

**ANL252**

**Python for Data Analytics**

**Group-based Assignment**

**JANUARY 2023 Presentation**

|  |  |
| --- | --- |
| **Name** | **PI No.** |
| Angel Ling Yong Eng | K2211248 |
| Lim Wei Quan | K2110655 |
| Kerry Kung Wei Jie | H2211486 |
| Ng Zi Ya | M2272807 |

Submission Date: **19 February 2023**

### Declaration Page

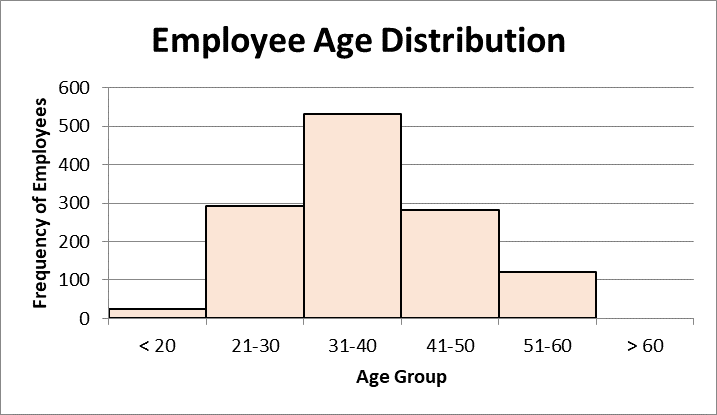
We, members of group 8, do hereby declare that we each contributed to this assignment and that we collectively agree to a shared grade.

|  |  |  |
| --- | --- | --- |
| **Name** | **Contribution** | **Signature** |
| Ng Zi Ya (Team Lead) | I did Q1a & 1b chart 2, did 1c together with the team | ZY |
| Angel Ling Yong Eng | I did Q1a & 1b chart 1, did 1c together with the team |  |
| Kerry Kung Wei Jie | I did Q1a & 1b chart 3, did 1c together with the team | Kerry |
| Lim Wei Quan | I did Q1a & 1b chart 4, did 1c together with the team |  |

### 

### Question 1 (a)

**Chart 1 (Histogram) - Angel**



**Table:**

|  |  |
| --- | --- |
| **Age Group** | **Frequency** |
| <20 | 24 |
| 21-30 | 293 |
| 31-40 | 531 |
| 41-50 | 281 |
| 51-60 | 121 |
| >60 | 0 |

**Extra Infomation:**

|  |  |
| --- | --- |
| MEAN | 37.0288 |
| MEDIAN | 36 |

**Description**:

The histogram shows the age distribution of the total of 1250 employees in the company.

**Interesting observation:**

We observe that the distribution of the graphs is **fairly normally distributed** (bell-shaped). The **majority of the employees are aged between 31-40 years old**. This is a pretty normal sight to see in a company as the majority of this age group are working for financial stability, career advancement, sense of purpose, and many more. There is a slight skewness to the right which can be further justified by the mean of the data being higher than the median, 37.0288 and 36 respectively. There are 24 data points of employees that are less than 20 years old. These 24 employees seem to be much younger as compared to the rest of the employees. This could imply that the company hires these young interns to help them with their day-to-day job. An interesting observation is that there is **no** employee who is more than 60 years old. Assuming that the retirement age in Singapore is 63, this could suggest that employee left their job earlier than expected.

**Chart 2 (Scatter Plot) - Zi Ya**

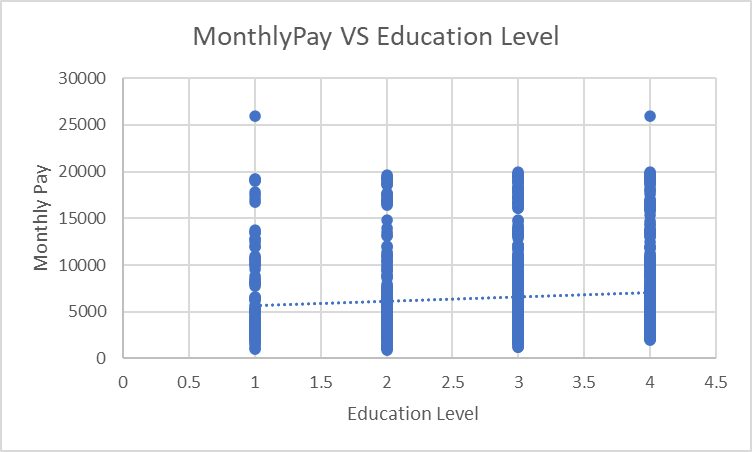


Table:

| Education Level: | Number of Employees: | Total MonthlyPay: |
| --- | --- | --- |
| 4 | 376 | $ 2,684,183.00 |
| 3 | 481 | $ 3,154,438.00 |
| 2 | 264 | $ 1,586,494.00 |
| 1 | 129 | $ 760,483.00 |

**Description:**

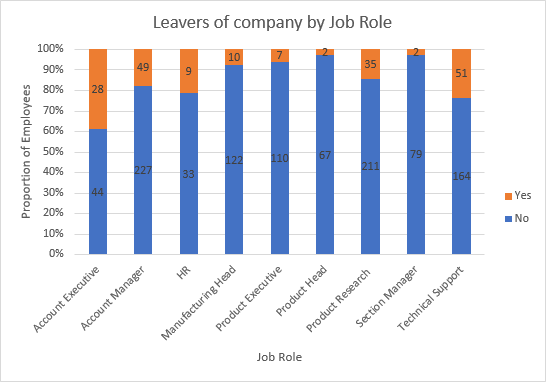
The scatter plot figure below shows the relationship between education level and monthly pay. We picked these two data to see if the level of education affects how much an employee earns monthly.

**Interesting Observation:**

The scatter plot below shows that more employees with postgraduate education earn more. The trend line shows a slight uphill pattern from left to right, indicating a positive relationship between Education Level and Monthly Pay. This means that if the value on the X-axis increases, the value on the y-axis also increases.

One interesting observation of the scatter plot below shows that an employee with other education is earning $ 25979 per month, which is only $20 different from the employee with a postgraduate education who earns $ 25999. This becomes a debating point for the social norms, where you will only get good pay if you get your degree. However, based on the scatter plot, it shows that the higher the education the more the monthly pay.

**Chart 3 (Bar Chart) - Kerry**

****

**Table:**

|  |  |  |
| --- | --- | --- |
|  | **Leavers** | |
| **Job Role** | **No** | **Yes** |
| Account Executive | 44 | 28 |
| Account Manager | 227 | 49 |
| HR | 33 | 9 |
| Manufacturing Head | 122 | 10 |
| Product Executive | 110 | 7 |
| Product Head | 67 | 2 |
| Product Research | 211 | 35 |
| Section Manager | 79 | 2 |
| Technical Support | 164 | 51 |

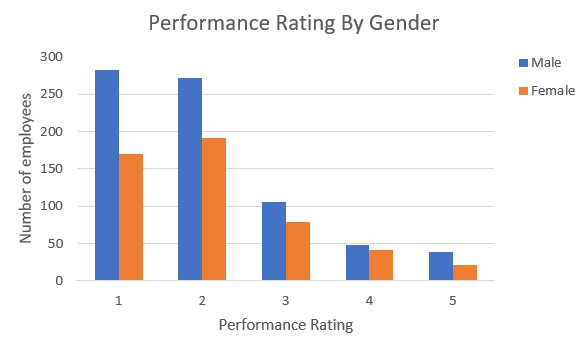
**Description**:

The bar chart above shows the proportion of employees who have left the company by each job role.

**Interesting observation:**

We have observed that there is a variation in the rate of employees leaving the company across different job roles, with an average of about 10% to 20%. The management should be concerned about job roles with a rate of leaving above 20% which are Account Executive, Technical Support, and HR. This could mean that there are underlying issues that need to be addressed. Among these three job roles, the Account Executive has the highest rate of employees leaving about 40% which is twice as much as the Account Manager. This could suggest that there might be an imbalance in workload between them, causing a higher number of Account Executives to leave. Conversely, the job roles with less than 10% of employees who have quitted, such as Product Executive, Product Head, and Section Manager, could be reviewed to identify what makes it more attractive for employees to stay. The management could compare the tasks and responsibilities between the different job roles to ensure that the employee retention rate remains high.

**Chart 4 (Bar Chart) - Wei Quan**

****

**Table:**

|  |  |  |
| --- | --- | --- |
| **Performance Rating** | **Male** | **Female** |
| 1 | 282 | 170 |
| 2 | 272 | 191 |
| 3 | 106 | 79 |
| 4 | 48 | 42 |
| 5 | 39 | 21 |

**Description**:

The chart above shows the performance rating by Gender

**Interesting observation:**

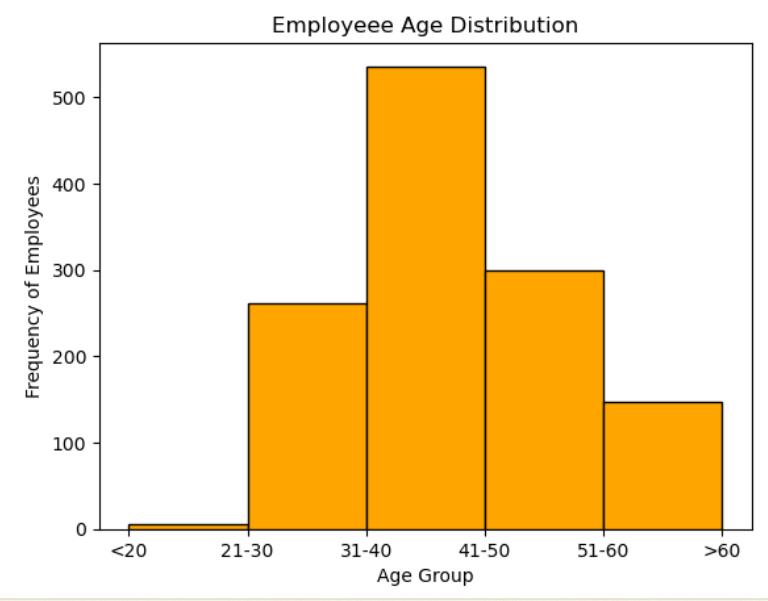
The total number of the male is 747 and the total number of females is 503. The above charts show some interesting observations about the performance of the female, almost the same as the performance of the male. For the performance rating under 1, 282 out of the 747 males score 1 which means that 37.75 % performed very badly in their job role. As compared to females, 170 out of the 503 females score 1 which means that only 33.79% were never performing well in their job role.

For the performance rating under 5, 39 out of the males score 5 which means that 5.22 % performed very well in their job role. Other than that, 21 out of the females also score 5 and which means that 4.17 % of the females also perform very well in their job roles.

Based on the observation, data show that gender discrimination does not happen in this company and women are as capable as men. It seemed that every employee had been treated fairly based on the performance rating rated by their respective manager.

### Question 1 (b)

**Chart 1 (Histogram) - Angel**



**Data:**

EmpID Age

0 1000 39

1 1001 41

2 1002 31

3 1003 34

4 1004 43

... ... ...

1245 2245 25

1246 2246 38

1247 2247 46

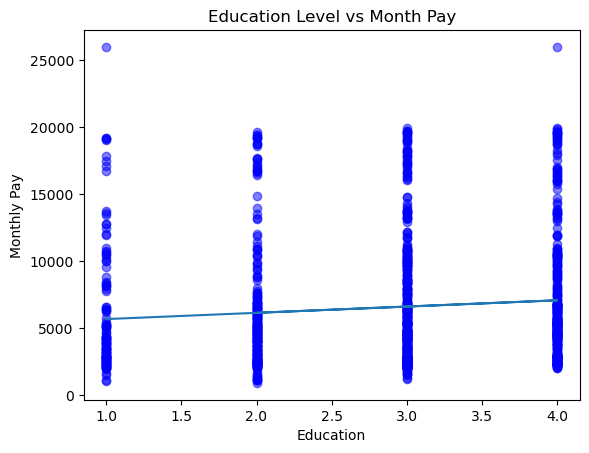
1248 2248 40

1249 2249 39

[1250 rows x 2 columns]

**Table:**

|  |  |
| --- | --- |
| **Age Group** | **Frequency** |
| <20 | 24 |
| 21-30 | 293 |
| 31-40 | 531 |
| 41-50 | 281 |
| 51-60 | 121 |
| >60 | 0 |

**Chart 2 (Scatter Plot) - Zi Ya**

**Table:**

**Education MonthlyPay**

**0 1 19197**

**1 4 2289**

**2 4 8161**

**3 4 6538**

**4 4 5675**

**... ... ...**

**1245 3 4851**

**1246 2 12061**

**1247 4 4615**

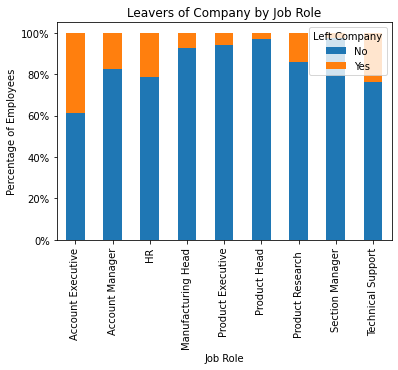
**1248 4 3617**

**1249 2 5151**

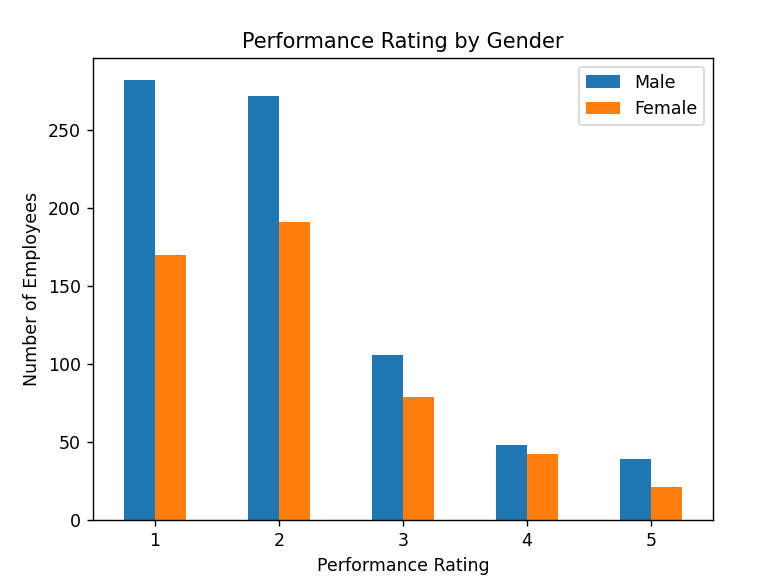
**[1250 rows x 2 columns]**

|  |  |  |
| --- | --- | --- |
| Education Level: | Number of Employees: | Total MonthlyPay: |
| 4 | 376 | $ 2,684,183.00 |
| 3 | 481 | $ 3,154,438.00 |
| 2 | 264 | $ 1,586,494.00 |
| 1 | 129 | $ 760,483.00 |

**Chart 3 (Bar Chart) - Kerry**

****

**Chart 4 (Bar Chart) - Wei Quan**

****

**Table:**

|  |  |  |
| --- | --- | --- |
| **Performance Rating** | **Male** | **Female** |
| 1 | 282 | 170 |
| 2 | 272 | 191 |
| 3 | 106 | 79 |
| 4 | 48 | 42 |
| 5 | 39 | 21 |

### 

### Question 1 (c)

**Output:**

Staff Directory:

1. Current staff strength.

2. Current staff annual renumeration.

3. Eligible long service award recipients.

4. Exit.

Choose option 1 to 4:

**Output from selecting option 1:**

**You have selected option 1: Current staff strength.**

Current Total Staff Strength: 1057

Breakdown of staff strength by business unit:

Business Development: 299

HR: 42

Product Development: 716

Breakdown of staff strength by gender:

Female: 434

Male: 623

Staff Directory:

1. Current staff strength.

2. Current staff annual renumeration.

3. Eligible long service award recipients.

4. Exit.

Choose option 1 to 4:

**Output from selecting option 2:**

**You have selected option 2: Current staff annual renumeration.**

Current total annual staff remuneration: $94,260,404.52

Current median monthly salary: $5,238.00

Staff Directory:

1. Current staff strength.

2. Current staff annual renumeration.

3. Eligible long service award recipients.

4. Exit.

Choose option 1 to 4:

**Output from selecting option 3:**

**You have selected option 3: Eligible long service award recipients.**

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

Staff Directory:

1. Current staff strength.

2. Current staff annual renumeration.

3. Eligible long service award recipients.

4. Exit.

Choose option 1 to 4:

**Output from selecting option 4:**

**You have selected option 4: Exit.**

Exit.

### 

### APPENDIX 1 – QUESTION 1 (B) CODES

**Output 1 Codes:**

*#import necessarily libraries into project*

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

*#retrieving the specific dataset that i am working on*

df = pd.read\_csv("GBA.csv", usecols = ["EmpID", "Age"])

*#converting rows x columns to list using tolist()*

empID\_list = df['EmpID'].values.tolist()

age\_list = df['Age'].values.tolist()

*#label the x and y axis*

plt.xlabel("Age Group")

plt.ylabel("Frequency of Employees")

*#set a title to the chart*

plt.title("Employeee Age Distribution")

*#define bin for the histogram*

bins = [10, 20, 30, 40, 50, 60]

plt.hist(age\_list, bins=bins, color = 'orange', edgecolor='black')

*#set each age grp in range in the x-axis*

labels = ['<20', '21-30', '31-40', '41-50', '51-60', '>60']

*#inputing the data and center the tick labels*

plt.xticks(bins, labels, ha='center')

plt.show()

\*tried to centre the x-axis label however unable to achieve it.

**Output 2 Codes:**

import pandas as pd

import matplotlib.pyplot as plt

comparison = pd.read\_csv("GBA.csv", usecols = ["Education", "MonthlyPay"])

x = list(comparison['Education'])

y = list(comparison['MonthlyPay'])

import numpy as np

plt.scatter(x, y, c="blue", alpha=0.5)

plt.title('Education Level vs Month Pay') #title

plt.xlabel('Education') #x label

plt.ylabel('Monthly Pay') #y label

z = np.polyfit(x, y, 1)

p = np.poly1d(z)

plt.plot(x, p(x))

plt.show()

**Output 3 Codes:**

import pandas as pd

import matplotlib.pyplot as plt

import matplotlib.ticker as mtick

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

# Which columns to plot and group them

grouped\_csv = csv.groupby('JobRole')['LeftCompany'].value\_counts(normalize=True).unstack()

# Create the 100% stacked bar chart

chart = grouped\_csv.plot(kind='bar', stacked=True)

chart.yaxis.set\_major\_formatter(mtick.PercentFormatter(xmax=1.0))

# Add title and labels

plt.title('Leavers of Company by Job Role')

plt.xlabel('Job Role')

plt.ylabel('Percentage of Employees')

plt.legend(title='Left Company', loc='upper right')

plt.show()

**Output 4 Codes:**

import pandas as pd

import matplotlib.pyplot as plt *#import matplotlib*

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

*#Create x and y values*

Male = [282, 272, 106, 48, 39]

Female = [ 170, 191, 79, 42, 21]

index = ['1', '2', '3', '4', '5']

df = pd.DataFrame({'Male': Male,

'Female': Female,},

index=index)

ax = df.plot.bar(rot=0)

#label the x and y axis

plt.xlabel("Performance Rating")

plt.ylabel("Number of Employees")

#set a title to the chart

plt.title("Performance Rating by Gender")

*#output chart*

plt.show()

### 

### APPENDIX 2 – QUESTION 1 (C) CODES

import pandas as pd

csv = pd.read\_csv("GBA.csv")

def option1():

*#Current Total Staff Strength*

no\_left\_company = (csv['LeftCompany'] == 'No').sum()

*#Grouping data in Business Unit and Gender*

bu = csv.groupby('BusinessUnit')

gender = csv.groupby('Gender')

*#Count number of staffs if Left Company = No*

countbu = bu['LeftCompany'].apply(lambda x: (x =='No').sum())

countgender = gender['LeftCompany'].apply(lambda x: (x =='No').sum())

print('Current Total Staff Strength:', no\_left\_company)

print('\nBreakdown of staff strength by business unit:')

for x, y in countbu.iteritems():

print(f'{x}: {y}')

print('\nBreakdown of staff strength by gender:')

for x, y in countgender.iteritems():

print(f'{x}: {y}')

def option2():

*# Select only currently employed employees and the relevant columns*

no\_left\_company = csv.loc[csv['LeftCompany'] == 'No', ['MonthlyPay', 'SalaryIncrement']]

*# Calculate the total annual staff remuneration (i.e., total salary plus any salary increments) for currently employed employees*

total\_remuneration = ((no\_left\_company['MonthlyPay']) +

(((no\_left\_company['MonthlyPay']) \* (no\_left\_company['SalaryIncrement'])/100)))

*# Calculate the median monthly salary (i.e., the middle value of the monthly pay column) for currently employed employees*

median\_monthly\_salary = no\_left\_company['MonthlyPay'].median()

*# Display results*

print(f'Current total annual staff remuneration: ${total\_remuneration.values.sum() \* 12:,.2f}')

print(f'Current median monthly salary: ${median\_monthly\_salary:,.2f}')

def option3():

print("Eligible long service award recipients.")

*# Filter WithCompany >= 35*

filter\_csv = csv.loc[csv['WithCompany'] >= 35, ['EmpID', 'JobRole', 'Age']]

filter\_csv = filter\_csv.sort\_values(by='Age', ascending=False)

*# List results*

print(filter\_csv.to\_string(index=False))

def option4():

print("Exit.")

*#UserInterface*

while True:

option = input('\nStaff Directory:\n1. Current staff strength.\n2. Current staff annual renumeration.\n3. Eligible long service award recipients.\n4. Exit.\nChoose option 1 to 4: ')

*# Loop for options 1-4 with bold text*

if option == '1':

print("\033[1mYou have selected option 1: Current staff strength.\033[0m")

option1()

elif option == '2':

print("\033[1mYou have selected option 2: Current staff annual renumeration.\033[0m")

option2()

elif option == '3':

print("\033[1mYou have selected option 3: Eligible long service award recipients.\033[0m")

option3()

elif option == '4':

print("\033[1mYou have selected option 4: Exit.\033[0m")

option4()

break

else:

print("\033[1mInvalid option. Please enter a number between 1 and 4.\033[0m")