A picture containing food, plate

Description automatically generated

**ANL252**

**Python for Data Analytics**

**Tutor-Marked Assignment 01**

**Koo Chi Hang**

**K2081853**

**13 August 2022**

**Question 1**

**Part A**

**Table

Description automatically generated**

(Figure 1: Bin Values of Annual Salaries)

**Chart, bar chart

Description automatically generated**

(Figure 2: Histogram of Annual Salaries in Excel)

We can observe from figure 2 that the data is skewed to the left with majority of the data clustering around the $50001 to $75000 range.

Table

Description automatically generated

(Figure 3: Descriptive statistics of annual salaries at a 95% confidence level)

From Table 3, we can observe that the mean annual salary of the staffs are at $69118.50 (corrected to 2.d.p). The lowest paid staff in the company earns $45046 and the highest paid staff earns $250000. At a 95% confidence level, the intervals for the annual salary is between $66803 to $72153 (corrected to whole numbers), which means there is a 95% chance that a randomly selected salary will land between $66803 to $72153.

Table

Description automatically generated

(Figure 4: Summarised table of average annual salaries)

Chart, bar chart

Description automatically generated

(Figure 5: Bar graph of average annual salaries in Excel)

With reference to figure 5, we can observe that workers in the manufacturing unit are paid the least on average when compared to the other departments. Even though the C-level unit consist of the highest paid workers, it is important to understand that there is only 1 staff in that department and hence the data might be extremely skewed.

**Part B**

import matplotlib.pyplot as plt

import pandas as pd

annual\_salaries = pd.read\_csv("/Users/chihang/Desktop/TMA\_Data.csv")

salaries\_list = list(annual\_salaries["Salary"])

rmin = min(salaries\_list)

rmax = max(salaries\_list)

plt.figure(figsize=(6,6), dpi=100)

plt.hist(salaries\_list, bins=10, range=(rmin, rmax), align="mid", orientation="vertical", rwidth=0.8)

plt.title("Histogram of Annual Salaries")

plt.xlabel("Annual Salaries")

plt.ylabel("Frequency")

plt.show()

Chart, bar chart, histogram

Description automatically generated

(Figure 6: Histogram of Annual Salaries in Python)

import matplotlib.pyplot as plt

Data = {'Unit': ['Admin','C-level','Engineering','IT','Manufacturing','Sales'],

'Average\_Salary': [77315.29,250000,95867,95550.76,59412.91,71426.59]}

df = pd.DataFrame(Data,columns=['Unit','Average\_Salary'])

plt.bar(df['Unit'], df['Average\_Salary'])

plt.title('Average annual salaries of workers according to the relevant departments', fontsize=12)

plt.xlabel('Unit', fontsize=12)

plt.ylabel('Average Salary in ($)', fontsize=12)

plt.show()

Chart, bar chart

Description automatically generated

(Figure 7: Bar graph of average annual salaries in Python)

**Part C**

#Part (c)

import pandas as pd

import numpy as np

import datetime

#Length service of all staffs known as tenure

Data = pd.read\_csv("/Users/chihang/Desktop/TMA\_Data.csv")

df = pd.DataFrame(Data,columns=['ID','Staff','JoinDate','LeftDate'])

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

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

end = pd.to\_datetime('2022-05-01')

df['LeftDate'] = df['LeftDate'].fillna(end)

df['Tenure(years)'] = round((((df['LeftDate'] - df['JoinDate']).dt.days)/365),1)

print(df)

#Average tenure

def Average(lst):

return sum(lst) / len(lst)

lst = df['Tenure(years)']

average = Average(lst)

print("Average tenure is", average, "years")

#Maximum tenure

maximum = max(df['Tenure(years)'])

print("Maximum tenure is", maximum, "years")

#Minimum tenure

minimum = min(df['Tenure(years)'])

print("Minimum tenure is", minimum, "years")

**Screenshot of output:**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Part D**

#part (d)

import pandas as pd

data = pd.read\_csv("/Users/chihang/Desktop/TMA\_Data.csv")

name\_list = list(data["Staff"])

user\_want\_to\_exit = True

while user\_want\_to\_exit:

staff\_name = str(input("Please Enter Full Name of Staff:"))

if staff\_name in name\_list:

print("Staff is a valid employee")

else:

print("Staff is an invalid employee")

while user\_want\_to\_exit:

terminate\_qns = str(input("Would you like to try again? (Y/N)"))

if terminate\_qns == "N":

print("Good Bye")

user\_want\_to\_exit = False

elif terminate\_qns == "Y":

break

else: print("Invalid input, please try again")