

**ANL252**

**Python for Data Analytics**

**Tutor-Marked Assignment**

**July 2022 Presentation**

**Name: Lee Pei Shi**

**SUSS PI No: Y2210991**

**ECA Submission Deadline: 14 August 2022, 2355hrs.**

**Question 1(a)**

|  |  |
| --- | --- |
| **Unit** | **Sum of Salary** |
| Admin | 541207 |
| C-Level | 250000 |
| Engineering | 958670 |
| IT | 3917581 |
| Manufacturing | 10040781 |
| Sales | 1571385 |
| **Table 1** | |

**Chart 1**

|  |  |
| --- | --- |
| **Unit** | **Average of Absence** |
| Admin | 7.857142857 |
| C-Level | 10 |
| Engineering | 9.4 |
| IT | 10.6097561 |
| Manufacturing | 10.06508876 |
| Sales | 12.86363636 |
| **Table 2** | |

**Chart 2**

The two charts represent the total salary and the average absence among different units.

Table 1 shows the sums of salary among the different units. From chart 1 it is clear that the ‘Manufacturing’ unit has the highest salary paid, with a high percentage of 58%. Followed by the ‘IT’ unit with 23%. Lastly, ‘C-Level’ accounts for 1% of the total salary.

Table 2 shows the average absence among different units. From chart 2, it can be seen that the average absence is roughly close with the highest percentage of 21% for the ‘Sales’ unit. 17% for ‘Manufacturing’ and ‘IT’ unit, 16% for ‘C-Level’ and ‘Engineering’ with just one per cent difference between the four. Lastly, the lowest average of absence from the ‘Admin’ unit.

**Question 1(b)**

# Import Libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from IPython.display import display

#Question 1(b)

df = pd.read\_csv('TMA\_Data.csv') #read data

#calculate total salary and avg satisfaction

T\_Salary = df.groupby("Unit")['Salary'].sum().reset\_index()

Avg\_Absence = df.groupby("Unit")['Absence'].mean().reset\_index()

#convert the DataFrame to a NumPy array

Salary = T\_Salary.loc[:,'Salary'].to\_numpy()

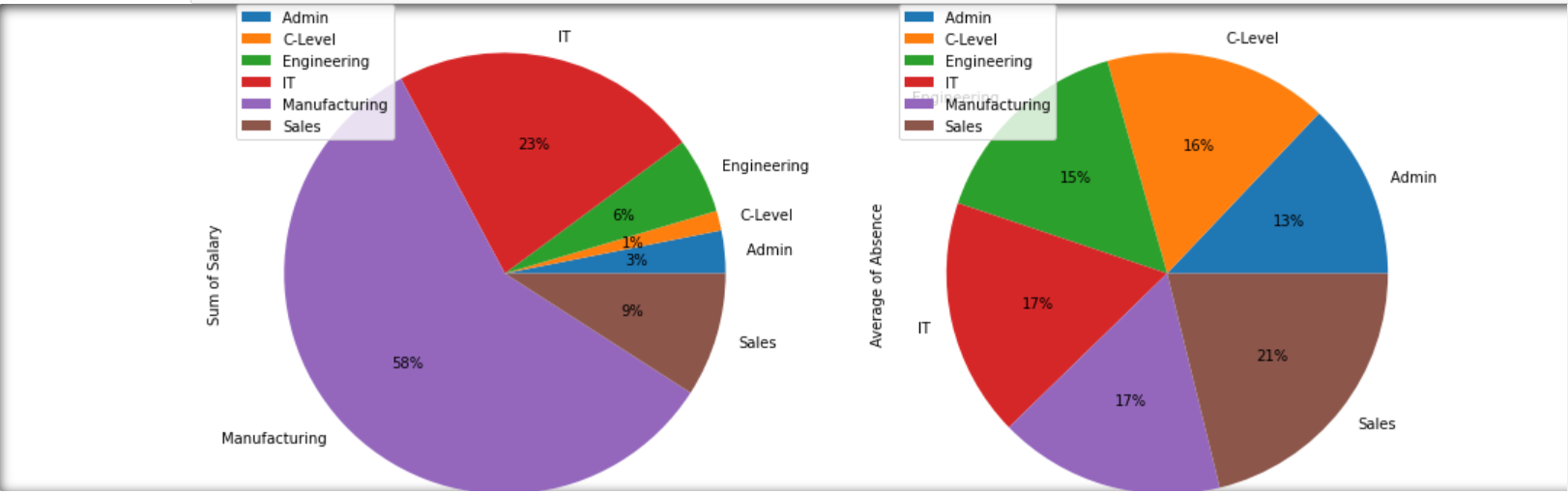
Absence = Avg\_Absence.loc[:,'Absence'].to\_numpy()

Unit = T\_Salary.loc[:,'Unit'].tolist()

df = pd.DataFrame({'Sum of Salary': Salary, 'Average of Absence': Absence}, index = (Unit))

plot = df.plot.pie(subplots=True, figsize=(15, 15),autopct='%.0f%%')

**Output**

****

**Question 1(c)**

#Question 1(c)

df = pd.read\_csv('TMA\_Data.csv')

df.fillna(value=pd.to\_datetime('2022-05-01'), inplace=True)

df[['JoinDate','LeftDate']] = df[['JoinDate','LeftDate']].apply(pd.to\_datetime)

df['ServiceLength'] = (df['LeftDate'] - df['JoinDate']).dt.days #days

df['ServiceLength'] = round((df['ServiceLength'] / 365),1) #convert to years, 1 decimal place

#compute min

DateDiff = df['ServiceLength']

print (min(DateDiff))

#compute max

DateDiff = df['ServiceLength']

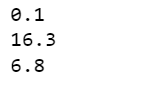
print (max(DateDiff))

#compute avg

DateDiff = round(np.mean(df['ServiceLength'], axis=0),1)

print (DateDiff)

**Output**



**Question 1(d)**

#Question 1(d)

df = pd.read\_csv('TMA\_Data.csv')

valid\_input = False

while valid\_input == False:

fullname = str(input("Please enter staff's full name or enter quit to exit:"))

if fullname == "quit":

break

else:

staff = df[df['Staff'] == fullname]

if staff.empty:

print ("This person is not an employee of this company")

else:

display(staff)

if staff.isnull().values.any() == True:

print ('This staff still employs in '+ staff['Unit'].to\_string(index=False))

else:

print ('This staff has left the company on '+ staff['LeftDate'].to\_string(index=False))

**Output**

