In [1]:
```
# Importing the necessary packages
import pandas as pd
import numpy as np
```

In [2]:
```
#importing the database to be used.
df = pd.read_csv("salaries.csv")
```

In [4]:
```
#printing out the first 5 rows
df.head()
```

Out[4]:

	rank	discipline	phd	service	sex	salary
0	Prof	B	56	49	Male	186960
1	Prof	A	12	6	Male	93000
2	Prof	A	23	20	Male	110515
3	Prof	A	40	31	Male	131205
4	Prof	B	20	18	Male	104800

In [5]:
```
#printing out the last 5 rows.
df.tail()
```

Out[5]:

	rank	discipline	phd	service	sex	salary
73	Prof	B	18	10	Female	105450
74	AssocProf	B	19	6	Female	104542
75	Prof	B	17	17	Female	124312
76	Prof	A	28	14	Female	109954
77	Prof	A	23	15	Female	109646

In [7]:
```
#getting the details of your database.
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 78 entries, 0 to 77
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   rank        78 non-null     object
 1   discipline  78 non-null     object
 2   phd         78 non-null     int64 
 3   service     78 non-null     int64 
 4   sex         78 non-null     object
 5   salary      78 non-null     int64 
dtypes: int64(3), object(3)
memory usage: 3.8+ KB
```

In [12]:
```
#knowing the shape of your database(rows,columns).
df.shape
```

Out[12]:
```
(78, 6)
```

In [13]:
```
#getting the index labels of your columns.
df.columns
```

Out[13]:
```
Index(['rank', 'discipline', 'phd', 'service', 'sex', 'salary'], dtype='object')
```

In [15]:
```
#getting the unique values
df.nunique()
```

Out[15]:
```
rank          3
discipline    2
phd          36
service      33
sex           2
salary       73
dtype: int64
```

In [17]:
```
#checking for null values.
df.isnull().sum()
```

Out[17]:
```
rank          0
discipline    0
phd           0
service       0
sex           0
salary        0
dtype: int64
```

In [90]:
```
#df.describe()
```

In [25]:
```
df.describe(include='object')
```

Out[25]:

	rank	discipline	sex
count	78	78	78
unique	3	2	2
top	Prof	B	Female
freq	46	42	39

In [47]:
```
#getting your numerical columns
num_df = df.select_dtypes(exclude='O')
num = num_df.columns.to_list()
```

In [48]:
```
print(num)
```

```
['phd', 'service', 'salary']
```

In [50]:
```
#getting your categorical columns
cat_df = df.select_dtypes(include='O')
cat = cat_df.columns.to_list()
```

In [51]:
```
print(cat)
```

```
['rank', 'discipline', 'sex']
```

In [63]:
```
#### getting the summary stat. of the salary column, grouped based on the rank in the dataframe
G = df.groupby('rank')
G[['salary']].agg(['mean','median','std'])
```

Out[63]:

	salary		
	mean	median	std
rank			
AssocProf	91786.230769	103613.0	18571.183714
AsstProf	81362.789474	78500.0	9381.245301
Prof	123624.804348	123331.5	24850.287853

In [75]:
```
#### getting the summary stat. of the phd column, grouped based on the rank in the dataframe
G = df.groupby('rank')
G[['phd']].agg(['mean','median','std'])
```

Out[75]:

	phd		
	mean	median	std
rank			
AssocProf	15.076923	13.0	5.589597
AsstProf	5.052632	4.0	2.738079
Prof	27.065217	24.5	10.185834

In [76]:
```
## getting the summary stat. of the service column, grouped based on the rank in the dataframe
G = df.groupby('rank')
G[['service']].agg(['mean','median','std'])
```

Out[76]:

	service		
	mean	median	std
rank			
AssocProf	11.307692	9	5.879124
AsstProf	2.210526	2	1.750522
Prof	21.413043	19	11.255766

In [77]:
```
## getting the summary stat. of the phd column, grouped based on the sex in the dataframe
G = df.groupby('sex')
G[['phd']].agg(['mean','median','std'])
```

Out[77]:

	phd		
	mean	median	std
sex			
Female	16.512821	17	9.784176
Male	22.897436	21	14.138032

In [78]:
```
## getting the summary stat. of the salary column, grouped based on the sex in the dataframe
G = df.groupby('sex')
G[['salary']].agg(['mean','median','std'])
```

Out[78]:

	salary		
	mean	median	std
sex			
Female	101002.410256	103750	25952.127317
Male	115045.153846	107300	29110.516397

In [85]:
```
## getting the summary stat. of the service column, grouped based on the sex in the dataframe
G = df.groupby('sex')
G[['service']].agg(['mean','median','std'])
```

Out[85]:

	service		
	mean	median	std
sex			
Female	11.564103	10	8.813252
Male	18.538462	19	13.999711

In [91]:
```
## getting the summary stat. of the service column, grouped based on the discipline in the dataframe
G = df.groupby('discipline')
G[['service']].agg(['mean','median','std'])
```

Out[91]:

	service		
	mean	median	std
discipline			
A	15.722222	16.0	12.584823
B	14.476190	12.5	11.867640

In [92]:
```
## getting the summary stat. of the salary column, grouped based on the discipline in the dataframe
G = df.groupby('discipline')
G[['salary']].agg(['mean','median','std'])
```

Out[92]:

	salary		
	mean	median	std
discipline			
A	98331.111111	93675	26000.183807
B	116331.785714	110581	27805.722287

In [93]:
```
# getting the summary stat. of the phd column, grouped based on the discipine in the dataframe.
G = df.groupby('discipline')
G[['phd']].agg(['mean','median','std'])
```

Out[93]:

	phd		
	mean	median	std
discipline			
A	21.527778	21.5	12.971367
B	18.142857	17.0	12.013349

In [112]:
```
#getting the minimum values grouped based on discipline
df.groupby(['discipline']).max()
```

Out[112]:

	rank	phd	service	sex	salary
discipline					
A	Prof	51	51	Male	155865
B	Prof	56	49	Male	186960

In [113]:
```
#getting the minimum values grouped based on discipline
df.groupby(['discipline']).min()
```

Out[113]:

	rank	phd	service	sex	salary
discipline					
A	AssocProf	2	0	Female	57800
B	AssocProf	1	0	Female	71065

In [119]:
```
#getting the maximum values grouped based on sex
df.groupby(['sex']).max()
```

Out[119]:

	rank	discipline	phd	service	salary
sex					
Female	Prof	B	39	36	161101
Male	Prof	B	56	51	186960

In [120]:
```
# getting the minimum values grouped based on discipline
df.groupby(['sex']).min()
```

Out[120]:

	rank	discipline	phd	service	salary
sex					
Female	AssocProf	A	2	0	62884
Male	AssocProf	A	1	0	57800

In [124]:
```
# getting the minimum values grouped based on salary.
df.groupby(['salary']).max()
```

Out[124]:

	rank	discipline	phd	service	sex
salary					
57800	Prof	A	51	51	Male
62884	AssocProf	A	25	22	Female
63100	AsstProf	A	7	6	Female
71065	AssocProf	B	12	9	Female
72500	AsstProf	A	3	1	Female
...	...	...	...	...	...
150750	Prof	B	22	21	Male
155865	Prof	A	45	43	Male
161101	Prof	B	24	15	Female
162200	Prof	B	35	33	Male
186960	Prof	B	56	49	Male

73 rows × 5 columns

In [ ]:
```

```