Logo, company name

Description automatically generated

**ANL252**

**Python for Data Analytics**

**Group-Based Assignment**

**January 2023 Presentation**

|  |  |
| --- | --- |
| **Student Name** | **PI No.** |
| Chan Huan Hueng Brandon | J1981618 |
| Tay Ban Chin Dave | Q2172629 |
| Teo Wei Liang Jack | B1971989 |

**Submission Date: 19 February 2023**

Declaration Page

We, members of GBA Group 1, do hereby declare that we each contributed to this assignment and that we collectively agree to a shared grade.

|  |  |  |
| --- | --- | --- |
| Name | Contribution | Signature |
| Chan Huan Hueng Brandon | I did question 1(a) |  |
| Tay Ban Chin Dave | I did question 1(b) | Shape  Description automatically generated with low confidence |
| Teo Wei Liang Jack | I did question 1(c) |  |

**Question 1**

**(a)**

**Analysis 1 – Satisfaction with Company**

|  |  |
| --- | --- |
| **Rating** | **Employees** |
| 1 | 235 |
| 2 | 247 |
| 3 | 386 |
| 4 | 382 |

Chart, pie chart

Description automatically generated

From Analysis 1 above, we observe the following:

* The majority (60%) of employees gave 3 and 4 high satisfaction ratings of the company
* This is a good sign that the company is doing a good job of creating a positive work environment and culture. The company can use this information to continue building on the company's strengths.
* The remaining 40% are not quite satisfied. This suggests that the company should focus on areas for improvement. The company should investigate causes of lower satisfaction ratings, such as salary, satisfaction with their managers, and other factors. By addressing these issues, the company will be able to increase overall employee satisfaction and retain top talent.

**Analysis 2 – Salary Increment and Performance Rating**

|  |  |
| --- | --- |
| **SalaryIncrement** | **PerformanceRating** |
| 7 | 3 |
| 13 | 4 |
| 6 | 3 |
| 8 | 3 |

… and so on for total 1250 rows of data, from two columns extracted from the GBA dataset.

Chart, scatter chart

Description automatically generated

From Analysis 2 above, we observe the following:

* Generally, the higher the salary increment, the higher the performance rating of the employees.
* There is a threshold for performance ratings and salary increments. Most employees have a performance rating of 3 and a salary increase ranging from 4% to 13%.
* Employees with salary increases below this range seem to perform worse, whereas those with salary increases above this range appear to perform better. This data could be used by the company to review its salary increment policies and consider raising the minimum increment to boost employee performance.

**Analysis 3 – Working Experience and Performance Rating**

|  |  |
| --- | --- |
| **WorkingExperience** | **PerformanceRating** |
| 21 | 3 |
| 5 | 4 |
| 10 | 3 |
| 6 | 3 |

… and so on for total 1250 rows of data, from two columns extracted from the GBA dataset.

Timeline

Description automatically generated

From Analysis 3 above, we observe the following:

1. Large majority of employees with top performance have around 28 years or less working experience.

2. Most employees with more than 28 years of experience have average performance.

3. One interesting finding is that employees with more than 28 years of experience have average performance ratings, implying that the company's work environment and engagement strategies need to be improved. This may involve addressing issues such as work-life balance, career development opportunities, and recognition and rewards. This way, company can improve employee performance while also retaining valuable expertise.

**Analysis 4 – Satisfaction with Job Role**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Count of Satisfaction Role** | **Job Role** |  |  |  |  |  |  |  |  |  |
| **Satisfaction with Role** | **Account Executive** | **Account Manager** | **HR** | **Manufacturing Head** | **Product Executive** | **Product Head** | **Product Research** | **Section Manager** | **Technical Support** | **Grand Total** |
| 1 | 11 | 63 | 10 | 23 | 22 | 12 | 44 | 16 | 46 | 247 |
| 2 | 16 | 42 | 14 | 27 | 19 | 14 | 49 | 17 | 44 | 242 |
| 3 | 25 | 81 | 9 | 45 | 38 | 25 | 72 | 20 | 61 | 376 |
| 4 | 20 | 90 | 9 | 37 | 38 | 18 | 81 | 28 | 64 | 385 |
| **Grand Total** | **72** | **276** | **42** | **132** | **117** | **69** | **246** | **81** | **215** | **1250** |

Chart, bar chart, treemap chart

Description automatically generated

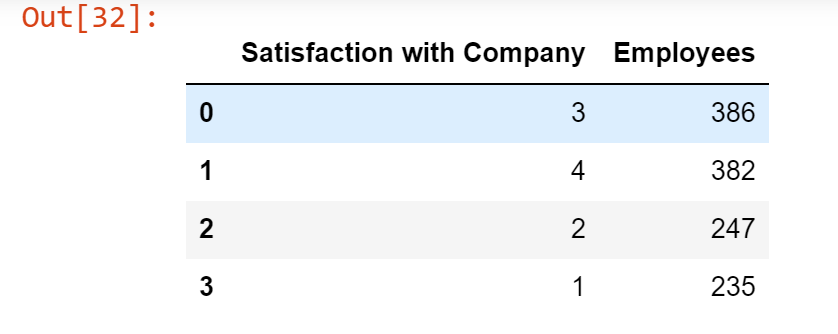
From Analysis 4 above, we observe the following:

* Overall, majority of employees are satisfied with their job roles.
* There are slightly more dissatisfied than satisfied HR employees. The company may need to investigate why more HR employees are dissatisfied and take steps to improve their job satisfaction.
* Compared to other job roles, there are relatively more Account Managers who are most dissatisfied. The company may need to address the reasons why Account Managers are the least satisfied in their roles. Understanding their specific concerns and responding to them may entail increasing their job responsibilities, providing additional training or support, or offering incentives or rewards to increase motivation and engagement.

**(b)**

**See Appendix 1**

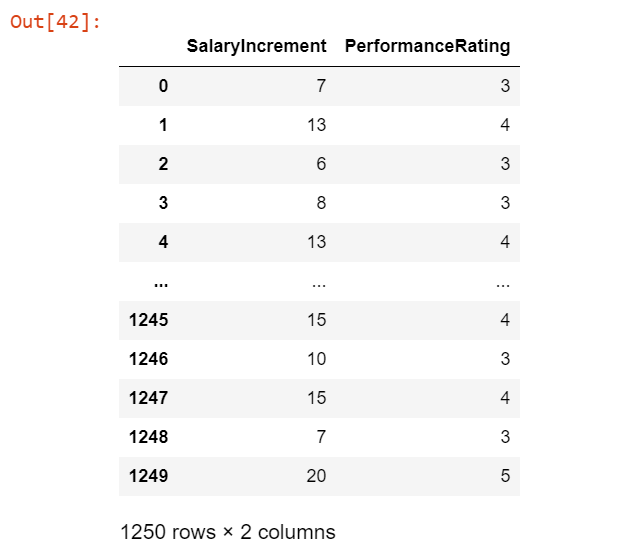
**Analysis 1 – Satisfaction with Company**



Chart, pie chart

Description automatically generated

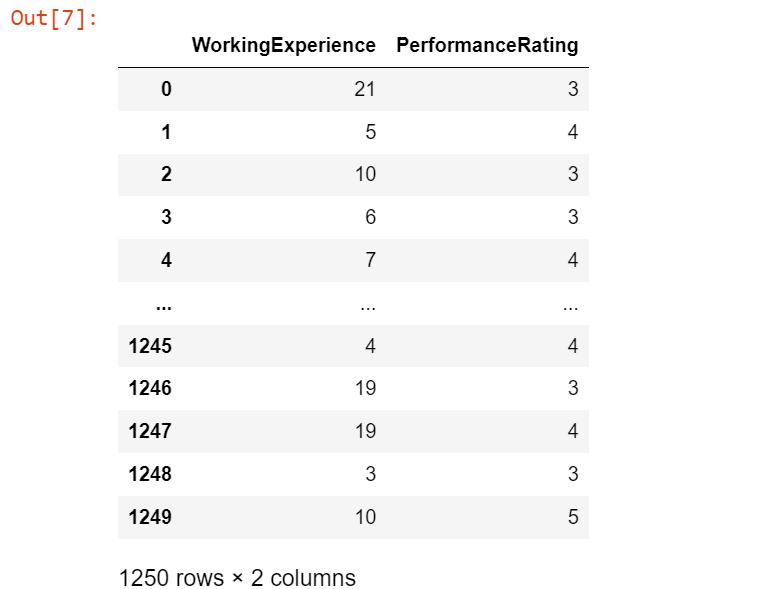
**Analysis 2 – Salary Increment and Performance Rating**



Chart, scatter chart

Description automatically generated

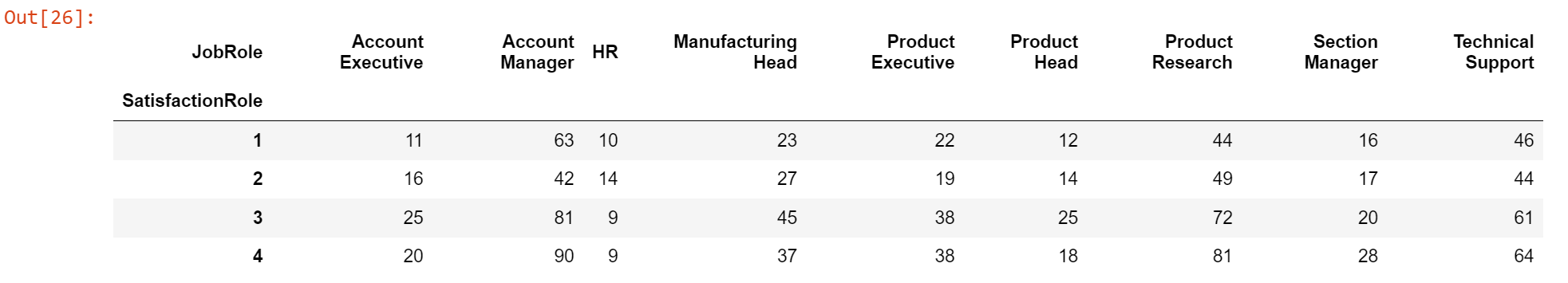
**Analysis 3 – Working Experience and Performance Rating**



Chart, scatter chart

Description automatically generated

**Analysis 4 – Satisfaction with Job Role**

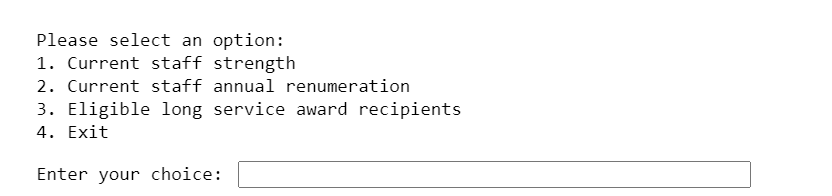


Chart, bar chart

Description automatically generated**(c)**

**See Appendix 2**

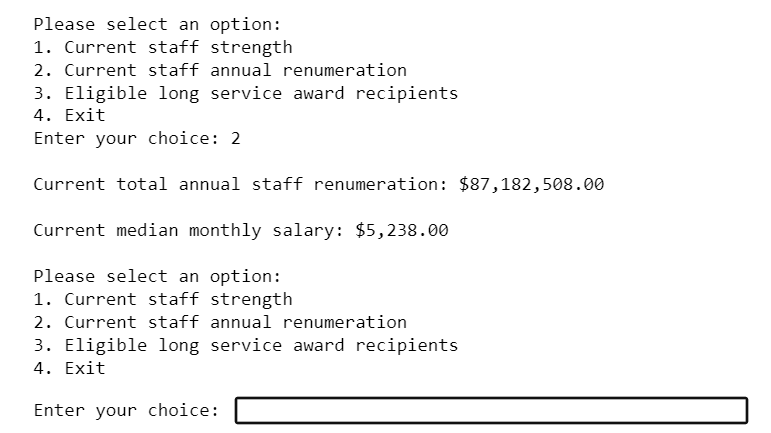
Upon running the program



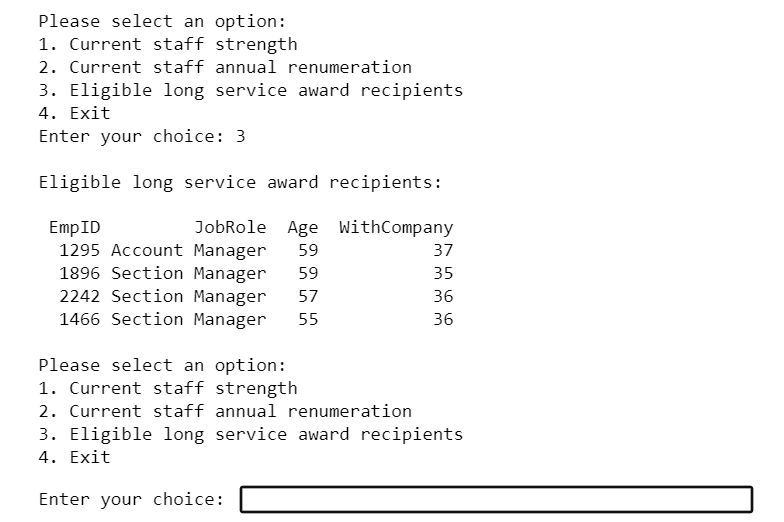
Input ‘1’

****

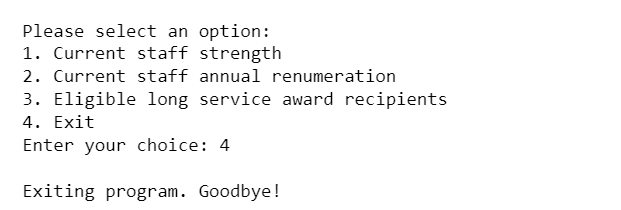
Input ‘2’



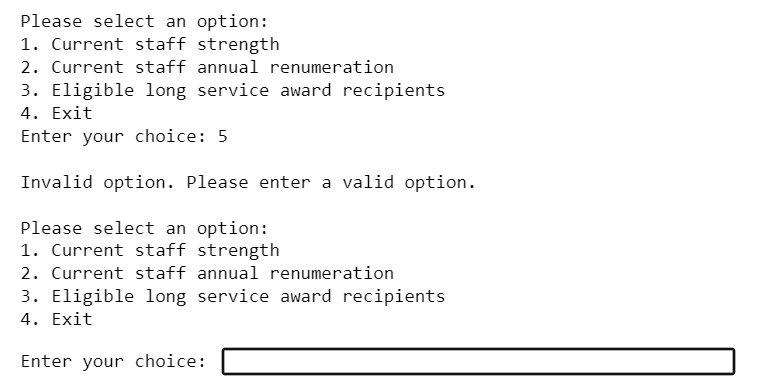
Input ‘3’

****

Input ‘4’



Any other input



**Appendix 1**

Remarks:

* Appending code for Question 1(b)
* This is a continuous code in text format copied from a jupyter notebook file
* Dataframes are created once in the earliest line required
* Libraries and packages are imported once in the earliest line required

Code:

import pandas as pd

# load the dataset into a pandas dataframe

empdata = pd.read\_csv('gba.csv')

# check for missing values

missing\_values = empdata.isnull().sum()

print("Number of missing values in each column: \n", missing\_values)

# Get the value count for each rating in the 'SatisfactionCompany' column

sat\_co = empdata['SatisfactionCompany'].value\_counts()

# Get the value count for each rating in the 'SatisfactionCompany' column

sat\_co = empdata['SatisfactionCompany'].value\_counts()

import matplotlib.pyplot as plt

# Get the value count for each rating in the 'SatisfactionCompany' column

sat\_co = empdata['SatisfactionCompany'].value\_counts()

# Plot the pie chart

plt.pie(sat\_co.values, labels=sat\_co.index, autopct='%1.1f%%')

# Add title to the chart

plt.title('Satisfaction with Company')

# Show the plot

plt.show()

# Select the columns to be used for the scatter plot

inc\_per = empdata[['SalaryIncrement', 'PerformanceRating']]

inc\_per

# Select the columns to be plotted

x = inc\_per['SalaryIncrement']

y = inc\_per['PerformanceRating']

# Plot the data as a scatter plot

plt.scatter(x, y)

# Add labels and title to the plot

plt.xlabel('Salary Increment')

plt.ylabel('Performance Rating')

plt.title('Scatter Plot of Salary Increment and Performance Rating')

# Show the plot

plt.show()

# Select the columns to be used for the scatter plot

wxp\_per = empdata[['WorkingExperience', 'PerformanceRating']]

wxp\_per

# Select the columns to be plotted

x = wxp\_per['WorkingExperience']

y = wxp\_per['PerformanceRating']

# Plot the data as a scatter plot

plt.scatter(x, y)

# Add labels and title to the plot

plt.xlabel('Working Experience')

plt.ylabel('Performance Rating')

plt.title('Scatter Plot of Working Experience and Performance Rating')

# Show the plot

plt.show()

# Flatten the 'SatisfactionRole' column

empdata['SatisfactionRole\_flat'] = empdata['SatisfactionRole'].apply(pd.Series).stack().reset\_index(drop=True)

# Create the pivot table

satrolepivot = pd.pivot\_table(empdata, values='SatisfactionRole\_flat', index='SatisfactionRole', columns='JobRole', aggfunc='count')

# Remove any rows or columns with all NaN values

satrolepivot.dropna(how='all', axis=0, inplace=True)

satrolepivot.dropna(how='all', axis=1, inplace=True)

# Print the pivot table

satrolepivot

# Create the stacked horizontal bar chart

ax = satrolepivot.plot(kind='barh', stacked=True, figsize=(10,8))

# Add individual values to the bars

for i in ax.containers:

ax.bar\_label(i, label\_type='center', fontsize=10)

# Add title and axis labels

plt.title('Job Role Satisfaction by Job Role')

plt.xlabel('Number of Employees')

plt.ylabel('Satisfaction Rating')

# Show the plot

plt.show()

**Appendix 2**

Remarks:

* Appending code for Question 1(c)
* This is a continuous code in text format copied from a jupyter notebook file
* Dataframe is created once in the earliest line required

import pandas as pd

# load the dataset into a pandas dataframe

empall = pd.read\_csv('gba.csv')

# filter out employees who have left the company

empdata = empall[empall['LeftCompany'] == 'No']

while True:

# prompt user for input

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")

user\_input = input("Enter your choice: ")

# execute code based on user input

if user\_input == '1':

# calculate total staff strength and breakdown by business unit and gender

staff\_by\_bu = empdata.groupby('BusinessUnit')['EmpID'].count().to\_frame()

staff\_by\_gender = empdata.groupby('Gender')['EmpID'].count().to\_frame()

# print information

print(f"\nCurrent total staff strength: {len(empdata)}\n")

print("Breakdown by business unit:\n")

print(staff\_by\_bu.to\_string())

print("\nBreakdown by gender:\n")

print(staff\_by\_gender.to\_string())

elif user\_input == '2':

# calculate total annual staff renumeration and median monthly salary

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

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

# print information

print(f"\nCurrent total annual staff renumeration: ${total\_renumeration:,.2f}\n")

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

elif user\_input == '3':

# filter employees with 35 or more years of service listed in the 'WithCompany' column of dataset, and sort by age

long\_service = empdata[empdata['WithCompany'] >= 35][['EmpID', 'JobRole', 'Age', 'WithCompany']]

long\_service = long\_service.sort\_values(by='Age', ascending=False)

# print information

print("\nEligible long service award recipients:\n")

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

elif user\_input == '4':

# exit program

print("\nExiting program. Goodbye!")

break

else:

# if user input is not one of the options above, prompt user to enter a valid option

print("\nInvalid option. Please enter a valid option.")