In [1]:

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

In [2]:

```
stack = pd.read_csv('stackoverflow-1.csv')
stack.head()
```

Out[2]:

	RespondentID	Country	Salary	YearsCodedJob	OpenSource	Hobby	CompanySizeNumbe
0	3	United Kingdom	113750.0	20	1	1	10000.
1	15	United Kingdom	100000.0	20	0	1	5000.
2	18	United States	130000.0	20	1	1	1000.
3	19	United States	82500.0	3	0	1	10000.
4	26	United States	175000.0	16	0	1	10000.

5 rows × 23 columns

Q1 :

In [3]:

```
stack.shape
```

Out[3]:

```
(6991, 23)
```

Q2 :

In [5]:

```
stack.isnull().sum()
```

Out[5]:

```
RespondentID                                0
Country                                     0
Salary                                      0
YearsCodedJob                               0
OpenSource                                  0
Hobby                                       0
CompanySizeNumber                          31
Remote                                      0
CareerSatisfaction                          0
Datascientist                               0
Database administrator                      0
Desktop applications developer              0
Developer with stats/math background        0
DevOps                                      0
Embedded developer                          0
Graphic designer                            0
Graphics programming                        0
Machine learning specialist                 0
Mobile developer                            0
Quality assurance engineer                  0
Systems administrator                       0
Web developer                               0
                                         6991
dtype: int64
```

Q3 :

In [6]:

```
stack['Country'].nunique()
```

Out[6]:

```
5
```

Q4 :

In [9]:

```
stack.Country.unique()
```

Out[9]:

```
array(['United Kingdom', 'United States', 'Germany', 'India', 'Canada'],
      dtype=object)
```

In [10]:

```
stack.Datascientist.unique()
```

Out[10]:

```
array([0, 1], dtype=int64)
```

In [11]:

```
stack.query(" Country =='India' & Datascientist==1").shape[0]
```

Out[11]:

```
43
```

Q5 :

In [12]:

```
stack.Remote.unique()
```

Out[12]:

```
array(['Not remote', 'Remote'], dtype=object)
```

In [13]:

```
stack.query(" Remote =='Remote' ").shape[0]
```

Out[13]:

```
718
```

Q6 :

In [14]:

```
stack.query("Country =='Germany' & YearsCodedJob >10 ").shape[0]
```

Out[14]:

```
181
```

Q7 :

In [15]:

```
stack.groupby('Country').agg({'Salary':np.sum}).sort_values('Salary',ascending=False)
```

Out[15]:

Country	Salary
United States	3.427196e+08
United Kingdom	6.853285e+07
Germany	4.978553e+07
Canada	3.604215e+07
India	7.697411e+06

Q8 :

In [23]:

```
stack.groupby(['Country','Remote'],as_index=False).agg({'RespondentID':np.size}).query(" Re
```

Out[23]:

	Country	Remote	RespondentID
1	Canada	Remote	43
3	Germany	Remote	49
5	India	Remote	66
7	United Kingdom	Remote	85
9	United States	Remote	475

Q9 :

In [25]:

```
round(stack[ stack['Mobile developer']==1 ].shape[0]/stack.shape[0]*100)
```

Out[25]:

19

Q10 :

In [26]:

```
round(stack[ stack['Web developer']==1 ].shape[0]/stack.shape[0]*100)
```

Out[26]:

72

Q11 :

In [27]:

```
stack[ (stack['Mobile developer']==1) & (stack['Web developer']==1) & (stack['Remote']=='Re
```

Out[27]:

115

Q12 :

In [31]:

```
stack.groupby('Country').agg({'Salary': np.max, 'YearsCodedJob':np.mean}).sort_values('Sala
```

Out[31]:

Country	Salary	YearsCodedJob
United States	197000.00	8.40
United Kingdom	193750.00	7.32
India	163134.80	3.70
Canada	151515.15	7.52
Germany	140000.00	7.06

Q13 :

In [32]:

```
def label(x):
    if x < 20 :
        return 'Low'
    elif 20<= x <=1000 :
        return 'Medium'
    else :
        return 'High'
```

In [33]:

```
stack['Company_label'] = stack['CompanySizeNumber'].apply(label)
```

In [34]:

```
stack.groupby('Company_label').size()
```

Out[34]:

```
Company_label
High      1530
Low       1156
Medium    4305
dtype: int64
```

Q14 :

In [35]:

```
stack[ (stack['Datascientist']==1) | (stack['Developer with stats/math background']==1) | (
```

Out[35]:

```
1129
```

Q15 :

In [36]:

```
# data["col"] =data["col"]._____(10000)
# Ans : fillna
```