

```
In [63]: import pandas as pd
import math
import matplotlib.pyplot as plt
```

```
In [64]: df = pd.read_csv("campus_r/Placement_Data_Full_Class.csv")
```

```
In [65]: df.head()
```

```
Out[65]:
```

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

```
In [66]: df.describe()
```

```
Out[66]:
```

	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
min	1.000000	40.890000	37.000000	50.000000	50.000000	51.210000	200000.000000
25%	54.500000	60.600000	60.900000	61.000000	60.000000	57.945000	240000.000000
50%	108.000000	67.000000	65.000000	66.000000	71.000000	62.000000	265000.000000
75%	161.500000	75.700000	73.000000	72.000000	83.500000	66.255000	300000.000000
max	215.000000	89.400000	97.700000	91.000000	98.000000	77.890000	940000.000000

```
In [71]: not_none = [True if not math.isnan(x) else False for x in df['salary']]
```

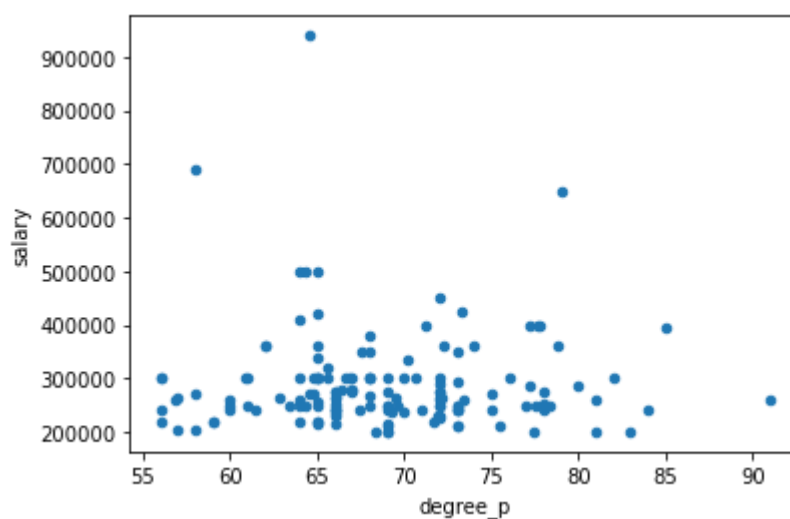
```
In [68]: df2 = df.loc[:,['gender', 'degree_p', 'salary']].loc[not_none,:]  
df2.head()
```

Out[68]:

	gender	degree_p	salary
0	M	58.00	270000.0
1	M	77.48	200000.0
2	M	64.00	250000.0
4	M	73.30	425000.0
7	M	66.00	252000.0

```
In [79]: degree_salary = df2.loc[:,['degree_p', 'salary']]  
degree_salary.plot.scatter(x="degree_p" , y="salary")
```

Out[79]: <matplotlib.axes.\_subplots.AxesSubplot at 0xcb93208>



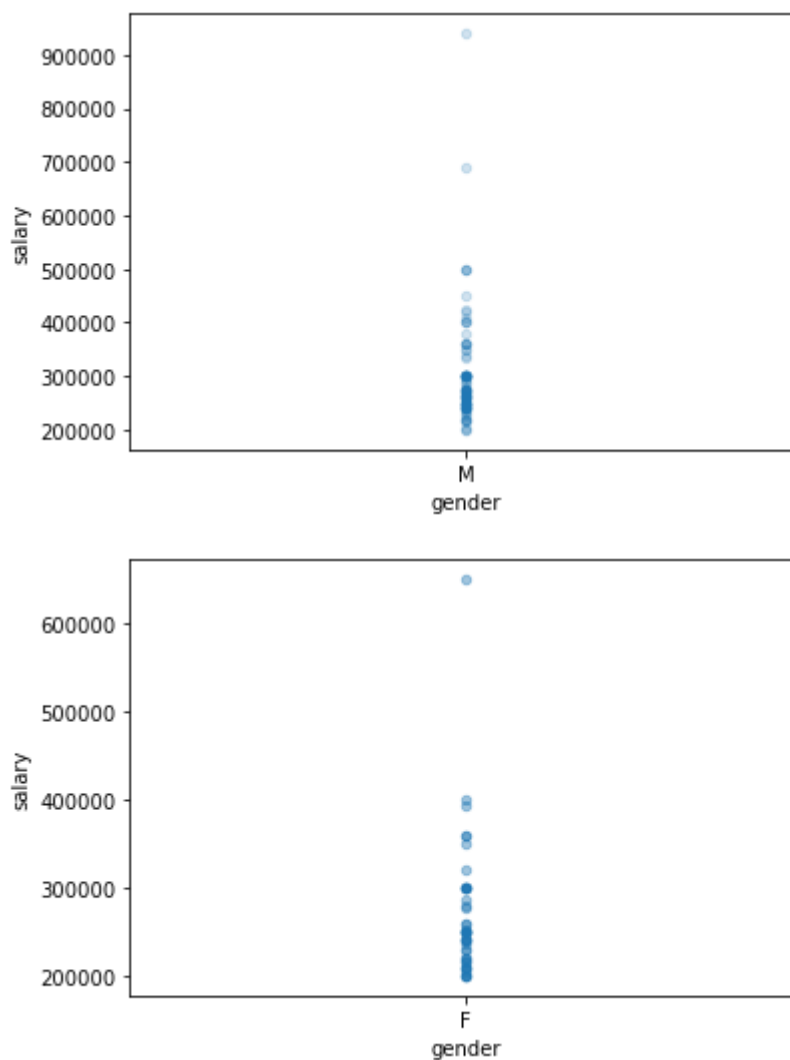
اینطور که به نظر میاد تراکم حقوق بیشتر حدود سیصد هزار دلار در سال هست در هر رده معدلی و ارتباط چندانی با معدل ندارد

```
In [81]: gender_salary = df2.loc[:,['gender', 'salary']]  
gender_salary['gender'].value_counts()
```

Out[81]: M 100  
F 48  
Name: gender, dtype: int64

```
In [97]: male_salary = df2.loc[df2['gender']=="M",['gender','salary']]
male_salary.plot.scatter(x="gender", y="salary", alpha=20/len(male_salary))
female_salary = df2.loc[df2['gender']=="F",['gender','salary']]
female_salary.plot.scatter(x="gender", y="salary", alpha=20/len(female_salary))
```

Out[97]: <matplotlib.axes.\_subplots.AxesSubplot at 0xcdcd0118>



در این مورد مقدار را بر تعداد تقسیم کردیم و آلفا را بر اساس آن تعیین کردیم تا تعداد بیشتر نمونه از آقایان باعث گمراهی نشود اما بازم گویا حقوق آقایان بیشتره