

Group-Based Assignment (GBA01)

For Module:

ANL252, Python for Data Analytics

Submitted by:

Shamini D/O Eliaperumal (E2170152)

Ng Kuo Xun (H2071354)

Nur Firdaus Bin Mohamed Riduan (B2170520)

Date of Submission:

19 February 2023



**Declaration Page**

We, members of group 9, do hereby declare that we each contributed to this assignment and that we collectively agree to a shared grade.

|  |  |  |
| --- | --- | --- |
| Name | Contribution | Signature |
| Shamini D/O Eliaperumal (Team Lead) | I did question 1(a), (b) and (c) | Shamini Eliaperumal |
| Ng Kuo Xun | I did question 1(a), (b) and (c) | Ng Kuo Xun |
| Nur Firdaus Bin Mohamed Riduan | I did question 1(a), (b) and (c) | Nur Firdaus |

**Question 1(a)**

Chart, scatter chart

Description automatically generated

Chart 1

Graphical user interface, text, application

Description automatically generated

Table 1

* The same average salary increase of 3% is given to employees with a performance rating of 1 or 2.
* The average salary increment percentage rises from 3 to 5 as the performance rating does. This suggests that bigger compensation increases are given to employees who receive better performance reviews.
* Between performance ratings 3 and 4, there is the biggest variation in the average salary increase.
* A performance rating of 4 receives an average salary increment percentage of 14.22%, which is over twice as much as the 7.25% obtained by individuals with a performance rating of 3.

Chart, bar chart, funnel chart

Description automatically generated

Chart 2

Table

Description automatically generated

Table 2

* We can see from the table that there are often more male current managers than female current managers.
* Irrespective of gender, there are more current managers with higher education levels (3 and 4) than there are with lower education levels (1 and 2).
* We can observe that there are more male current managers than female current managers across all educational levels.
* According to a research published in Personnel Psychology in 2018, men were more likely to be selected or evaluated as leaders overall, partly because they were more assertive and therefore spoke up more. Women were more likely to be perceived as leaders only after having more in-depth interactions. (Hudson, 2018)

Chart, line chart

Description automatically generated

Chart 3

Table

Description automatically generated

Table 3

* Above table shows the average of an employee’s job grade and monthly pay by their number of years of working experiences.
* We can infer that from working experience years 0 to 20, there is a steady increase in the job grade and monthly pay.
* But from working experience years 21 to 40 we can see fluctuations both up and down in both job grade and monthly pay.
* After a certain point in an employee's career, we may infer that having more years of experience doesn't always convert into a better monthly salary or a higher job grade.
* The data indicates that at a certain juncture in a career, salary remains fairly stagnant.

Chart

Description automatically generated

Chart 4

Table

Description automatically generated

Table 4

* We can infer that the Product Development business unit has 79 employees who rate their work life balance as high (rating of 4), followed by Business Development with 45 ratings of 4. This is not significant considering the total number of employees in the unit.
* Overall, we can see that the employees have a relatively good work-life balance, with a significant number of employees giving a rating of 3.
* Additionally, it appears that the Product Development business unit may have the most significant room for improvement, as they have the highest number of respondents rating their work-life balance as 1 and 2.

**Question 1(b)**

**Chart 1:**

**Chart, line chart

Description automatically generated**

**Chart 2:**

**Chart, bar chart

Description automatically generated**

**Chart 3:**

**A picture containing text, wall

Description automatically generated**

**Chart, line chart

Description automatically generated**

**Chart 4:**

**Chart, waterfall chart

Description automatically generated**

**Question 1(c)**

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

1

Option selected --> 1. Current staff strength

Current total staff strength: 1057

Business Development Female 129

Male 170

HR Female 12

Male 30

Product Development Female 293

Male 423

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

2

Option selected --> 2. Current staff annual renumeration

Current total annual staff runumeration expenditure = $ 87,182,508

Current median monthly salary = $ 5,238

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

3

Option selected --> 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

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

9

Invalid option, please try again.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Please select an option:

1. Current staff strength

2. Current staff annual renumeration

3. Eligible long service award recipients

4. Exit

4

Thank you and have a pleasant day ahead :)

# References

Hudson, M. (2018, October 5). *Why Are There So Few Female Leaders?* Retrieved from Scientific American: https://www.scientificamerican.com/article/why-are-there-so-few-female-leaders/

# Appendix

**Question (1b)**

**Chart 1:**

import pandas as pd

import matplotlib.pyplot as plt

import matplotlib as mpl

import numpy as np

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

filtered1 = df[(df['PerformanceRating'] == 1)]

filtered2 = df[(df['PerformanceRating'] == 2)]

filtered3 = df[(df['PerformanceRating'] == 3)]

filtered4 = df[(df['PerformanceRating'] == 4)]

filtered5 = df[(df['PerformanceRating'] == 5)]

avgsal1 = round(filtered1["SalaryIncrement"].mean(),2)

avgsal2 = round(filtered2["SalaryIncrement"].mean(),2)

avgsal3 = round(filtered3["SalaryIncrement"].mean(),2)

avgsal4 = round(filtered4['SalaryIncrement'].mean(),2)

avgsal5 = round(filtered5['SalaryIncrement'].mean(),2)

data = {'Performance Rating': ['1', '2', '3', '4', "5"],

'Average of Salary Increment (%)': [ avgsal1, avgsal2, avgsal3, avgsal4, avgsal5]

}

df2 = pd.DataFrame(data)

df2['Average of Salary Increment (%)'] = df2['Average of Salary Increment (%)'].map('{:.2f}%'.format)

style\_object = df2.style.set\_table\_styles([{'selector': 'th', 'props': [('border', '1px solid black')]},

{'selector': 'td', 'props': [('border', '1px solid black')]},

{'selector': 'th:first-child', 'props': [('display', 'none')]},

{'selector': 'td:first-child', 'props': [('display', 'none')]}]).set\_properties(subset=["Performance Rating"], \*\*{'text-align': 'center'})

mpl.rcParams['figure.facecolor'] = 'white'

style\_object.set\_properties(\*\*{'background-color': 'white',

'color': 'black',

'border-color': 'black'})

display(style\_object)

y = avgsal1, avgsal2, avgsal3, avgsal4, avgsal5

x = 1.0, 2.0, 3.0, 4.0, 5.0

plt.scatter(x,y)

slope, intercept = np.polyfit(x, y, 1)

plt.plot(x, slope\*np.array(x) + intercept)

plt.xlim(0.5, 5.5)

plt.title('Relationship between Performance Rating and Salary Increment')

plt.xlabel('Salary Increment')

plt.ylabel('Performance Rating')

plt.grid(True)

plt.show()

**Chart 2:**

import pandas as pd

import matplotlib.pyplot as plt

import matplotlib as mpl

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

filtered1 = df[(df['Gender'] == 'Male') & (df['Education'] == 1)]

filtered2 = df[(df['Gender'] == 'Male') & (df['Education'] == 2)]

filtered3 = df[(df['Gender'] == 'Male') & (df['Education'] == 4)]

filtered4 = df[(df['Gender'] == 'Male') & (df['Education'] == 3)]

filtered5 = df[(df['Gender'] == 'Female') & (df['Education'] == 1)]

filtered6 = df[(df['Gender'] == 'Female') & (df['Education'] == 2)]

filtered7 = df[(df['Gender'] == 'Female') & (df['Education'] == 4)]

filtered8 = df[(df['Gender'] == 'Female') & (df['Education'] == 3)]

male1 = len(filtered1)

male2 = len(filtered2)

male3 = len(filtered3)

male4 = len(filtered4)

female1 = len(filtered5)

female2 = len(filtered6)

female3 = len(filtered7)

female4 = len(filtered8)

data = {'Education': ['3', '4', '2', '1'],

'Female': [ female4, female3, female2, female1],

'Male' : [male4, male3, male2, male1] ,

'Grand Total' : [male4+ female4, male3+ female3, male2+ female2, male1+ female1]

}

df2 = pd.DataFrame(data)

style\_object = df2.style.set\_table\_styles([{'selector': 'th', 'props': [('border', '1px solid black')]},

{'selector': 'td', 'props': [('border', '1px solid black')]},

{'selector': 'th:first-child', 'props': [('display', 'none')]},

{'selector': 'td:first-child', 'props': [('display', 'none')]}]).set\_properties(subset=["Education"], \*\*{'text-align': 'center'})

mpl.rcParams['figure.facecolor'] = 'white'

style\_object.set\_properties(\*\*{'background-color': 'white',

'color': 'black',

'border-color': 'black'})

display(style\_object)

a = [female1, female2, female3, female4]

b = [male1, male2, male3, male4]

df3 = pd.DataFrame({'Female' : a,'Male' : b})

fig, ax = plt.subplots()

fig.set\_size\_inches(10,6)

df3.plot.barh(stacked=True, ax=ax);

ax.set\_title("Relationships beetween No. of Current Managers by Education Level and Gender")

box = ax.get\_position()

ax.set\_position([box.x0, box.y0 + box.height \* 0.1,

box.width, box.height \* 0.9])

ax.legend(loc='upper center', bbox\_to\_anchor=(0.5, -0.05),

fancybox=True, shadow=True, ncol=5)

plt.show()

**Chart 3:**

import pandas as pd

import matplotlib.pyplot as plt

import matplotlib as mpl

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

new\_df = df[['WorkingExperience', 'JobGrade', 'MonthlyPay']]

new\_df.to\_csv('Chart3.csv', index=False)

table = round(df.groupby('WorkingExperience')[['JobGrade', 'MonthlyPay']].mean(), 2).reset\_index()

style\_object = table.style.format({'JobGrade': '{:.2f}', 'MonthlyPay': '{:.2f}'})\

.set\_table\_styles([{'selector': 'th', 'props': [('border', '1px solid black')]},

{'selector': 'td', 'props': [('border', '1px solid black')]},

{'selector': 'th:first-child', 'props': [('display', 'none')]},

{'selector': 'td:first-child', 'props': [('display', 'none')]}])\

.set\_properties(subset=["JobGrade"], \*\*{'text-align': 'center'})

mpl.rcParams['figure.facecolor'] = 'white'

style\_object.set\_properties(\*\*{'background-color': 'white',

'color': 'black',

'border-color': 'black'})

display(style\_object)

fig, ax1 = plt.subplots()

ax2 = ax1.twinx()

ax1.plot(table.index, table['MonthlyPay'], label='MonthlyPay', color='orange')

ax1.set\_xlabel('WorkingExperience')

ax1.set\_ylabel('Average Value (MonthlyPay)', color='orange')

ax2.plot(table.index, table['JobGrade'], label='JobGrade', color='blue')

ax2.set\_ylabel('Average Value (JobGrade)', color='blue')

plt.title('Relationship between Working Experience, Job Grade and Monthly Pay')

lines, labels = ax1.get\_legend\_handles\_labels()

lines2, labels2 = ax2.get\_legend\_handles\_labels()

ax2.legend(lines + lines2, labels + labels2, loc='best')

plt.show()

**Chart 4:**

import pandas as pd

import matplotlib.pyplot as plt

import math

import matplotlib as mpl

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

new\_df = df[['BusinessUnit', 'WorkLifeBalance']]

new\_df.to\_csv('Chart4.csv', index=False)

counts = pd.pivot\_table(new\_df, index='BusinessUnit', columns= 'WorkLifeBalance', aggfunc=len, fill\_value=0)

totals = counts.sum()

counts = pd.concat([counts, pd.DataFrame(totals).T])

counts = counts.rename({0: 'Total'}, axis=0)

style\_object = counts.style.set\_table\_styles([{'selector': 'th', 'props': [('border', '1px solid black')]},

{'selector': 'td', 'props': [('border', '1px solid black')]},

{'selector': 'th:first-child', 'props': [('display', 'table-cell')]},

{'selector': 'td:first-child', 'props': [('display', 'table-cell')]}])

style\_object.set\_properties(\*\*{'background-color': 'white',

'color': 'black',

'border-color': 'black'})

display(style\_object)

counts = counts.drop('Total', axis=0)

counts.plot(kind='bar')

plt.xlabel('Business Unit')

plt.ylabel('Count')

plt.show()

**Question (1c)**

import pandas as pd

from tabulate import tabulate

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

while True:

print()

print('Please select an option:')

print('1. Current staff strength')

print('2. Current staff annual renumeration')

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

print('4. Exit')

print()

option = input()

if option == '1':

one = employee\_data[employee\_data['LeftCompany'] == 'No']

one = pd.DataFrame(one[['BusinessUnit','Gender','LeftCompany']])

rows = len(one.axes[0])

print("Option selected --> 1. Current staff strength")

print()

print('Current total staff strength: ', rows)

print()

one\_BUU = one.groupby(['BusinessUnit','Gender'])['BusinessUnit'].count()

print(one\_BUU.to\_string(header=False))

print()

print("\n \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* \n")

elif option == '2':

two = employee\_data

employee\_data['AnnualPay'] = employee\_data['MonthlyPay']\*12

two = two[two['LeftCompany'] == 'No']

two\_annual = two["AnnualPay"].sum()

two\_median = two["MonthlyPay"].median()

print("Option selected --> 2. Current staff annual renumeration")

print()

print("Current total annual staff runumeration expenditure = $ {:>10,.0f}".format(two\_annual))

print("Current median monthly salary = $ {:>10,.0f}".format(two\_median))

print()

print("\n \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* \n")

elif option == '3':

pd.set\_option('display.max\_rows', None)

three = employee\_data[employee\_data['WithCompany'] >= 35]

three = three[['EmpID','JobRole','Age']]

three = three.sort\_values(['Age','EmpID'], ascending = [False, True])

print("Option selected --> 3. Eligible long service award recipients")

print()

print('Eligible long service award recipients:')

print()

print(tabulate(three, showindex=False, headers=three.columns, numalign="left"))

print()

print("\n \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* \n")

elif option == '4':

print()

print('Thank you and have a pleasant day ahead :)')

break

else:

print()

print('Invalid option, please try again.')

print()

print("\n \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* \n")