Teacher of the Year Award

Name:Khaloud Alkhaled

Email: khalkhaled@tvtc.gov.sa

Data Science Camp

Objective:

- Main Target.
- Visualization with python code explaining
- The Result.
- My point of view.
- Conclusion.

Main Target:

"Teacher of the year award"

This reward based on the students interaction during the class in the three impotent subjects in the school base on the number of students in each subject classes in the Educational Stages.

So, I explore the school data to figure out the students and teachers interactions in classes, which will help us to identified the teachers who will be rewarded.

Exploring Data Analysis:

```
In [5]: df.shape

Out[5]: (480, 17)
```

The data that I used contain 480 students.

```
In [9]: df.isnull().values.any()

Out[9]: False
```

The data is clean .

Exploring Data Analysis:

```
In [10]: print(np.mean(df.raisedhands))
         print(np.max(df.raisedhands))
         print(np.min(df.raisedhands))
         print(np.std(df.raisedhands))
```

```
46.775
100
0
30.74714417632964
```

The output showing us that the minimum is zero of raising hands out of 480 students

Handling outlier:

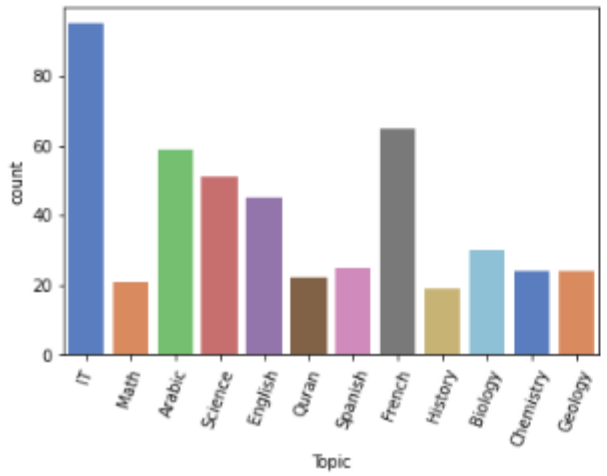
```
In [54]: df["Topic"].value_counts()

Out[54]: IT           95
         French       65
         Arabic       59
         Science      51
         English      45
         Biology      30
         Spanish      25
         Chemistry    24
         Geology      24
         Quran        22
         Math         21
         History      19
         Name: Topic, dtype: int64
```

I start to handle the outlier by figure out the subjects which has the max of the students, the output showing us that IT, Arabic , and French has the max of the students by using **value_counts** function

Handling outlier:

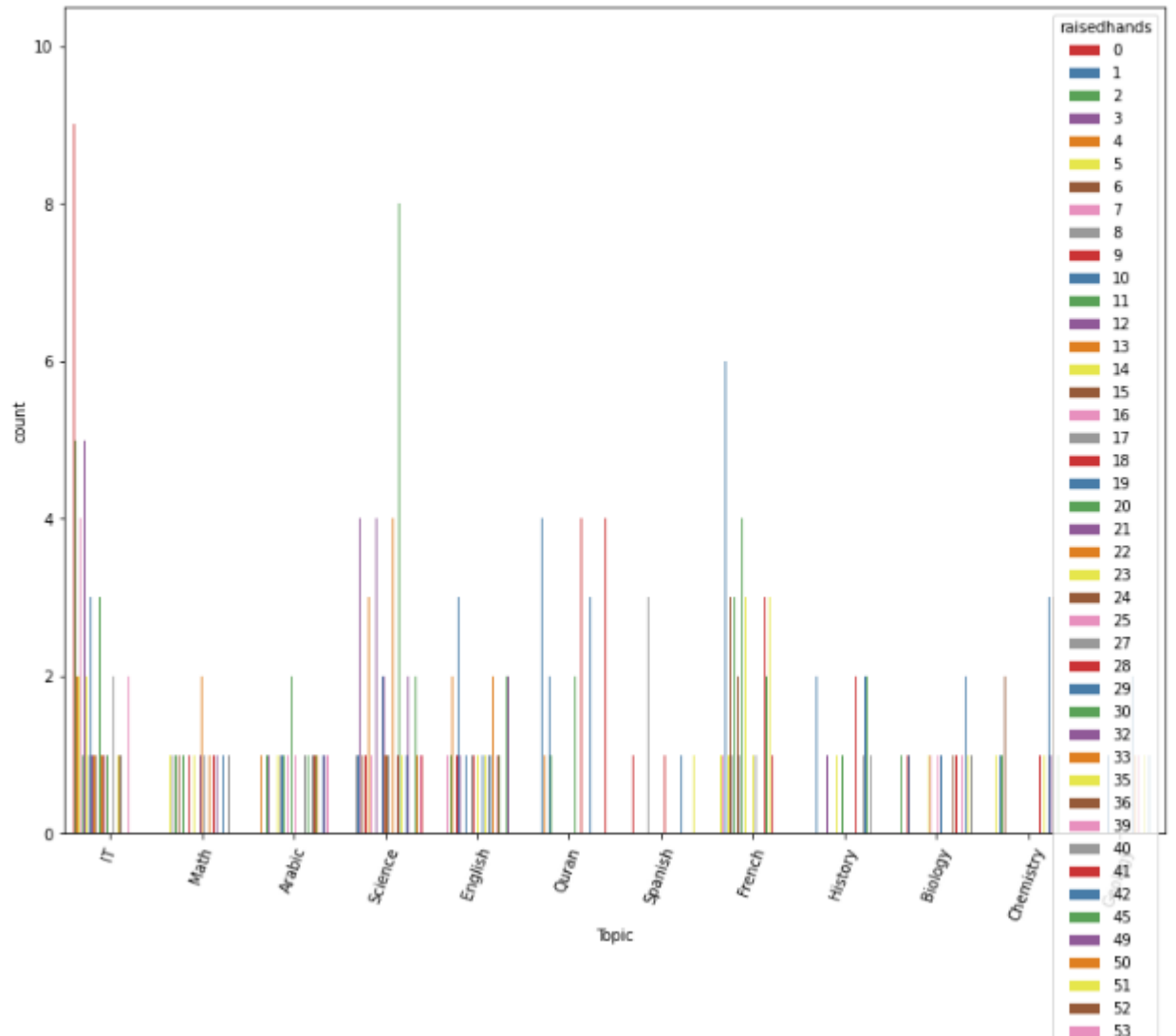
```
In [20]: fig = sns.countplot(x="Topic", data=df, palette="muted");
         fig.set_xticklabels( fig.get_xticklabels() ,  rotation=70)
         plt.show()
```



The chart showing that IT, Arabic , and French has the max of the students by using counplot function

Information about the most interactive class in the school:

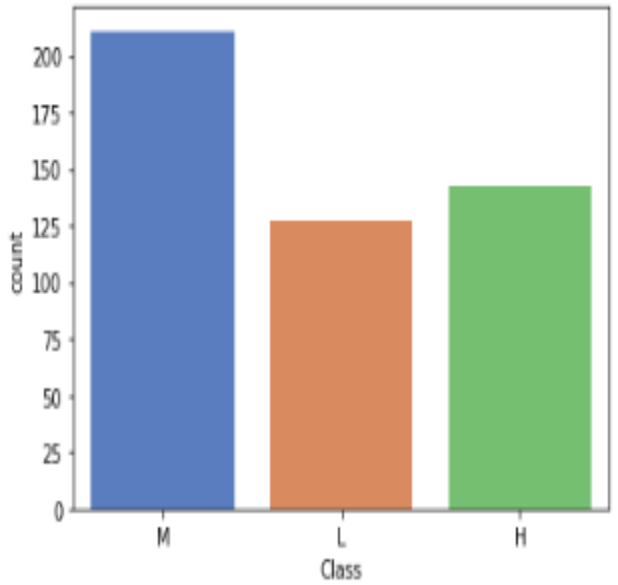
```
In [35]: plt.subplots(figsize=(13,10))
         fig = sns.countplot(x="Topic", data=df, palette="Set1",hue="raisedhands");
         fig.set_xticklabels( fig.get_xticklabels() ,  rotation=70)
         plt.show()
```



This chart showing us the raise hands in each subject for all the classes and even the countitiy .
Moreover, this chart showing us that the IT is the one. By using also the countplot function

My point of view:

```
In [62]: fig = sns.countplot(x="Class", data=df, palette="muted");
         plt.show()
```

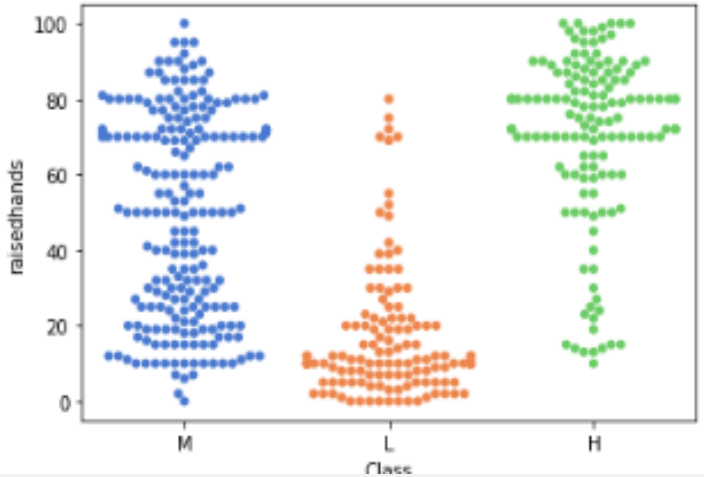


I used The **countplot** function to show how is the students level in the school and how can I make my condition to have a right decision, the most students in the school have medium marks. So, I focus in those students.

My point of view:

```
In [54]: Raised_hand = sns.swarmplot(x="Class", y="raisedhands", data=df,  palette="muted")
         plt.show()
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 11.0% of the points cannot be pl
aced; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 8.5% of the points cannot be pla
ced; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
```



I used The **swarmplot** function to show the relationship between the class and raise hands , it shows that the students who raise their hands more then 60 times got high score and the students who raise their hand less than 20 times got low marks. So, I decide to build my decision on the students who is higher than 20 times.

Information about the most interactive class in the school:

```
In [31]: df.query(  "Topic == 'French' and raisedhands > 20" )[["raisedhands"]]
```

Out[31]:

	raisedhands
156	70
162	70
163	60
164	55
165	72
166	51
167	80
168	60
169	30
170	40
171	60
174	50
176	60
178	80
179	40
180	60
181	50
316	60
317	50
318	60
319	50
320	60
321	50
323	30
324	24
325	27
327	30
328	80

The French subject comes with 31 outputs.

However, the school printable mentions that the comparative should be the counties also. Which means over 20 times. So, I do the comparative to find out the winner teacher.

Information about the most interactive class in the school:

```
In [39]: df.query(  "Topic == 'IT' and raisedhands > 20" )[["raisedhands"]]
```

Out[39]:

	raisedhands
3	30
4	40
5	42
9	70
15	30
16	36
18	69
19	70
20	60
37	45
43	30
44	33
61	50
62	80
68	70
69	39
70	22
71	29
84	50
86	70
91	80
92	50
93	55
94	80
95	100
99	50
100	50
101	70
107	70

The IT subject comes with 39 outputs.

Information about the most interactive class in the school:

```
In [27]: df.query( "Topic == 'Arabic' and raisedhands > 20" )[["raisedhands"]]
```

Out[27]:

	raisedhands
27	25
29	30
30	35
182	85
183	25
185	87
186	85
187	80
188	75
189	85
190	23
192	95
193	81
194	53
196	92
197	83
198	27
199	45
201	45
202	25
203	22
204	29
205	72
206	67
253	39
362	90
363	80

The Arabic subject comes with 27 outputs.

My point of view:

I think the way that the school followed to select the good teacher was not rational because the interaction is not the only way to use to identify who is the good teacher, but there are many aspects that could be used like students Grade/Mark.

Conclusion:

In conclusion, based on the data analysis that we have shown , the reward will go to IT teachers.