

**ANL252 (Online)**

**Python for Data Analytics**

# **Tutor-Marked Assignment**

**July 2022 Presentation**

**Submitted by:**

|  |  |
| --- | --- |
| **Name** | **PI No.** |
| **Chan Teck Fong** | **W2070448** |
| Click to enter NAME. | Click to enter PI NO. |
| Click to enter NAME. | Click to enter PI NO. |
| Click to enter NAME. | Click to enter PI NO. |

**Tutorial Group: ­­­­­­­­­­­­ TV 09**

**Instructor’s Name: Dr. Munish Kumar**

**Submission Date: 12/08/2022**

**Question 1**

**a)** The given dataset was collected from 250 staffs from an organization. The dimension of dataset comprises of 15 variables in columns and 250 rows, type of variable generally divided into quantitative/ numeric data and categorical/ qualitative data (Bevans, 2022). Let’s review the relationship or correlational of variables Unit, Gender and Salary from dataset using MS Excel.

i. Chart 1: Staff’s Gender by Business Unit

Using pivot table to summarize variables per Table-1 and plot stack bar chart per Figure-1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Unit** | **Gender** | |  |
| **Unit** | **F** | **M** | **Grand Total** |
| Manufacturing | 105 | 64 | 169 |
| IT | 18 | 23 | 41 |
| Sales | 11 | 11 | 22 |
| Engineering | 5 | 5 | 10 |
| Admin | 4 | 3 | 7 |
| C-Level | 1 |  | 1 |
| **Grand Total** | **144** | **106** | **250** |

Table-1: Summarize table (Unit vs Gender)

Figure-1: Staff’s Gender by Business Unit

Refer to Figure-1, there are total of 6 sub-variables “Unit” which are Manufacturing, IT, Sales, Engineering, Admin, C-Level. Main workforce of the organization came from Manufacturing business unit, follow by supporting roles such as IT, subsequently Sales, Engineering, Admin and lastly from C-level. From Table-1, it revealed that 169 staffs (67.6%) are engaged in Manufacturing business unit, followed by 41 staffs (16.4%) in IT, 22 staffs (8.8%) in Sales.

Another observation made was female-male workforce ratio. There are 144 female staffs (57.6%) and 106 male staffs (42.4%) in total workforce of the organization. For Manufacturing business unit, 105 female staffs be the dominant in comparison with 64 male staff.

ii. Chart 2: Average of Salary by Business Unit

Using pivot table to summarize variables of “Salary”, average the salary according to the “Unit” (refer to Table-2) and plot stack bar chart (refer to Figure-2).

|  |  |
| --- | --- |
| **Unit** | **Average of Salary** |
| C-Level | 250000 |
| Engineering | 95867 |
| IT | 95551 |
| Admin | 77315 |
| Sales | 71427 |
| Manufacturing | 59413 |
| **Grand Total** | **69119** |

Table-2: Summarize table (Unit vs Average Salary)

Figure-2: Average Salary by Business Unit

Refer to Figure-2, there are total of 6 sub-variables “Unit” which are C-Level, Engineering, IT, Admin, Sales, Manufacturing. Staff from “C-Level drawn the highest average salary at $250,000, follow by Engineering $95,867, IT $95,551, Admin $77,315, Sales $71,427 and lastly Manufacturing $59,413.

Another observation made was the salary drawn for staffs in the organization. The average salary of staffs in the organisation at the range from $59,413 (lowest average) to $250,000 (highest average). Salaries vary drastically between different business unit.

**b)** In order to generate summarised table and plot same chart of Staff’s Gender by Business Unit as per part a using Python, below please find the codes:

i. Python coding for Chart 1: Staff’s Gender by Business Unit

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv("C:\\Users\\Candice\\Desktop\\Python practice\\TMA\_Data.csv")

df1 = df.groupby(['Unit', 'Gender'])['Gender'].count().unstack().fillna(0)

## Create another column as total to sort stack bar in ascending order

df1['total'] = df1.sum(axis=1)

df1=df1.sort\_values('total', ascending=False)

print(df1) ## Summarised table for Staff’s Gender by Business Unit is produced

## To drop column total so that when plotting stack bar chart, the total figure will be excluded

df1.drop('total', inplace=True, axis=1)

df1.plot(kind='bar', stacked=True)

plt.title("Staff's Gender by Business Unit")

plt.xlabel("Business Unit")

plt.ylabel("No. of Staffs")

plt.xticks(rotation=0, ha='center')

## Staff’s Gender by Business Unit Chart is produced

plt.show()

ii. Python coding for Chart 2: Average Salary by Business Unit

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

pd.set\_option("display.max\_columns", None)

df = pd.read\_csv("C:\\Users\\Candice\\Desktop\\Python practice\\TMA\_Data.csv")

df1 = df.groupby(['Unit']) ['Salary'].sum(numeric\_only=False)

search\_product\_df = df.groupby(['Unit'])

## find the average of salary per business unit

def findavg(itemname):

selected\_items = search\_product\_df.get\_group(itemname)

unit\_count=selected\_items['Salary'].count()

unit\_total\_sal=selected\_items['Salary'].sum()

unit\_avg=round(unit\_total\_sal/unit\_count,2)

return unit\_avg

## Assign average value to dictionaries

data = {'C-Level':(findavg('C-Level')),'Engineering':(findavg('Engineering')), 'IT':(findavg('IT')),'Admin':(findavg('Admin ')),'Sales':(findavg('Sales')), 'Manufacturing':(findavg('Manufacturing'))}

units = list(data.keys())

print(units)

values = list(data.values())

print(values)

figure = plt.figure(figsize = (20,10))

<Figure size 1440x720 with 0 Axes>

from tabulate import tabulate

Table = pd.DataFrame({"Unit":['C-Level', 'Engineering', 'IT', 'Admin', 'Sales', 'Manufacturing'], "Average of Salary":[250000.0, 95867.0, 95550.76, 77315.29, 71426.59, 59412.91]})

Table

## Summarised table for Unit & Average of Salary is produced

plt.bar(units, values, color = 'blue', width = 0.5)

plt.xlabel("Business Unit")

plt.ylabel("Average salary")

plt.title("Average of Salary by Business Unit")

## Average of Salary by Business Unit Chart is produced

plt.show()

**c)** Python coding to find length of service of all staffs, statistical analysis of minimum, maximum and average length of service are per below:

import pandas as pd

pd.set\_option("display.max\_columns", None)

df = pd.read\_csv("C:\\Users\\Candice\\Desktop\\Python practice\\TMA\_Data.csv")

## Find out missing date at "LeftDate" and replace with 1-May-2022

df['LeftDate'] = df['LeftDate'].fillna('1-May-2022')

## Find out length of service and round to 1 decimal place

i=0

while i <(len(df.index)):

JoinDate=pd.to\_datetime((df.loc[i,"JoinDate"]),dayfirst=False)

LeftDate=pd.to\_datetime((df.loc[i,"LeftDate"]),dayfirst=False)

Duration=round((len(pd.date\_range(start=JoinDate,end=LeftDate,freq='D'))/365),1)

df.at[i,"length of service"]=Duration

i=i+1

print(df[['Staff', 'JoinDate', 'LeftDate', 'length of service']]) ## Length of service for all staffs is produced

## Analysis of employee's length of service (in years)

print("Minimum length of service is :\033[4m\033[1m",(df['length of service'].min()),"\033[0m years")

Minimum length of service is : **0.1**  years

print("Maximum length of service is :\033[4m\033[1m",(df['length of service'].max()),"\033[0m years")

Maximum length of service is : **16.3**  years

print("Average length of service is :\033[4m\033[1m",(round(df['length of service'].mean(),1)),"\033[0m years")

Average length of service is : **6.8**  years

**d)** Developed a simple interactive user input coding for checking the staff’s name availability using Python.

import pandas as pd

df = pd.read\_csv("C:\\Users\\Candice\\Desktop\\Python practice\\TMA\_Data.csv")

while True:

keyinname=input("Enter the full name of employee to enquire or type END to exit:\n")

if keyinname == "END" or keyinname =="end" or keyinname =="End": ## Request user to exit from the loop

print("\nThank you for using ME :)")

break

test=df[df['Staff'].str.match(keyinname)] ## to check the column of Staff for name matching

if test.empty:

print("\033[1m",keyinname,"\033[0m" , "cannot be found in the system.\n")

else:

if (pd.isnull(test['LeftDate'].iloc[0])) == True:

print(keyinname + " is still around.\n") ## if the name can be found from the Staff column

else:

print(keyinname + " is no more with us!\n") ## if the name cannot be found from the Staff column

### Testing Case result

Enter the full name of employee to enquire or type END to exit:

Adaline Deacon

Adaline Deacon is no more with us!

Enter the full name of employee to enquire or type END to exit:

Aaminah Mcmillan

Aaminah Mcmillan is still around.

Enter the full name of employee to enquire or type END to exit:

END

Thank you for using ME :)

**References**

Bevans, R. (2022, July 21). *Types of Variables in Research & Statistics.* https://www.scribbr.com/methodology/types-of-variables/