DataLink -: https://drive.google.com/file/d/1MRpw9qRtUewILQyTOvNKuZbkN3R45dGY/view?usp=sharing (https://drive.google.com/file/d/1MRpw9qRtUewILQyTOvNKuZbkN3R45dGY/view?usp=sharing)

Exploratory Data Analysis

- What is the avg age of the employees
- What is the Median Age of the employees in the organization
- what is the avg age of the employees working in Accounts department
- What is the avg age are of the employees working in Hr department
- What is the max & min age of the employees in the orangization
- What is the avg age of the employees working in SDE department in USA
- What is the avg age of the employees working in Admin department in Canada
- What is the name of most senior employee with respect to age?
- What is the name of the most senior employee with respect to working duration?
- What is the salary of the employee whose name is Maria Williams
- What is the salary of the employee whose Virginia Hogan and age is 30
- What are the different departments, this organization has?
- What are total frequency of the employees with respect to each department?
- What percentage of employee work in SDE deparment
- What is the average rating of an employee with respect to each department
- what is the average salary of an employee with respect to each country
- What is the average salary of SDE profile with respect to each country
- What is the average salary of the employee who take taken the loan
- What are the top 5 domain name used in Employee email id's

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
df = pd.read_csv('Employees.csv')
df.head()
```

Out[2]:

	Id	Name	Age	Base Pay	OverTime Pay	Benefits	Total Pay benefits	Loan	Department	Duration	Rating
0	1	Marcia Stephens	22	120000	60000.0	15000.0	195000.0	No	SDE	7	8.2
1	2	Karen Lopez	22	120000	60000.0	15000.0	195000.0	No	SDE-T	6	9.3
2	3	Heather Reilly	22	60000	30000.0	7500.0	97500.0	No	SDE	5	6.6
3	4	Cindy Leblanc	32	70000	35000.0	8750.0	113750.0	No	Admin	1	8.0
4	5	Colin Richardson	22	120000	60000.0	15000.0	195000.0	Yes	SDE	3	5.7

In [3]:
```
df.shape
```

Out[3]:
```
(1000, 14)
```

1. What is the avg age of the employees

In [4]:
```
df['Age'].mean()
```

Out[4]:
```
31.836
```

2. What is the Median Age of the employees in the organization

In [5]:
```
df['Age'].median()
```

Out[5]:
```
30.0
```

3. what is the avg age of the employees working in Accounts department

In [6]:
```
df.groupby('Department').get_group('Accounts')['Age'].mean()
```

Out[6]:
```
31.23076923076923
```

4. What is the avg age are of the employees working in Hr department

In [7]:
```
df.groupby('Department').get_group('Hr')['Age'].mean()
```

Out[7]:
```
31.28846153846154
```

5. What is the max & min age of the employees in the orangization

In [8]:
```
print('minimum age=',df.Age.min())
print('maximum age=',df.Age.max())
```

```
minimum age= 22
maximum age= 55
```

6. What is the avg age of the employees working in SDE department in USA

In [9]:
```
df.groupby('Department').size()
```

Out[9]:
```
Department
Accounts      65
Admin         53
Hr            52
Logistics    158
SDE          497
SDE-T        175
dtype: int64
```

In [10]:
```
df.groupby('Department').get_group('SDE')['Age'].mean()
```

Out[10]:
```
31.613682092555333
```

7. What is the avg age of the employees working in Admin department in Canada

In [11]:
```
df.groupby(['Department','Country']).size()
```

Out[11]:
```
Department  Country
Accounts    Canada             17
            Russia             14
            USA                21
            United Kingdom     13
Admin       Canada             11
            Russia             17
            USA                15
            United Kingdom     10
Hr          Canada             12
            Russia             17
            USA                13
            United Kingdom     10
Logistics   Canada             37
            Russia             39
            USA                42
            United Kingdom     40
SDE         Canada            123
            Russia            119
            USA               132
            United Kingdom    123
SDE-T       Canada             49
            Russia             39
            USA                51
            United Kingdom     36
dtype: int64
```

In [12]:
```
df.groupby(['Department','Country']).get_group(('Admin','Canada'))['Age'].mean()
```

Out[12]:
```
32.18181818181818
```

8. What is the name of most senior employee with respect to age?

In [24]:
```
df.groupby('Age')['Name'].size()
```

Out[24]:
```
Age
22    333
27    107
30    105
32     99
37    109
40     74
43     67
48     28
50     41
55     37
Name: Name, dtype: int64
```

In [88]:
```
df.groupby('Age')['Name'].get_group(df['Age'].max())
```

Out[88]:
```
15           Megan Chambers
35              Erik Sutton
42           Michael Cooper
97             Frank Taylor
124         Harold Johnston
132           Audrey Barber
173          Kayla Franklin
186          Crystal Bryant
204          Denise Manning
213              Mark Hardy
239       Dr. Ryan Mitchell
290               Dawn Vega
320          Benjamin Brown
353    Mr. Nathan Underwood
387            Nathan Brown
419           Jessica Miles
457             Jamie Moody
466              Megan Rowe
468           Ashley Lopez
490           Danny Franco
527          Larry Shepard
561              John Smith
678           Aaron Douglas
688           Joseph Spears
696          Anthony Barnes
711           Hannah Forbes
715             Lauren York
729             Kara Murphy
741             Robert Ball
759         Charlotte Ramos
772             Austin Cook
889            Wendy Norton
922            Taylor Garza
940             Karen White
989             Karen Lucas
990            Patrick Lane
991       Elizabeth Wheeler
Name: Name, dtype: object
```

In [18]:
```
# filterr = df['Name'].where(df['Age']==df['Age'].max())
# filterr
# df[df['Name']==filterr]
```

. . .

9. What is the name of the most senior employee with respect to working duration?

In [28]:
```
df.groupby('Duration')['Name'].size()
```

Out[28]:
```
Duration
1    138
2    143
3    153
4    140
5    153
6    122
7    151
Name: Name, dtype: int64
```

In [29]:
```
df['Duration'].max()
```

Out[29]:
```
7
```

In [34]:
```
duration_sr = df.groupby('Duration')['Name'].get_group(df['Duration'].max())
duration_sr
```

Out[34]:
```
0      Marcia Stephens
9          Steven Lynn
16        Steven Hicks
28       James Wilkins
32         Troy Thomas
              ...
964     Joseph Johnson
973      Rachel Harris
975      Melissa Leach
989        Karen Lucas
998         Cindy Knox
Name: Name, Length: 151, dtype: object
```

10. What is the salary of the employee whose name is Maria Williams

In [37]:
```
df['Total Pay benefits'].loc[df['Name']=='Maria Williams']
```

Out[37]:
```
808    156000.0
Name: Total Pay benefits, dtype: float64
```

11. What is the salary of the employee whose Virginia Hogan and age is 30

In [38]:
```
df['Total Pay benefits'].loc[(df['Name']=='Virginia Hogan') & (df['Age']==30)]
```

Out[38]:
```
406    195000.0
Name: Total Pay benefits, dtype: float64
```

12. What are the different departments, this organization has?

In [40]:
```
df['Department'].unique()
```

Out[40]:
```
array(['SDE', 'SDE-T', 'Admin', 'Logistics', 'Hr', 'Accounts'],
      dtype=object)
```

13. What are total frequency of the employees with respect to each department?

In [41]:
```
df['Department'].value_counts()
```

Out[41]:
```
SDE          497
SDE-T        175
Logistics    158
Accounts      65
Admin         53
Hr            52
Name: Department, dtype: int64
```

14. What percentage of employee work in SDE deparment

In [43]:
```
df['Department'].value_counts()/len(df['Department'])*100
```

Out[43]:
```
SDE          49.7
SDE-T        17.5
Logistics    15.8
Accounts      6.5
Admin         5.3
Hr            5.2
Name: Department, dtype: float64
```

15. What is the average rating of an employee with respect to each department

In [44]:
```
df.groupby('Department').Rating.mean()
```

Out[44]:
```
Department
Accounts     7.750769
Admin        8.060377
Hr           8.150000
Logistics    7.955063
SDE          7.894769
SDE-T        7.652571
Name: Rating, dtype: float64
```

16. what is the average salary of an employee with respect to each country

In [45]:
```
df.groupby('Country')['Total Pay benefits'].mean()
```

Out[45]:
```
Country
Canada             152776.104418
Russia             151662.244898
USA                152999.087591
United Kingdom     154487.068966
Name: Total Pay benefits, dtype: float64
```

17. What is the average salary of SDE profile with respect to each country

In [87]:
```
sde = df.groupby(['Department']).get_group('SDE')
sde.groupby('Country')['Total Pay benefits'].mean()
```

Out[87]:
```
Country
Canada             150953.252033
Russia             150455.882353
USA                151765.151515
United Kingdom     155656.504065
Name: Total Pay benefits, dtype: float64
```

19. What is the average salary of the employee who take taken the loan

In [83]:
```
df.groupby('Loan').get_group('Yes')['Total Pay benefits'].mean()
```

Out[83]:
```
151902.5787965616
```

In [82]:
```
df_loan = df.loc[df['Loan']=='Yes']
df_loan['Total Pay benefits'].mean()
```

Out[82]:
```
151902.5787965616
```

20. What are the top 5 domain name used in Employee email id's

In [74]:
```
def domain(x):
    if x:
        if x:= x.split('@')[-1]:
                return x.split('.')[0]
```

In [75]:
```
df['domain'] = df['Email'].map(domain)
```

In [79]:
```
df['domain'].value_counts().head()
```

Out[79]:
```
outlook    213
hotmail    205
yahoo      205
gmail      198
reddit     179
Name: domain, dtype: int64
```

In [ ]:
```

```