ANL252

Python for Data Analytics

Group-based Assignment

Singapore University of Social Sciences

T05

|  |  |
| --- | --- |
| **Name:** | **PI No.:** |
| Daniel Ong Song Hng | H2181372 |
| Nur Jannah Binte Yazit | W1981876 |

Submitted On: 19 February 2023

**Declaration Page**

We, members of group 11 , do hereby declare that we each contributed to this assignment and that we collectively agree to a shared grade.

|  |  |  |
| --- | --- | --- |
| **Name:** | **Contribution:** | **Signature:** |
| Daniel Ong Song Hng (Team Leader) | I did question 1(a) and 1(c) |  |
| Nur Jannah Binte Yazit | I did question 1(a) and 1(b) |  |

# Table of Contents

[List of Figures 4](#_Toc127716781)

[List of Tables 4](#_Toc127716782)

[Question 1(a) 5](#_Toc127716783)

[Analysis 1 5](#_Toc127716784)

[Analysis 2 7](#_Toc127716785)

[Analysis 3 9](#_Toc127716786)

[Analysis 4 11](#_Toc127716787)

[Question 1(b) 13](#_Toc127716788)

[Output for Figure 1 13](#_Toc127716789)

[Output for Figure 2 16](#_Toc127716790)

[Output for Figure 3 18](#_Toc127716791)

[Output for Figure 4 19](#_Toc127716792)

[Question 1(c) 21](#_Toc127716793)

[Output (for correct options): 21](#_Toc127716794)

[Output (for incorrect option): 24](#_Toc127716795)

[References 25](#_Toc127716796)

[Appendix A: Code for Figure 1 26](#_Toc127716797)

[Appendix B: Code for Figure 2 29](#_Toc127716798)

[Appendix C: Code for Figure 3 31](#_Toc127716799)

[Appendix D: Code for Figure 4 33](#_Toc127716800)

[Appendix E: Code Qn 1(c) 36](#_Toc127716801)

# List of Figures

[Figure 1: Employees’ Satisfactory Rate with the Company, their Role and Co-workers in Percentage 6](#_Toc127716802)

[Figure 2: No. of Employees who Left the Company 7](#_Toc127716803)

[Figure 3: Employees’ Age Group 9](#_Toc127716804)

[Figure 4: Frequencies of Employees’ Job Grade and Performance Rating 11](#_Toc127716805)

# List of Tables

[Table 1: Employees’ Satisfactory Rate with the Company, their Role and Co-workers 5](#_Toc127716806)

[Table 2: Employees’ Satisfactory Rate with the Company, their Role and Co-workers in Percentage 5](#_Toc127716807)

[Table 3: No. of Employees who Left the Company 7](#_Toc127716808)

[Table 4: Employees’ Age Group 9](#_Toc127716809)

[Table 5: Frequencies of Employees’ Job Grade and Performance Rating 11](#_Toc127716810)

# Question 1(a)

## Analysis 1

#### Table 1: Employees’ Satisfactory Rate with the Company, their Role and Co-workers

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Ratings** | | | |
| **1** | **2** | **3** | **4** |
| **SatisfactionCompany** | 235 | 247 | 386 | 382 |
| **SatisfactionRole** | 247 | 242 | 376 | 385 |
| **SatisfactionCoworkers** | 230 | 258 | 388 | 374 |

#### Table 2: Employees’ Satisfactory Rate with the Company, their Role and Co-workers in Percentage

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Ratings** | | | |
| **1** | **2** | **3** | **4** |
| **SatisfactionCompany** | 19% | 20% | 31% | 31% |
| **SatisfactionRole** | 20% | 19% | 30% | 31% |
| **SatisfactionCoworkers** | 18% | 21% | 31% | 30% |



### Figure 1: Employees’ Satisfactory Rate with the Company, their Role and Co-workers in Percentage

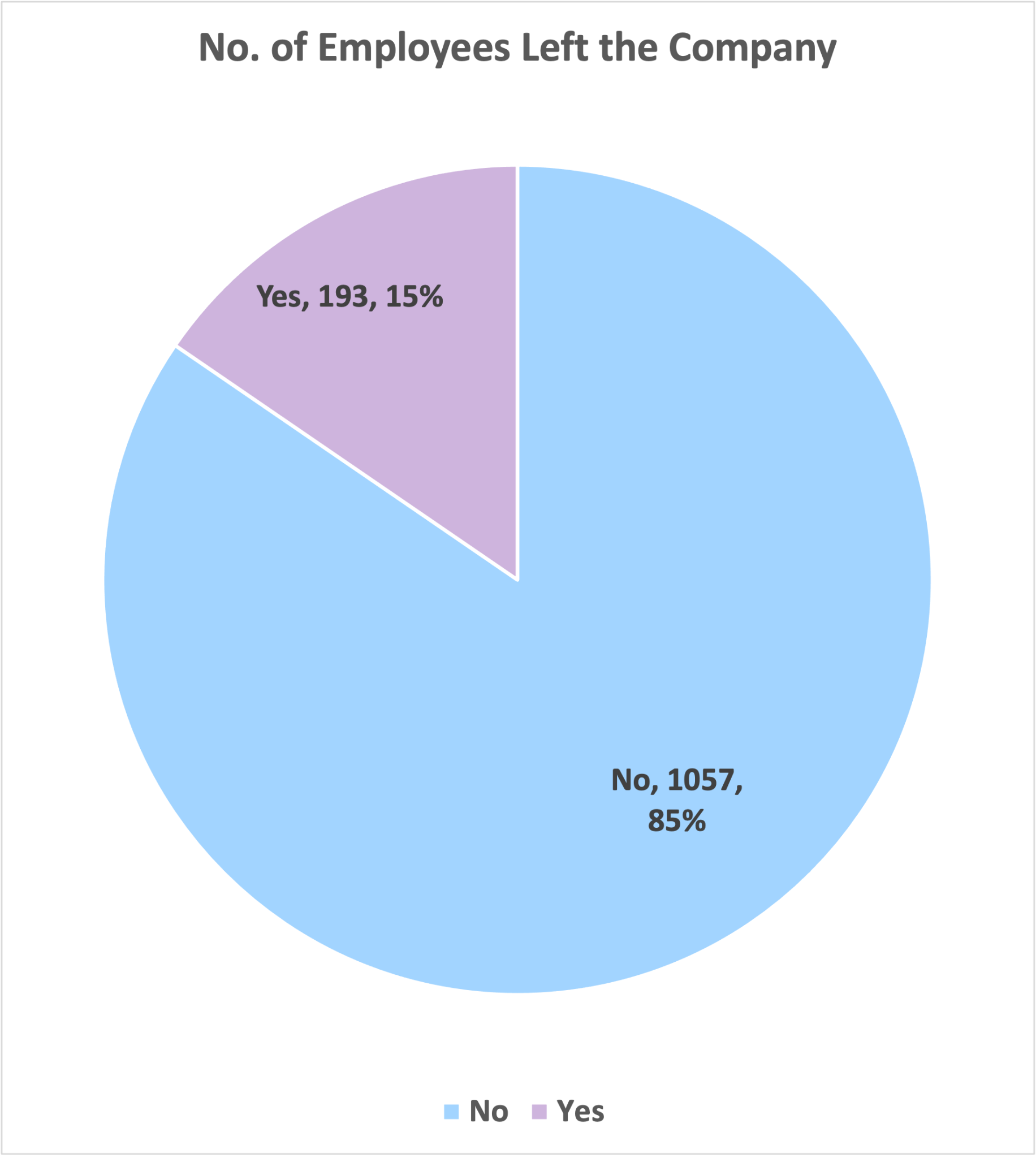
For the first chart, we will analyse the employees’ satisfaction rate with the company, their current roles, and their co-workers using a **stacked bar chart**. By utilising the stacked bar chart, we can combine more than one variable that shares the same y-axis to have a holistic view of the analysis.

From ***Figure 1*** above, we can deduce that most employees, at least 60%, feel satisfied. However, the company would want to focus on 39% of the employees with a satisfaction rate of 1 and 2. The company can improve their relationships with these individuals and make the work environment more enjoyable. It would have a significant impact on their satisfaction levels. Furthermore, every company should strive to maintain a high employee satisfaction rate to improve productivity and the culture of effort.

## Analysis 2

#### Table 3: No. of Employees who Left the Company

|  |  |
| --- | --- |
| **No. of Employees Left the Company** | |
| **No** | 1057 |
| **Yes** | 193 |



### Figure 2: No. of Employees who Left the Company

Secondly, we can use a **pie chart** to find the percentage of employees who left the company. Pie charts allow us to analyse a variable clearly and identify the most frequently occurring items.

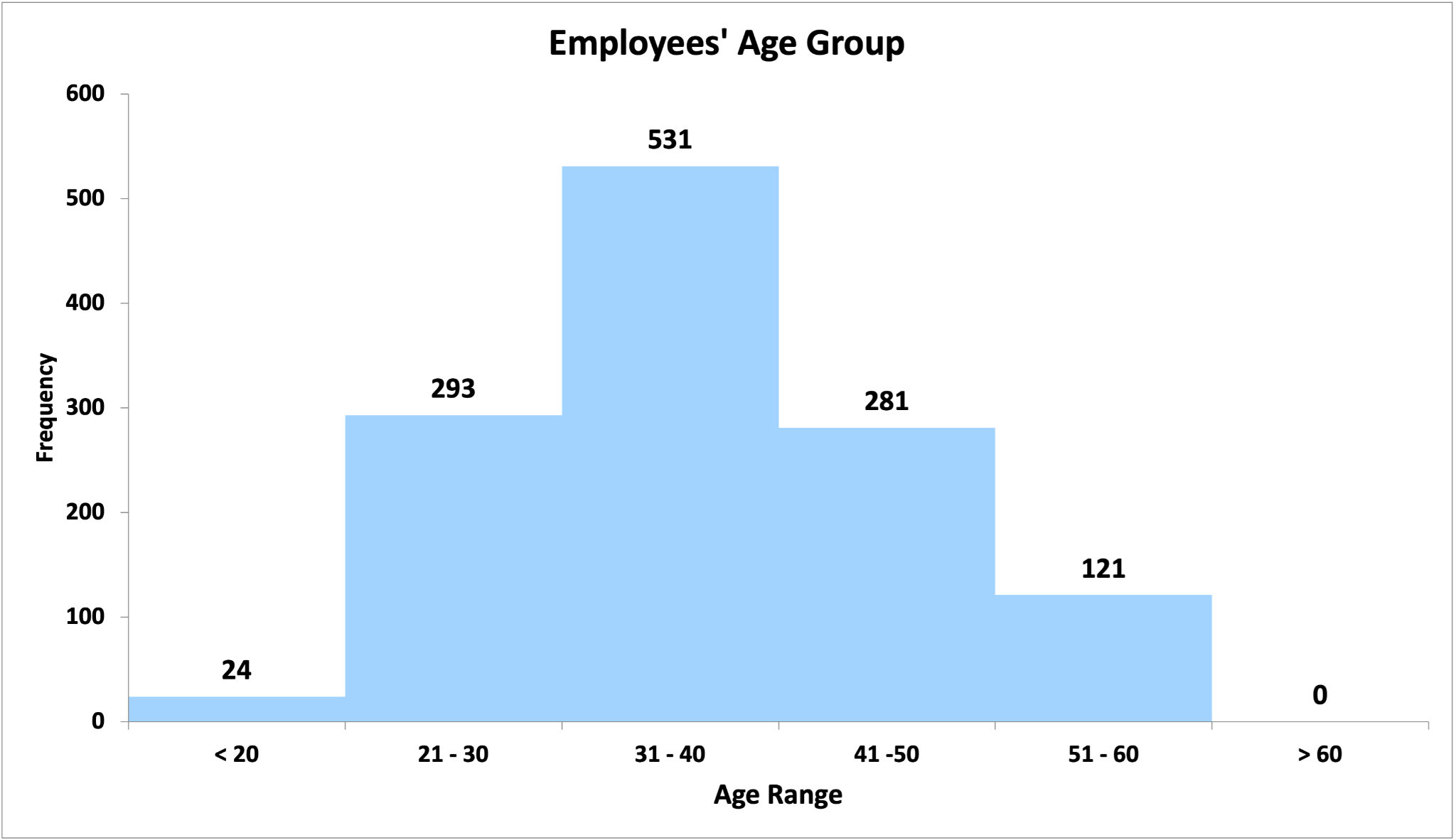
From ***Figure 2*** above, only 15% of the employees left the company. Though the percentage may be small, it is 193 employees out of 1250. From this, we can assume that the majority stayed in the company since they were satisfied with the company, their current role, and their co-workers from Figure 1 above. There could be several reasons for the 15% of them leaving. Firstly, they may be the ones that rated low in satisfaction with the company (as per Figure 1). There is also the possibility of them having a better offer from another company. Thirdly, they would be interested in exploring and entering new industries. The company should maintain a low employee turnover rate, as this will be seen as a positive impression on newcomers and depicts a positive work environment.

## 

## Analysis 3

#### Table 4: Employees’ Age Group

|  |  |
| --- | --- |
| **Age Range** | **Frequency** |
| 19 | 5 |
| 20 | 19 |
| 30 | 293 |
| 40 | 531 |
| 50 | 281 |
| 60 | 121 |
| More | 0 |



### Figure 3: Employees’ Age Group

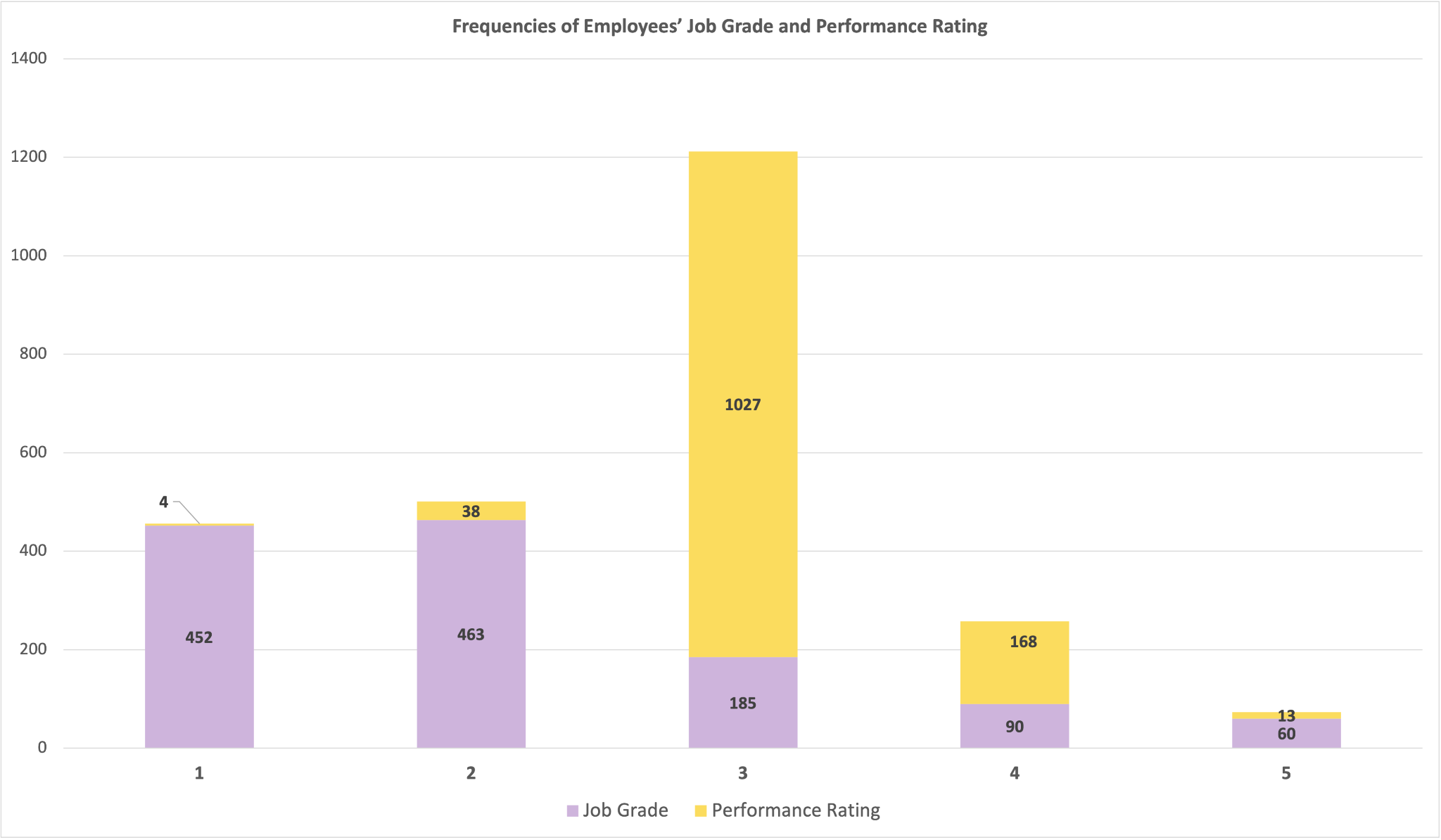
Next, we use a **histogram** to analyse the employees’ age group. A histogram would be suitable for continuous data sets such as age ranges, as the peak of the histogram indicates where the frequency is highest or most frequent.

From ***Figure 3*** above, we can examine that the majority of the employees, 531 of them, are in their 30s. It is also worth noting that no one at the company is older than 60, indicating that employees retire by that age. Moreover, a handful of them (24 employees) are younger than 20. They could either be interns or fresh graduates at the company. Lastly, there is a similar number of employees in their 20s and 40s.

## Analysis 4

#### Table 5: Frequencies of Employees’ Job Grade and Performance Rating

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Ratings** | | | | |
| **1** | **2** | **3** | **4** | **5** |
| **Job Grade** | **452** | **463** | **185** | **90** | **60** |
| **Performance Rating** | **4** | **38** | **1027** | **168** | **13** |



### Figure 4: Frequencies of Employees’ Job Grade and Performance Rating

As a final analysis, we analysed the frequency of employees’ job grades and appraisal performance ratings using a **stacked bar chart**. As aforementioned, a stacked bar would allow us to examine the data that shares the same x- or y-axis. In this case, Job Grade and Performance Rating have the same numeral ratings, where 5 is the highest score.

From ***Figure 4*** above, only a few employees scored a 5 for their Job Grade and Performance Rating. Those employees are likely to hold high positions at the company. The majority of the employees, 1027, scored a median Performance Rating of 3, while only 4 scored the lowest Performance Rating of 1. As for Job Grades, most employees scored 1 and 2.

# Question 1(b)

## Output for Figure 1

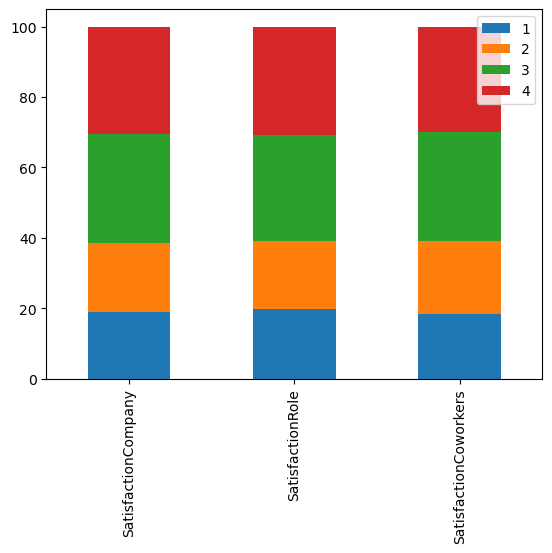
|  |  |  |  |
| --- | --- | --- | --- |
| SatisfactionCompany | SatisfactionRole | SatisfactionCoworkers |  |
| 0 | 2 | 3 | 3 |
| 1 | 3 | 1 | 2 |
| 2 | 1 | 3 | 1 |
| 3 | 3 | 3 | 1 |
| 4 | 1 | 4 | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| SatisfactionCompany | SatisfactionRole | SatisfactionCoworkers |  |
| 1 | 235 | 247 | 230 |
| 2 | 247 | 242 | 258 |
| 3 | 386 | 376 | 388 |
| 4 | 382 | 385 | 374 |

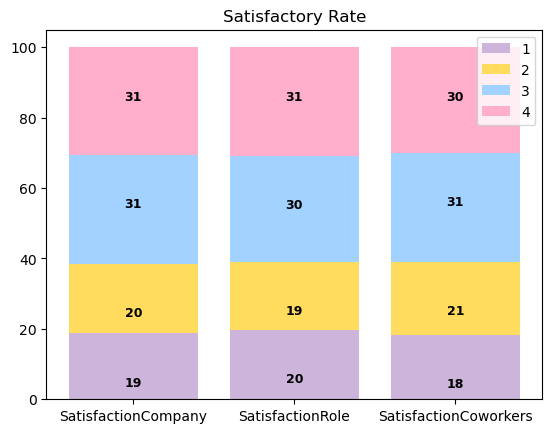
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 |  |
| SatisfactionCompany | 235 | 247 | 386 | 382 |
| SatisfactionRole | 247 | 242 | 376 | 385 |
| SatisfactionCoworkers | 230 | 258 | 388 | 374 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 |  |
| SatisfactionCompany | 18.80 | 19.76 | 30.88 | 30.56 |
| SatisfactionRole | 19.76 | 19.36 | 30.08 | 30.80 |
| SatisfactionCoworkers | 18.40 | 20.64 | 31.04 | 29.92 |

<AxesSubplot:>



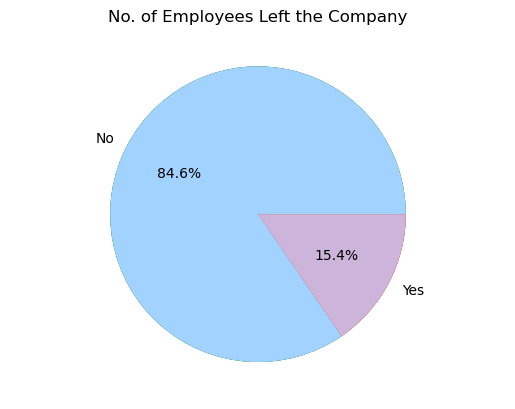
<matplotlib.legend.Legend at 0x7ff4cf52cc10>



## Output for Figure 2

|  |  |
| --- | --- |
| LeftCompany |  |
| 0 | No |
| 1 | No |
| 2 | Yes |
| 3 | No |
| 4 | No |

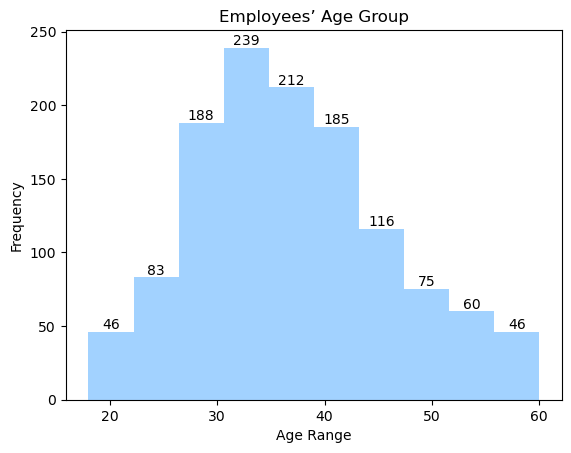
|  |  |
| --- | --- |
| LeftCompany |  |
| No | 1057 |
| Yes | 193 |



## 

## Output for Figure 3

|  |  |
| --- | --- |
| Age |  |
| 0 | 39 |
| 1 | 41 |
| 2 | 31 |
| 3 | 34 |
| 4 | 43 |

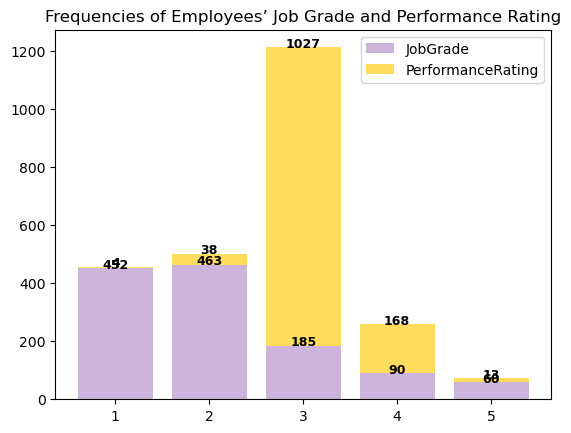


## Output for Figure 4

|  |  |  |
| --- | --- | --- |
| JobGrade | PerformanceRating |  |
| 0 | 5 | 3 |
| 1 | 1 | 4 |
| 2 | 3 | 3 |
| 3 | 2 | 3 |
| 4 | 2 | 4 |

|  |  |  |
| --- | --- | --- |
| JobGrade | PerformanceRating |  |
| 1 | 452 | 4 |
| 2 | 463 | 38 |
| 3 | 185 | 1027 |
| 4 | 90 | 168 |
| 5 | 60 | 13 |

<matplotlib.legend.Legend at 0x7fd34e6a7820>



# 

# Question 1(c)

## Output (for correct options):

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

Enter your choice: 1

Total staff strength:- 1250

Staff by unit and gender:-

BusinessUnit Gender

Business Development Female 159

Male 218

HR Female 17

Male 34

Product Development Female 327

Male 495

Name: EmpID, dtype: int64

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

Enter your choice: 2

Total annual renumeration expenditure:- 98227176

Median monthly salary:- 5001.0

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

Enter your choice: 3

Eligible long service award recipients:-

EmpID JobRole Age

1295 Account Manager 59

1896 Section Manager 59

1711 Product Executive 58

2242 Section Manager 57

1466 Section Manager 55

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

Enter your choice: 4

Thank You and Goodbye.

## Output (for incorrect option):

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

Enter your choice: 5

Invalid choice. Please try again.

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

Enter your choice: 4

Thank You and Goodbye.

# References

Alex. (2021, January 22). *Stacked Bar Charts with Labels in Matplotlib*. PythonCharts. https://www.pythoncharts.com/matplotlib/stacked-bar-charts-labels/

Jie Jenn. (2021, July 27). *Plot a histogram graph using matplotlib for beginners* [Video]. YouTube. https://www.youtube.com/watch?v=woBnzolCH5Q

# Appendix A: Code for Figure 1

# import the necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

# read the dataset

df = pd.read\_csv('GBA.csv')

# retrieving data that are only relavant to the employees' satisfaction rate - which include: SatisfactionCompany, SatisfactionRole and SatisfactionCoworkers

# print the first 5 dataset

satisfactionData = df[["SatisfactionCompany", "SatisfactionRole", "SatisfactionCoworkers"]]

satisfactionData.head()

# determine the frequencies of each variable

frequencies\_satisfaction = {}

for i in satisfactionData.columns:

frequencies\_satisfaction[i] = satisfactionData[i].value\_counts()

# create the table

plotdata\_satisfaction = pd.DataFrame(frequencies\_satisfaction)

plotdata\_satisfaction

# switching the row and columns

plotdata\_satisfaction = plotdata\_satisfaction.transpose()

plotdata\_satisfaction

# create stacked bar chart in percentage

plotdata\_satisfaction2 = plotdata\_satisfaction.div(plotdata\_satisfaction.sum(axis=1), axis=0)\*100

plotdata\_satisfaction2

# plot the stacked bar chart

plotdata\_satisfaction2.plot(kind = "bar", stacked = True)

# plot the stacked bar chart

# cited from Alex(2021), https://www.pythoncharts.com/matplotlib/stacked-bar-charts-labels/

fig, ax = plt.subplots()

# change stacked bar chart colour

colors = ['#CDB4DB', '#FFDC5E', '#A2D2FF', '#FFAFCC']

bottom = np.zeros(len(plotdata\_satisfaction2))

for i, col in enumerate(plotdata\_satisfaction2.columns):

ax.bar(plotdata\_satisfaction2.index, plotdata\_satisfaction2[col], bottom=bottom, label=col,

color=colors[i])

bottom += np.array(plotdata\_satisfaction2[col])

# adding data labels to each bar with a negative offset

y\_offset = -15

# adding labels in the charts within the bar

for bar in ax.patches:

ax.text(

# to add the labels in the center of the bar (width)

bar.get\_x() + bar.get\_width() / 2,

# to add the labels in the center of the bar (vertical)

bar.get\_height() + bar.get\_y() + y\_offset,

# bar height

round(bar.get\_height()),

# to change the font and colour of the label

ha="center",

color="black",

weight="bold",

size=9

)

# add stacked bar chart title

ax.set\_title("Satisfactory Rate")

# show stacked bar chart

ax.legend()

# Appendix B: Code for Figure 2

# import the necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

# read the dataset

df = pd.read\_csv('GBA.csv')

# retrieving data of no. of employees left the company

# print the first 5 dataset

leftData = df[["LeftCompany"]]

leftData.head()

# determine the frequencies of each variable

frequencies\_left = {}

for i in leftData.columns:

frequencies\_left[i] = leftData[i].value\_counts()

# create the table

plotdata\_left = pd.DataFrame(frequencies\_left)

plotdata\_left

# plot pie chart

sizes = [1057, 193]

labels = "No", "Yes"

plt.pie(sizes,

labels = labels)

# adding title to the pie chart

plt.title("No. of Employees Left the Company")

plt.axis("equal")

#to show percentage

plt.pie(sizes,

autopct = "%1.1f%%")

# change colour

plt.pie(sizes,

colors = ["#A2D2FF", "#CDB4DB"])

plt.show()

# Appendix C: Code for Figure 3

# import the necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

# read the dataset

df = pd.read\_csv('GBA.csv')

# retrieving data of employees' age group

# print the first 5 dataset

ageData = df[["Age"]]

ageData.head()

# plot the historgram

# cited from Jie Jenn (2021), https://www.youtube.com/watch?v=woBnzolCH5Q

fig, ax = plt.subplots()

# change histogram colour

hist = ax.hist(ageData, color = '#A2D2FF')

# adding labels to the histogram

plt.xlabel("Age Range")

plt.ylabel("Frequency")

plt.title("Employees’ Age Group")

# to add the total sum / data frequencies (label) onto the histogram and let hist be the variable for the histogram

labels = [hist[0][i] for i in range(len(hist[0]))]

rects = ax.patches

for rect, label in zip(rects, labels):

height = rect.get\_height()

ax.text(

rect.get\_x() + rect.get\_width() / 2,

height + 0.01,

int(label),

ha = 'center',

va = 'bottom'

)

# show histogram

plt.show()

# Appendix D: Code for Figure 4

# import the necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

# read the dataset

df = pd.read\_csv('GBA.csv')

# retrieving data of employees' job grade and their appraisal

# print the first 5 dataset

gradeData = df[["JobGrade", "PerformanceRating"]]

gradeData.head()

# determine the frequencies for job grade and appraisal

frequencies\_grade = {}

for i in gradeData.columns:

frequencies\_grade[i] = gradeData[i].value\_counts()

# create the table

plotdata\_grade = pd.DataFrame(frequencies\_grade)

plotdata\_grade

# plot the stacked bar chart

# cited from Alex(2021), https://www.pythoncharts.com/matplotlib/stacked-bar-charts-labels/

fig, ax = plt.subplots()

# change stacked bar chart colour

colors = ['#CDB4DB', '#FFDC5E']

bottom = np.zeros(len(plotdata\_grade))

for i, col in enumerate(plotdata\_grade.columns):

ax.bar(plotdata\_grade.index, plotdata\_grade[col], bottom=bottom, label=col,

color=colors[i])

bottom += np.array(plotdata\_grade[col])

# adding labels in the charts within the bar

for bar in ax.patches:

ax.text(

# to add the labels in the center of the bar (width)

bar.get\_x() + bar.get\_width() / 2,

# to add the labels in the center of the bar (vertical)

bar.get\_height() + bar.get\_y(),

# bar height

round(bar.get\_height()),

# to change the font and colour of the label

ha="center",

color="black",

weight="bold",

size=9

)

# add stacked bar chart title

ax.set\_title("Frequencies of Employees’ Job Grade and Performance Rating")

# show stacked bar chart

ax.legend()

# Appendix E: Code Qn 1(c)

# Import pandas

import pandas as pd

# Load the dataset

df = pd.read\_csv('GBA.csv')

# Define function to calculate total staff strength by business unit and gender

def get\_staff\_strength(df):

total\_staff = df['EmpID'].nunique()

staff\_by\_unit\_gender = df.groupby(['BusinessUnit', 'Gender'])['EmpID'].nunique()

return (total\_staff, staff\_by\_unit\_gender)

# Define function to calculate total annual staff renumeration expenditure and median monthly salary

def get\_renumeration(df):

total\_renumeration = df['MonthlyPay'].sum() \* 12

median\_salary = df['MonthlyPay'].median()

return (total\_renumeration, median\_salary)

# Define function to get eligible long service award recipients

# Assume Long service receipients been with company at least 10 years

def get\_long\_service\_recipients(df):

long\_service\_recipients = df[df['WithCompany'] >= 35].sort\_values('Age', ascending=False)[['EmpID', 'JobRole', 'Age']]

return long\_service\_recipients

# Define function to display menu options and get user input

def display\_menu():

print('\nPlease select an option:')

print('1. Current staff strength')

print('2. Current staff annual renumeration')

print('3. Eligible long service award recipients')

print('4. Exit')

choice = input('\nEnter your choice: ')

return choice

# Main program loop

while True:

choice = display\_menu()

if choice == '1':

total\_staff, staff\_by\_unit\_gender = get\_staff\_strength(df)

print('\nTotal staff strength:-', total\_staff)

print('\nStaff by unit and gender:-\n')

print(staff\_by\_unit\_gender)

elif choice == '2':

total\_renumeration, median\_salary = get\_renumeration(df)

print('\nTotal annual renumeration expenditure:-', total\_renumeration)

print('\nMedian monthly salary:-', median\_salary)

elif choice == '3':

long\_service\_recipients = get\_long\_service\_recipients(df)

print('\nEligible long service award recipients:-\n')

print(long\_service\_recipients.to\_string(index=False))

elif choice == '4':

break

else:

print('\nInvalid choice. Please try again.')

# close the system

print("Thank You and Goodbye.")