

**ANL 252 (Online)**

**PYTHON FOR DATA ANALYTICS**

# **Tutor-Marked Assignment**

**July 2022 Presentation**

**Submitted by:**

|  |  |
| --- | --- |
| **Name** | **PI No.** |
| **CHANG GUO RUI, JOEL** | **Q1910355** |

**Tutorial Group: ­­­­­­­­­­­­ T 09**

**Instructor’s Name: Mr. MUNISH KUMAR**

**Submission Date: 14/08/2022**

Question 1a part 1)

|  |  |
| --- | --- |
| **Unit** | **Count of Unit** |
| C-Level | 1 |
| Manufacturing | 169 |
| IT | 41 |
| Sales | 22 |
| Engineering | 10 |
| Admin | 7 |

The pie chart shows the percentage total of each department in the company. Up to 70% of the company consist of employees from the manufacturing department, followed by IT. Based on these statistics, it could be determined that this is a manufacturing company producing IT related products for its consumers. Examples of big companies like this in Singapore could be companies like Micron or GlobalFoundaries which produces storage devices or semiconductors. The chart shows a 0% under the C-level department because there is only one employee in that department. This could be translated as the employee could in the top-level management position like the CEO or Managing director of the company.

Question 1a part 2)

|  |  |  |
| --- | --- | --- |
| **Average of Salary** | | |
| **Unit** | **Gender** | **Total** |
| Admin | F | 80159.5 |
|  | M | 73523 |
| C-Level | F | 250000 |
| Engineering | F | 99044.6 |
|  | M | 92689.4 |
| IT | F | 87905.38889 |
|  | M | 101534.087 |
| Manufacturing | F | 59655.22857 |
|  | M | 59015.34375 |
| Sales | F | 75900.27273 |
|  | M | 66952.90909 |

The bar chart shows the average salary count for male and female in each department. The purpose of this graph shows if there was any gender inequality towards the salary in the department. Based on the results shown, the female group had a higher average in the admin, engineering, and sales department. Whereas the male had a higher average of salary in the IT department. Although the head of the company is a female, the statistics based on the graph does not show a huge difference between the salaries of male and female. The data might not be accurate as there is more female employees than male employees from the data provided. Therefore, it is not ideal to use this graph representation for comparison of salary.

Question 1b part 1)

**Input [1]:**

import matplotlib.pyplot as plt

import pandas as pd

df = pd.read\_csv('C:/Users/chimp/OneDrive/Desktop/TMA\_Data.csv')

unit = df["Unit"]

unit.value\_counts()

**Output [1]:**

Manufacturing 169

IT 41

Sales 22

Engineering 10

Admin 7

C-Level 1

Name: Unit, dtype: int64

**Input [2]:**

sizes = [68, 16, 9, 4, 3, 0.04]

labels = 'Manufacturing', 'IT', 'Sales', 'Engineering', 'Admin', 'C-Level'

plt.pie(sizes,

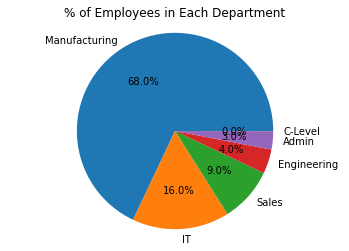
labels = labels,

autopct='%1.1f%%')

plt.title('% of Employees in Each Department')

plt.axis('equal')

plt.show()



Question 1b part 2)

**Input [1]:**

import matplotlib.pyplot as plt

import csv

import pandas as pd

import numpy as np

df = pd.read\_csv("C:/Users/chimp/OneDrive/Desktop/TMA\_Data.csv")

df.index.name = "emp\_id"

plotdata = pd.DataFrame(df.groupby(['Unit', 'Gender']).agg({'Salary': 'mean'}))

display(plotdata)

plotdata.plot(kind="bar")

plt.title("Average Salary of Male/Female in each Department")

plt.xlabel("Unit")

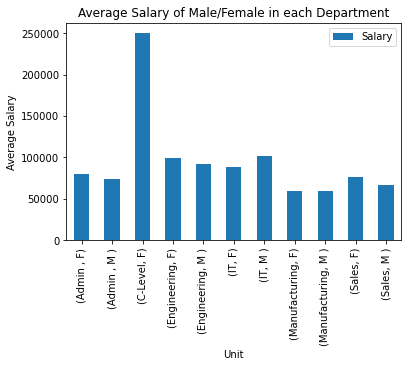
plt.ylabel("Average Salary")

**Output [1]:**

|  | **Salary** |
| --- | --- |
| **Unit** | **Gender** |  |
| **Admin** | **F** | 80159.500000 |
| **M** | 73523.000000 |
| **C-Level** | **F** | 250000.000000 |
| **Engineering** | **F** | 99044.600000 |
| **M** | 92689.400000 |
| **IT** | **F** | 87905.388889 |
| **M** | 101534.086957 |
| **Manufacturing** | **F** | 59655.228571 |
| **M** | 59015.343750 |
| **Sales** | **F** | 75900.272727 |
| **M** | 66952.909091 |

Out[1]:

Text(0, 0.5, 'Average Salary')



Question 1c)

**Input [1]:**

import pandas as pd

import numpy as np

import datetime

from datetime import datetime

from datetime import date

df = pd.read\_csv('C:/Users/chimp/OneDrive/Desktop/TMA\_Data.csv')

df2 = df[df['LeftDate'].isna()]

**Input [2]:**

*#a date to substract dates in csv*

defined\_date = pd.to\_datetime('5/1/2022')

df2['JoinDate'] = pd.to\_datetime(df2['JoinDate'])

diff=(defined\_date - df2['JoinDate']).dt.days

minmum=diff.min()/365

maximum=diff.max()/365

average=diff.mean()/365

print('The minimum service length of Staff (as of 1 May 2022) is ' + "{:.1f}".format(minmum) +' years')

print('The longest service length of Staff (as of 1 May 2022) is ' + "{:.1f}".format(maximum) +' years')

print('The average service length of Staff (as of 1 May 2022) is ' + "{:.1f}".format(average) +' years')

**Output:**

The minimum service length of Staff (as of 1 May 2022) is 0.5 years

The longest service length of Staff (as of 1 May 2022) is 16.3 years

The average service length of Staff (as of 1 May 2022) is 8.4 years

Question 1d)

**Input [\*]:**

import csv

while True:

with open("C:/Users/chimp/OneDrive/Desktop/TMA\_Data.csv", "r") as f:

reader = csv.reader(f)

StaffName = input("Enter Staff Name: ")

for line\_num, content in enumerate(reader):

if (content[1] == StaffName) and (content[8] == "") :

print("Staff works here")

elif (content[1] == StaffName) and (content[8] != "") :

print("Staff had resigned")

**Output:**

Enter Staff Name: Aairah Hastings

Staff works here

Enter Staff Name: Ace Potter

Staff had resigned

Enter Staff Name: Joel Chang

Enter Staff Name: 