

## Summary of the project:

To study the Gender of the students per year and the graduation students ration per year and the Saudi students comparing to non Saudi.

## Data of the project :

The data is in excel file format I use pandas to convert it to csv format

The data is clean

`data.info()`

shows no null data and the data is clean

`data.info` shows the following :

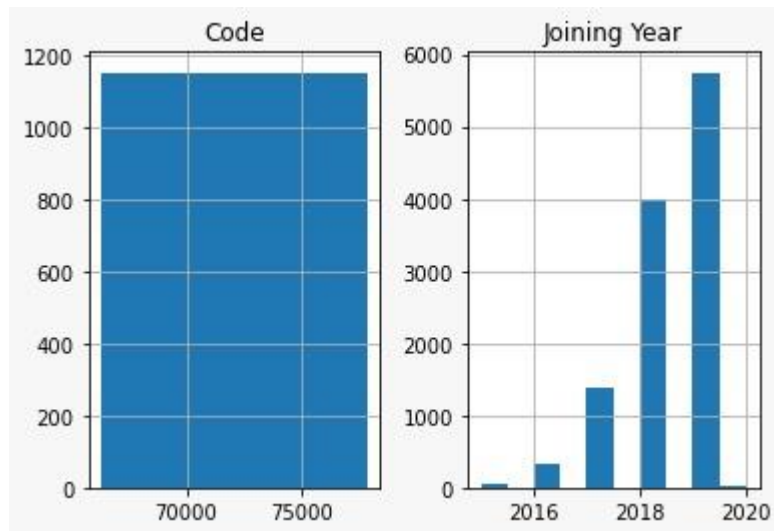
Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	Code	11522 non-null	int64
1	First Name	11522 non-null	object
2	Last Name	11522 non-null	object
3	Father Name	11522 non-null	object
4	Mother Name	11522 non-null	object
5	Gender	11522 non-null	object
6	Nationality	11522 non-null	object
7	Birth Date	11522 non-null	object
8	Level	11522 non-null	object
9	City	11522 non-null	object
10	ID	11522 non-null	object
11	Joining Day	11522 non-null	int64
12	Joining Month	11522 non-null	int64
13	Joining Year	11522 non-null	int64
14	Current Class	11522 non-null	object

From the Data.shape I have  
(11522, 15)

```
print(data['Joining Year'].mean())>>> 2018.312098593994
```

linear plotting for  
x=code  
y=joining year



the future work

to convert all non-numerical values to numerical value and start plotting