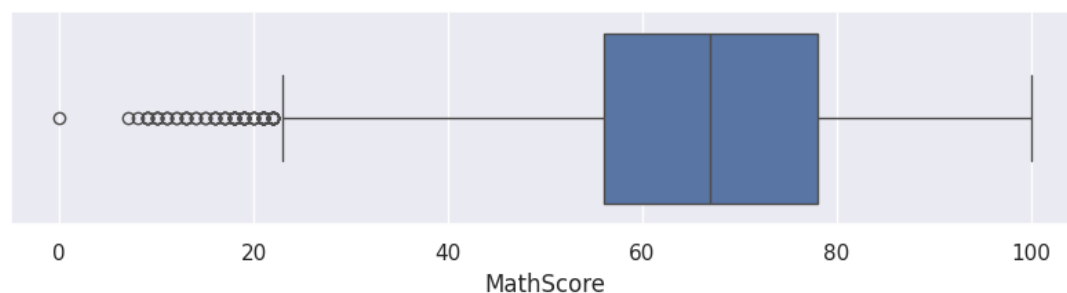
```
<Axes: ylabel='ParentMaritalStatus'>
```
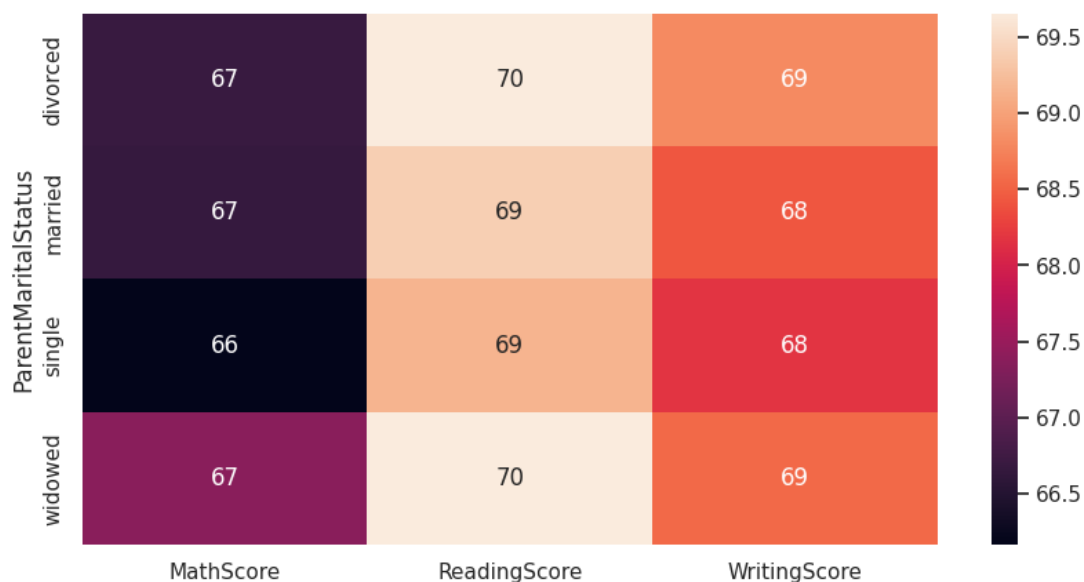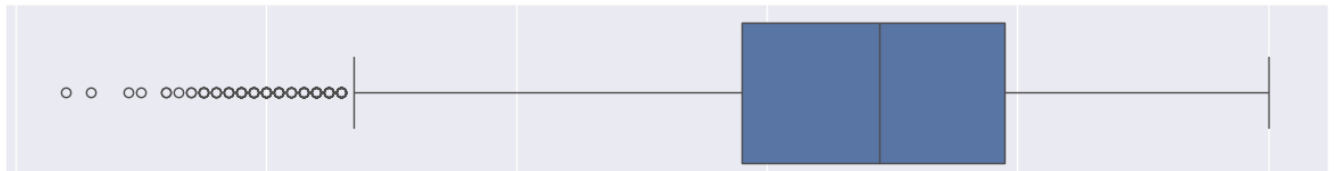


```
#from this above heatmap we concluded that the parent marital status do not have any significant impact or negligible impact on the
```
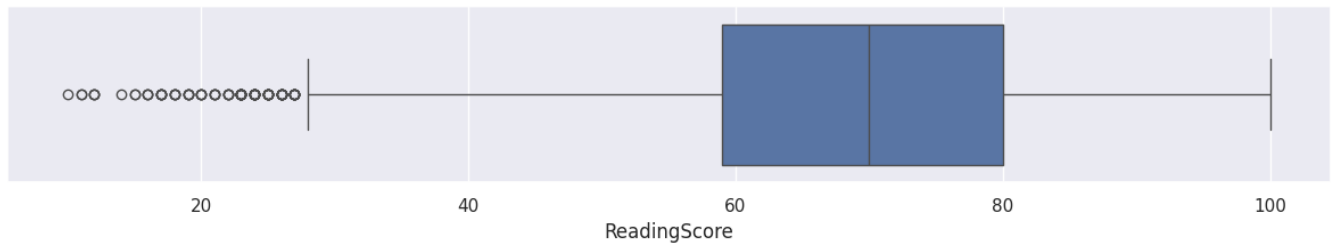
```
sb.boxplot(x = "MathScore", data = df)
sb.set(rc = {"figure.figsize" : (15,2)})
```



```
sb.boxplot(x = "WritingScore", data = df)
sb.set(rc = {"figure.figsize" : (15,2)})
```
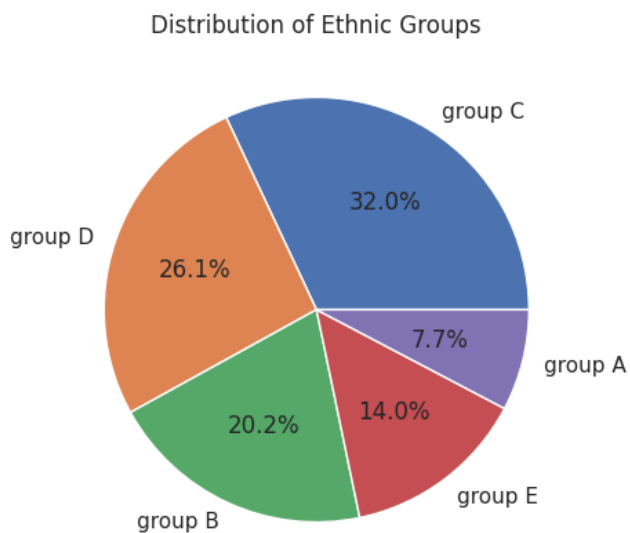
```
sb.boxplot(x = "ReadingScore", data = df)
sb.set(rc = {"figure.figsize" : (15,2)})
```



```
#from the above boxplots of all three subjects we can see that the minimum range of Math is much at lower side compared to other tw
```

```
dt = df["EthnicGroup"].value_counts()
```

```
plt.pie(dt,labels = dt.index,autopct='%1.1f%%')
plt.title("Distribution of Ethnic Groups")
sb.set(rc = {"figure.figsize": (20,5)})
```


Distribution of Ethnic Groups

```
# from the above pie chart we conclude that group c contributes the highest number of students followed by group d, group b, group
```