

# EDA of placement data

First the given dataset was explored.

```
df = pd.read_csv('Placement_Data_Full_Class.csv')
df.head()
```

✓ 0.0s Python

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	270000.0
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000.0
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	250000.0
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	NaN
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	425000.0

We can see among quantitative features, we have ssc\_p, hsc\_p, degree\_p, etest\_p and mba\_p. Among categorical features, we have gender, ssc\_b, hsc\_b, hsc\_s, degree\_t, workex and specialization. Further, we noticed that each of the categorical data have only two categories.

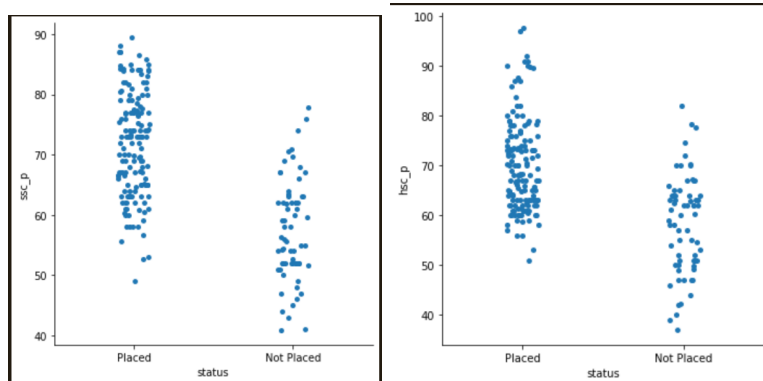
Salary Column has NaN values for people who have not been placed. In data cleanup, we replace NaN values with 0:

```
df['salary'] = df['salary'].fillna(0)
df.head()
```

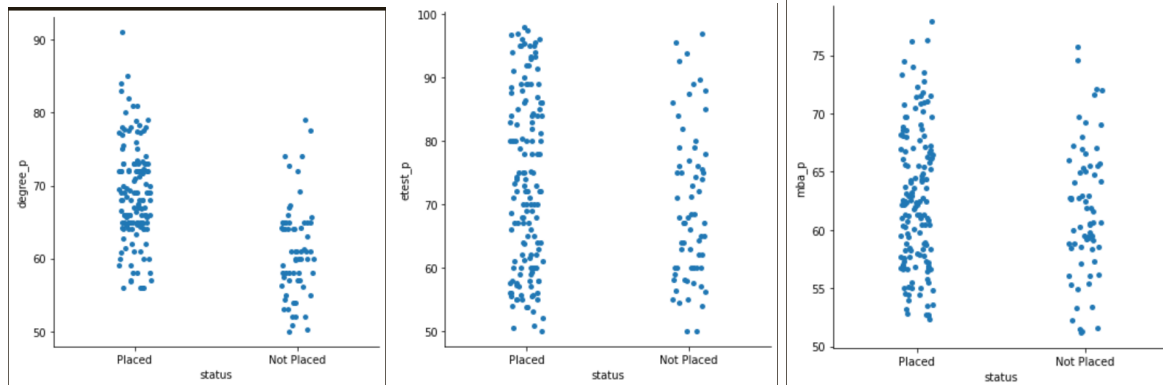
✓ 0.0s Python

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed	270000.0
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000.0
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed	250000.0
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed	0.0
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed	425000.0

To explore the effects of different quantitative factors on placement status:



We can say that people with ssc\_p or hsc\_p or degree\_p below 60 % have a lesser chance of getting placed.

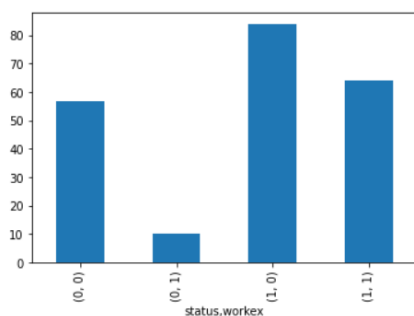


etest\_p or employability test score seems to have less impact on placement status, seeing as people scoring high were not placed and people scoring less were placed in good numbers. Likewise, mba\_p does not seem to have a strong impact on placement status, and a decisive assumption cannot be drawn from the scatter plot.

**To explore the effects of different categorical features on placement status:**

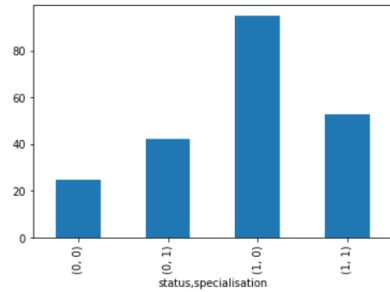
Next we explored the placement status based on several factors.

#### 1. WorkExp



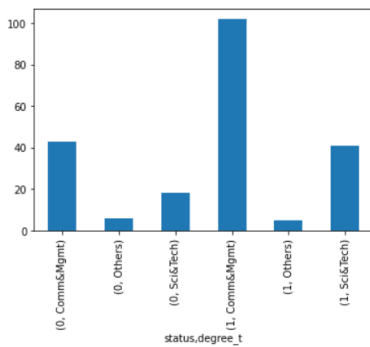
We observed that students having some work experience have a higher chance of getting placed than students having no experience.

#### 2. MBA Specialization



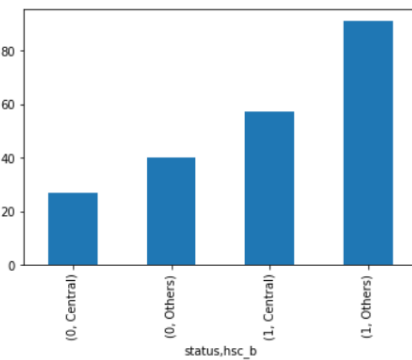
Here, we observed that a majority number of students having specialisation in Mkt&Fin are placed.

### 3. Degree Type



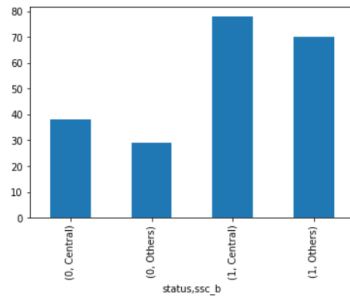
Here we see that chances of students getting placed is higher for students having degree type Comm&Mgmt.

### 4. HSC board



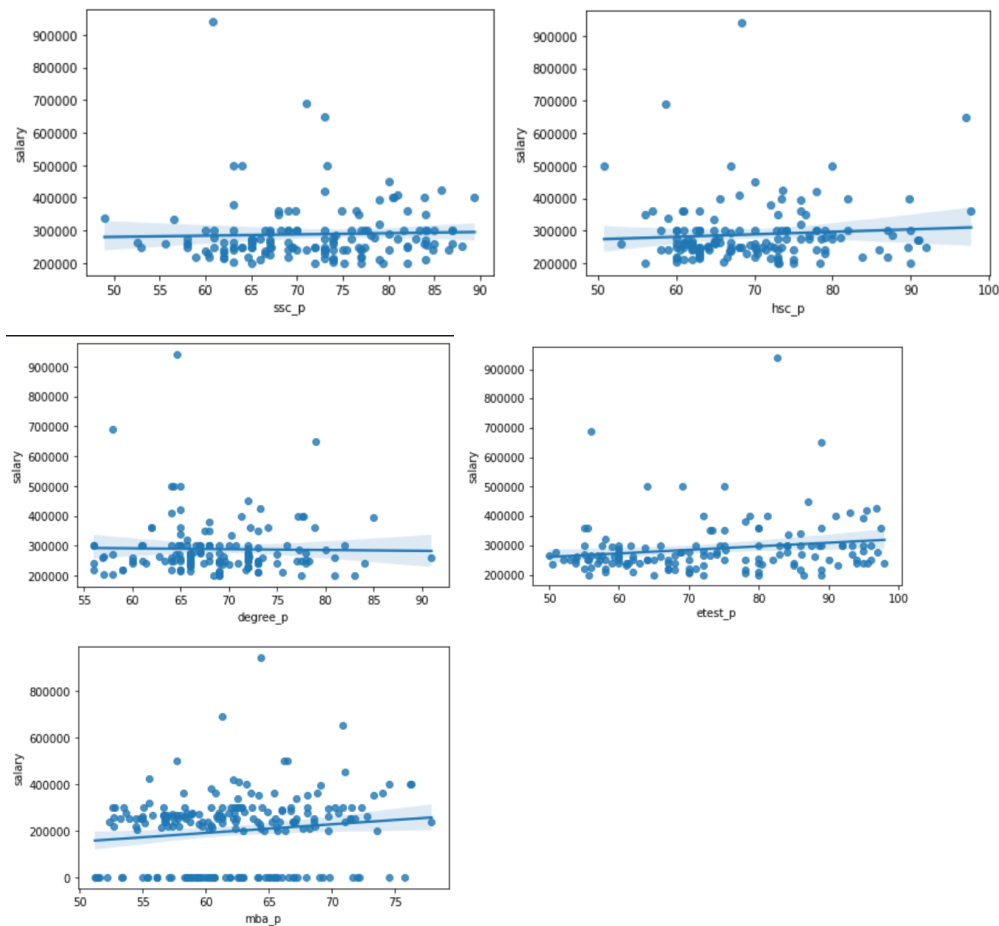
Here we observed that placement chances of students belonging to other boards are very high.

### 5. SSC board



It is clearly seen that students belonging to central board are getting placed more in number.

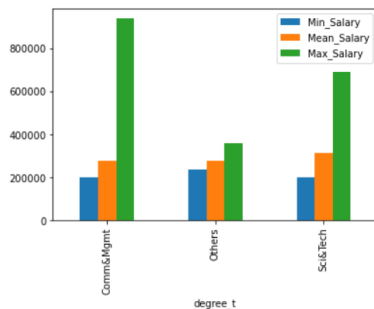
**To explore the effects of different quantitative factors on salary received:**



There is a slight increase in salary received for people with higher etest\_p, following which we can say that better etest\_p means higher salary received.

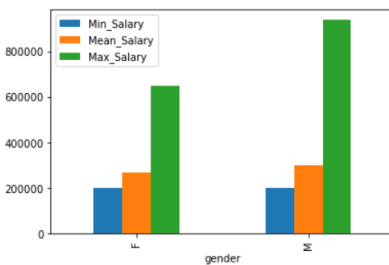
## To explore the effects of different categorical factors on placement status:

We explored the highest, average and lowest salaries of the students based on their degree type.



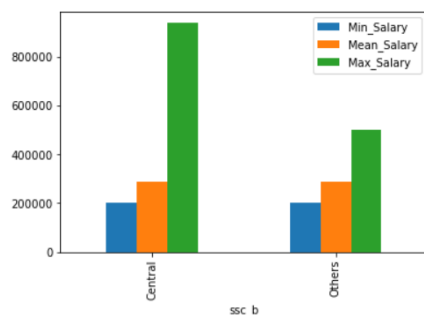
Clearly, we can observe that the maximum salary of degree type Comm&Mgmt is the highest one, the minimum salary of degree type Sci&Tech is the lowest one. The maximum salary paid to the Comm&Mgmt students are very much higher than the other degree type students. Furthermore, there is not much difference in their mean salaries.

Next, we tried to find out the pattern in salaries according to the gender of students.



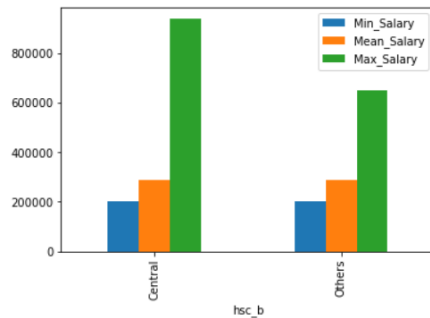
So, here we found that the difference between the highest salary paid to male and female students are much more. Males are paid a much higher maximum salary than females. Average salary of male students is also a little bit higher.

We also tried to explore the pattern of salaries on the basis of Secondary Education Boards.



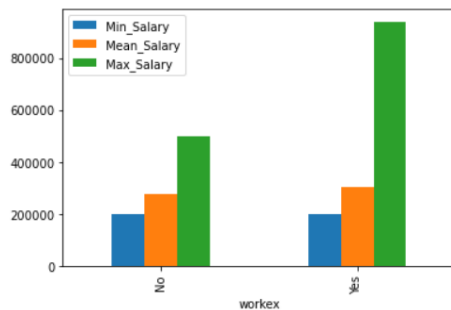
Clearly, central board secondary students are paid higher maximum, minimum and average salary than others.

Similarly, we found out some interesting pattern on salary according to the Higher Secondary Education Board students.



Here, we see that the maximum salary offered to the students of central board students are way higher than the students of other boards. There is not much difference between the average and lowest salary offered to them.

Now, we came to know that work experience played a major role in the salaries that are offered to the students.



Here, it is observed that if a student has some work experience then he/she is offered a very high maximum salary. The average salary is also higher for students having some experience.

**To find correlation matrix of given dataset:**



From the above heatmap/correlation matrix, we can clearly see that etest\_p and mba\_p have the highest effect on salary received, as we concluded from our scatter plots above.