

```
In [1]: greetings = "Assalam-o-Alaikum!"
print(greetings)
```

Assalam-o-Alaikum!

## Import Libraries

```
In [22]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
```

## Import Dataset

```
In [5]: df = pd.read_csv("StudentsPerformance.csv")
df
```

```
Out[5]:
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
...	...	...	...	...	...	...	...	...
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

1000 rows × 8 columns

```
In [6]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column                                Non-Null Count  Dtype  
---  -
 0   gender                                1000 non-null   object  
 1   race/ethnicity                        1000 non-null   object  
 2   parental level of education           1000 non-null   object  
 3   lunch                                 1000 non-null   object  
 4   test preparation course               1000 non-null   object  
 5   math score                           1000 non-null   int64   
 6   reading score                         1000 non-null   int64   
 7   writing score                         1000 non-null   int64   
dtypes: int64(3), object(5)
memory usage: 62.6+ KB
```

```
In [7]: df.describe()
```

```
Out[7]:
```

	math score	reading score	writing score
count	1000.00000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000
std	15.16308	14.600192	15.195657
min	0.00000	17.000000	10.000000
25%	57.00000	59.000000	57.750000
50%	66.00000	70.000000	69.000000
75%	77.00000	79.000000	79.000000
max	100.00000	100.000000	100.000000

## Analysis Questions:

1. What is the average score in math for all students?

```
In [10]: avg maths score = df["math score"].agg("mean").round(2)
```

```
avg_maths_score = df["math score"].agg("mean").round(2)
print(" Average Maths Score for all Students = " + str(avg_maths_score))
```

Average Maths Score for all Students = 66.09

## 2. Is there any relationship between parental level of education and student scores in reading?

```
In [18]: relation = df.groupby("parental level of education")["reading score"].agg("mean").to_frame().round(2).reset_index()
relation.columns = ["Levels Of Education", "Average Score"]
relation
```

```
Out[18]:
```

	Levels Of Education	Average Score
0	associate's degree	70.93
1	bachelor's degree	73.00
2	high school	64.70
3	master's degree	75.37
4	some college	69.46
5	some high school	66.94

```
In [28]: groups = df.groupby('parental level of education')['reading score'].apply(list)
anova_result = stats.f_oneway(*groups)
print("ANOVA p-value:", anova_result.pvalue.round(5))
```

ANOVA p-value: 0.0

```
In [30]: if anova_result.pvalue < 0.05:
        print("Yes, there is a relationship between parental level of education and student scores in reading ")
    else:
        print("No, there is no relationship between parental level of education and student scores in reading")
```

Yes, there is a relationship between parental level of education and student scores in reading

```
In [19]: df.head(5)
```

```
Out[19]:
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

## 3. Which group (A, B, C, D, E) has the highest average score in writing?

```
In [47]: highest_score = df.groupby("race/ethnicity")["writing score"].agg("mean").round(2).to_frame().reset_index().sort_index()
highest_score.columns = ["Groups", "Average Score"]
highest_score = highest_score.iloc[0]
print("'" + highest_score["Groups"] + "'", "has the highest average score in writing = " + str(highest_score["Average Score"]))
```

'Group E' has the highest average score in writing = 71.41

## 4. How many students have completed the test preparation course?

```
In [54]: prep_course = df[df["test preparation course"] == "completed"]
prep_course = prep_course["test preparation course"].count()
print("The Total Students Who have Completed their test preparation course = " + str(prep_course))
```

The Total Students Who have Completed their test preparation course = 358