

ZZBU6511 Predictive Analytics

Assignment 2 Data Exploration
Globex Pharma Employee Survey

Name: Amanda Wright

Due Date: 27th May 2024

Executive Summary

Attached is an exploratory analysis of potential indicators of employee discontent at Globex Pharma.

Combining the Globex Employee Survey 2023 and literature on common reasons for employee turnover in Australia, analysis has been completed on employee pay, benefits (stock option levels) and work life balance as possible causes of discontent at Globex Pharma.

We have concluded that pay within certain roles may be inequitable between gender, the stock option levels of employees were inconsistent across usual merit-based criteria such as tenure or performance, and work life balance in the lower levels of the organization may be contributing to discontent.

Based on these indicators, we have provided recommendations to review salary bandings in specific roles, initiate a senior level review of criteria for awarding stock option levels and communicated to the wider organisation upon completion of review, and an operational review of workflow distribution and salaries in the lower income roles.

Table of Contents

- Executive Summary
- Introduction
- Research Questions
- Conclusion
- Recommendations
- References
- Appendices

Introduction

Australian Pharmaceutical giant Globex Pharma, recognising the impacting costs of employee attrition, has requested recommendations to improve employee retention of their one thousand employees after increasing talk of dissatisfaction within the employee ranks.

Focusing on employee retention has a top line benefit and a bottom-line benefit to every business, particularly when 90% of businesses were affected by staff shortages in 2023(“Solving the skills shortage crisis | Ai Group,” n.d.).

According to an IPSOS survey from Dec 2023 on Australians in the Workforce (2222 *Australia Mini Report 260124 ONLINE VERSION.pdf*, no date), the top reasons why employees leave their jobs are:

- Pay and Compensation
- Bonus Schemes or Merit Based Benefits
- Not feeling valued or recognised
- Work Life Balance/Workload

Using these four key reasons as a guide in conjunction with the Globex Pharma Employee Listening Survey 2023, we here at Wright Consultancy will provide recommendations based on three potential factors which may be causing employee discontent at Globex Pharma.

Research Questions

We will focus our analysis on the following three questions:

1. Is there an imbalance of monthly incomes within job roles?
2. How are stock option levels of an employee determined - is it a merit-based system awarded on specific criteria?
3. What, if anything, do the employees who rated their Work Life Balance as “Bad” have in common?

Q1. Is there an imbalance of monthly incomes within job roles?

Key Findings

Analysis of the distribution of monthly income by nine job roles: (Fig. 1):

- The spread between the incomes is reasonably compact and comparable in the lower income roles and middle-income roles (Fig.1)
- The spread of the incomes of the “Research Directors” ranges from \$10,000 to \$20,000 a month, the widest spread in the organization.
- There are 61 Research Directors – 21 Female, 30 Male (Fig. 2)

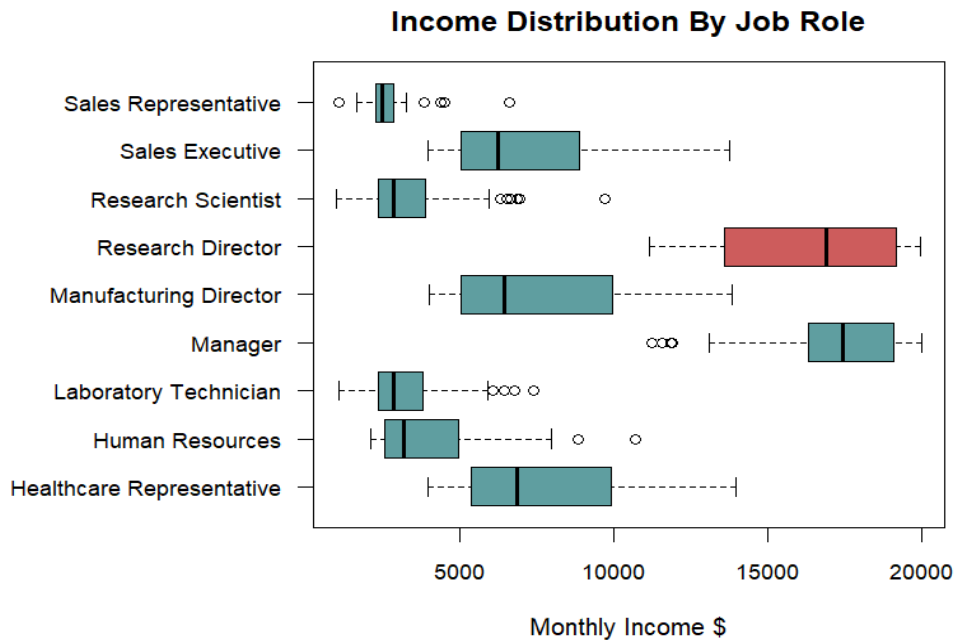


Fig. 1 Distribution of Incomes by Job Role

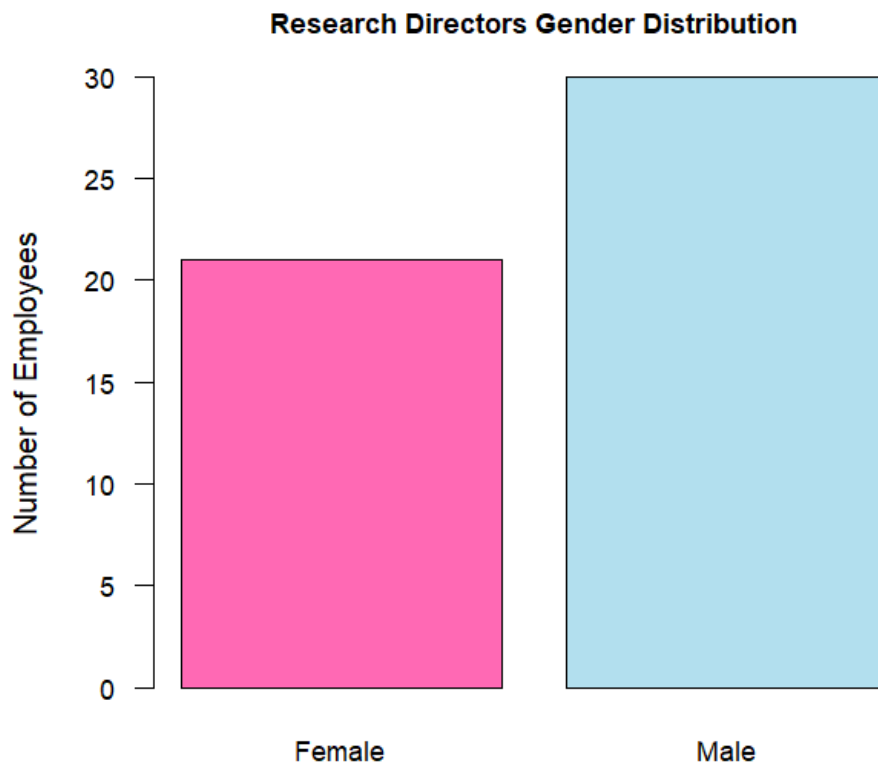


Fig. 2 Research Director Population at Globex



Fig. 3 Histogram of Male and Female Research Director Monthly Income Distributions

- “Female Research Directors” are on average younger, have been in the role much longer than “Male Research Directors” (Fig.4), however have significantly lower distribution of monthly incomes(Fig.3)



Fig. 4 Median Years in Role and Age Research Directors

Q2. How are stock option levels of an employee determined - is it a merit-based system awarded on specific criteria?

Background

As described by human resource literature, stock options:

- Typically, are awarded as a component of an overall remuneration package. (“Everything You Need to Know About Stock Options and RSUs,” n.d.)
- Ultimately designed to improve retention of employees particularly when wage competition is high in the market (“Why Do Companies Continue to Use Stock Option Incentives?” 2024)

According to the IPSOS survey (2222 *Australia Mini Report 260124 ONLINE VERSION.pdf*, no date) employee benefits are in the top reasons employees left their jobs.

There are four stock option levels at Globex – None, Some, High and Very High.

Key Findings

- With 43.2% of employees having no stock options (Fig. 5), analysis of possible merit-based criteria for earning stock options were inconclusive from the data provided.
- Age, Monthly Income, Years at Company, Performance Rating and Job Role are not criteria aligned with stock option levels from our analysis.

Percentage Employees By Stock Option Level

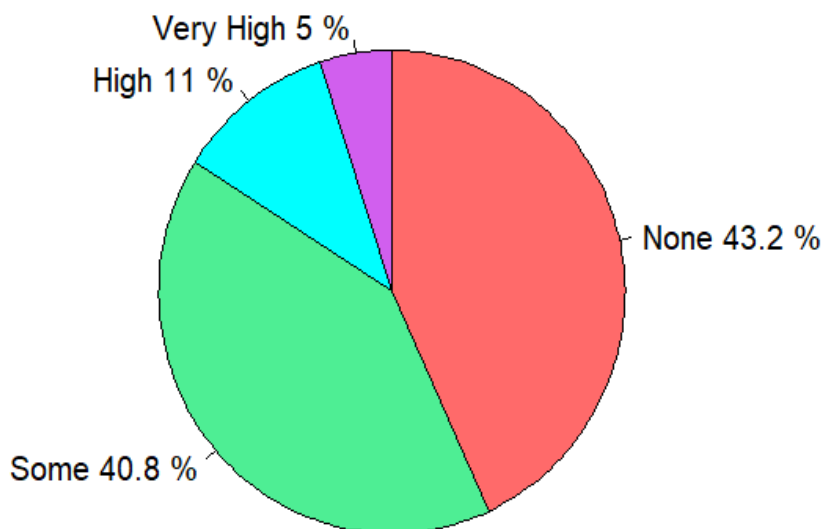


Fig. 5 Source: Globex Pharma Employee Survey 2023

- Employees in every job role have stock option levels ranging from None to Very High, with no roles excluded (Fig. 6)
- Proportions with no stock options and those with options varied across roles (Fig. 7)

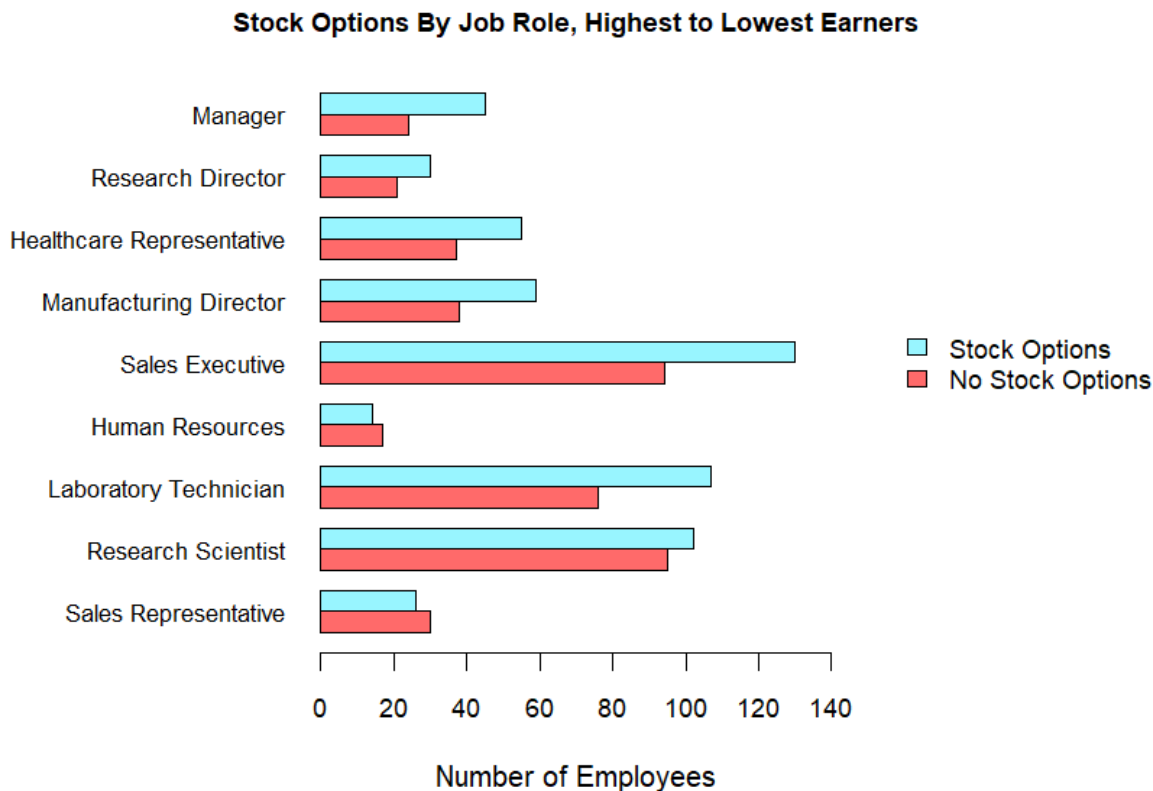


Fig. 6 Number of employees with and without stock options, by job role

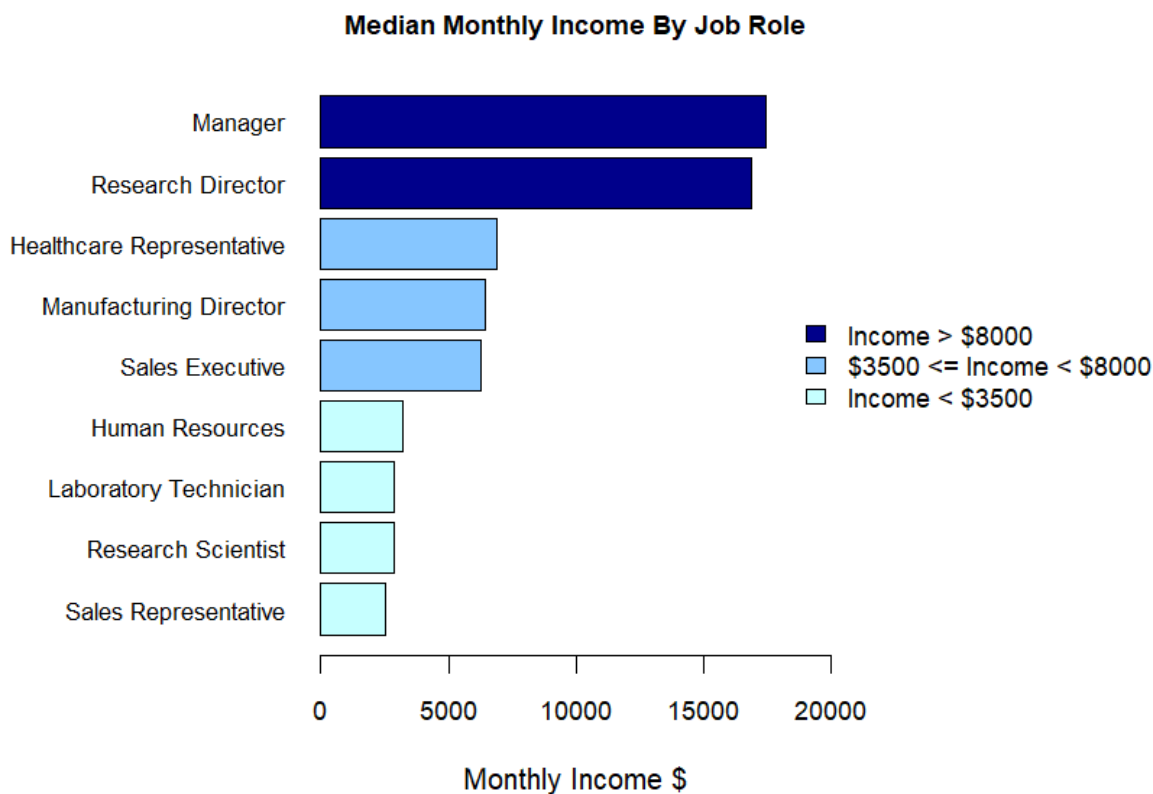


Fig. 7 Median Monthly Income by Job Role grouped by Senior Management, Middle Management, Entry Level Role Source: Globex Pharma Employee Survey 2023

- There is a group of high monthly income earners (>\$15000 a month) with no stock options at all, all are Managers or Research Directors (Fig. 8 and Appendix A)

Dept	JobRole	Income > \$15000, No Stock Option
Human Resources	Manager	3
Research & Development	Manager	10
Sales	Manager	8
Research & Development	Research Director	12

Fig.8 Subset of High-Income Earners Salary > \$15,000 a month, Stock Option Level = "None"

Q3. What, if anything, do the employees who rated their Work Life Balance as “Bad” have in common?

Background

We know that work life balance is important based on IPSOS survey (2222 *Australia Mini Report 260124 ONLINE VERSION.pdf*, no date).

There are four ratings in the Globex survey of employees – Bad, Satisfactory, Good and Very Good.

Key Findings

- The median monthly income of the employees who rated their work life balance as “Bad” was significantly lower than those who rated otherwise (Fig 9)

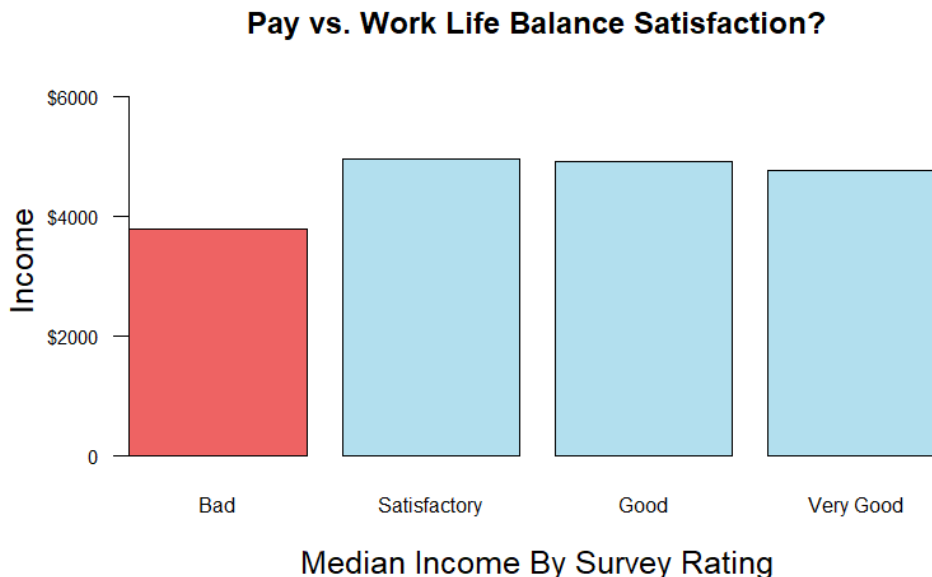


Fig. 9 Median monthly income by rating of “Work Life Balance” Source: Globex Employee Survey 2023 Refer Appendix A for detailed Boxplot

- Reviewing in order of decreasing income apart from Healthcare Representative, the proportion of employees who rated “Work Life Balance” as “Bad” increases as the monthly income level decreases (Fig. 10).

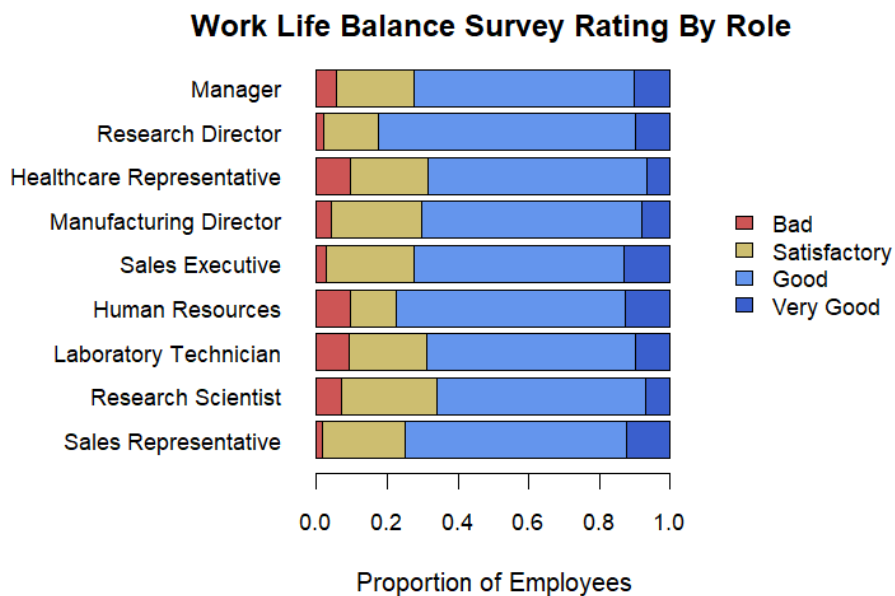


Fig. 10 “Work Life Balance” ratings options = “Bad”, “Satisfactory”, “Good”, “Very Good”, Ordered by Highest Monthly Income to Lowest Monthly Income - Source Globex Employee Survey 2023

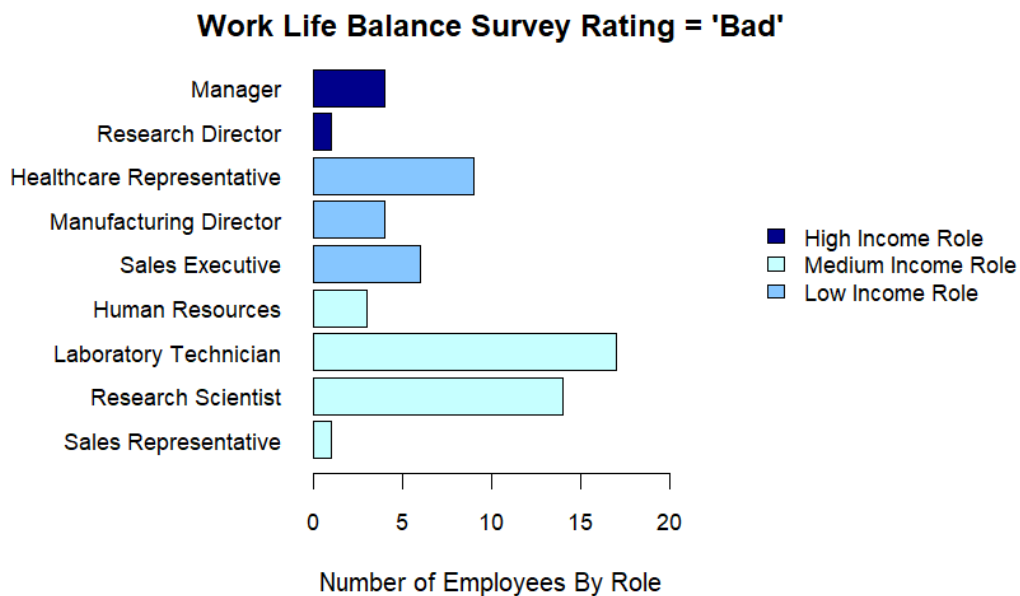


Fig. 11 “Work Life Balance Rating = Bad, Ordered By Decreasing Monthly Income

- “Laboratory Technicians” and “Research Scientists” have the largest total number of employees rating “Work Life Balance” as “Bad” (Fig. 11).
- “Laboratory Technicians” and “Research Scientists” are two of the largest cohorts of employees in the organisation (Fig. 12).

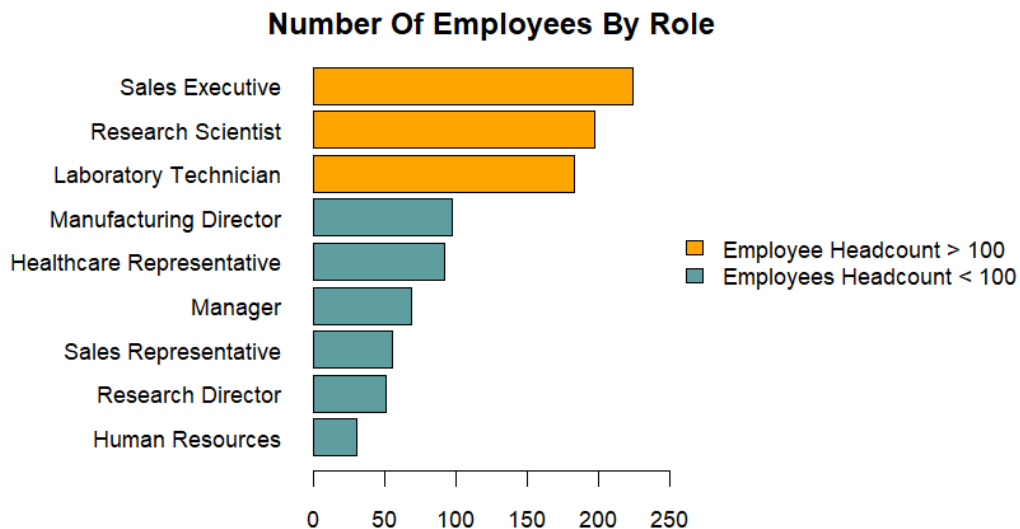


Fig. 12 Number of Employees By Role Source Globex Survey 2023

Conclusions

Monthly Income

There appears to be an imbalance in the “Research Director” role employee salaries particularly between male and female research directors.

Given pay being a key factor for employees' decision to leave an organisation (2222 *Australia Mini Report 260124 ONLINE VERSION.pdf*, no date), our findings here could be a cause of discontent, particularly due to the females being in the role for longer than their male counterparts.

Stock Option Levels

Intuitively one may expect job role, monthly income, age, years with the company or performance rating to indicate the level of stock options for the employee, however this was not the case given the data we had.

As an employee it may seem unclear how stock is awarded.

If the employees at Globex perception is in line with our analysis here this could lead to employee discontent.

Work Life Balance

With survey ratings of “Satisfactory” or higher given by 94% of employees, on the surface work life balance appears not to be a main indicator of employee discontent.

The analysis indicates the proportion and the overall numbers of employees rating “Bad” increase as monthly income decreases.

Given 52% of employees who rated “Bad” were in two of the lowest two median monthly income job roles, this may indicate these employees are feeling undervalued or inadequately compensated for the work they do at the lower levels in the organisation.

There may be an uneven distribution of workload between the hierarchical levels at Globex based on the analysis which may have a knock-on effect across other departments and job roles.

Recommendations

Recommendation 1

Review and recalculate the “Research Director” desired salary bandings (upper and lower bounds) ensuring they are appropriate for this role.

Potentially if there is a hierarchy within this role, consider a “Junior Research Director” / “Senior Research Director” division in the title as well for clarity.

Recommendation 2

Review the promotion process within the “Research Director” role.

Ensure it includes all cross functional stakeholders including Human Resources and is inclusive of not only female employees but all employees.

Recommendation 3

A senior management led review of criteria for awarding stock options at Globex, including a deeper survey of employees on their attitudes to stock options specifically.

Upon review completion, an announcement of the criteria in company literature and company communication channels for clarity, and a follow up survey in 6 months, compare results.

Recommendation 4

Review the salary bandings for job roles “Laboratory Technician” and “Research Scientist”, including comparison to industry averages for those roles.

Recommendation 5

Review the position description/expectations of “Laboratory Technician” and “Research Scientist” roles compared to monthly income.

References

2222 Australia Mini Report 260124 ONLINE VERSION.pdf (no date). Available at:

<https://www.ipsos.com/sites/default/files/ct/news/documents/2024-02/2222%20Australia%20Mini%20Report%20260124%20ONLINE%20VERSION.pdf> (Accessed: 13 May 2024).

Everything You Need to Know About Stock Options and RSUs [WWW Document], n.d.

URL <https://hbr.org/2021/08/everything-you-need-to-know-about-stock-options-and-rsus> (accessed 5.14.24).

Solving the skills shortage crisis | Ai Group [WWW Document], n.d. URL
<https://www.aigroup.com.au/resourcecentre/research-economics/economics-intelligence/2023/solving-the-skills-shortage-crisis/> (accessed 5.27.24).

Why Do Companies Continue to Use Stock Option Incentives? [WWW Document], 2024. . Stanford Graduate School of Business. URL
<https://www.gsb.stanford.edu/insights/why-do-companies-continue-use-stock-option-incentives> (accessed 5.14.24).

Appendices

Appendix A – Boxplot Income Distribution By Stock Option Level

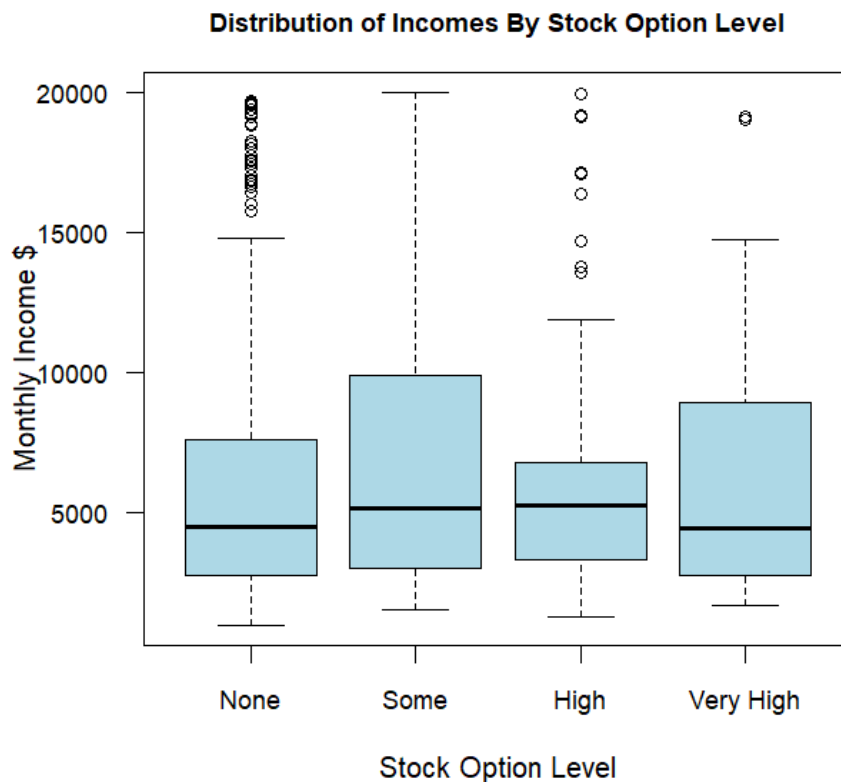


Fig. 13 Boxplot of Monthly Income Distribution By Stock Option Level

Appendix B – Boxplot Income Distribution By Work Life Balance

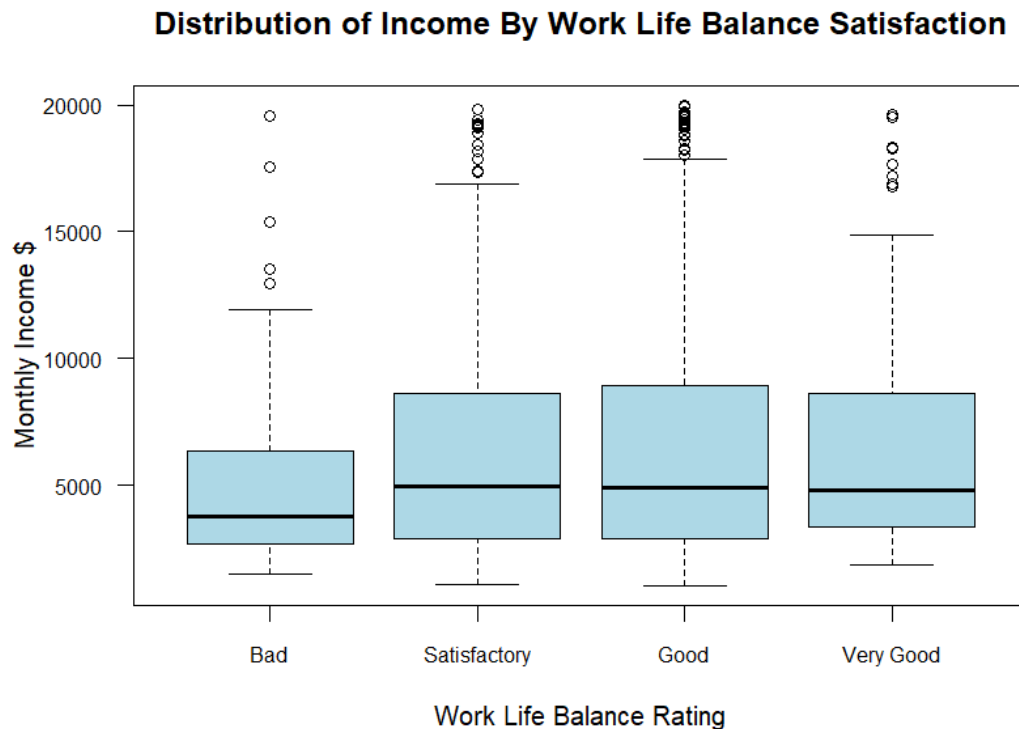


Fig. 14 Boxplot of Monthly Income Distribution By Work Life Balance Rating

Appendix C

R-Code

```
# Import employees csv file, treat any columns with strings as factors
employees_df <- read.csv('employees.csv', stringsAsFactors = T)
survey_dict <- read.csv('Dictionary.csv', header = FALSE, col.names =
c('SurveyQuestion', 'QuestionDetail'))

# RESEARCH QUESTION ONE - HOW DOES INCOME VARY BY JOB?
# Set Plot parameters
par(
  # Specify the margins: bottom, left, top, right
  mar = c(6, 11, 3, 3),
  cex.axis = 0.9
)

# Calculate the median income for each job role
income_by_job <- aggregate(
  MonthlyIncome ~ JobRole,
  employees_df,
  FUN = median
)

# Plot data for distribution of income by Job Role
boxplot(
  MonthlyIncome ~ JobRole,
  data = employees_df,
```

```

horizontal = TRUE,
las = 1,
ylab = "",
col = c(rep("cadetblue",5),rep("indianred",1),rep("cadetblue",3)),
main = "Income Distribution By Job Role",
xlab = "Monthly Income $"
)

# Filter the data to "Research Director" Job Role Only
rdirector_data <- subset(employees_df,employees_df$JobRole == 'Research
Director')

# Create a summary of the data for Research Directors
summary(rdirector_data$MonthlyIncome)

#Plot Distribution of Gender amongst Research Directors
plot(
  rdirector_data$Gender,
  ylab = "Number of Employees",
  col = c("hotpink","lightblue2"),
  main = "Research Directors Gender Distribution"
)

# Create 1 x 2 parameters for next plots
par(
  mfrow = c(1,2),
  mar = c(5,4,4,1),
  cex.axis = 0.6,
  las = 1,
  cex.main = 0.9
)

# Create a histogram of distribution of Monthly Income for Female R.
Directors
hist(
  rdirector_data$MonthlyIncome[rdirector_data$Gender == "Female"],
  main = "Female Research Director Wage",
  xlim = c(10000, 20000),
  ylim = c(0,12),
  breaks = 8,
  col = "hotpink",
  xlab = "Monthly Income $",
  ylab = "Number of Women"
)

# Create a histogram of distribution of Monthly Income for Male R.
Directors
hist(
  rdirector_data$MonthlyIncome[rdirector_data$Gender == "Male"],
  main = "Male Research Director Wage",
  xlim = c(10000, 20000),
  ylim = c(0,12),
  breaks = 8,
  col = "lightblue1",
  xlab = "Monthly Income $",
  ylab = "Number Of Men"
)

# Calculate the median years in current role and age by Gender for Research
Directors
years_by_gender <- aggregate(

```

```

    cbind(YearsInCurrentRole, Age) ~ Gender,
    data = rdirector_data,
    FUN = median
)

# Bar Plot of Median Years in Current Role By Gender
barplot(
  YearsInCurrentRole ~ Gender,
  data = years_by_gender,
  col = c("hotpink", "lightblue"),
  main = "Years In Role"
)

# Bar Plot of Median Age in Current Role By Gender
barplot(
  Age ~ Gender,
  data = years_by_gender,
  ylim = c(0, 50),
  col = c("hotpink", "lightblue"),
  main = "Age"
)

# RESEARCH QUESTION 2 - HOW DO STOCK OPTION LEVELS GET AWARDED?
# Proportions of Employees With Stock Options
prop_stockoption <- prop.table(
  table(employees_df$StockOptionLevel)
)

# Set Up Pie Piece Labels with Stock Option Level and % of Employees from
table above
pie_labels <- paste(c("None", "Some", "High", "Very High"),
  paste(prop_stockoption * 100, "%"))

# Reset parameters for plots and margin
par(
  mfrow = c(1, 1),
  mar = c(6, 6, 3, 6),
  cex.main = 0.9
)

# Create Pie Chart % employees at each stock option level
pie(
  # Use the data from the proportion table
  prop_stockoption,
  # Use the Pie Labels set up earlier
  labels = pie_labels,
  # Set up title and resize
  main = "Percentage Employees By Stock Option Level",
  # Start wedge at 0 degrees
  clockwise = TRUE,
  # Colours of Wedges
  col = c('indianred1', 'seagreen2', 'cyan', 'mediumorchid2')
)

# Grab all employees who have stock level > "None"/1
stockholders <- employees_df[employees_df$StockOptionLevel >=1,]

# Count the number of employees who have stock option levels >"None"/1 in
each role, rename columns
stockholder_frequencies <- setNames(aggregate(
  stockholders$StockOptionLevel ~ stockholders$JobRole,

```

```

    data = stockholders,
    FUN = length),
    c('JobRole', 'NumWithStock'))
)
stockholder_frequencies

# Order Median Income Highest to Lowest
income_by_job <- income_by_job[order(income_by_job$MonthlyIncome),]
income_by_job

# Merge Number Employees Stock Options and Median Monthly Income By Role
merged_data1 <- merge(stockholder_frequencies, income_by_job, by =
"JobRole")

# Grab all employees who have No Stock Options
no_stock_data <- employees_df[employees_df$StockOptionLevel == 0,]

# Count the number of employees who have No Stock Option By Role, rename
columns
nostock_frequencies <- setNames(aggregate(
  no_stock_data$StockOptionLevel ~ no_stock_data$JobRole,
  data = no_stock_data,
  FUN = length),
  c('JobRole', 'NumWithNoStock'))
)
nostock_frequencies

# Make a dataframe of Number of Employees By Role who have stock > "None",
and who have "None"
# Ordered By Median Monthly Income highest to lowest
merged_data2 <- merge(merged_data1, nostock_frequencies, by = "JobRole")
merged_data2
stockoption_byrole <- merged_data2[order(merged_data2$MonthlyIncome),]

par(
  mar = c(6, 10, 3, 10),
  cex.main = 0.9,
  cex.axis = 0.9
)

# Generate Bar plot show Num Employees with and without stock, each job
role
barplot(
  # Grab the two columns stock and no stock for the y axis, by Job Role on
x axis
  cbind(NumWithNoStock, NumWithStock) ~ JobRole,
  data = stockoption_byrole,
  # side by side bars, horizontal bars
  beside = TRUE,
  horiz = TRUE,
  # Vertical labels on x axis, Label for y axis, remove z axis label
  las = 1,
  xlab = "Number of Employees",
  ylab = "",
  # Colour the bars
  col = c("indianred1", "cadetblue1"),
  # Change the Y axis to be clearer
  xlim = c(0,140),
  # Title for the Plot
  main = "Stock Options By Job Role, Highest to Lowest Earners",
  cex.names = 0.85

```



```

)

legend(
  # Put at the top
  "right",
  # Outside plot margin, set position
  xpd = TRUE,
  inset = c(-.65, 0),
  # Specify the labels in the legend
  legend = c("Stock Options", "No Stock Options" ),
  # Specify the colours in the legend
  fill = c("cadetblue1", "indianred1"),
  # Change size text in legend
  cex = .9,
  # Remove box from legend
  bty = "n"
)
# income_by_job <- income_by_job[order(income_by_job$MonthlyIncome),]
# income_by_job

# Create Bar Plot of Job Role By Income Hight to Low
barplot(
  income_by_job$MonthlyIncome,
  names.arg = income_by_job$JobRole,
  main = "Median Monthly Income By Job Role",
  horiz = TRUE,
  las = 1,
  cex.names = 0.85,
  xlab = "Monthly Income $",
  ylab = "",
  xlim = c(0, 20000),
  col = ifelse(income_by_job$MonthlyIncome < 3500, "#c6ffff",
               ifelse(income_by_job$MonthlyIncome >= 8000, "darkblue",
               "#86c6ff")),
  cex.axis = 0.9
)
legend = legend(
  "right",
  xpd = TRUE,
  inset = c(-.65, 0),
  cex = 0.9,
  bty = "n",
  legend = c("Income > $8000", "$3500 <= Income < $8000", "Income < $3500"),
  fill = c("darkblue", "#86c6ff", "#c6ffff"),
)

par(
  # Specify the margins: bottom, left, top, right
  mar = c(6, 11, 3, 3),
  cex.axis = 0.9
)

# Distrubtion of Incomes By Stock Option Level Boxplot
boxplot(
  employees_df$MonthlyIncome ~ employees_df$StockOptionLevel ,
  data = employees_df,
  # data = employees_df[employees_df$StockOptionLevel == 0,],
  xlab = "Stock Option Level",
  names = c("None", "Some", "High", "Very High"),
  ylab = "Monthly Income $",

```

```

col = "lightblue",
main = "Distribution of Incomes By Stock Option Level"
)

# Filter the outliers from boxplot, income >15000 and no stock option level
highincome_nostock <- employees_df[employees_df$StockOptionLevel == 0 &
employees_df$MonthlyIncome > 15000,]

# Group Outliers by Dept and Name and count frequencies
highincome_nostock <- setNames(aggregate(
  MonthlyIncome ~ Department + JobRole,
  data = highincome_nostock,
  FUN = length),
  c('Dept', 'JobRole', 'NumNoStockHighIncome'))
)

# Write to an Excel Table
library(writexl)
write_xlsx(highincome_nostock, 'HighIncomeNoStock.xlsx')
print(highincome_nostock)

### RESEARCH QUESTION THREE WORKLIFE BALANCE = BAD

# Show distribution of Monthly Income by Work Life Balance Rating 1-4
boxplot(
  MonthlyIncome ~ WorkLifeBalance,
  data = employees_df,
  # Format colours, axis text size and titles
  col = "lightblue",
  cex.axis = .9,
  las = 1,
  main = "Distribution of Income By Work Life Balance Satisfaction",
  xlab = "Work Life Balance Rating",
  ylab = "Monthly Income $",
  # Names for each boxplot
  names = c('Bad', 'Satisfactory', 'Good', 'Very Good'))
)

# Create a table of breakdown of work life balance rating for each job role
worklife_by_job <- table(
  employees_df$WorkLifeBalance,
  employees_df$JobRole
)

# Calculate the median income for each job role
income_by_job <- aggregate(
  MonthlyIncome ~ JobRole,
  employees_df,
  FUN = median
)

# Order highest to lowest income
income_by_job <- income_by_job[order(-income_by_job$MonthlyIncome),]
income_by_job

# Calculate the Median Monthly Income By Work Life Balance Level
income_by_workbalance <- aggregate(
  MonthlyIncome ~ WorkLifeBalance,
  data = employees_df,
  FUN = median
)

```

```

)
# Rename Columns appropriately
colnames(income_by_workbalance) <-
c("WorkLifeBalance", "MedianMonthlyIncome")

# Order the Table of work_life_by_job from highest to lowest income
worklife_by_job <-
worklife_by_job[, income_by_job$JobRole[order(income_by_job$MonthlyIncome)]]
worklife_by_job

# Create parameters for plots
par(
  # Specify the margins: bottom, left, top, right
  mar = c(10, 11, 2, 12)
)

# Generate bar plot
barplot(
  # Calculate the proportion of Work Life Balance rating in each Job Role
  prop.table(worklife_by_job, margin = 2),
  # Labels in line with the axis
  las = 1,
  # Change axis text size and name size
  cex.axis = 0.9,
  cex.names = 0.9,
  # Horizontal Bar Plot not vertical
  horiz = TRUE,
  # Colours for each Work Life Balance Rating Bad - Very Good
  col = c('indianred3', 'lightgoldenrod3', 'cornflowerblue', 'royalblue3'),
  # Labels and Title
  xlab = "Proportion of Employees",
  main = "Work Life Balance Survey Rating By Role"
)

# Create a legend
legend(
  "right",
  # Put legend outside the plot area
  xpd = TRUE,
  # Inset legend
  inset = c(-.65, 0),
  # Change size text
  cex = 0.9,
  # Remove border from legend
  bty = "n",
  # Labels
  legend = c('Bad', 'Satisfactory', 'Good', 'Very Good'),
  # Colours for each label
  fill = c('indianred3', 'lightgoldenrod3', 'cornflowerblue', 'royalblue3')
)

# Create new parameters for plot
par(
  # Specify the margins: bottom, left, top, right
  mar = c(6, 6, 3, 6)
)

# Generate barplot median monthly income by rating of work life balance
rating
barplot(
  MedianMonthlyIncome ~ WorkLifeBalance,

```

```

data = income_by_workbalance,
# Labels, fontsize, titles
names.arg = c('Bad', 'Satisfactory', 'Good', 'Very Good'),
cex.names = 0.9,
cex.axis = 0.7,
yaxt = "n",
ylab = "Monthly Income $",
xlab = 'Survey Rating for Work Life Balance',
ylim = c(0,6000),
main = 'Pay vs. Work Life Balance Satisfaction?',
col = c('indianred2', 'lightblue2', 'lightblue2', 'lightblue2'),
)

# Generate own ticks for Y-axis
axis(
  # Y axis
  side = 2,
  # Tick marks
  at = c(0, 2000, 4000, 6000),
  # labels and axis size
  las = 1,
  cex.axis = 0.75,
  labels = c(0, "$2000", "$4000", "$6000")
)

# Create frequency of "Bad" Work Life Balance Rating for each role
frequencybad_byrole <- aggregate(
  Department ~ JobRole,
  data = employees_df[employees_df$WorkLifeBalance == 1,],
  FUN = length
)

# Rename Columns
colnames(frequencybad_byrole)[colnames(frequencybad_byrole) ==
'Department'] <- "NumEmployees"

# Parameters for plots
par(
  # Specify the margins: bottom, left, top, right
  mar = c(10, 11, 2, 12)
)

# Order By Frequency lowest to highest
frequencybad_byrole <-
frequencybad_byrole[order(frequencybad_byrole$NumEmployees),]
frequencybad_byrole

# Merge data of frequency bad ratings and median monthly incomes for each
job role
merged_data <- merge(frequencybad_byrole, income_by_job)
# Order by income level
merged_data <- merged_data[order(merged_data$MonthlyIncome),]
merged_data

# Bar Plot
barplot(
  # Frequency of Employees Bad Work Life Balance by Job Role
  merged_data$NumEmployees,
  names.arg = merged_data$JobRole,
  # Horizontal Plot, not vertical
  horiz = TRUE,

```

```

# Colour bar less than $3500 lightblue, medium blue middle income, dark
blue
# for high income
col = ifelse(merged_data$MonthlyIncome < 3500, "#c6ffff",
             ifelse(merged_data$MonthlyIncome > 8000, "darkblue",
"#86c6ff"))
),
# Format labels, axis, fontsize, titles
las = 1,
xlim = c(0,20),
ylab = "",
cex.names = .9,
cex.axis = .9,
main = "Work Life Balance Survey Rating = 'Bad'",
xlab = "Number of Employees By Role"

)

# Create and format legend
legend = legend(
  "right",
  xpd = TRUE,
  inset = c(-1, 0),
  cex = 0.9,
  bty = "n",
  legend = c("High Income Role", "Medium Income Role", "Low Income Role"),
  fill = c("darkblue", "#c6ffff", "#86c6ff")
)

# Calculate the number of employees in each job role
num_by_role <- setNames(
  aggregate(
    Department ~ JobRole,
    data = employees_df,
    FUN = length
  ), c('JobRole', 'NumEmployees'))

# Order number of employees in each job role
num_by_role <- num_by_role[order(num_by_role$NumEmployees),]

# BarPlot
barplot(
  # Frequency of employees in each job role
  num_by_role$NumEmployees,
  names.arg = num_by_role$JobRole,
  # data = num_by_role,
  horiz = TRUE,
  # Format labels, axis, colours and title
  las = 1,
  cex.axis = 0.9,
  cex.names = 0.9,
  xlim = c(0, 250),
  col = ifelse(num_by_role$NumEmployees > 100, "orange", "cadetblue"),
  main = "Number Of Employees By Role"
)

# Create and format legend
legend = legend(
  "right",
  xpd = TRUE,
  inset = c(-1, 0),

```

```
cex = 0.9,  
bty = "n",  
legend = c("Employee Headcount > 100", "Employees Headcount < 100"),  
fill = c("orange", "cadetblue")  
)
```