

IT350 - Data Analytics

Lab Assignment 1

Name-Bhagyashri Bhamare.

Roll No-181IT111

Dataset-1(Employment in different industries)

Data set contains positions open for the jobs in different industries. dataset contains job id number, salary, and industries. data set is available in the

```
In [98]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
mobility = pd.read_csv("4.csv")
mobility.head(7)
```

Out[98]:

	Series_reference	Salary	STATUS	Subject	Group	Series_title_1	Series_title_2	Series_title_3
0	BDCQ.SEA1AA	80078.0	F	Business Data Collection - BDC	Industry by employment variable	Filled jobs	Agriculture, Forestry and Fishing	Actual
1	BDCQ.SEA1AA	78324.0	F	Business Data Collection - BDC	Industry by employment variable	Filled jobs	Agriculture, Forestry and Fishing	Actual
2	BDCQ.SEA1AA	85850.0	F	Business Data Collection - BDC	Industry by employment variable	Filled jobs	Agriculture, Forestry and Fishing	Actual
3	BDCQ.SEA1AA	90743.0	F	Business Data Collection - BDC	Industry by employment variable	Filled jobs	Agriculture, Forestry and Fishing	Actual
4	BDCQ.SEA1AA	81780.0	F	Business Data Collection - BDC	Industry by employment variable	Filled jobs	Agriculture, Forestry and Fishing	Actual
5	BDCQ.SEA1AA	79261.0	F	Business Data Collection - BDC	Industry by employment variable	Filled jobs	Agriculture, Forestry and Fishing	Actual
6	BDCQ.SEA1AA	87793.0	F	Business Data Collection - BDC	Industry by employment variable	Filled jobs	Agriculture, Forestry and Fishing	Actual

(Fig-dataset-1)

Frequency Distribution

1)GROUPS

```
In [65]: country = mobility.groupby("Group")
country["Group"].count()
```

```
Out[65]: Group
Age by employment variable      836
Industry by employment variable 1520
Region by employment variable   1824
Sex by employment variable      152
Territorial authority by employment variable 7669
Name: Group, dtype: int64
```

Fig-2(frequency distribution of group)

2)Series_reference

```
In [66]: country = mobility.groupby("Series_reference")
country["Series_reference"].count()

Out[66]: Series_reference
BDCQ.SEA1AA      38
BDCQ.SEA1BA      38
BDCQ.SEA1CA      38
BDCQ.SEA1DA      38
BDCQ.SEA1EA      38
..
BDCQ.SEE3073A    38
BDCQ.SEE3074A    38
BDCQ.SEE3075A    38
BDCQ.SEE3076A    38
BDCQ.SEE3999A    29
Name: Series_reference, Length: 318, dtype: int64
```

Fig-3(frequency distribution of Series_reference)

3)Status

```
gender= mobility.groupby("STATUS")
gender["STATUS"].count()

STATUS
C      31
F    11970
Name: STATUS, dtype: int64
```

Fig-3(frequency distribution of Status)

4)

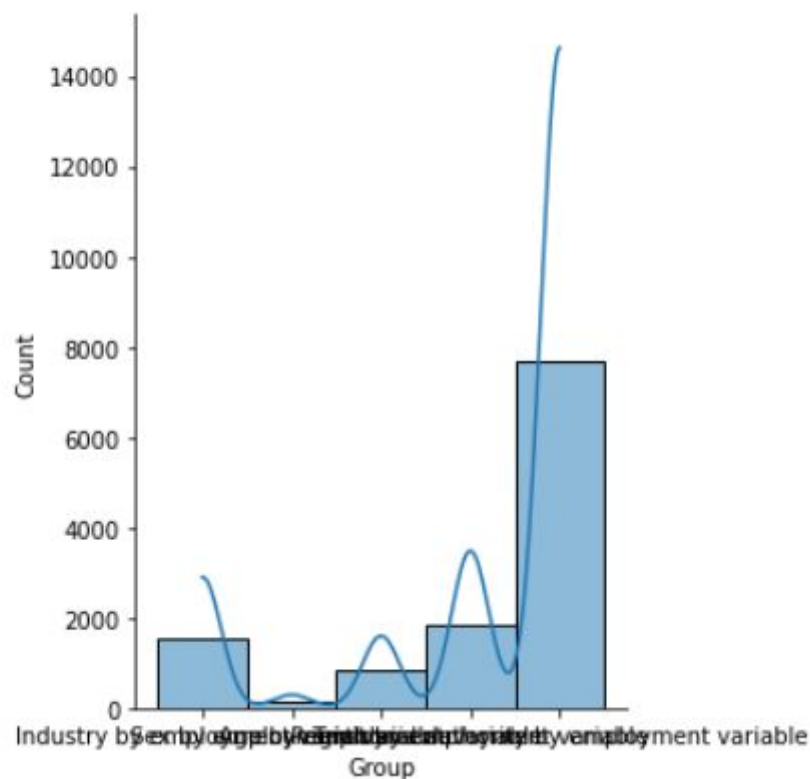
```
In [9]: country = mobility.groupby("Industry")
country["Industry"].count()
```

```
Out[9]: Industry
15-19          76
20-24          76
25-29          76
30-34          76
35-39          76
...
Westland District    114
Whakatane District  114
Whanganui District   114
Whangarei District   114
Wholesale Trade      76
Name: Industry, Length: 116, dtype: int64
```

1)Histogram of number of employ in industry

```
In [68]: sns.displot(mobility.Group, bins = 15 , kde = True)
```

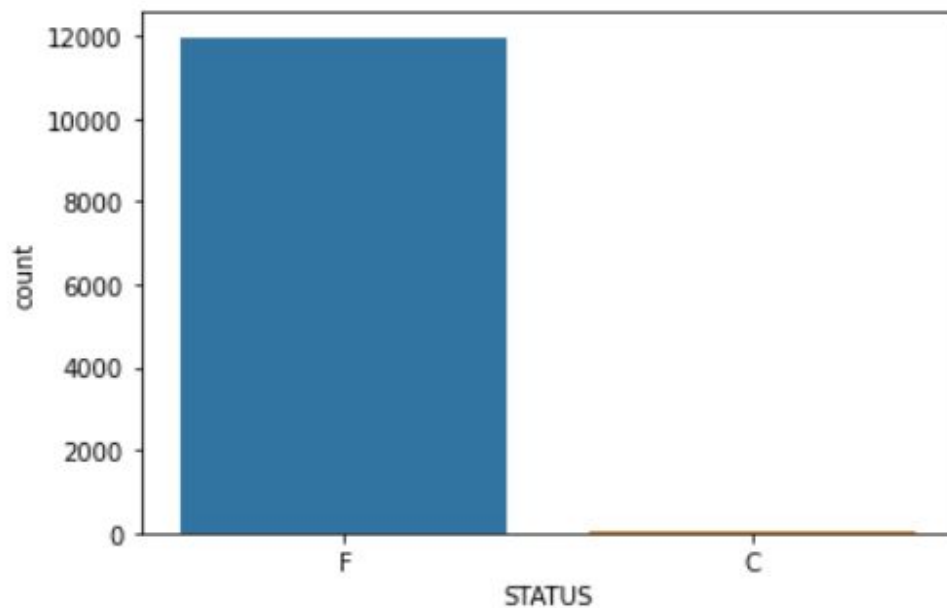
```
Out[68]: <seaborn.axisgrid.FacetGrid at 0x19027a65220>
```



2)Histogram of status

```
sns.countplot(x=mobility.STATUS,data=mobility)
```

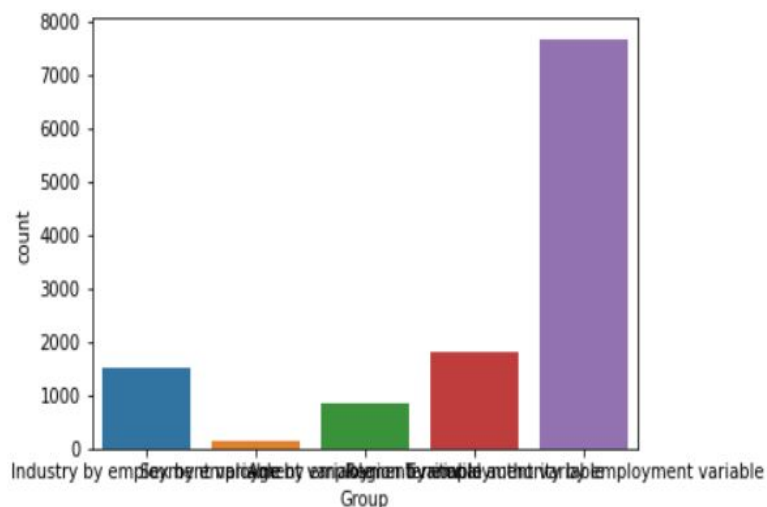
```
<AxesSubplot:xlabel='STATUS', ylabel='count'>
```



3)Histogram of Group

```
In [27]: sns.countplot(x=mobility.Group,data=mobility)
```

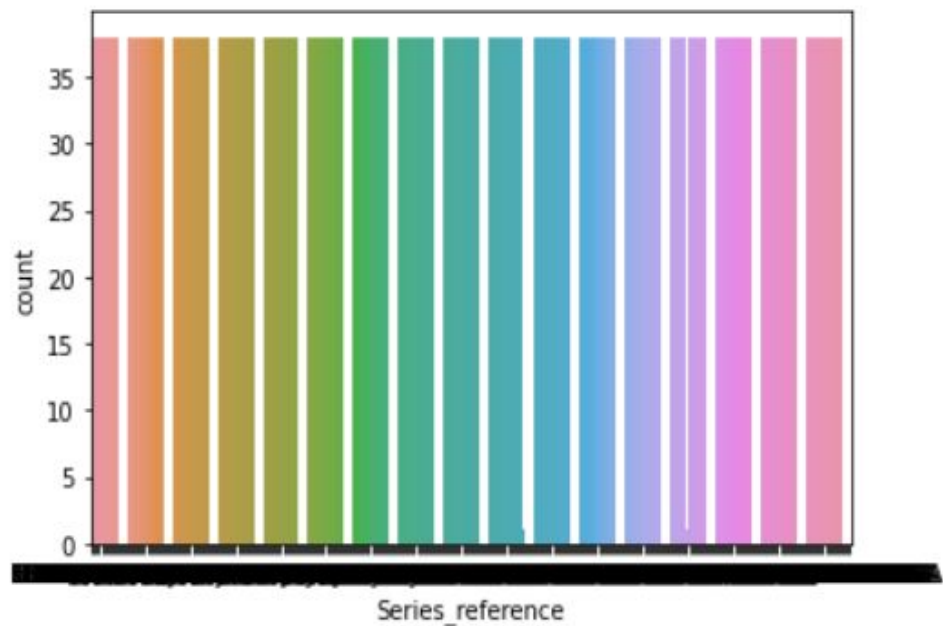
```
Out[27]: <AxesSubplot:xlabel='Group', ylabel='count'>
```



4)Histogram of job id

```
In [11]: sns.countplot(x=mobility.Series_reference,data=mobility)
```

```
Out[11]: <AxesSubplot:xlabel='Series_reference', ylabel='count'>
```



Statistics analysis of the data

```
In [111]: mobility.describe()
```

```
Out[111]:
```

	Salary
count	1.197000e+04
mean	4.909224e+04
std	1.557601e+05
min	-3.829897e+00
25%	6.904644e+02
50%	6.995000e+03
75%	2.610650e+04
max	2.134716e+06

Analysis-

From the above data we can see that there are 11,970 employee in different industries. From frequency distribution of the 'status' we can see 12,001 jobs are available in which 11,970 are filled. We can see that most of the jobs are available in Agriculture, Forestry and Fishing. Average salary of all the employees is 49,092.24. More than 50% vacancy are paying more than 6,995. Max salary of employee is 21,34,716 and min salary is 31,190. Employee are also divided into Age by employment variable, Industry by employment Region by employment variable, Sex by employment variable, and Territorial authority by employment variable. Most of employee belongs to Industry by employment variable. Most of the job openings are available in Industry groups. Most of the jobs are filled in the Fishing industry. Only jobs are available in areas outside territorial authority because most people prefer jobs in their locality. Professional, Scientific and Technical Services field is paying max salary to employee and Electricity, Gas, Water and Waste Services field is paying min salary to employee. In general Employee belongs to BDCQ.SEA1 series are getting more pay than BDC Q.SEE. Average pay of a employee belongs to BDCQ.SEA1 series is 61500.54 and average pay of the employee belongs to the BDC Q.SEE series is getting 35000.62. From histogram we can see that their equal number of jobs are available every field. The above data set helps to find the number of job application are open in the particular field also helps to find the Average pay in the employee in the particular sector

