tamrin5

May 28, 2020

[32]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

[33]:
```
database=pd.read_csv("datasets_596958_1073629_Placement_Data_Full_Class.csv")
```

[34]:
```
database
```

[34]:
```
     sl_no gender  ssc_p    ssc_b  hsc_p    hsc_b     hsc_s  degree_p  \
0        1      M  67.00   Others  91.00   Others  Commerce     58.00
1        2      M  79.33  Central  78.33   Others   Science     77.48
2        3      M  65.00  Central  68.00  Central      Arts     64.00
3        4      M  56.00  Central  52.00  Central   Science     52.00
4        5      M  85.80  Central  73.60  Central  Commerce     73.30
..     ...    ...    ...      ...    ...      ...       ...       ...
210    211      M  80.60   Others  82.00   Others  Commerce     77.60
211    212      M  58.00   Others  60.00   Others   Science     72.00
212    213      M  67.00   Others  67.00   Others  Commerce     73.00
213    214      F  74.00   Others  66.00   Others  Commerce     58.00
214    215      M  62.00  Central  58.00   Others   Science     53.00

      degree_t workex  etest_p specialisation  mba_p      status    salary
0     Sci&Tech     No     55.0         Mkt&HR  58.80      Placed  270000.0
1     Sci&Tech    Yes     86.5        Mkt&Fin  66.28      Placed  200000.0
2    Comm&Mgmt     No     75.0        Mkt&Fin  57.80      Placed  250000.0
3     Sci&Tech     No     66.0         Mkt&HR  59.43  Not Placed       NaN
4    Comm&Mgmt     No     96.8        Mkt&Fin  55.50      Placed  425000.0
..         ...    ...      ...            ...    ...         ...       ...
210  Comm&Mgmt     No     91.0        Mkt&Fin  74.49      Placed  400000.0
211   Sci&Tech     No     74.0        Mkt&Fin  53.62      Placed  275000.0
212  Comm&Mgmt    Yes     59.0        Mkt&Fin  69.72      Placed  295000.0
213  Comm&Mgmt     No     70.0         Mkt&HR  60.23      Placed  204000.0
214  Comm&Mgmt     No     89.0         Mkt&HR  60.22  Not Placed       NaN

[215 rows x 15 columns]
```

[35]:
```
database.head()
```

[35]:
```
   sl_no gender  ssc_p    ssc_b  hsc_p    hsc_b     hsc_s  degree_p  \
0      1      M  67.00   Others  91.00   Others  Commerce     58.00
1      2      M  79.33  Central  78.33   Others   Science     77.48
2      3      M  65.00  Central  68.00  Central      Arts     64.00
3      4      M  56.00  Central  52.00  Central   Science     52.00
4      5      M  85.80  Central  73.60  Central  Commerce     73.30

    degree_t workex  etest_p specialisation  mba_p      status    salary
0   Sci&Tech     No     55.0         Mkt&HR  58.80      Placed  270000.0
1   Sci&Tech    Yes     86.5        Mkt&Fin  66.28      Placed  200000.0
2  Comm&Mgmt     No     75.0        Mkt&Fin  57.80      Placed  250000.0
3   Sci&Tech     No     66.0         Mkt&HR  59.43  Not Placed       NaN
4  Comm&Mgmt     No     96.8        Mkt&Fin  55.50      Placed  425000.0
```

[36]:
```
database.describe()
```

[36]:
```
            sl_no       ssc_p       hsc_p    degree_p     etest_p       mba_p  \
count  215.000000  215.000000  215.000000  215.000000  215.000000  215.000000
mean   108.000000   67.303395   66.333163   66.370186   72.100558   62.278186
std     62.209324   10.827205   10.897509    7.358743   13.275956    5.833385
min      1.000000   40.890000   37.000000   50.000000   50.000000   51.210000
25%     54.500000   60.600000   60.900000   61.000000   60.000000   57.945000
50%    108.000000   67.000000   65.000000   66.000000   71.000000   62.000000
75%    161.500000   75.700000   73.000000   72.000000   83.500000   66.255000
max    215.000000   89.400000   97.700000   91.000000   98.000000   77.890000

              salary
count     148.000000
mean   288655.405405
std     93457.452420
min    200000.000000
25%    240000.000000
50%    265000.000000
75%    300000.000000
max    940000.000000
```

[37]:
```
df=database.sort_values('salary')
```

[38]:
```
df2=df.iloc[:,[7,14]]
```

[39]:
```
df2=df2['degree_p']>60.0
df2
```

[39]:
```
107     True
1       True
135     True
50      True
```

```
44      True
       …
198     True
201    False
206    False
208     True
214    False
Name: degree_p, Length: 215, dtype: bool
```

[40]:
```
db9=(database['degree_p']>60.0)&(database['salary']>300000)
db20=database.loc[db9,:]
db20.count()
```

[40]:
```
sl_no             27
gender            27
ssc_p             27
ssc_b             27
hsc_p             27
hsc_b             27
hsc_s             27
degree_p          27
degree_t          27
workex            27
etest_p           27
specialisation    27
mba_p             27
status            27
salary            27
dtype: int64
```

pas natije migirim ke 27 ta az anha in khososiat ra darand va darsad ziadist chon man balay 300000 ra gereftam hoghogh bala

[41]:
```
DB=database.loc[database.gender=='F']
DB2=database.loc[database.gender=='M']
DB2['salary'].value_counts()
```

[41]:
```
300000.0    16
250000.0    11
240000.0    10
265000.0     6
275000.0     5
260000.0     5
270000.0     4
360000.0     4
400000.0     3
500000.0     3
```

```
220000.0     3
350000.0     2
276000.0     2
200000.0     2
420000.0     1
264000.0     1
340000.0     1
425000.0     1
336000.0     1
225000.0     1
218000.0     1
280000.0     1
204000.0     1
255000.0     1
236000.0     1
411000.0     1
252000.0     1
231000.0     1
285000.0     1
450000.0     1
295000.0     1
233000.0     1
690000.0     1
290000.0     1
380000.0     1
940000.0     1
216000.0     1
268000.0     1
Name: salary, dtype: int64
```

[42]:
```
DB['salary'].value_counts()
```

[42]:
```
250000.0    7
300000.0    6
240000.0    5
200000.0    4
210000.0    4
216000.0    2
220000.0    2
360000.0    2
230000.0    2
260000.0    2
236000.0    1
218000.0    1
280000.0    1
393000.0    1
278000.0    1
```

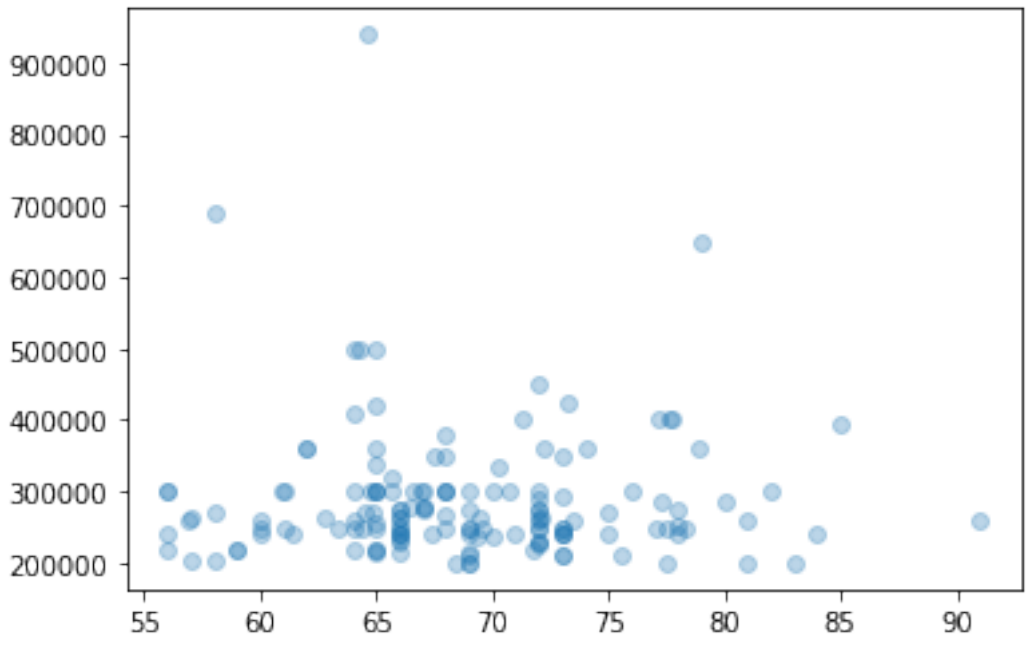
```
350000.0    1
400000.0    1
252000.0    1
650000.0    1
204000.0    1
320000.0    1
287000.0    1
Name: salary, dtype: int64
```

[43]:
```
plt.scatter(database['gender'],database['salary'], alpha=0.3)
```

[43]:
```
<matplotlib.collections.PathCollection at 0x7f6dbea30c10>
```

[44]:
```
plt.scatter(database['degree_p'],database['salary'], alpha=0.3)
```

[44]:
```
<matplotlib.collections.PathCollection at 0x7f6dbe99b890>
```


900000
800000
700000
600000
500000
400000
300000
200000
55
60
65
70
75
80
85
90

60 80

[ ]: