

+ Code

+ Text

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

```
#reading data set
df=pd.read_csv('/content/drive/MyDrive/data set/Placement_Data_Full_Class.csv')
df.head()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	speci
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	

```
df.tail()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	speci
210	211	M	80.6	Others	82.0	Others	Commerce	77.6	Comm&Mgmt	N		
211	212	M	58.0	Others	60.0	Others	Science	72.0	Sci&Tech	N		
212	213	M	67.0	Others	67.0	Others	Commerce	73.0	Comm&Mgmt	Ye		
213	214	F	74.0	Others	66.0	Others	Commerce	58.0	Comm&Mgmt	N		
214	215	M	68.0	Central	58.0	Others	Science	58.0	Comm&Mgmt	N		

```
shape=df.shape
print(shape)
```

```
(215, 15)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215 entries, 0 to 214
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sl_no                 215 non-null   int64
1   gender                215 non-null   object
2   ssc_p                 215 non-null   float64
3   ssc_b                 215 non-null   object
4   hsc_p                 215 non-null   float64
5   hsc_b                 215 non-null   object
6   hsc_s                 215 non-null   object
7   degree_p              215 non-null   float64
8   degree_t              215 non-null   object
9   workex                215 non-null   object
10  etest_p               215 non-null   float64
11  specialisation         215 non-null   object
12  mba_p                 215 non-null   float64
13  status                 215 non-null   object
14  salary                 148 non-null   float64
dtypes: float64(6), int64(1), object(8)
memory usage: 25.3+ KB
```

```
df.describe()
```

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	salary
count	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	148.000000
mean	108.000000	67.303395	66.333163	66.370186	72.100558	62.278186	288655.405405
std	62.209324	10.827205	10.897509	7.358743	13.275956	5.833385	93457.452420

```
df.columns
```

```
Index(['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'],
      dtype='object')
```

```
max 215 000000 80 400000 67 700000 61 000000 68 000000 77 800000 940000 000000
```

```
#to find missing values
```

```
df.isnull().sum()
```

```
sl_no      0
gender      0
ssc_p      0
ssc_b      0
hsc_p      0
hsc_b      0
hsc_s      0
degree_p   0
degree_t   0
workex     0
etest_p    0
specialisation 0
mba_p      0
status     0
salary    67
dtype: int64
```

```
null_values=[columns for columns in df.columns if df[columns].isnull().sum()>1]
print(null_values)
```

```
['salary']
```

▼ OBSERVATION:

- It has been found that 'salary' field in the dataset contains 64 null values

```
df.salary.value_counts()
```

```
300000.0    22
250000.0    18
240000.0    15
260000.0     7
360000.0     6
200000.0     6
265000.0     6
220000.0     5
275000.0     5
210000.0     4
400000.0     4
270000.0     4
216000.0     3
350000.0     3
500000.0     3
252000.0     2
236000.0     2
230000.0     2
280000.0     2
218000.0     2
204000.0     2
276000.0     2
255000.0     1
285000.0     1
340000.0     1
690000.0     1
233000.0     1
290000.0     1
650000.0     1
264000.0     1
225000.0     1
940000.0     1
```

```

393000.0    1
380000.0    1
420000.0    1
425000.0    1
336000.0    1
231000.0    1
268000.0    1
450000.0    1
287000.0    1
411000.0    1
320000.0    1
278000.0    1
295000.0    1
Name: salary, dtype: int64

```

```

## missing values have to be replaced with mode of the corresponding column
df.salary.mode()

```

```

0    300000.0
Name: salary, dtype: float64

```

```

df.salary=df['salary'].fillna(300000.0)
df.salary.isnull().sum()
#thus missing values are replaced

```

```

0

```

```

#Dealing with categorical values
df['gender']=df['gender'].map({'F':0,'M':1})
df.head()

```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	speci
0	1	1	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	
1	2	1	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	
2	3	1	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	
3	4	1	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	
4	5	1	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	

```

df.ssc_b.unique()
#ssc_b is also a categorical feature

```

```

array(['Others', 'Central'], dtype=object)

```

```

df['ssc_b']=df['ssc_b'].map({'Others':0,'Central':1})
df['hsc_b']=df['hsc_b'].map({'Others':0,'Central':1})
df.head()

```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	special
0	1	1	67.00	0	91.00	0	Commerce	58.00	Sci&Tech	No	55.0	
1	2	1	79.33	1	78.33	0	Science	77.48	Sci&Tech	Yes	86.5	
2	3	1	65.00	1	68.00	1	Arts	64.00	Comm&Mgmt	No	75.0	
3	4	1	56.00	1	52.00	1	Science	52.00	Sci&Tech	No	66.0	
4	5	1	85.80	1	73.60	1	Commerce	73.30	Comm&Mgmt	No	96.8	

```

df['hsc_s'].unique()
array(['Commerce', 'Science', 'Arts'], dtype=object)

```

```

df['hsc_s']=df['hsc_s'].map({'Commerce':0,'Science':1,'Arts':2})
df.head()

```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisat
0	1	1	67.00	0	91.00	0	0	58.00	Sci&Tech	No	55.0	Mkt
1	2	1	79.33	1	78.33	0	1	77.48	Sci&Tech	Yes	86.5	Mkt

```
df['degree_t'].unique()
```

```
array(['Sci&Tech', 'Comm&Mgmt', 'Others'], dtype=object)
```

```
df['degree_t']=df['degree_t'].map({'Sci&Tech':0,'Comm&Mgmt':1,'Others':2})
```

```
df.head()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisator
0	1	1	67.00	0	91.00	0	0	58.00	0	No	55.0	Mkt&HF
1	2	1	79.33	1	78.33	0	1	77.48	0	Yes	86.5	Mkt&Fir
2	3	1	65.00	1	68.00	1	2	64.00	1	No	75.0	Mkt&Fir
3	4	1	56.00	1	52.00	1	1	52.00	0	No	66.0	Mkt&HF
4	5	1	85.80	1	73.60	1	0	73.30	1	No	96.8	Mkt&Fir

```
df['workex'].unique()
```

```
array(['No', 'Yes'], dtype=object)
```

```
df['workex']=df['workex'].map({'No':0,'Yes':1})
```

```
df.head()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisator
0	1	1	67.00	0	91.00	0	0	58.00	0	0	55.0	Mkt&HF
1	2	1	79.33	1	78.33	0	1	77.48	0	1	86.5	Mkt&Fir
2	3	1	65.00	1	68.00	1	2	64.00	1	0	75.0	Mkt&Fir
3	4	1	56.00	1	52.00	1	1	52.00	0	0	66.0	Mkt&HF
4	5	1	85.80	1	73.60	1	0	73.30	1	0	96.8	Mkt&Fir

```
df['specialisation'].unique()
```

```
array(['Mkt&HR', 'Mkt&Fin'], dtype=object)
```

```
df['specialisation']=df['specialisation'].map({'Mkt&HR':0, 'Mkt&Fin':1})
```

```
df.head()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisator
0	1	1	67.00	0	91.00	0	0	58.00	0	0	55.0	0
1	2	1	79.33	1	78.33	0	1	77.48	0	1	86.5	1
2	3	1	65.00	1	68.00	1	2	64.00	1	0	75.0	1
3	4	1	56.00	1	52.00	1	1	52.00	0	0	66.0	0
4	5	1	85.80	1	73.60	1	0	73.30	1	0	96.8	1

```
df['status'].unique()
```

```
array(['Placed', 'Not Placed'], dtype=object)
```

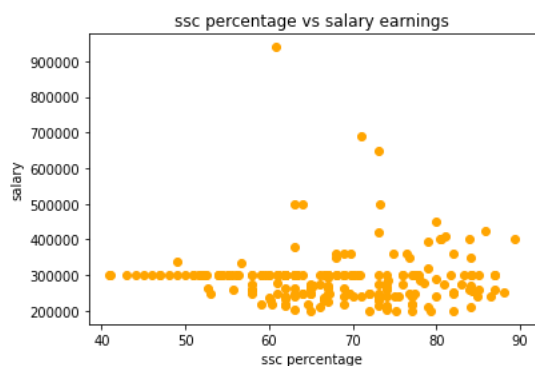
```
df['status']=df['status'].map({'Placed':1,'Not Placed':0})
```

```
df.head()
```

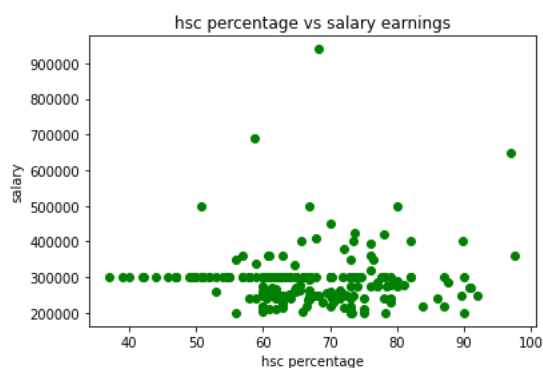
	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation
0	1	1	67.00	0	91.00	0	0	58.00	0	0	55.0	0
1	2	1	70.00	1	70.00	0	1	77.00	0	1	96.0	1

DATA VISUALISATION

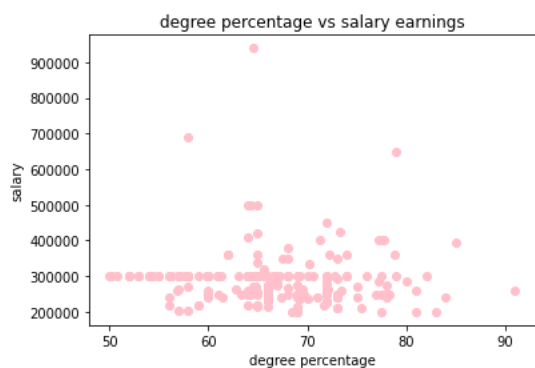
```
plt.scatter(df['ssc_p'],df['salary'],c='orange')
plt.xlabel('ssc percentage')
plt.ylabel('salary')
plt.title('ssc percentage vs salary earnings')
plt.show()
```



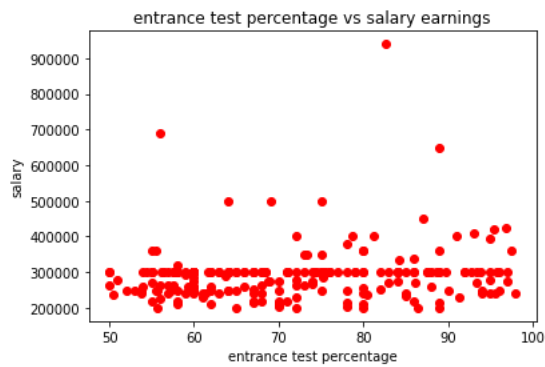
```
plt.scatter(df['hsc_p'],df['salary'],c='green')
plt.xlabel('hsc percentage')
plt.ylabel('salary')
plt.title('hsc percentage vs salary earnings')
plt.show()
```



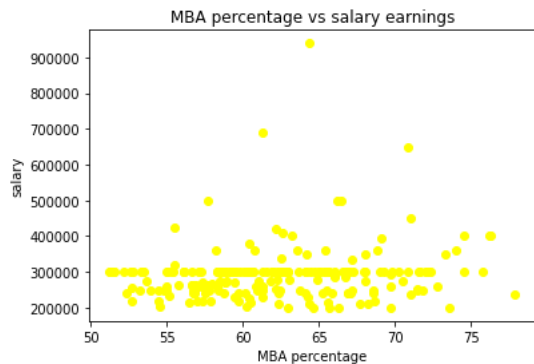
```
plt.scatter(df['degree_p'],df['salary'],c='pink')
plt.xlabel('degree percentage')
plt.ylabel('salary')
plt.title('degree percentage vs salary earnings')
plt.show()
```



```
plt.scatter(df['etest_p'],df['salary'],c='red')
plt.xlabel('entrance test percentage')
plt.ylabel('salary')
plt.title('entrance test percentage vs salary earnings')
plt.show()
```



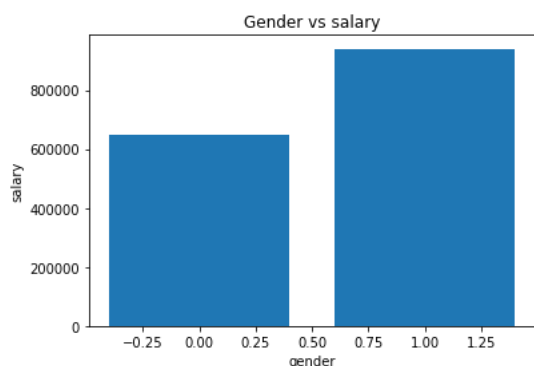
```
plt.scatter(df['mba_p'],df['salary'],c='yellow')
plt.xlabel('MBA percentage')
plt.ylabel('salary')
plt.title('MBA percentage vs salary earnings')
plt.show()
```



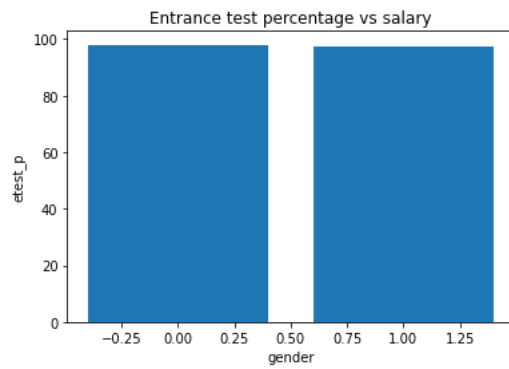
▼ OBSERVATION:

- *The person whose score is around 60 in sslc is earning higher salary
- *The person whose score is around 65 in hsc is earning higher salary
- *The person whose score is around 65 in degree is earning higher salary
- *The person whose score is around 65 in MBA is earning higher salary

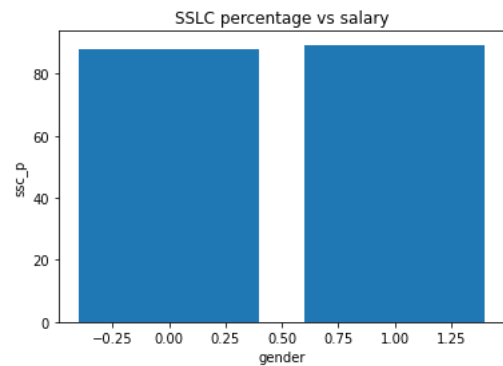
```
plt.bar(df['gender'],df['salary'])
plt.xlabel('gender')
plt.ylabel('salary')
plt.title('Gender vs salary')
plt.show()
```



```
plt.bar(df['gender'],df['etest_p'])  
plt.xlabel('gender')  
plt.ylabel('etest_p')  
plt.title('Entrance test percentage vs salary')  
plt.show()
```



```
plt.bar(df['gender'],df['ssc_p'])  
plt.xlabel('gender')  
plt.ylabel('ssc_p')  
plt.title('SSLC percentage vs salary')  
plt.show()
```



▼ OBSERVATION:

*Men is earning higher salary than women