

**ANL252  
Python for Data Analytics**

**Tutor-Marked Assignment  
July 2022 Presentation**

**Submitted by:**

|  |  |
| --- | --- |
| **Name** | **PI Number** |
| Wong Hang Sin | Q2071045 |

**Tutorial Group :** T09

**Submission Date :** 14/08/22

**Question 1(a)**

**Chart 1:** Line graph showing sum of salary expenses for each unit  
 **Table 1:**

|  |  |
| --- | --- |
| **Unit** | **Sum of Salary** |
| Admin | 541207 |
| C-Level | 250000 |
| Engineering | 958670 |
| IT | 3917581 |
| Manufacturing | 10040781 |
| Sales | 1571385 |
| **Total Sum** | **17279624** |

From chart 1 and table 1, we find out that the organization has paid in total, $17,279,624 in salary to its staff. Specifically, the highest salary expenditure is for its manufacturing unit with a total of $10,040,781, which is approximately 58.11% of total salary paid. On the other hand, the lowest is $250,000 for the C-Level unit, approximately only 1.45% which is significantly lower.

**Chart 2:** Bar graph showing average days of absence for each unit  
 **Table 2:**

|  |  |
| --- | --- |
| **Unit** | **Average Absence Days** |
| Admin | 7.86 |
| C-Level | 10.00 |
| Engineering | 9.40 |
| IT | 10.61 |
| Manufacturing | 10.07 |
| Sales | 12.86 |
| **Total Average** | **10.31** |

From chart 2 and table 2, we discover that on average, the organisation’s staff were absent for approximately 10.31 days. Specifically, its sales staff are absent the most with an average of approximately 12.86 days, which is 24.73% higher than the organisation average. On the other hand, admin staff were absent the least with an average of only approximately 7.86 days, which is 23.76% lower than the organisation average.

**Question 1(b)**

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt

#Import the dataset  
TMA\_Data = pd.read\_csv('TMA\_Data.csv')

#Plot Table 1   
TMA\_Data.groupby(['Unit'])['Salary'].sum()

#Plot Graph 1  
plt.title("Unit Salary Expense")  
plt.ylabel("Total Salary")  
TMA\_Data.groupby(['Unit'])['Salary'].sum().plot(kind='line')  
plt.show()

#Plot Table 2  
round(TMA\_Data.groupby(['Unit'])['Absence'].mean(),2)

#Plot Graph 2  
plt.title("Unit Average Days of Absence")  
plt.ylabel("Absence Days")  
TMA\_Data.groupby(['Unit'])['Absence'].mean().plot(kind='bar')  
plt.show()

**Question 1(c)**

#fill null leftDate with 1st May 2022  
TMA\_Data.LeftDate = TMA\_Data.LeftDate.fillna('01/05/2022')

#convert date field into datetime  
TMA\_Data['LeftDate'] = pd.to\_datetime(TMA\_Data['LeftDate'])  
TMA\_Data['JoinDate'] = pd.to\_datetime(TMA\_Data['JoinDate'])

#calculate number of days elapsed between JoinDate and LeftDate  
#divided by 365 to convert from days to years, rounded off to 1 significant figure  
TMA\_Data['length\_of\_service'] = round((TMA\_Data.LeftDate - TMA\_Data.JoinDate).dt.days/365,1)

#display the minimum, maximum and mean length of service   
print('Minimum: {} years'.format(round(TMA\_Data['length\_of\_service'].min(), 1)))  
print('Maximum: {} years'.format(round(TMA\_Data['length\_of\_service'].max(), 1)))  
print('Mean: {} years'.format(round(TMA\_Data['length\_of\_service'].mean(), 1)))

**Required**  
For length of service,   
Mininum: 0.1 years  
Maximum: 16.0 years  
Average: 6.6 years

**Question 1(d)**

#store all staff names in a list  
staff\_names = list (TMA\_Data.Staff.str.lower())  
def search\_staff(name):  
   
 #remove trailing newline characters  
 name = name.lower().rstrip()   
 if name in staff\_names:  
 print("{} is/was a staff of the organization".format(name))  
 else:  
 print("{} is not a staff of the organization".format(name))

play = True  
  
while play:  
 name = input("Please enter staff name: ")  
 search\_staff(name)  
 print('\nEnter 1 to Search for another user : \n Enter 2 to exit: ')  
 choice = input()  
   
 if choice == '1':  
 play = True  
 elif choice == '2':  
 play = False  
 print('\nThank You.')  
 else:  
 print('\nInvalid Input. Restarting Process')