In [1]:

```
#Import The Differnt Libaries
import pandas as wd
import numpy as kp
import matplotlib
from matplotlib import pyplot as plt
import seaborn as sns
```

In [2]:

```
#Read The Data From the CSV File
#Import the dataset
x = wd.read_csv('Employee_monthly_salary.csv')
print(x)
```

```
      EmpID              Name Gender Date_of_Birth  Age   Join_Date  \
0     19575      Keven Norman      M    03-09-1994   25  02-12-2019
1     19944    Kristin Werner      F    23-06-1994   26  13-01-2020
2     20055      Avery Barber      M    27-02-1996   24  14-11-2019
3     20058      Boris Gibson      M    29-09-1993   26  13-01-2020
4     20332         Leif Mack      M    01-05-1991   29  04-06-2018
...     ...               ...    ...           ...  ...         ...
1797  18835  Darius Wilkerson      M    14-01-1991   29  21-08-2017
1798  19066     Erick Ballard      M    29-08-1992   27  25-09-2017
1799  21644   Lawerence Downs      M    05-07-1991   29  01-04-2019
1800  19673     Abdul Watkins      M    19-08-1972   47  26-12-2017
1801  19790   Chase Fernandez      M    20-03-1993   27  22-01-2018

      Tenure_in_org_in_months   GROSS  Net_Pay  Deduction  \
0                           7   74922    71494       3428
1                           6   44375    39971       4404
2                           8   82263    77705       4558
3                           6   44375    40164       4211
4                          25  235405   143963      91442
...                       ...     ...      ...        ...
1797                       34   88934    88734        200
1798                       33  133224   133024        200
1799                       15   72547    71246       1301
1800                       30  227176   220778       6398
1801                       29  114641   114441        200

      Deduction_percentage                            Designation  \
0                     4.58    Product Operations Analyst.Associate.
1                     9.92  Platform Operations Engineer.Associate.
2                     5.54  Platform Operations Engineer.Associate.
3                     9.49  Platform Operations Engineer.Associate.
4                    38.84                Software Engineer.Senior.
...                    ...                                     ...
1797                  0.22           Technical Solutions Engineer..
1798                  0.15                   Software Engineer II..
1799                  1.79            Business Operations Analyst..
1800                  2.82              Manager..Account Management
1801                  0.17                           Order Analyst..

                                  Department
0                  IT Product Management & Ops
1                          Platform Operations
2                          Platform Operations
3                          Platform Operations
4                 Enterprise Access Engineering
...                                         ...
1797                         AmaTec - EMEA TSE
1798              GSS EPIC Engineering (HC COGS)
1799                     Marketing - Operations
1800                               Americas- AMG
1801  Finance - Customer Revenue Operations G&A

[1802 rows x 13 columns]
```

In [8]:

```
#Assign it to a variable called Emp.
Emp= wd.read_csv('Employee_monthly_salary.csv')
print(Emp)
```

```
      EmpID              Name Gender Date_of_Birth  Age   Join_Date  \
0     19575      Keven Norman      M    03-09-1994   25  02-12-2019
1     19944    Kristin Werner      F    23-06-1994   26  13-01-2020
2     20055      Avery Barber      M    27-02-1996   24  14-11-2019
3     20058      Boris Gibson      M    29-09-1993   26  13-01-2020
4     20332         Leif Mack      M    01-05-1991   29  04-06-2018
...     ...               ...    ...           ...  ...         ...
1797  18835  Darius Wilkerson      M    14-01-1991   29  21-08-2017
1798  19066     Erick Ballard      M    29-08-1992   27  25-09-2017
1799  21644   Lawerence Downs      M    05-07-1991   29  01-04-2019
1800  19673     Abdul Watkins      M    19-08-1972   47  26-12-2017
1801  19790   Chase Fernandez      M    20-03-1993   27  22-01-2018

      Tenure_in_org_in_months   GROSS  Net_Pay  Deduction  \
0                           7   74922    71494       3428
1                           6   44375    39971       4404
2                           8   82263    77705       4558
3                           6   44375    40164       4211
4                          25  235405   143963      91442
...                       ...     ...      ...        ...
1797                       34   88934    88734        200
1798                       33  133224   133024        200
1799                       15   72547    71246       1301
1800                       30  227176   220778       6398
1801                       29  114641   114441        200

      Deduction_percentage                              Designation  \
0                     4.58    Product Operations Analyst.Associate.
1                     9.92  Platform Operations Engineer.Associate.
2                     5.54  Platform Operations Engineer.Associate.
3                     9.49  Platform Operations Engineer.Associate.
4                    38.84                Software Engineer.Senior.
...                    ...                                     ...
1797                  0.22           Technical Solutions Engineer..
1798                  0.15                   Software Engineer II..
1799                  1.79            Business Operations Analyst..
1800                  2.82              Manager..Account Management
1801                  0.17                           Order Analyst..

                                  Department
0                  IT Product Management & Ops
1                          Platform Operations
2                          Platform Operations
3                          Platform Operations
4                Enterprise Access Engineering
...                                        ...
1797                         AmaTec - EMEA TSE
1798             GSS EPIC Engineering (HC COGS)
1799                    Marketing - Operations
1800                             Americas- AMG
1801  Finance - Customer Revenue Operations G&A

[1802 rows x 13 columns]
```

In [4]:

```
#How Many Empolye Working
print(Emp.count()['EmpID'])
```

```
1802
```

In [21]:

```
#Select only the Age column.
print(Emp['Age'])
```

```
0       25
1       26
2       24
3       26
4       29
        ..
1797    29
1798    27
1799    29
1800    47
1801    27
Name: Age, Length: 1802, dtype: int64
```

In [27]:

```
#Filter Employes Who's age is more than 25


y=Emp.loc[Emp['Age']>25]
print(y)
```

```
      EmpID              Name Gender Date_of_Birth  Age   Join_Date  \
1     19944    Kristin Werner      F    23-06-1994   26  13-01-2020
3     20058      Boris Gibson      M    29-09-1993   26  13-01-2020
4     20332         Leif Mack      M    01-05-1991   29  04-06-2018
5     20849    Fredric Peters      M    29-04-1988   32  07-11-2019
6     20852     Sherry Bowman      F    17-11-1992   27  13-01-2020
...     ...               ...    ...           ...  ...         ...
1797  18835  Darius Wilkerson      M    14-01-1991   29  21-08-2017
1798  19066     Erick Ballard      M    29-08-1992   27  25-09-2017
1799  21644   Lawerence Downs      M    05-07-1991   29  01-04-2019
1800  19673     Abdul Watkins      M    19-08-1972   47  26-12-2017
1801  19790   Chase Fernandez      M    20-03-1993   27  22-01-2018

      Tenure_in_org_in_months   GROSS  Net_Pay  Deduction  \
1                           6   44375    39971       4404
3                           6   44375    40164       4211
4                          25  235405   143963      91442
5                           8   87300    57351      29949
6                           6   44375    39939       4436
...                       ...     ...      ...        ...
1797                       34   88934    88734        200
1798                       33  133224   133024        200
1799                       15   72547    71246       1301
1800                       30  227176   220778       6398
1801                       29  114641   114441        200

      Deduction_percentage                              Designation  \
1                     9.92  Platform Operations Engineer.Associate.
3                     9.49  Platform Operations Engineer.Associate.
4                    38.84               Software Engineer.Senior.
5                    34.31                  LMS Administrator II..
6                    10.00  Platform Operations Engineer.Associate.
...                    ...                                     ...
1797                  0.22           Technical Solutions Engineer..
1798                  0.15                   Software Engineer II..
1799                  1.79            Business Operations Analyst..
1800                  2.82              Manager..Account Management
1801                  0.17                          Order Analyst..

                                  Department
1                         Platform Operations
3                         Platform Operations
4                 Enterprise Access Engineering
5                        Learning & Development
6                         Platform Operations
...                                       ...
1797                          AmaTec - EMEA TSE
1798              GSS EPIC Engineering (HC COGS)
1799                     Marketing - Operations
1800                               Americas- AMG
1801  Finance - Customer Revenue Operations G&A
```

```
[1610 rows x 13 columns]
```

In [30]:

```
# It shows True or False weather above 25 age exist or not in data
print(Emp['Age']>=25)
```

```
0        True
1        True
2       False
3        True
4        True
        ...
1797     True
1798     True
1799     True
1800     True
1801     True
Name: Age, Length: 1802, dtype: bool
```

In [10]:

```
# This describes the basic stat behind the dataset used
print(Emp.describe())
```

```
              EmpID          Age  Tenure_in_org_in_months         GROSS  \
count   1802.000000  1802.000000              1802.000000  1.802000e+03
mean   15653.334628    31.516648                48.376804  1.758266e+05
std     5716.864347     5.225084                35.639826  1.055131e+05
min     1934.000000    21.000000                 6.000000  2.560000e+02
25%    10873.750000    28.000000                23.000000  1.042170e+05
50%    17158.000000    31.000000                39.000000  1.503300e+05
75%    20434.750000    35.000000                66.000000  2.211358e+05
max    23062.000000    56.000000               190.000000  1.024965e+06

             Net_Pay      Deduction  Deduction_percentage
count    1802.000000    1802.000000           1802.000000
mean   122454.117092   53372.527747             25.643940
std     58592.231163   57094.380849             13.261775
min       256.000000       0.000000              0.000000
25%     84099.750000   16333.500000             14.892500
50%    109753.000000   35534.500000             24.410000
75%    145898.000000   72298.500000             34.742500
max    660810.000000  775212.000000             86.400000
```

In [11]:

```
#What is the type of the columns?
print(type(Emp))
```

```
<class 'pandas.core.frame.DataFrame'>
```

In [12]:

```
#What are columns are present in the data set
print(Emp.columns)
```

```
Index(['EmpID', 'Name', 'Gender', 'Date_of_Birth', 'Age', 'Join_Date',
       'Tenure_in_org_in_months', 'GROSS', 'Net_Pay', 'Deduction',
       'Deduction_percentage', 'Designation', 'Department'],
      dtype='object')
```

In [13]:

```
#It shows details roles and salary in organization
print(Emp[['Name','GROSS','Department','Designation']])
```

```
                 Name   GROSS                                Department  \
0         Keven Norman   74922               IT Product Management & Ops
1       Kristin Werner   44375                       Platform Operations
2         Avery Barber   82263                       Platform Operations
3         Boris Gibson   44375                       Platform Operations
4            Leif Mack  235405             Enterprise Access Engineering
...                ...     ...                                       ...
1797  Darius Wilkerson   88934                       AmaTec - EMEA TSE
1798     Erick Ballard  133224            GSS EPIC Engineering (HC COGS)
1799   Lawerence Downs   72547                    Marketing - Operations
1800     Abdul Watkins  227176                          Americas- AMG
1801   Chase Fernandez  114641  Finance - Customer Revenue Operations G&A

                                Designation
0       Product Operations Analyst.Associate.
1     Platform Operations Engineer.Associate.
2     Platform Operations Engineer.Associate.
3     Platform Operations Engineer.Associate.
4                   Software Engineer.Senior.
...                                       ...
1797          Technical Solutions Engineer..
1798                  Software Engineer II..
1799           Business Operations Analyst..
1800             Manager..Account Management
1801                         Order Analyst..

[1802 rows x 4 columns]
```

In [14]:

```
# Total Count of Male and Female
print(Emp['Gender'].value_counts())
```

```
M    1303
F     499
Name: Gender, dtype: int64
```

In [15]:

```
#How many Employes work in each department
print(Emp['Department'].value_counts())
```

```
Enterprise Applications                       89
Finance - Customer Revenue Operations G&A     85
IT DevOps                                     84
Platform Operations                           74
Corporate Systems                             71
                                              ..
Media & Carrier Specialists                    1
APJ Channel Pre-Sales                          1
Training - Sales Enablement                    1
Media Global Industry Marketing                1
Syscomm                                        1
Name: Department, Length: 154, dtype: int64
```

In [16]:

```
#On which date how many employe joined
print(Emp['Join_Date'].value_counts())
```

```
08-07-2019    42
02-07-2018    37
04-07-2016    27
03-07-2017    21
20-06-2016    17
              ..
17-10-2006     1
13-07-2009     1
27-06-2008     1
28-06-2010     1
09-05-2013     1
Name: Join_Date, Length: 705, dtype: int64
```

In [17]:

```
#Highest Salary and Lowest Salary
print("Highest Salary:",Emp['Net_Pay'].max())
print("Lowest Salary:",Emp['Net_Pay'].min())
```

```
Highest Salary: 660810
Lowest Salary: 256
```

In [18]:

```
#How Many Department exist in the company
y = Emp['Department'].drop_duplicates()
#print(y)
print(y.value_counts())
```

```
IT Product Management & Ops          1
GSS EPIC CIAM                        1
APJ PS - SA/IN                       1
Service Delivery Americas Media      1
Amatec - BOCC/EMM                    1
                                    ..
Global Channel - Programs Support    1
Technology                           1
Service Delivery EMEA                1
LUNA/OPEN Engineering                1
Syscomm                              1
Name: Department, Length: 154, dtype: int64
```

In [19]:

```
#How Many Designation exist in the company
z = Emp['Designation'].drop_duplicates()
#print(z)
print(z.count())
```

```
390
```

In [7]:

```
f = Emp['Join_Date'].sort_values()
print(f)
```

```
774     01-01-2019
372     01-02-2010
806     01-02-2016
1242    01-02-2016
475     01-02-2018
           ...
21      31-10-2019
22      31-10-2019
23      31-10-2019
24      31-10-2019
1730    31-12-2018
Name: Join_Date, Length: 1802, dtype: object
```

In [ ]:

```
y=Emp.loc[Emp['Age']>25]
print(y)
```

In [ ]: