Importing Necessary Libraries

```
import pandas as pd
import numpy as np
```

```
C:\Users\aakas\AppData\Local\Temp\ipykernel_16520\2162656668.py:1:
DeprecationWarning:
Pyarrow will become a required dependency of pandas in the next major
release of pandas (pandas 3.0),
(to allow more performant data types, such as the Arrow string type,
and better interoperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at
https://github.com/pandas-dev/pandas/issues/54466

  import pandas as pd
```

Reading Dataset

```
salary = pd.read_csv("assignment3-part-1-dataset1.csv")

#store = pd.read_csv("assignment3-part-1-dataset2.csv")

salary.head()
```

```
   Unnamed: 0                Company Name       Job Title  Salaries
Reported  \
0           0                     Mu Sigma  Data Scientist
105
1           1                          IBM  Data Scientist
95
2           2  Tata Consultancy Services  Data Scientist
66
3           3           Impact Analytics  Data Scientist
40
4           4                   Accenture  Data Scientist
32

    Location     Salary
0  Bangalore   648573.0
1  Bangalore  1191950.0
2  Bangalore   836874.0
3  Bangalore   669578.0
4  Bangalore   944110.0
```

Replacing ? with NaN

```
salary.replace("?", np.nan,inplace=True)
```

Checking Missing Values

```
salary.isnull().sum()

Unnamed: 0            0
Company Name          0
Job Title             0
Salaries Reported     0
Location              0
Salary                0
dtype: int64
```

Grouping dataset by Job Title

```
salary.groupby("Job Title")

<pandas.core.groupby.generic.DataFrameGroupBy object at
0x000001E13C1BB140>

salary.head()

   Unnamed: 0               Company Name       Job Title  Salaries
Reported  \
0           0                   Mu Sigma  Data Scientist
105
1           1                        IBM  Data Scientist
95
2           2  Tata Consultancy Services  Data Scientist
66
3           3           Impact Analytics  Data Scientist
40
4           4                  Accenture  Data Scientist
32

    Location     Salary
0  Bangalore   648573.0
1  Bangalore  1191950.0
2  Bangalore   836874.0
3  Bangalore   669578.0
4  Bangalore   944110.0
```

Finding Mean

```
mean = salary.groupby("Job Title")["Salary"].mean()
print(mean)

Job Title
Associate Machine Learning Engineer     4.643720e+05
Data Analyst                            6.164699e+05
Data Engineer                           1.309051e+06
Data Science                            3.649053e+05
Data Science Associate                  1.203913e+06
```

```
Data Science Consultant                   2.671464e+06
Data Science Lead                         4.068310e+06
Data Science Manager                      4.619021e+06
Data Scientist                            1.411330e+06
Data Scientist - Trainee                  6.105120e+05
Junior Data Scientist                     5.963231e+05
Lead Data Scientist                       1.852189e+06
Machine Learning Associate                2.951140e+05
Machine Learning Consultant               7.064010e+05
Machine Learning Data Analyst             3.613780e+05
Machine Learning Data Associate           2.758410e+05
Machine Learning Data Associate I         2.585960e+05
Machine Learning Data Associate II        3.832130e+05
Machine Learning Developer                5.811190e+05
Machine Learning Engineer                 7.971884e+05
Machine Learning Scientist                1.701180e+05
Machine Learning Software Engineer        1.397347e+06
Senior Data Scientist                     1.766130e+06
Senior Machine Learning Engineer          1.473436e+06
Software Engineer - Machine Learning      1.566780e+06
Name: Salary, dtype: float64
```

Finding median

```
median = salary.groupby("Job Title")["Salary"].median()
print(median)
```

```
Job Title
Associate Machine Learning Engineer       464372.0
Data Analyst                              508150.5
Data Engineer                             792683.0
Data Science                              240780.0
Data Science Associate                   1203913.0
Data Science Consultant                  2671464.0
Data Science Lead                        4068310.0
Data Science Manager                     4619021.0
Data Scientist                            914480.0
Data Scientist - Trainee                  610512.0
Junior Data Scientist                     554963.0
Lead Data Scientist                      1664364.0
Machine Learning Associate                295114.0
Machine Learning Consultant               706401.0
Machine Learning Data Analyst             361378.0
Machine Learning Data Associate           275841.0
Machine Learning Data Associate I         258596.0
Machine Learning Data Associate II        383213.0
Machine Learning Developer                581119.0
Machine Learning Engineer                 627048.5
Machine Learning Scientist                170118.0
Machine Learning Software Engineer       1397347.0
```

```
Senior Data Scientist                      1733388.0
Senior Machine Learning Engineer           1335445.0
Software Engineer - Machine Learning       1566780.0
Name: Salary, dtype: float64
```

Finding mode

```
mode = salary.groupby("Job Title")["Salary"].apply(lambda
x:x.mode().iloc[0])
print(mode)

# lambda x: This defines an anonymous function (lambda function) that
takes one argument x.
# In this context, x represents each group of salaries within each job
title.

# x.mode(): Inside the lambda function, x is a Series object
containing all the salaries within a specific job title group.
# The mode() function is called on this Series object to compute the
mode, i.e., the most frequently occurring value.

# .iloc[0]: After calculating the mode, .iloc[0] is used to retrieve
the first value from the resulting Series.
# This is necessary because the mode() function may return multiple
values if there are ties for the most frequent value.
# By selecting the first value, we ensure that only one mode value is
returned.
```

```
Job Title
Associate Machine Learning Engineer      464372.0
Data Analyst                             338792.0
Data Engineer                            515940.0
Data Science                             180000.0
Data Science Associate                  1203913.0
Data Science Consultant                 2671464.0
Data Science Lead                       4068310.0
Data Science Manager                    4619021.0
Data Scientist                           600000.0
Data Scientist - Trainee                 610512.0
Junior Data Scientist                    616492.0
Lead Data Scientist                     1520967.0
Machine Learning Associate               295114.0
Machine Learning Consultant              186475.0
Machine Learning Data Analyst            361378.0
Machine Learning Data Associate          275841.0
Machine Learning Data Associate I        258596.0
Machine Learning Data Associate II       383213.0
Machine Learning Developer               410952.0
Machine Learning Engineer                128988.0
Machine Learning Scientist                62160.0
```

```
Machine Learning Software Engineer      1397347.0
Senior Data Scientist                   2474429.0
Senior Machine Learning Engineer         229416.0
Software Engineer - Machine Learning    1521236.0
Name: Salary, dtype: float64
```

Finding minimum value

```
minimum = salary.groupby("Job Title")["Salary"].min()
print(minimum)
```

```
Job Title
Associate Machine Learning Engineer      464372.0
Data Analyst                              10814.0
Data Engineer                             33120.0
Data Science                              60840.0
Data Science Associate                  1203913.0
Data Science Consultant                 2671464.0
Data Science Lead                       4068310.0
Data Science Manager                    4619021.0
Data Scientist                            48000.0
Data Scientist - Trainee                 610512.0
Junior Data Scientist                     60840.0
Lead Data Scientist                     1520967.0
Machine Learning Associate               295114.0
Machine Learning Consultant              186475.0
Machine Learning Data Analyst            361378.0
Machine Learning Data Associate          275841.0
Machine Learning Data Associate I        258596.0
Machine Learning Data Associate II       383213.0
Machine Learning Developer               410952.0
Machine Learning Engineer                 21628.0
Machine Learning Scientist                62160.0
Machine Learning Software Engineer      1397347.0
Senior Data Scientist                    324089.0
Senior Machine Learning Engineer         229416.0
Software Engineer - Machine Learning    1521236.0
Name: Salary, dtype: float64
```

Finding Maximum Values

```
maximum = salary.groupby("Job Title")["Salary"].max()
print(maximum)
```

```
Job Title
Associate Machine Learning Engineer     4.643720e+05
Data Analyst                            3.900962e+07
Data Engineer                           1.190400e+08
Data Science                            2.000000e+06
Data Science Associate                  1.203913e+06
```

```
Data Science Consultant                    2.671464e+06
Data Science Lead                          4.068310e+06
Data Science Manager                       4.619021e+06
Data Scientist                             1.661404e+08
Data Scientist - Trainee                   6.105120e+05
Junior Data Scientist                      1.498750e+06
Lead Data Scientist                        2.839138e+06
Machine Learning Associate                 2.951140e+05
Machine Learning Consultant                1.226327e+06
Machine Learning Data Analyst              3.613780e+05
Machine Learning Data Associate            2.758410e+05
Machine Learning Data Associate I          2.585960e+05
Machine Learning Data Associate II         3.832130e+05
Machine Learning Developer                 7.512860e+05
Machine Learning Engineer                  6.518917e+06
Machine Learning Scientist                 2.780760e+05
Machine Learning Software Engineer         1.397347e+06
Senior Data Scientist                      3.654010e+06
Senior Machine Learning Engineer           3.110514e+06
Software Engineer - Machine Learning       1.612324e+06
Name: Salary, dtype: float64
```

Finding Standard Deviation

```
std = salary.groupby("Job Title")["Salary"].std()
std.replace(np.NaN, 0, inplace=True)
print(std)
```

```
Job Title
Associate Machine Learning Engineer        0.000000e+00
Data Analyst                               1.292116e+06
Data Engineer                              6.009190e+06
Data Science                               3.388020e+05
Data Science Associate                     0.000000e+00
Data Science Consultant                    0.000000e+00
Data Science Lead                          0.000000e+00
Data Science Manager                       0.000000e+00
Data Scientist                             5.140558e+06
Data Scientist - Trainee                   0.000000e+00
Junior Data Scientist                      3.931792e+05
Lead Data Scientist                        5.017356e+05
Machine Learning Associate                 0.000000e+00
Machine Learning Consultant                7.352864e+05
Machine Learning Data Analyst              0.000000e+00
Machine Learning Data Associate            0.000000e+00
Machine Learning Data Associate I          0.000000e+00
Machine Learning Data Associate II         0.000000e+00
Machine Learning Developer                 2.406525e+05
Machine Learning Engineer                  7.047460e+05
Machine Learning Scientist                 1.526757e+05
```

```
Machine Learning Software Engineer      0.000000e+00
Senior Data Scientist                   7.833905e+05
Senior Machine Learning Engineer        9.506370e+05
Software Engineer - Machine Learning    6.440894e+04
Name: Salary, dtype: float64
```