

1. Introduction

This is an educational data set which is collected from Kaggle site, which has been designed for the Knowing and visualizing the student's performance.

Student Academic Performance system is used to check the performance of the student through their classes and semesters.

In this system we can analyze the student performance and accuracy of student's answering in the class and their activities.

In much country like Kuwait, Jordan, Venezuela, Iran, Saudi Arabia and USA and they also used the Student Academic Performance system for the predicting the marks and their performance based on their class and semester.

With this kind of system we can also maintain the other topics like student's grade, parent's satisfaction effects on their performance, also maintain the marks etc.

2. Problem statement

The goal of this project was to examine a number of ML algorithms that were capable of adjusting to categorical data and attempt to predict student performance.

Some parts about our problem that make it unique are:

There are 3 classes and most of our data is categorical data and not purely quantitative.

By studying this dataset we can answer these questions like..

1. Can parents make effect on student performance?
2. How many students are active in the classes?
3. Can we compare the semester for the knowing which semester has better student and performance?
4. Does any student view any announcements or not?

And we can get many answers by study this dataset..,

Our goal with this was to perform some initial data visualization and to determine which classifier handles this data the best.

3.Plagiarism report

4. Understand data

4.1 Dataset Source:

Dataset URL: www.kaggle.com

Kaggle is the place to do data science projects. It contains information About student academic Performance date till Dec 2016.

4.2 Columns explanation:

- The number of rows and columns in the dataset :

```
>>> print(df.shape)
(480, 19)
```

- The First six values in the dataset

```
==== RESTART: S:\Marwadi college\sem-4\Project\claaa.py =====
Student_ID gender Nationality ... StudentAbsenceDays Student Marks Class
0 STDN00001      M      KW ...           Under-7           77      M
1 STDN00002      M      KW ...           Under-7           68      M
2 STDN00003      M      KW ...           Above-7           42      L
3 STDN00004      M      KW ...           Above-7           49      L
4 STDN00005      M      KW ...           Above-7           72      M

[5 rows x 19 columns]
>>> |
```

- The names of variables :

```

NAME2202: Name      IS NOT DEFINED
>>> print(df.columns)
Index(['Student_ID', 'gender', 'NationalITy', 'PlaceofBirth', 'StageID',
      'GradeID', 'SectionID', 'Topic', 'Semester', 'Relation', 'raisedhands',
      'VisITedResources', 'AnnouncementsView', 'Discussion',
      'ParentAnsweringSurvey', 'ParentschoolSatisfaction',
      'StudentAbsenceDays', 'Student Marks', 'Class'],
      dtype='object')
>>> |

```

- Columns

1 Gender :

student's gender (nominal: 'Male' or 'Female')

2 Nationality

student's nationality (nominal: 'Kuwait', 'Lebanon', 'Egypt', 'SaudiArabia', 'USA', 'Jordan', 'Venezuela', 'Iran', 'Tunis', 'Morocco', 'Syria', 'Palestine', 'Iraq', 'Lybia')

3 Place of birth:

student's Place of birth (nominal: 'Kuwait', 'Lebanon', 'Egypt', 'SaudiArabia', 'USA', 'Jordan', 'Venezuela', 'Iran', 'Tunis', 'Morocco', 'Syria', 'Palestine', 'Iraq', 'Lybia')

4 Educational Stages:

educational level student belongs (nominal: 'lowerlevel', 'MiddleSchool', 'HighSchool')

5 Grade Levels:

grade student belongs (nominal: 'G-01', 'G-02', 'G-03', 'G-04', 'G-05', 'G-06', 'G-07', 'G-08', 'G-09', 'G-10', 'G-11', 'G-12')

6 Section ID

classroom student belongs (nominal: 'A', 'B', 'C')

7 Topic:

course topic (nominal: 'English', 'Spanish', 'French', 'Arabic', 'IT', 'Math', 'Chemistry', 'Biology', 'Science', 'History', 'Quran', 'Geology')

8 Semester:

school year semester (nominal: 'First', 'Second')

9 Parent responsible for student (nominal:'mom','father')

10 Raised hand:

how many times the student raises his/her hand on classroom (numeric:0-100)

11- Visited resources:

how many times the student visits a course content(numeric:0-100)

12 Viewing announcements:

how many times the student checks the new announcements(numeric:0-100)

13 Discussion groups:

how many times the student participate on discussion groups (numeric:0-100)

14 Parent Answering Survey:

parent answered the surveys which are provided from school or not (nominal:'Yes','No')

15 Parent School Satisfaction:

the Degree of parent satisfaction from school(nominal:'Yes','No')

16 Student Absence Days:

the number of absence days for each student (nominal: above-7, under-7)

4.3 Basic questions regarding dataset analysis:

1. How did I obtain the data set?

: Kaggle is the place to do data science projects. It contains information About student academic Performance date till Dec 2016.

2. How was the data originally collected?

: The data is originally collected from Kaggle website. So that we can say that our data is original and valid.

3. How many rows and how many columns are in the data set?

□ : There are 480 rows and 19 columns in the our data set.

4. What are the columns in the data set?

```
NAMEError: name 'data' is not defined
>>> print(df.columns)
Index(['Student_ID', 'gender', 'NationalITy', 'PlaceofBirth', 'StageID',
      'GradeID', 'SectionID', 'Topic', 'Semester', 'Relation', 'raisedhands',
      'VisITedResources', 'AnnouncementsView', 'Discussion',
      'ParentAnsweringSurvey', 'ParentschoolSatisfaction',
      'StudentAbsenceDays', 'Student Marks', 'Class'],
      dtype='object')
>>> |
```

4.4 Data Preprocessing:

Data cleansing or data cleaning is the process of identifying and removing (or correcting) inaccurate records from a dataset, table, or database and refers to recognising unfinished, unreliable, inaccurate or non-relevant parts of the data and then restoring, remodelling, or removing the dirty or crude data. It may be performed as batch processing through scripting or interactively with data wrangling tools.

Benefits of Data Cleaning

Here are several key benefits that come out of the data cleaning process:

1. It removes major errors and inconsistencies that are inevitable when multiple sources of data are getting pulled into one dataset.
2. Using tools to cleanup data will make everyone more efficient since they'll be able to quickly get what they need from the data.
3. Fewer errors means happier customers and fewer frustrated employees.
4. The ability to map the different functions and what your data is intended to do and where it is coming from your data.

We have picked up our data set from the well-known portal [kaggle.com](https://www.kaggle.com).

Some of the columns which are available in our dataset are SectionID, Topic, semester etc .which having blank values .so we have filled values at the place of that blank value.

Find the null value in over dataset and fill the data.

```
import pandas as pd

r=pd.read_csv("S:\Marwadi college\sem-4\Project\s1.csv")

ans=r['SectionID'].describe()

r["SectionID"].fillna(method='ffill',inplace=True)

ans=r['Topic'].describe()

r["Topic"].fillna(ans[2],inplace=True)

ans=r['Semester'].describe()

r["Semester"].fillna(ans[2],inplace=True)

ans=r['raisedhands'].describe()

r["raisedhands"].fillna(int(ans[1]),inplace=True)

ans=r['VisITedResources'].describe()

r["VisITedResources"].fillna(int(ans[1]),inplace=True)

ans=r['AnnouncementsView'].describe()

r["AnnouncementsView"].fillna(int(ans[1]),inplace=True)

ans=r['Discussion'].describe()

r["Discussion"].fillna(int(ans[1]),inplace=True)

r.to_csv("data.csv", sep=',', index=False)
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Student_I	gender	NationalIT	PlaceofBir	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhan	VisiTedRe	Announce	Discussion	ParentAns	Parentsch	StudentAI	Student N	Class
2	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	15	16	2		Yes	Good	Under-7	77	M
3	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	20	20	3		Yes	Good	Under-7	68	M
4	STDN0000	M	KW	KuwaIT	lowerleve	G-04			F	Father	10	7	0		No	Bad	Above-7	42	L
5	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT		Father			5		No	Bad	Above-7	49	L
6	STDN0000	M	KW	KuwaIT	lowerleve	G-04		IT	F	Father	40	50	12		No	Bad	Above-7	72	M
7	STDN0000	F	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	42	30	13		Yes	Bad	Above-7	69	M
8	STDN0000	M	KW	KuwaIT	MiddleSci	G-07	A		F	Father			0		No	Bad	Above-7	59	L
9	STDN0000	M	KW	KuwaIT	MiddleSci	G-07	A	Math	F	Father	50	10	15		Yes	Good	Under-7	65	M
10	STDN0000	F	KW	KuwaIT	MiddleSci	G-07		Math	F	Father	12	21			Yes	Good	Under-7	64	M
11	STDN0001	F	KW	KuwaIT	MiddleSci	G-07	B		F	Father	70	80	25		Yes	Good	Under-7	69	M
12	STDN0001	M	KW	KuwaIT	MiddleSci	G-07	A	Math	F	Father			30		Yes	Good	Under-7	83	H
13	STDN0001	M	KW	KuwaIT	MiddleSci	G-07	B	Math	F	Father	19	6			Yes	Good	Under-7	76	M
14	STDN0001	M	KW	KuwaIT	lowerleve	G-04		IT	F	Father	5	1	0		No	Bad	Above-7	24	L
15	STDN0001	M	lebanon	lebanon	MiddleSci	G-08	A	Math	F	Father	20	14	12		No	Bad	Above-7	24	L
16	STDN0001	F	KW	KuwaIT	MiddleSci	G-08	A		F	Father		70	44	60	No	Bad	Above-7	93	H
17	STDN0001	F	KW	KuwaIT	MiddleSci	G-06	A	IT		Father	30		22	66	Yes	Good	Under-7	68	M
18	STDN0001	M	KW	KuwaIT				IT	F	Father				80	No	Bad	Above-7	71	M
19	STDN0001	M	KW	KuwaIT	MiddleSci	G-07	A	Math	F	Father	55		35		No	Bad	Above-7	69	M
20	STDN0001	F	KW	KuwaIT	MiddleSci	G-07		IT	F	Mum	69	15	36	96	Yes	Good	Under-7	72	M
21	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	B	IT	F	Mum		50	40		Yes	Good	Under-7	80	H
22	STDN0002	F	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Mum	60	60	33	90	No	Bad	Above-7	75	M
23	STDN0002	F	KW	KuwaIT	MiddleSci	G-07	B		F	Father	10	12	4		No	Bad	Under-7	79	M
24	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	15		2	90	No	Bad	Under-7	63	M
25	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	2	0	2	50	No	Bad	Above-7	43	L
26	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	B	IT		Father	0	2	3	70	Yes	Good	Above-7	23	L
27	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	8	7	30	40	Yes	Good	Above-7	36	L

(Before Data Cleaning)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Student_I	gender	NationalIT	PlaceofBir	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhan	VisiTedRe	Announce	Discussion	ParentAns	Parentsch	StudentAI	Student N	Class
2	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	15	16	2	43	Yes	Good	Under-7	77	M
3	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	20	20	3	43	Yes	Good	Under-7	68	M
4	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	10	7	0	43	No	Bad	Above-7	42	L
5	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	46	55	5	43	No	Bad	Above-7	49	L
6	STDN0000	M	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	40	50	12	43	No	Bad	Above-7	72	M
7	STDN0000	F	KW	KuwaIT	lowerleve	G-04	A	IT	F	Father	42	30	13	43	Yes	Bad	Above-7	69	M
8	STDN0000	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	46	55	0	43	No	Bad	Above-7	59	L
9	STDN0000	M	KW	KuwaIT	MiddleSci	G-07	A	Math	F	Father	50	10	15	43	Yes	Good	Under-7	65	M
10	STDN0000	F	KW	KuwaIT	MiddleSci	G-07	A	Math	F	Father	12	21	38	43	Yes	Good	Under-7	64	M
11	STDN0001	F	KW	KuwaIT	MiddleSci	G-07	B	IT	F	Father	70	80	25	43	Yes	Good	Under-7	69	M
12	STDN0001	M	KW	KuwaIT	MiddleSci	G-07	A	Math	F	Father	46	55	30	43	Yes	Good	Under-7	83	H
13	STDN0001	M	KW	KuwaIT	MiddleSci	G-07	B	Math	F	Father	19	6	38	43	Yes	Good	Under-7	76	M
14	STDN0001	M	KW	KuwaIT	lowerleve	G-04	B	IT	F	Father	5	1	0	43	No	Bad	Above-7	24	L
15	STDN0001	M	lebanon	lebanon	MiddleSci	G-08	A	Math	F	Father	20	14	12	43	No	Bad	Above-7	24	L
16	STDN0001	F	KW	KuwaIT	MiddleSci	G-08	A	IT	F	Father	46	70	44	60	No	Bad	Above-7	93	H
17	STDN0001	F	KW	KuwaIT	MiddleSci	G-06	A	IT		Father	30	55	22	66	Yes	Good	Under-7	68	M
18	STDN0001	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	46	55	38	80	No	Bad	Above-7	71	M
19	STDN0001	M	KW	KuwaIT	MiddleSci	G-07	A	Math	F	Father	55	55	35	43	No	Bad	Above-7	69	M
20	STDN0001	F	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Mum	69	15	36	96	Yes	Good	Under-7	72	M
21	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	B	IT	F	Mum	46	50	40	43	Yes	Good	Under-7	80	H
22	STDN0002	F	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Mum	60	60	33	90	No	Bad	Above-7	75	M
23	STDN0002	F	KW	KuwaIT	MiddleSci	G-07	B	IT	F	Father	10	12	4	43	No	Bad	Under-7	79	M
24	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	15	55	2	90	No	Bad	Under-7	63	M
25	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	2	0	2	50	No	Bad	Above-7	43	L
26	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	B	IT	F	Father	0	2	3	70	Yes	Good	Above-7	23	L
27	STDN0002	M	KW	KuwaIT	MiddleSci	G-07	A	IT	F	Father	8	7	30	40	Yes	Good	Above-7	36	L

(After Data Cleaning)

5.Methodology:

Python:-

- Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable.
- Python's simple, easy to learn syntax emphasizes readability reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.
- Python was developed by Guido van Rossum in the late 1980s.
- Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, Smalltalk, and Unix shell and other scripting languages.
- Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).
- Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.
- Python supports both the standard library and the interpreter are available free of charge, in both binary and source form.
- It provides Optical facilities for data analysis and displays either directly at the computer or printing at the papers.

Python of Features

- Easy-to-learn - Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- Easy-to-read - Python code is more clearly defined and visible to the eyes.
- Easy-to-maintain - Python's source code is fairly easy-to-maintain.
- A broad standard library - Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- Interactive Mode - Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- Portable - Python can run on a wide variety of hardware platforms and has the same interface on all platforms.

- Extendable - You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- Databases - Python provides interfaces to all major commercial databases.
- GUI Programming - Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- Scalable - Python provides a better structure and support for large programs than shell scripting.

5.2 Library:-

1. pandas:-

- Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.
- Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

2. pandasql:-

- pandasql allows you to query pandas DataFrames using SQL syntax. It works similarly to sqldf in R. pandasql seeks to provide a more familiar way of manipulating and cleaning data for people new to Python or pandas.
- pandasql uses the pandas.io.sql module to transfer data between DataFrames and SQLite databases. Operations are performed in SQL, the results returned, and the database is then torn down.

3. matplotlib.pyplot:-

- matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure.
- Matplotlib function provides many types of graphs like line plot, histograms, bar charts, pie charts, tables ,scatter plots etc.

4. numpy:-

- numpy is the most basic yet a powerful package for scientific computing and data manipulation in python.
- It is an open source library available in python. It helps as to do the mathematical and scientific operation and used extensively in data science.

5.re:

- A regular expression is a special sequence of characters that helps you match or find other strings or sets of strings, using a specialized syntax held in a pattern.
- The Python module re provides full support for Perl-like regular expressions in Python. The re module raises the exception re.error if an error occurs while compiling or using a regular expression.
- re module use many type of function like match(),search(),sub() etc

6.sklearn

- Simple and efficient tools for data mining and data analysis.Accessible to everybody, and reusable in various contexts.Built on NumPy, SciPy, and matplotlib.Open source, commercially usable - BSD license

5.3 Alogorithm

Linear Regression

- In statistics, Regression analysis is one of the most important fields in statistics as well as machine learning and the Regression analysis searches for the relationships among dependent variable and one or more independent variables.
- For example, we can observe several employees of any company and we will try to understand how their salaries depend on the features, such as experience as well as level of education, role, and so on. Regression analysis is contains various methods and Linear regression is one of them.
- With the use of Regression we can find how several variables are related as well as also used for forecast a response.
- Linear regression fits straight line to some data in order to capture the linear relationship between the dependent variable and one or more independent variables.

5. Data visualization

6.1 which type of topic are maximum percentage

```
import pandas as pd

import matplotlib.pyplot as mp

r=pd.read_csv("S:\Marwadi college\sem-4\Project\s1.csv")

j=r["Topic"].value_counts()

l=dict(j)

s=""

x=l.keys()

y=l.values()

print(x,y)

mp.pie(y,labels=x,autopct="% 1.2f%% ")

mp.show()
```

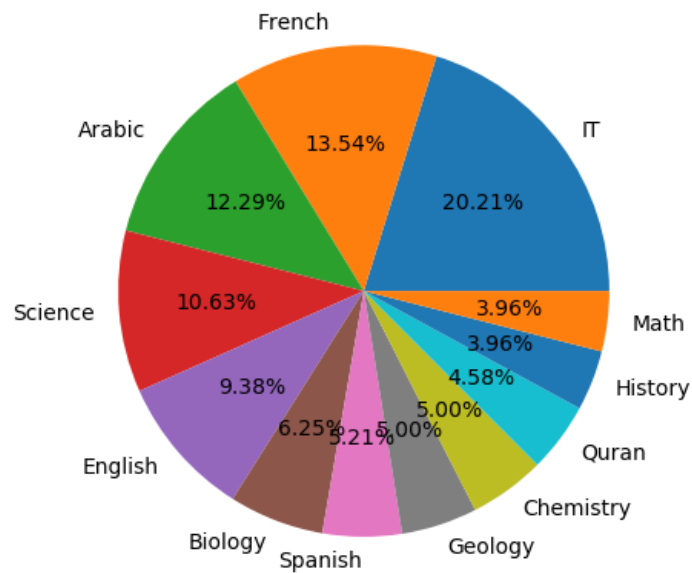


Fig:6.1

Output:

```
dict_keys(['IT', 'French', 'Arabic', 'Science', 'English', 'Biology', 'Spanish', 'Chemistry', 'Geology', 'Quran', 'Math', 'History']) dict_values([97, 65, 59, 51, 45, 30, 25, 24, 24, 22, 19, 19])
```

Conclusion: Above pie chart represent the percentage

6.2 How many student who raised hands in topic?

```
import pandas as pd
import re
import matplotlib.pyplot as plt
r=pd.read_csv("S:\Marwadi college\sem-4\Project\s1.csv")
j=r["Topic"].value_counts()
d1=dict(j)
plt.xlabel("Topic")
plt.ylabel("raisedhands")
plt.title("No of passholders per Trip-route category")
plt.bar(d1.keys(),d1.values(),width=0.3,color="purple")
plt.show()
```

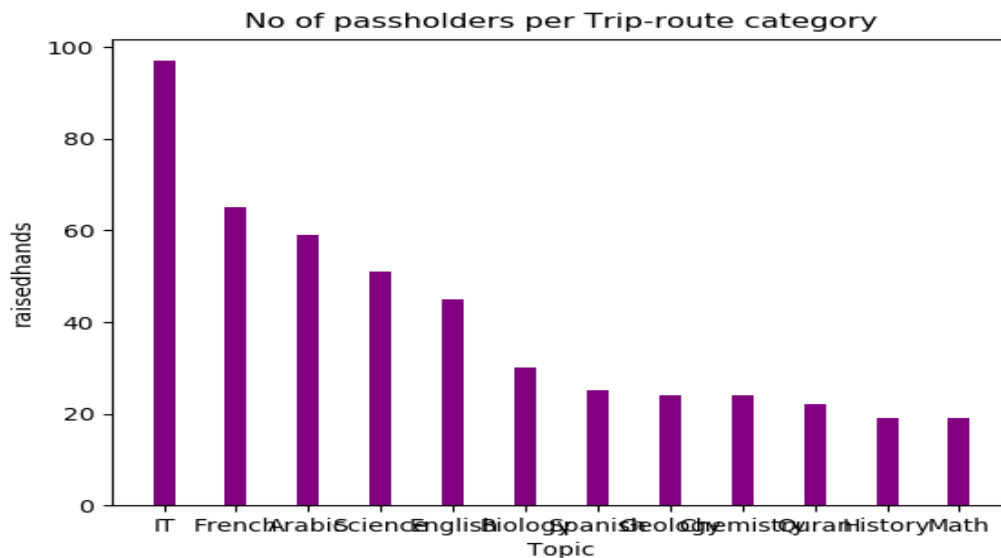


Fig:6.2

Conclusion: Above bar graph represent how many student who has raised hands in the topic.

6.3 How many student who raised hands based on nationality?

```
import pandas as pd
import re
import matplotlib.pyplot as plt
r=pd.read_csv("S:\Marwadi college\sem-4\Project\s1.csv")
j=r["NationalITy"].value_counts()
d1=dict(j)
plt.xlabel("NationalITy")
plt.ylabel("raisedhands")
plt.title("nationality wise raise hand")
plt.bar(d1.keys(),d1.values(),width=0.3,color="purple")
plt.show()
```

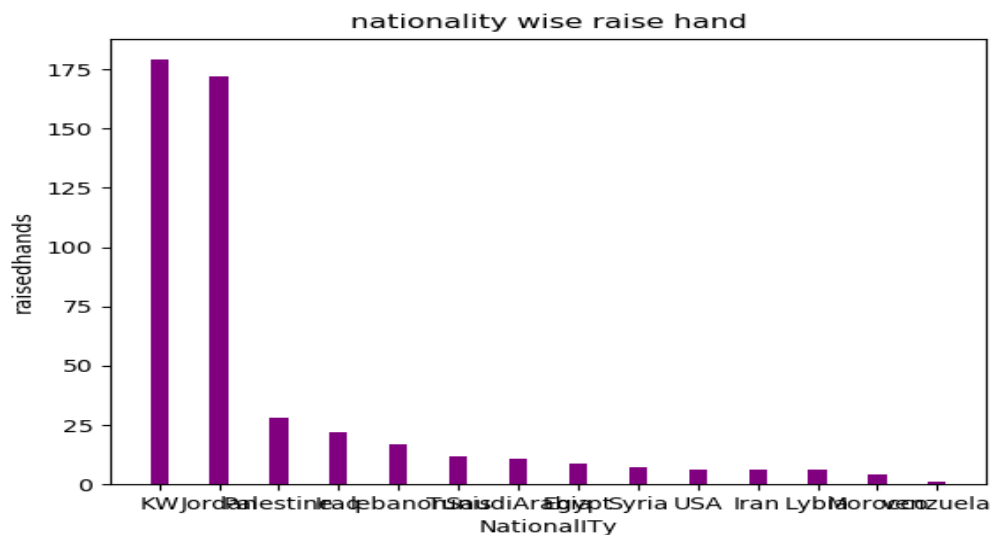


Fig:6.3

Conclusion: Above bar graph represent the nationality wise student raised hands.

6.4 How the ration of student semester wise?

```
import pandas as pd
import matplotlib.pyplot as mp
r=pd.read_csv("S:\Marwadi college\sem-4\Project\s1.csv")
j=r["Semester"].value_counts()
l=dict(j)
s=""
x=l.keys()
y=l.values()
print(x,y)
mp.pie(y,labels=x,autopct="%1.2f%%")
mp.show()
```

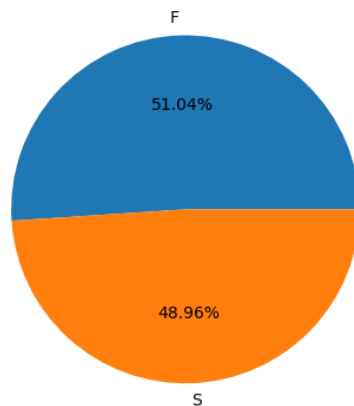


Fig:6.4

Output:

```
dict_keys(['F', 'S']) dict_values([245, 235])
```

Conclusion: above pie chart show the ratio of the semester

6.5 how many types of gender are available in the dataset?

```
import pandas as pd
from collections import Counter
r=pd.read_csv("S:\Marwadi college\sem-4\Project\s1.csv")
j=r["gender"].value_counts()
l=dict(j)
s=""
for i in l.keys():
s=s+i+" "
print("There are ",len(l)," Types of gender available in dataset::",s)
```

output:

There are 2 Types of gender available in dataset:: M F

Conclusion: Above part represent the gender available in dataset.

6.6 how many types of Topic are available in the dataset?

```
import pandas as pd
from collections import Counter
r=pd.read_csv("S:\Marwadi college\sem-4\Project\s1.csv")
j=r["Topic"].value_counts()
l=dict(j)
s=""
for i in l.keys():
s=s+i+" "
print("There are ",len(l)," Types of Topic available in dataset::",s)
```

output:

There are 12 Types of Topic available in dataset:: IT French Arabic Science English
Biology Spanish Geology Chemistry Quran Math History

Conclusion: Above part represent the topics which are available in dataset.

6.7 Student's Nationality and which subject they select

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as mp
r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
mp.figure(figsize=(10,4))
sns.countplot(x='Topic',data=r,hue='NationalITy')
mp.show()
```

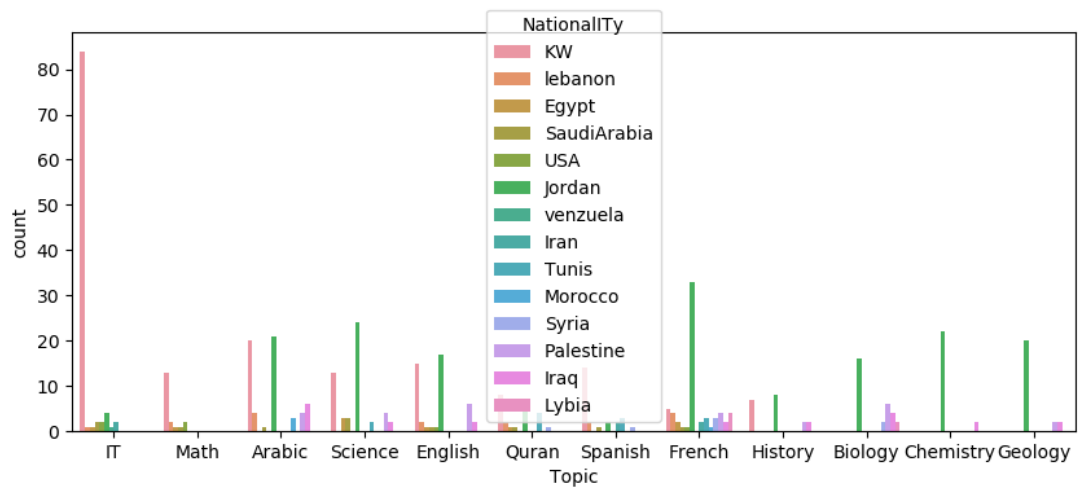


Fig:6.7

Conclusion: Above count plot represent nationality wise student select the topic (subject).

6.8 How many students in each class

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp
r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
sns.countplot(x='Class',data=r)
mp.show()
```

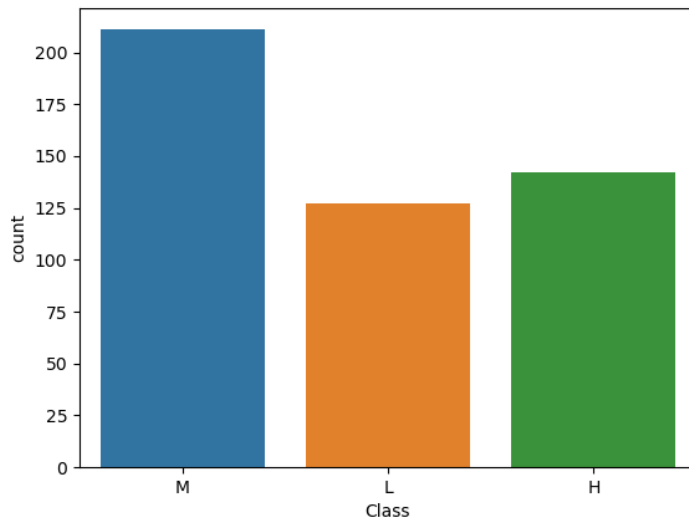


Fig:6.8

Conclusion: Above count plot represent class wise how many student.

6.9 Parents School Satisfaction and how it impact class performance?

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp

r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')

sns.countplot(x='ParentschoolSatisfaction',data=r,hue='Class')

mp.show()
```

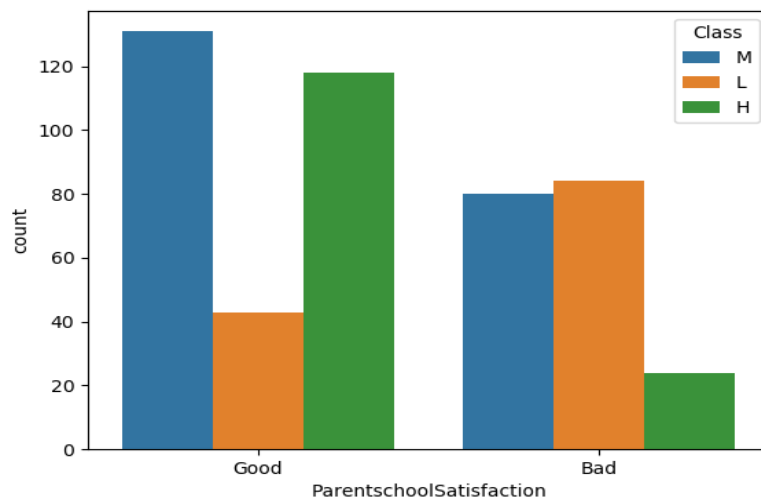


Fig:6.9

Conclusion: Above graph represent the parents school satisfaction which impact on performance.

6.10 Student who raised hands in comparison to their class of performance

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp
r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
raised_hands=sns.boxplot(x='Class',y='raisedhands',data=r)
mp.show()
```

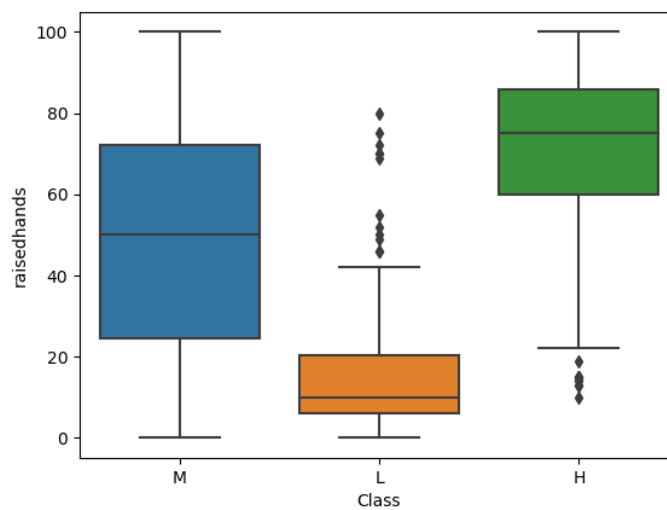


Fig:6.10

Conclusion: Above box plot represent the student who raised hands in class performace.

6.11 Student who took part in comparison to their class of performance

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp
r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
dis=sns.boxplot(x='Class',y='Discussion',data=r)
mp.show()
```

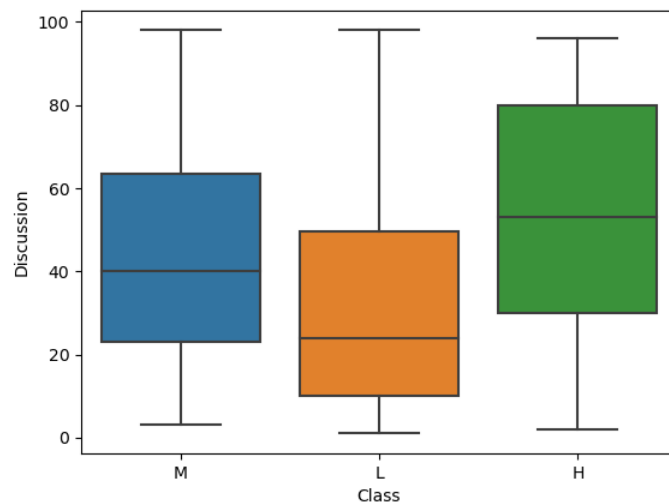


Fig:6.11

Conclusion: Above box plot represent the comparison of the student at class level.

6.12 Students who viewed announcement in comparison to their class of performance

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp
r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
anc=sns.boxplot(x='Class',y='AnnouncementsView',data=r)
mp.show()
```

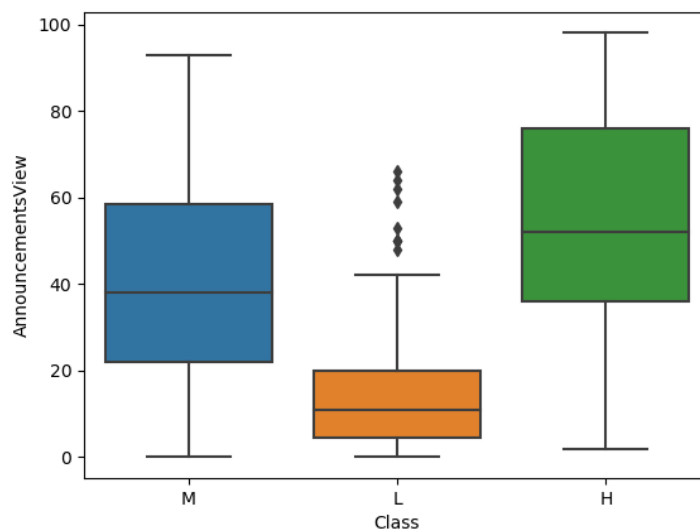


Fig:6.12

Conclusion: Above plot represent the comparison of student who viewed the announcements

6.13 Count the student absence days

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp
r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
sns.countplot(x='StudentAbsenceDays',hue='Class',data=r)
mp.show()
```

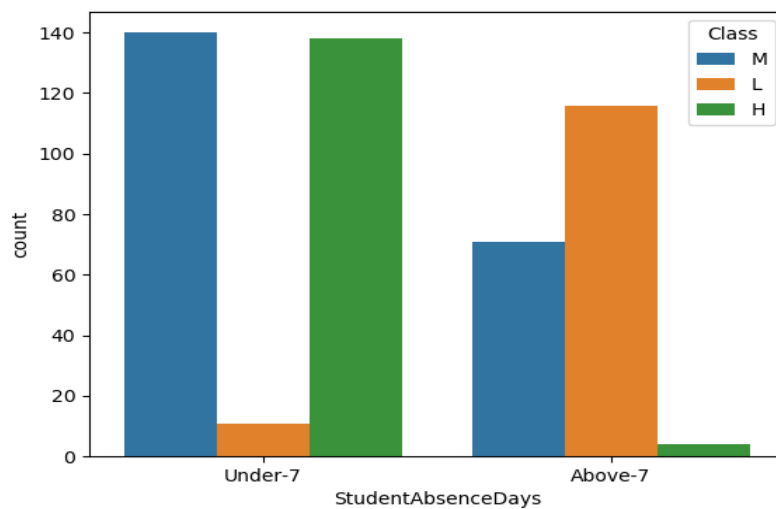


Fig:6.13

Conclusion: Above count plot count the total absence day through their class.

6.14 Gender Comparison With Parents Relationship

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

data = pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
plot = sns.countplot(x='Class', hue='Relation', data=data, order=['L', 'M', 'H'], palette='Set1')
plot.set(xlabel='Class', ylabel='Count', title='Gender comparison')
plt.show()
```

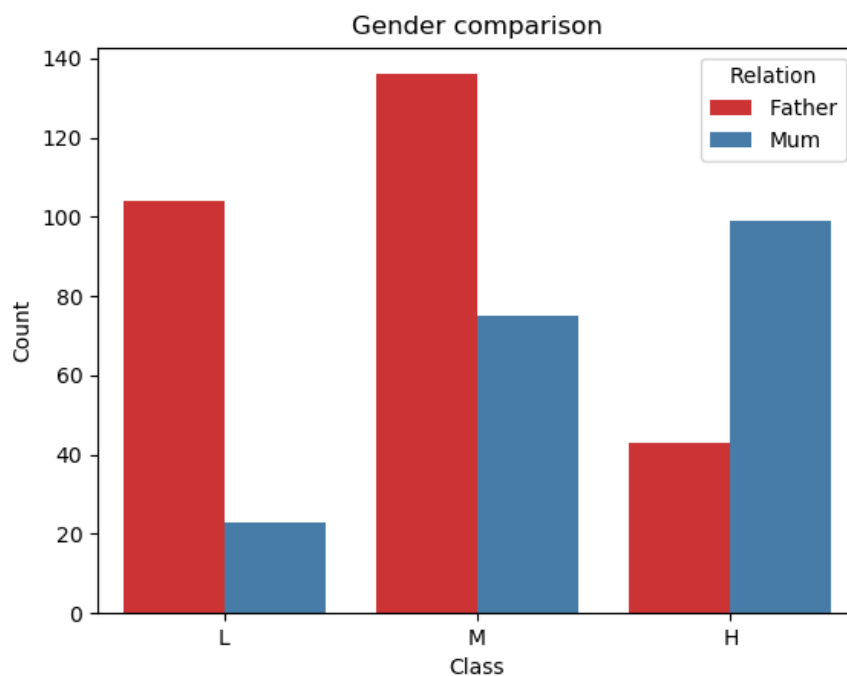


Fig:6.14

Conclusion: Above plot represent the comparison of gender through with parent's relationship.

6.15 count the relation of class wise

```
import pandas as pd
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
fig= plt.subplots(figsize=(15,7))
sns.countplot(x='Relation', hue='Class', data=df,hue_order = ['L', 'M', 'H'],palette="pastel" )
plt.show()
```

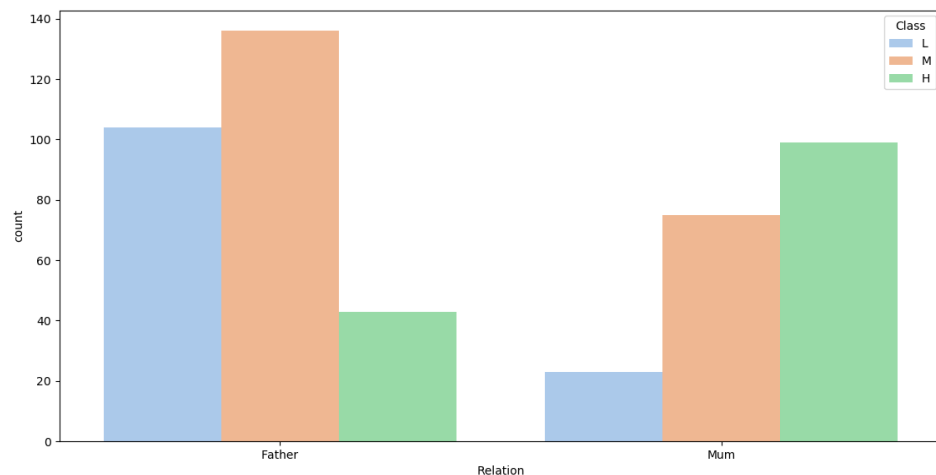


Fig:6.15

Conclusion: Above plot represent count of relation through their class.

6.16 GradeId wise count the student

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp

r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')

sns.countplot(x='GradeID', data=r, order=['G-02', 'G-04', 'G-05', 'G-06', 'G-07', 'G-08', 'G-09', 'G-10', 'G-11', 'G-12'])

mp.show()
```

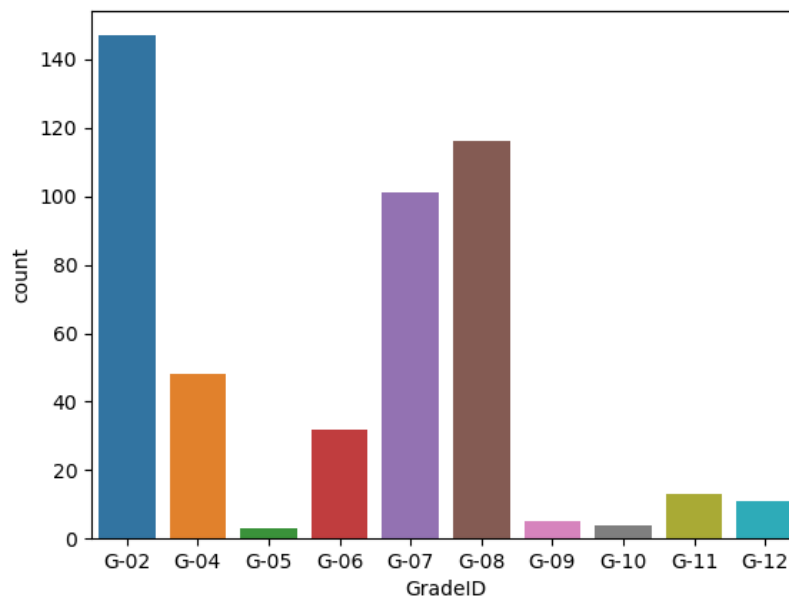


Fig:6.16

Conclusion: Above graph represent the how many students in each grade id.

6.17 Which Semester student has better performance on each class

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp

r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')

sem = sns.countplot(x='Class', hue='Semester', order=['L', 'M', 'H'], data=r)
sem.set(xlabel='Class', ylabel='Count', title='Semester comparison')
mp.show()
```

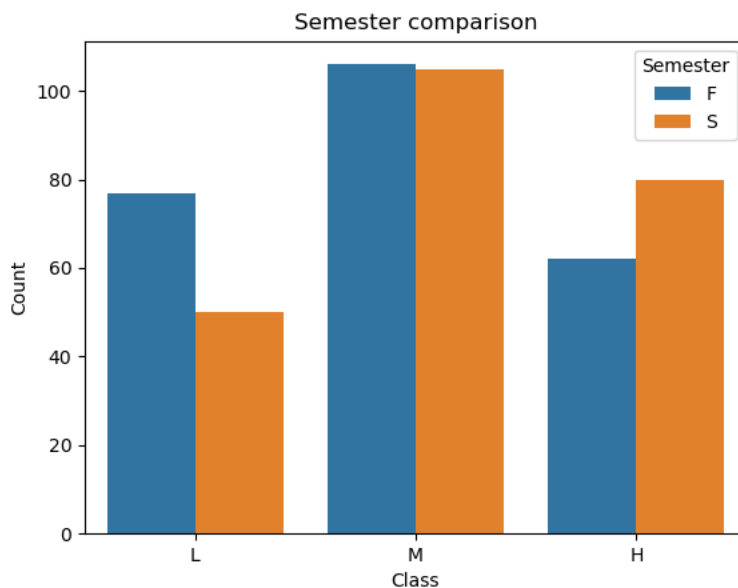


Fig:6.17

Conclusion: Above graph represent semester comparison based on the classes and performance.

6.18 different stage of education among classes

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as mp

r=pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
mp.figure(figsize=(15,5))

sns.countplot(data = r, x ='StageID', hue = 'Class', hue_order= ['H', 'M', 'L']).set_title("Graph showing number of stages among different classes", fontsize = 20)

mp.show()
```

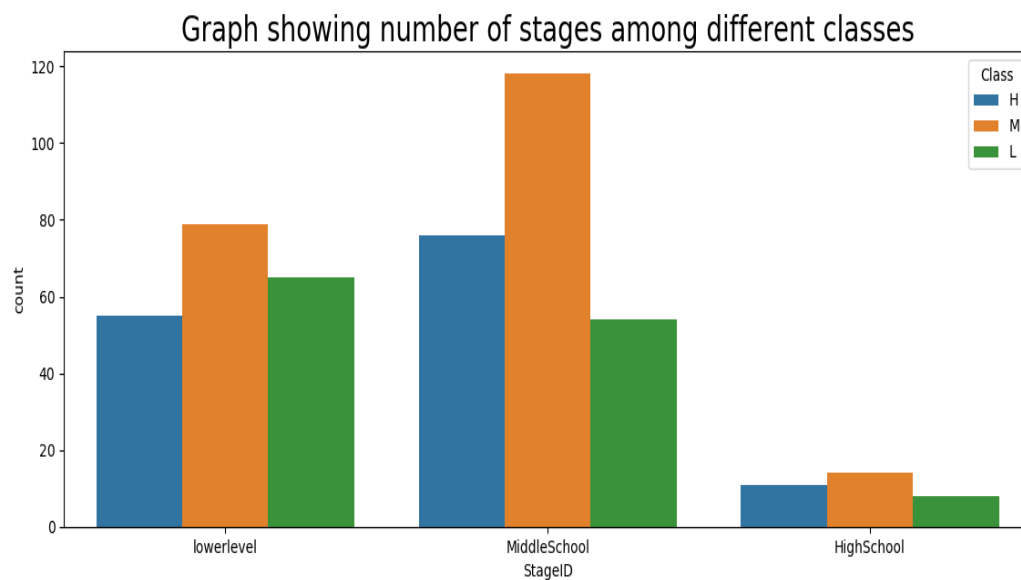


Fig:6.18

Conclusion: Above graph represent the number of stages among the classes

6.19 Analysis on students marks with raisehnads using linear regression.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.linear_model import LinearRegression
data = pd.read_csv('S:\Marwadi college\sem-4\Project\s1.csv')
data.shape
raisedhands=data['raisedhands']
Discussion=data['Student Marks']
X = data.iloc[:, 10].values
x=X.reshape(-1, 1)
Y = data.iloc[:, 17].values
y=Y.reshape(-1, 1)
lr = LinearRegression()
lr.fit(x,y)
r_sq=lr.score(x,y)
print('coefficient of determination:', r_sq)
print('intercept:', lr.intercept_)
print('slope:', lr.coef_)
Y_pred = lr.predict(x)
print('predicted response:', Y_pred, sep='\n')
plt.scatter(x, y)
plt.plot(x, Y_pred, color='red')
plt.show()
```

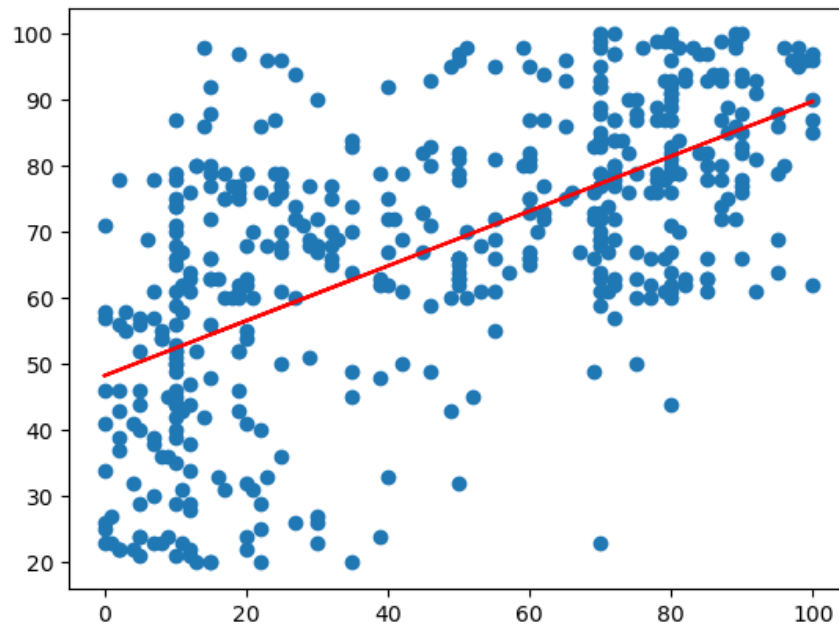


Fig:6.19

Output:

```

coefficient of determination: 0.38082411507947334
intercept: [48.27604736]
slope: [[0.41539306]]
predicted response:
[[54.50694329]
 [56.58390859]
 [52.42997798]
 [67.38412819]
 [64.89176982]
 [65.72255501]]

```

Conclusion: above graph represent the students marks with raise hande in LinearRegression

7. Conclusion

- ❖ We can get various information about Student Academic Performance using this dataset .
- ❖ From above analysis we can get idea which Topic maximum percentage of the student choose.
- ❖ We can also find those student who raise hand in the Topic.
- ❖ We can also find those students who raise hand with his nationality.
- ❖ From above analysis ratio of student his semester wise.
- ❖ From above analysis we can get idea which nationality student choose his subject.
- ❖ We can also find how many student in each class.
- ❖ We can also find the student who took part in class of performance.
- ❖ From above analysis student who viewed announcement in compare to their class.
- ❖ We can also count student absence days.

8. Bibliography

- ❖ www.kaggle.com
- ❖ www.tutorialspoint.com
- ❖ www.coursera.com

