

INDIVIDUAL ASSIGNMENT

APD2F2019CS(IS)

HAND OUT DATE: 4 2021

HAND IN DATE: 22 2021

WEIGHTAGE:

INSTRUCTIONS TO CANDIDATES:

**1 Submit your assignment at the administrative counter.**

**2 Students are advised to underpin their answers with the use of references (cited using the American Psychological Association (APA) Referencing).**

**3 Late submission will be awarded zero (0) unless Extenuating Circumstances (EC) are upheld.**

**4 Cases of plagiarism will be penalized.**

**5 The assignment should be bound in an appropriate style (comb bound or stapled).**

**6 Where the assignment should be submitted in both hardcopy and softcopy, the softcopy of the written assignment and source code (where appropriate) should be on a CD in an envelope / CD cover and attached to the hardcopy.**

**7 You must obtain 50% overall to pass this module.**

**STUDENT NAME: Wong Horng Woei**

**TP NUMBER: TP055241**

**LECTURER NAME: Liew Yee Jing**

Table of Contents

[1.0 Introduction 6](#_Toc88491181)

[2.0 Assumption 6](#_Toc88491182)

[3.0 Install Packages, Load Libraries & Data Import 7](#_Toc88491183)

[3.1 Install Packages 7](#_Toc88491184)

[3.2 Load Libraries 7](#_Toc88491185)

[3.3 Data Import 8](#_Toc88491186)

[4.0 Data Exploration 10](#_Toc88491187)

[5.0 Data Pre-processing 18](#_Toc88491188)

[6.0 Data Manipulation 21](#_Toc88491189)

[7.0 Questions and Analysis 27](#_Toc88491190)

[7.1 Question 1: Why did the employees left the company? 27](#_Toc88491191)

[7.1.1 Analysis 1: Determine the relationship between Year and Termination Reason 27](#_Toc88491192)

[7.1.2 Analysis 2: Determine the relationship between Department and Termination Reason 30](#_Toc88491193)

[7.1.3 Analysis 3: Determine the Age of employees that resign in Customer Service department 33](#_Toc88491194)

[7.1.4 Analysis 4: Determine the Duration of Service in Customer Service department with employees that are 21 years old. 36](#_Toc88491195)

[7.1.5 Analysis 5: Determine the Age of employees that retire in Meats, and Produce department 38](#_Toc88491196)

[7.1.6 Analysis 6: Determine the Duration of Service in Meats, and Product department with employees over 60 years old. 40](#_Toc88491197)

[7.1.7 Analysis 7: Determine the relationship between Age and Duration of Service 42](#_Toc88491198)

[7.1.8 Conclusion on Question 1 44](#_Toc88491199)

[7.2 Question 2: What are the Age Class that get the most layoff? 45](#_Toc88491200)

[7.2.1 Analysis 1: Determine the relationship between Year and Termination Reason 45](#_Toc88491201)

[7.2.2 Analysis 2: Determine the Layoff in City 48](#_Toc88491202)

[7.2.3 Analysis 3: Determine the relationship between Job Position and Employee Count 50](#_Toc88491203)

[7.2.4 Conclusion for Question 2 54](#_Toc88491204)

[7.3 Question 3: Which job is in high demand in a crowded city? 55](#_Toc88491205)

[7.3.1 Analysis 1: Determine the Active Employee in each City 55](#_Toc88491206)

[7.3.2 Analysis 2: Determine the Active Employee in the store 57](#_Toc88491207)

[7.3.3 Analysis 3: Determine the jobs that have the most employee in Vancouver 61](#_Toc88491208)

[7.3.4 Conclusion for Question 3 63](#_Toc88491209)

[7.4 Question 4: Is there gender inequality in the organization? 64](#_Toc88491210)

[7.4.1 Analysis 1: Determine the relationship between Year and Gender 64](#_Toc88491211)

[7.4.2 Analysis 2: Determine the relationship between HireYear and Gender 66](#_Toc88491212)

[7.4.3 Analysis 3: Determine the relationship between Status and Gender within 2011-2015 68](#_Toc88491213)

[7.4.4 Analysis 4: Determine the relationship between Department and Gender 70](#_Toc88491214)

[7.4.5 Conclusion on Question 4 72](#_Toc88491215)

[7.5 Question 5: What is the organization of the store? 73](#_Toc88491216)

[7.5.1 Analysis 1: Determine the relationship between Store Number and Department 73](#_Toc88491217)

[7.5.2 Analysis 2: Determine the Highest Employee Count in Store with Job Position 75](#_Toc88491218)

[7.5.3 Analysis 3: Determine the Lowest Employee Count in Store with Job Title 78](#_Toc88491219)

[7.5.4 Analysis 4: Determine the store that has Store Manager 81](#_Toc88491220)

[7.5.5 Conclusion for Question 5 84](#_Toc88491221)

[8.0 Extra Features 85](#_Toc88491222)

[8.1 fct\_reorder 85](#_Toc88491223)

[8.2 ggthemes 88](#_Toc88491224)

[8.3 Population Pyramid Chart 91](#_Toc88491225)

[9.0 Conclusion 94](#_Toc88491226)

[References 95](#_Toc88491227)

# 1.0 Introduction

The main goal of the documentation is to investigate the hidden issue in human resources management and create a few questions and answered by using analysis. The data given for this assignment is related to employees’ job information and attribution. The human resource department manager provided the dataset of employees within an organization from British Columbia, Canada for analysis to provide meaningful visualization for decision making. In Addition, a dataset file is provided to import and read on R Studio.

In the dataset, the data analyst is responsible to analyze why employees left the company and determine the relationship between several variables and parameters which can benefit the data analyst to tell the story. After dataset is imported into R Studio, data analyst began on data exploration to get familiarize with the dataset. Then, the analysis was also conducted to study the relationship among the attributes in the pre-processing section and answer the questions that was created . On the bright side, there are no NA (not available) data in the dataset as the appropriate pre-processing techniques were applied to check the NA data.

# 2.0 Assumption

According to the dataset given, various assumptions were made for the dataset to enhance the value and reliability of the analysis and results. Firstly, the data that given are the latest and up to date to reflect the employee status in the organization. Then, the current year is 2016, these data provided are employees information based on real-life events from the past years. Lastly, the accuracy and consistency of the data’s standard is maintained for quality data analysis and results.

# 3.0 Install Packages, Load Libraries & Data Import

## 3.1 Install Packages

Text

Description automatically generated

Figure 1:Install Packages

Figure 1 above shows that these packages will be installed with “install.packages()” function in order to do analysis on this dataset in R Studio. Before beginning the analysis, 7 packages will be used to install, which are dplyr, ggplot2, magrittr, tibble, stringr, forcats, and ggthemes. dplyr is mainly used for data manipulation and transformation. Next, ggplot2 is used to visualize the graphs. Magrittr is to allow piping function to be used on later analysis and tibble allows to build columns sequentially. Then, stringr is used to replace wrongly spelled characters in the dataset. The forcats function will often use as to reorder the graphs and ggthemes to have a better visual of the graph. After all packages are downloaded and installed successfully, the packages must be loaded into library by using the library() function in order for the packages to work correctly.

## 3.2 Load Libraries

Text

Description automatically generated

Figure 2: Load Libraries

After all packages are downloaded and installed successfully, the packages must be loaded into library by using the library() function in order for the packages to work correctly.

## 3.3 Data Import

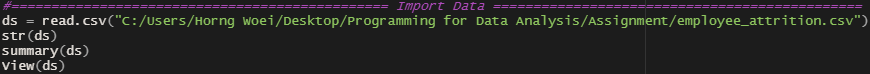


Figure 3:Import Dataset

A dataset is provided and required to import into R Studio. The read.csv() function is used to allow user to insert the full file path of the file to be loaded into R Studio. However, whenever the file path is being copied directly from the windows file explorer into the read.csv(), user have to replace the backslash(\) to slash(/) OR double backslashes(\\) so that the dataset can be imported successfully.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 4: str() Function

Once the dataset is imported, data type can be displayed by using str() function, and the summary() function will display every variable a set of descriptive statistics, depending on the type of variable.

A black screen with white text

Description automatically generated with low confidence

Figure 5: summary() Function

For instance, numerical variable will show the range, quartiles, median, and mean. Factor variables will display frequencies. Character variables will not receive any information from summary() except for length and class. In addition, summary() will also display the number of missing values for numerical and factor variables. Then, the View() function allows R to invoke a spreadsheet style data viewer for user to view the data in the next tab.

# 4.0 Data Exploration

A simple data exploration is conducted to have a clearer overview of the dataset as it is the initial step in data analysis before starting to create questions and analyze.

**First Exploration**

Text

Description automatically generated

Figure :Source Code for Employee Status by Year

Chart, bar chart, histogram

Description automatically generated

Figure : Bar Chart for Employee Status by Year

First exploration is using bar chart to view the employee status compare by year, where the X-Axis is Year and number of employees as Y-Axis to find out the total number of employees getting terminated or still active throughout from year 2006 to year 2015.

**Second Exploration**

Graphical user interface, text

Description automatically generated

Figure : Source Code for Length of Service

Chart, bar chart, histogram

Description automatically generated

Figure : Bar Chart for Length of Service

This bar chart shows the length of service with numbers of employees. The highest length of service would be 13 years which have 2885 employees meanwhile the lowest length of service is 26 years which only has 14 employees.

**Third Exploration**

Text

Description automatically generated

Figure : Source Code for City Name

Chart, bar chart

Description automatically generated

Figure : Bar Chart for City Name

The third exploration will be displaying city name with employee count. Vancouver is clearly seen to have the most employees that has a total of 11211 employees as it is the most populous city in British Columbia, Canada (DigiMarConCanadaWest, 2021). The least employee in the city is Blue River that only has 9 employees in total.

**Fourth Exploration**

Text

Description automatically generated

Figure : Source Code for Department

Chart, bar chart

Description automatically generated

Figure : Bar chart for Department

This exploration in department shows that “Meats” department has 10269 employees and also the most employees are in that department, follow by “Dairy” department and “Produce” department. The department that has the lowest employee count is from “Legal” department.

**Fifth Exploration**

Text

Description automatically generated

Figure : Source Code for Job Title

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Job Title of Employee

The bar chart shows that there are 47 jobs and the “Meat Cutter” job has 9984 employees that is the highest among all other job title, follow by “Dairy Person” that has 8590 employees and “Produce Clerk” with 8237 employees.

**Sixth Exploration**

Text

Description automatically generated

Figure : Source Code for Gender

Chart, pie chart

Description automatically generated

Figure : Pie Chart for Gender

This pie chart is showing that the female has more employees compare to male. Female has a total of 25898 employees, and male only has a total of 23755 employees.

**Seventh Exploration**

Text

Description automatically generated

Figure : Source Code for Termination Type

Chart, pie chart

Description automatically generated

Figure : Pie Chart for Termination Type

This chart shows the employee count by termination type from a pie chart. 1270 employees that got terminated are categorized as voluntary type. The other 215 terminated employees are categorized as involuntary type, which means they are left the company not willingly.

**Eighth Exploration**

Text

Description automatically generated

Figure : Source Code for Termination Reason

Chart, pie chart

Description automatically generated

Figure : Pie Chart for Termination Reason

The pie chart displayed 3 termination reason which are layoff, resignation, and retirement. Among these termination reason, retirement has 885 employees which also has the most employees are terminated voluntary. Resignation has 385 employees that are also categorized as voluntary termination. Lastly, the layoff that is categorized as involuntary termination has 215 employees that are terminated from that group.

# 5.0 Data Pre-processing

After exploring the dataset, data pre-processing will be next as data analyst now should have acknowledged whether the data has any empty data or weird data. Data pre-processing is essential to conduct before starting analysis in R. If data analyst neglected this part, it may cause incorrect data happen in R and might affect the analysis results, leading to inaccurate result.



Figure : Check for Missing Values & Results

First of all, the source code on the left of Figure 22 is to check for missing values (NA) in the dataset, then the output is on the right which to prove that this dataset has no missing data in it. Hence, it is not necessary to replace any data since it has no missing data.

Graphical user interface, text

Description automatically generated

Figure : Spelling Errors

First of all, it is noticeable that there are some spelling errors of data in some column shown in Figure 23 above, the “Resignaton” from termreason\_desc should be “Resignation”. The “CHief Information Officer” in job\_title column should be “Chief Information Officer” and “Accounts Receiveable Clerk” should be “Accounts Receivable Clerk”. Then, “Accounts Receiveable” should replace to “Accounts Receivable” in department\_name column.

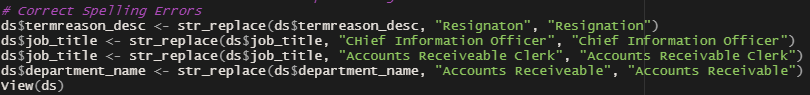


Figure : Source Code for Correct Spelling Errors

It is important to correct the spelling errors to avoid any missing data, as it might confuse the data analyst later when conducting analysis. stringr package is needed to be loaded into the library, then use str\_replace() function, this allows to replace matched patterns in a string. After correcting the spelling errors, View() function is used to check the dataset again for changes shown in Figure 25 below.

Graphical user interface

Description automatically generated

Figure : Corrected Spelling Errors

Text

Description automatically generated with medium confidence

Figure 26:Remove Meaningless Attributes

Next, remove the meaningless attributes so that it doesn’t get in the way when doing analysis later. The recorddate\_key is considered meaningless as the “STATUS\_YEAR” can replace that attribute with the same year. Then, the birthdate\_key will also be remove, it is also meaningless for gender\_short because there are 2 attributes for gender already, which is gender\_short and gender\_full. gender\_short is showing gender in short form which consist of “M” and “F”, and gender\_full has gender in full words, “Male” and “Female”, so the gender\_short will be dropped instead of gender\_full. Assigning “NULL” to the specific column in the data frame to drop the columns, this approach is a simple and also efficient way to drop data frame columns.

A black screen with white text

Description automatically generated with low confidence

Figure 27:Rename columns

After removing meaningless attributes, renaming the columns is optional but it is to let data analyst read the data columns easier. The rename() function changes the column name of a data frame based on the older name. Then, use View() function to view the results. Results will be shown in Figure 28.



Figure 28: Result for renamed attributes

# 6.0 Data Manipulation

Data manipulation step starts with the process of rearrange the data to make it easier for the reader to read. Renaming, converting, adding new columns are considered as data manipulation and will be demonstrated below.

Text

Description automatically generated

Figure 29: Add new column – JobPosition

A new column is added for JobPosition as the JobTitle has too many variables, it is recommended to group the variables to make it easier for data analyst to analyze them. Source Code above shows that mutate() and case\_when() function is used. The mutate() function by creating new variables and preserve the existing ones. The case\_when() function allows user to use multiple “if” and “else if” statements. Then if no cases match, the last else statement will set data to “Unknown” and .after will be adding the column after JobTitle. After that, entering “new\_ds” in R console will show the newly added column “JobPosition” at the R console below at Figure 30.

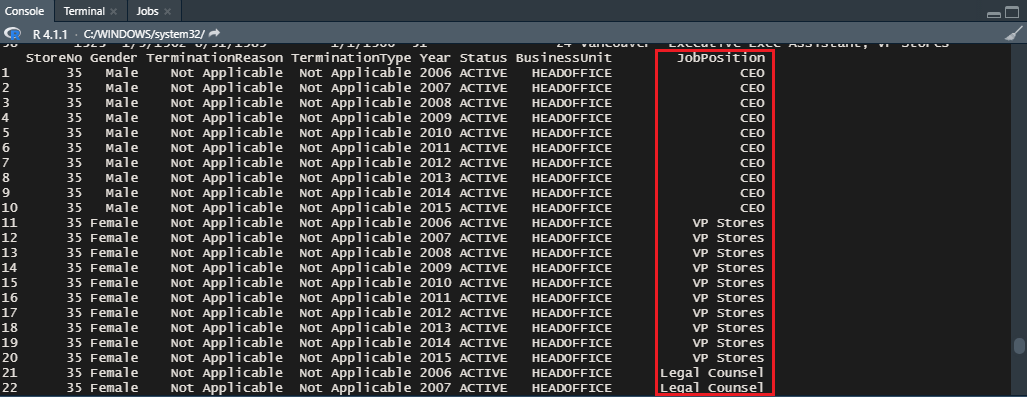


Figure 30:Result for check JobPosition column

Text

Description automatically generated

Figure 31: Source code to convert date format

After adding new column for JobPosition, HireDate and TerminationDate is used to convert to date format instead of character data type. The source code from Figure 31 shows that a new data frame is assigned under “newcol\_ds” and mutate() function is used and as.Date() to convert the HireDate and TerminationDate into date data type under a format starting with YYYY,MM,DD. The results are shown below with str() function.

Text

Description automatically generated

Figure 32: Result for convert date format

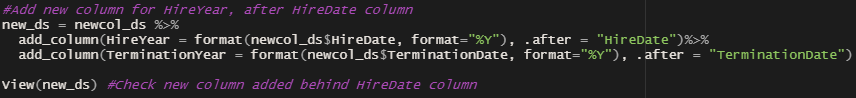


Figure 33: Source code to add new column for HireYear and Termination Year

In Figure 33, the source code explains that adding new columns for HireYear and TerminationYear with different approach, which is the add\_column() function from tibble library. The function starts with inserting the data frame to append to, then select from the HireDate and TerminationDate and extract the format of the Year. Then, add those columns after HireDate and TerminationDate. Results are shown below on Figure 34 after View() function is called.

Calendar

Description automatically generated with medium confidence

Figure 34: Result for add new column for HireYear and Termination Year



Figure 35: Source code to add new column for AgeClass

In the source code on Figure 35, mutate() function is used to create a new column for “Age” in the data frame to group the “Age” variables as the “Age” variable has too much data, so it is difficult to do analysis on big data. The “AgeClass” uses the cut() function to cut the “Age” from 18 to 24 and labeled as “19-24”, 24 to 34 labeled as “25-34”, 34 to 44 labeled as “35-44” age group, 44 to 54 are labeled as “45-54” age range, 54 to 64 are “55-64”, and lastly for number 64 to 70 are labeled as “65+” age class. After adding new age class column, the View() function is used again for checking the newly added column like Figure 36 below.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 36: Results for add new column for AgeClass

Text

Description automatically generated

Figure 37: Source code to convert to factor

In Figure 37 above shows the source code that to be converted from character(chr) and some integer(int) data types into factor datatypes. This is because changing to factor allows to store data values for level indication. After the converting is done, it is recommended to use str() again to check for the latest changes in data types.

Text

Description automatically generated

Figure 38: Results for convert to factor

# 7.0 Questions and Analysis

## 7.1 Question 1: Why did the employees left the company?

This question is created to find out the reason behind why the employees left the company. This study is important to the Human Resources as they could understand on what affected the employees to leave the organization. By understanding the reason and situation, Human Resources could do decision making to assist the organization.

### 7.1.1 Analysis 1: Determine the relationship between Year and Termination Reason

#### 7.1.1.1 Source Code

Text

Description automatically generated

Figure : Source Code for Year and Termination Reason

Explanation:

A bar chart is selected to display and compare the year in which the column name “Year” and the termination reason which the column name is “TerminationReason”. Filter function is used to filter out the “Not Applicable” from TerminationReason variable as the “Not Applicable” attribute is for active employee only, so it is necessary to filter out. For mappings, Year will be on X-Axis and filling color will be TerminationReason. The labs() function is then used for title, x-axis, y-axis, and fill labeling. Then, geom\_text() is used for displaying the label of the employee count on year of each termination reason categories. By adding position = position\_stack() function, the labels will be in the middle of each bar, giving the reader easier to read the values.

#### 7.1.1.2 Output of the graph

Chart, histogram

Description automatically generated

Figure : Bar Chart for Year and Termination Reason

Explanation:

A bar chart with X-Axis showing the Year and the Y-Axis is the employee count. This chart is to display and determine the termination reason of employee on each year. Red color is indicating the layoff termination, green color is showing the resignation termination, and blue color is indicating retirement termination.

#### 7.1.1.3 Outline the findings based on results

I found out there are a total of 1485 employees that got terminated and it shows that there were 142 layoffs in 2014 and 73 layoffs in 2015 which are involuntary terminates. However, there are voluntary terminates such as 385 employees are categorized as resignation and total of 885 employees categorized for retirement from year 2006 to year 2015. Human Resource might be foreseeing that the company could not support too many employees, so some employees get layoff in the recent year of 2014 and 2015.

The total termination are 1485 employees according to Figure 21 from data exploration above and it shows that there were 142 layoffs in 2014 and 73 layoffs in 2015 which are involuntary terminates. However, there are voluntary terminates such as 385 employees are categorized as resignation and total of 885 employees categorized for retirement from year 2006 to year 2015. Human Resource might be foreseeing that the company could not support too many employees, so some employees get layoff in the recent year of 2014 and 2015.

### 7.1.2 Analysis 2: Determine the relationship between Department and Termination Reason

#### 7.1.2.1 Source Code

Text

Description automatically generated

Figure : Source Code for Department and Termination Reason

Explanation:

A bar chart is chosen to show and compare the Department in which the column name “Department” and the termination reason which the column name is “TerminationReason”. Filter function is used to filter the “TERMINATED” from Status variable as the “TERMINATED” attribute is needed for the analysis only, “ACTIVE” employees are not needed, so it is necessary to filter and take only terminated employee. For mappings, Department will be on X-Axis and Y-Axis will be the employee count, filling color will be TerminationReason. The theme() function is used for adjusting the X-axis text to 90 degrees for cleaner and tidier to read. The labs() function is then used for title, x-axis, y-axis, and fill labeling.

#### 7.1.2.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Department and Termination Reason

Explanation:

A bar chart with X-Axis is the Department and the Y-Axis is the employee count. This chart is to display and determine that each department has termination reason. Red color is indicating the layoff termination, green color is showing the resignation termination, and blue color is indicating retirement termination.

#### 7.1.2.3 Outline the findings based on results

By looking at the bar plot, a few things stick out quite obviously. Customer Service has much larger proportion of resignation compared to other departments, and retirement is high in Meats and Produce department. Other than that, it is noticeable that every department has employee retire too. Hence, it is worth for much deeper analysis to search for interesting findings.

### 7.1.3 Analysis 3: Determine the Age of employees that resign in Customer Service department

#### 7.1.3.1 Source Code

Text

Description automatically generated

Figure : Source Code for Age of employees that resign in Customer Service Department

Explanation:

In the source code above, the bar chart will display the age in which the column name “Age” and the employee count. Filter function is used to filter the “Resignation” from TerminationReason variable as the “Resignation” attribute is needed for the analysis only. Another attribute that is needed to filter is “Customer Service” from Department variable, as the graph from Figure 42 shows that customer service has the most resignation for employees. For mappings, Age will convert into factor to count the age and it will be on X-Axis and Y-Axis will be the count of employee. The theme\_stata() from ggthemes library is applied for better visual of the graph. Also, geom\_text() is used for displaying the label of the employee count on age for clearer understandings.

#### 7.1.3.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Age of employees that resign in Customer Service Department

Explanation:

A bar chart with X-Axis is the Age and the Y-Axis is the numbers of employees. This is to display the numbers of employee that resign at what age in the Customer Service department.

#### 7.1.3.3 Outline the findings based on results

It shows that age 21 has 58 employees resign in the Customer Service position. This voluntary termination tells that employee leave the company themselves. According to (Oberholtzer, 2017), the customer service job is lack of compensation, the employees does not foresee getting a raise in the future. Then, it has no growth for customer service as it will never get promoted or entrusted with more responsibilities, it is a job that has no room for improvement. Next, employees that work for customer service might get lack of recognition as they interact with a huge amount of customer every single day and get high call-in volumes. Therefore, these 3 reasons might lead them to resign for customer service job.

### 7.1.4 Analysis 4: Determine the Duration of Service in Customer Service department with employees that are 21 years old.

#### 7.1.4.1 Source Code

Text

Description automatically generated

Figure : Source Code for the Duration of Service in Customer Service department with employees that are 21 years old.

Explanation:

In the source code above, the bar chart will display the Service Duration(Year) in which the column name “ServiceDuration” and the employee count. Filter function is used to filter the “Resignation” from TerminationReason variable. Another attribute that is needed to filter is “Customer Service” from Department variable, as the graph from Figure 42 shows that customer service has the most resignation for employees. Then, Age of “21” is filtered to get the employees that are 21 years old along with the other 2 filters which are “Resignation” and “Customer Service”. For mappings, ServiceDuration will convert into factor, and it will be on X-Axis and Y-Axis will be the count of employee. The theme\_stata() from ggthemes library is applied for clearer visual of the graph. Also, geom\_text() is used for displaying the label of the employee count on age.

#### 7.1.4.2 Output of the graph

Chart, bar chart, waterfall chart

Description automatically generated

Figure : Bar Chart for the Duration of Service in Customer Service department with employees that are 21 years old.

Explanation:

A bar chart with X-Axis of the Service Duration(Years) and the Y-Axis is the numbers of employees. “0” indicates that employee has less than a year working experience, where “1” shows that employees has at least 1 year working experience already.

#### 7.1.4.3 Outline the findings based on results

This plot shows that 54 employees did work for at least 1 year before they resign from the Customer Service department. Meanwhile, 4 employees work less than a year and resigned. This shows that employees that are 21 years old did not work for more than 1 year in the Customer Service department, they could be struggling to work as they could be also studying at the same time, or they are not satisfied with the lack of compensation and recognition given.

### 7.1.5 Analysis 5: Determine the Age of employees that retire in Meats, and Produce department

#### 7.1.5.1 Source Code

Text

Description automatically generated

Figure : Source Code for Age of employees that retire in Meats, and Produce department

Explanation:

The bar chart will display the age in which the column name “Age” and the employee count. Filter function is used to filter the “Retirement” from TerminationReason variable as the “Retirement” attribute is needed for the analysis only. Another attribute that is needed to filter is “Meats” and “Produce” Department variable, as the graph from Figure 42 shows that customer service has the most retirement for employees. For mappings, Age will convert into factor to count the age and it will be on X-Axis and Y-Axis will be the count of employee. A theme from ggthemes library is applied for clearer visual of the graph. Also, geom\_text() is used for displaying the label of the employee count on age. Next, the facet\_wrap() function is used for separating Departments into multiple panels.

#### 7.1.5.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Age of employees that retire in Meats, and Produce department

Explanation:

A bar chart with X-Axis of the Age and the Y-Axis is the numbers of employees. A total of 2 classes are separated to compare the findings which are Meats department and Produce department to find out the total of employee that retire at the specific age.

#### 7.1.5.3 Outline the findings based on results

This plot shows 2 departments that has the most retirement compared to other departments. The "Meats" department has 247 employees retire at the age of 65, and 70 employees retire at the age of 60. Meanwhile in the "Produce" department, 169 employees retire at age of 65 and 137 employees retire at age of 60. According to Canada Pension Plan (CPP) retirement pension, employees need to be at least 60 years old to retire (Government of Canada, 2021).

### 7.1.6 Analysis 6: Determine the Duration of Service in Meats, and Product department with employees over 60 years old.

#### 7.1.6.1 Source Code

Text

Description automatically generated

Figure : Source Code for Duration of Service in Meats, and Product department with employees that over 60 years old.

Explanation:

The bar chart will display the Service Duration(Years) in which the column name “ServiceDuration” and the employee count. Filter function is used to filter the “Retirement” from TerminationReason variable as the “Retirement” attribute is needed for the analysis only. Another attribute that is needed to filter is “Meats” and “Produce” from Department variable, as the graph from Figure 42 shows that meats and produce department has the most retirement for employees. Age for 60 and 65 will also need to filter as the retirement are mostly elderly from the Figure 48. For mappings, Duration of Service will convert into factor to count the age and it will be on X-Axis and Y-Axis will be the count of employee. The theme\_stata from ggthemes library is applied for clearer visual of the graph. Also, geom\_text() is used for displaying the label of the employee count on age. Then, the facet\_wrap() function is used for separating Departments into multiple panels.

#### 7.1.6.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Duration of Service in Meats, and Product department with employees that over 60 years old.

Explanation:

A bar chart with X-Axis of the Service Duration(Years) and the Y-Axis is the numbers of employees. A total of 2 classes are separated to compare the findings which are Meats department and Produce department.

#### 7.1.6.3 Outline the findings based on results

It shows that in "Meats" and "Produce" department, these retirement employees are working in this company for at least 8 years and above. These retirement employees are considered loyal employees as they work long enough to gain experience and knowledge about this company. This include sacrificing their own time and interest to put more energy into the organization.

### 7.1.7 Analysis 7: Determine the relationship between Age and Duration of Service

#### 7.1.7.1 Source Code

Text

Description automatically generated

Figure : Source Code for Age and Duration of Service

Explanation:

In the source code above, the box plot is chosen and to display the “Age” and the “ServiceDuration”. For mappings, ServiceDuration will be on X-Axis and Y-Axis will be the Age, filling with Status to show Active and Terminated employees. A theme from ggthemes library is applied for clearer visual of the graph. Also, geom\_text() is used for displaying the label of the employee count on age. Then, the facet\_grid() function is used for 2 discrete variables which is the “ACTIVE” and “TERMINATED” in Status variable.

#### 6.1.7.2 Output of the graph

Chart, box and whisker chart

Description automatically generated

Figure : Box Plot for Age and Duration of Service

Explanation:

A boxplot with X-Axis of the Service Duration(Years) and the Y-Axis is the Age of employee. Each box represents different status and different color. The blue color represent active employee and red color indicates terminated employee.

#### 6.1.7.3 Outline the findings based on results

This boxplot shows that the terminated employees have a high average age as compared to active employees. Meanwhile Service Duration shows not much difference between active and terminated.

### 7.1.8 Conclusion on Question 1

According to the findings from the 7 analyses, a simple conclusion is made such that only a few attributes are affected to justify why the employees left the company. It is clear that in termination reason retirement has the most count compared to layoff and resignation. However, the findings also showed that resignation in Customer Service department also stands out. In Customer Service, most employees that resign are 21 years old. This could be due to they could be struggling to work as they could also be studying at the same time, or they are not satisfied with the lack of compensation and recognition given. Employees that are at least 60 years old and above has the most retirement numbers. According to (Government of Canada, 2021), the Canada Pension Plan (CPP) retirement pension requires employees need to be at least 60 years old to retire.

The takeaway on the findings from the 7 analyses is that Human Resource should pay attention to resignations on customer service as they play an important role in helping to resolve any problem that customer has. It is recommended build a program where employees hit target on the amount of customer service, a cash voucher is given to the employee to make them feel exciting and motivated at the same time. Next, to invest in the employees’ growth such as setting up training session, giving them a program that consist of developing the interacting skills. In addition, adding in an extra company culture where employees that goes beyond the expectations or gain huge achievement, celebrate and recognize them in a timely manner (Oberholtzer, 2017).

## 7.2 Question 2: What are the Age Class that get the most layoff?

This question is created to find out which range of age will get the most layoff. This study is important to the Human Resources as they could understand on which city that are needed to pay attention and decide what range of age best suit to hire.

### 7.2.1 Analysis 1: Determine the relationship between Year and Termination Reason

#### 7.2.1.1 Source Code

Text

Description automatically generated

Figure : Source Code for Year and Termination Reason

Explanation:

The bar chart will display the year in which the column name “Year” and the employee count. Filter function is used to filter the “Layoff” from TerminationReason variable. For mappings, Year be on X-Axis and Y-Axis will be the count of employee, fill by TerminationReason. The theme\_clean() from ggthemes library is applied for clearer visual of the graph. Also, geom\_text() is used for displaying the label of the employee count on layoff, position = position\_stack() is used to arrange the text in the middle of the bar.

#### 7.2.1.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Year and Termination Reason

Explanation:

A bar chart with Year on the X-Axis and the Y-Axis is the numbers of employees. This graph is to determine the employee that gotten layoff in which year.

#### 7.2.1.3 Outline the findings based on results

Layoff termination type seems to have it on year 2014 and 2015 only. In the year of 2014, there were a total of 142 employees getting layoff, and 73 employees got layoff in year 2015. The company might have sudden business change or employees are over hired and leading the company failed to support them. According to (macrotrends, 2021), there has been a decline on year 2013 and increase on year 2014, then decline again on year 2015 in Figure 55 below, this proves that if inflation rate is lower, the unemployment rate will be higher which might be affected. Therefore, it is worth looking into these termination reason in between these 2 years.

Chart, line chart

Description automatically generated

Figure : Canada Inflation Rate from 2013 to 2015

### 7.2.2 Analysis 2: Determine the Layoff in City

#### 7.2.2.1 Source Code

Text

Description automatically generated

Figure :Source Code for Layoff in City

Explanation:

The bar chart will display the “City” and the employee count. Filter function is used to filter the “Layoff” from TerminationReason variable as the “Layoff” attribute is needed to focus on this question only. The count() function is used for city to count the variable rows, then mutate() it on city with fct\_reorder() function to reorder the factor levels by sorting along with n, which is the employee count. For mappings, city will be on Y-Axis and the employee count will be on X-Axis. A theme from ggthemes library is applied for clearer visual of the graph. Also, geom\_text() is used for displaying the label of the employee count on each city.

#### 7.2.2.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Layoff in City

Explanation:

A bar chart with X-Axis is the Employee Count and the Y-Axis is the city. This is to display the city that the numbers of employee got layoff.

#### 7.2.2.3 Outline the findings based on results

In this graph, it is clearly to be seen that Fort Nelson is the city that got the most layoff of employees, which has total of 39 layoff employees, follow by 31 layoff employees in New Westminister and 30 layoff employees from Grand Forks. As for the least layoff employees is Blue River that has only 1 layoff. It is essential to do deeper analysis to find out the age class that got layoff.

### 7.2.3 Analysis 3: Determine the relationship between Job Position and Employee Count

#### 7.2.3.1 Source Code

Text

Description automatically generated

Figure : Source Code for Job Position and Employee Count

Explanation:

The bar chart will display the age class in which the column name “AgeClass” and the employee count. Filter function is used to filter the “Fort Nelson” from City variable as the city has the most layoff in the previous analysis. Another filter is needed to filter the “Layoff” from TerminationReason. For mappings, age class be on X-Axis and Y-Axis will be the number of employees. The ylim() is used to add limit from 0 to 15 counts for cleaner graph display. A theme from ggthemes library is applied for clearer visual of the graph. Then, geom\_text() is used for displaying the label of the employee count on layoff. Lastly, although facet\_grid() is used for multiple discrete variables, but it is applied here to just show the city name.

#### 7.2.3.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Job Position and Employee Count

Explanation:

A bar chart where X-Axis is the Age Class and the Y-Axis is the Employee Count. This is to display the city that the numbers of employee got layoff. The Age Class are separated into different bars for better comparison.

#### 7.2.3.3 Outline the findings based on results

Results shown that in Fort Nelson, employees that are most likely to get layoffs are around the range of 25-34 in the age, which has a total of 12 layoff employees. According to (Statistics Canada, 2021) that in 2016, the statistics from Figure 60 shows the range from 19-24 has total 220 population, 25-34 age has 575 total population, 35-44 age has 480 population, age range of 45-54 has total of 500 population, and 55-64 has 480 population. Overall, it is almost similar to the graph, so that it can be considered the more population, the more layoff in the age class.

Table

Description automatically generated with medium confidence

Figure : Fort Nelson population in 2016

### 7.2.4 Conclusion for Question 2

In conclusion for the 3 analyses from this question, layoff only happens in year 2014 and 2015, this could be due to the company might have sudden business change or employees are over hired and leading the company failed to support them. In addition, the most layoff city would be in Fort Nelson. The analysis also shows that Age ranges from 25-34 has the most layoff in Fort Nelson city while the age range from 55-64 has least layoff. Human resource should take note on these specific cities and avoid hiring in the stated age range as if the organization layoff too many employees may leave a bad review for the organization and cause that no applicants will apply for the job. Therefore, it is important to hiring employees carefully to avoid having bad reputations for the organization.

## 7.3 Question 3: Which job is in high demand in a crowded city?

This question is created to investigate which job are in demand in a popular city. This study is important to the Human Resources as they could understand on which city that they need to pay attention to and decide what kind of job scope they should look into.

### 7.3.1 Analysis 1: Determine the Active Employee in each City

#### 7.3.1.1 Source Code

Text

Description automatically generated

Figure : Source Code for Active Employee in each City

The bar chart will display the city in which the column name “City” and the employee count. Firstly, it is essential to avoid duplication when active employees are required to filter for analysis, due to Figure 7 from data exploration above. The first filter() function is used to filter the “ACTIVE” from Status variable to select all active employee in the Status column. Then, second filter() is used to get year of 2015 so that it will get the latest active employee. The distinct() function will take the top EmployeeID and remove the rest from the duplicated ones, resulting there will only be 1 unique EmployeeID on the year 2015. Then, the count() function is used for city to count the variable rows, then mutate() it on city with fct\_reorder() function to reorder the factor levels by sorting along with n, which is the employee count. For mappings, employee count is X-Axis and Y-Axis will be the city. The stat = “identity” is also used in geom\_bar() to allow user provide X-Axis and Y-Axis values. A theme from ggthemes library is applied for clearer visual of the graph. Then, geom\_text() is used for displaying the label of the employee count.

#### 7.3.1.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Active Employee in each City

Explanation:

A bar chart with X-Axis is the Employee Count and the Y-Axis is the city. This is to display the active employee that are in each city.

#### 7.3.1.3 Outline the findings based on results

From the graph above, Vancouver has the most active employee which total of 1097, where the Bella Bella has the least that has only 13 employees. According to (WorldAtlas, 2021), Vancouver is one of the largest cities in Canada.

### 7.3.2 Analysis 2: Determine the Active Employee in the store

#### 7.3.2.1 Source Code

Text

Description automatically generated

Figure : Source Code for the Active Employee in Store

Explanation:

The first section of source code is showing that first filter() function is used to get year of 2015 so that it will get the latest active employee. Then, second filter() is used to filter the “ACTIVE” from Status variable to select all active employee in the Status column. Third filter() function is used to filter the “Vancouver” from City as it has the most employee. The distinct() function will take the top EmployeeID and remove the rest from the duplicated ones, resulting there will only be 1 unique EmployeeID on the year 2015. For mappings, BusinessUnit is X-Axis and Y-Axis will be the employee count. A theme\_economist() from ggthemes library is applied for clearer visual of the graph. Then, geom\_text() is used for displaying the label of the employee count.

The second section of the source code is having the similar code as the first section as it is essential to get the unique count and avoid duplication. The first filter() function is used to get year of 2015 so that it will get the latest active employee. Then, second filter() is used to filter the “ACTIVE” from Status variable to select all active employee in the Status column. Third filter() function is used to filter the “Vancouver” from City as it has the most employee. In addition, STORES are required to filter in as the second section is to find the store number. For mappings, Store Number is X-Axis and Y-Axis will be the employee count. A theme\_economist() from ggthemes library is applied for clearer visual of the graph. Then, geom\_text() is used for displaying the label of the employee count.

#### 7.3.2.2 Output for graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Active Employee in Vancouver by Business Unit

Explanation:

This bar chart shows that with X-Axis is the Business Unit and the Y-Axis is the Employee Count, to determine the active employee in Vancouver .

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Active Employee in Vancouver by Stores

Explanation:

This bar chart shows that with X-Axis is the Store Number and the Y-Axis is the Employee Count, to find out the active employee in Vancouver by each Stores .

#### 7.3.2.3 Outline the findings based on results

In my findings, it is clear that Store Number 42 has the most employee which has a total of 370 workers working inside, it can be assumed that it is one of the largest stores in Vancouver compared to the other stores in Vancouver.

### 7.3.3 Analysis 3: Determine the jobs that have the most employee in Vancouver

#### 7.3.3.1 Source Code

Text

Description automatically generated

Figure : Source Code for jobs that has the most employee in Vancouver

Explanation:

The source code is showing that the first filter() function is used to get year of 2015 so that it will get the latest active employee. Then, second filter() is used to filter the “ACTIVE” from Status variable to select all active employee in the Status column. Third filter() function is used to filter the “Vancouver” from City as it has the most employee. Furthermore, a Store Number “42” is required for filter as need to find the popular job in Vancouver. The distinct() function will take the top EmployeeID and remove the rest from the duplicated ones, resulting there will only be 1 unique EmployeeID on the year 2015. Next, the count() function is used for JobTitle to count the variable rows, then mutate() it on JobTitle with fct\_reorder() function to reorder the factor levels by sorting along with n, which is the employee count. For mappings, JobTitle is X-Axis and Y-Axis will be the employee count. The stat = “identity” is also used in geom\_bar() to allow user provide X-Axis and Y-Axis values. A theme\_economist() from ggthemes library is applied for clearer visual of the graph. Then, geom\_text() is used for displaying the label of the employee count.

#### 7.3.3.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Job Title in Vancouver by Store No.42

Explanation:

A bar chart with Job Title as X-Axis and Employee Count as Y-Axis to find out the most popular job in Store Number 42 that are based in Vancouver.

#### 7.3.3.3 Outline for findings based on results

From the bar chart above shows, there highest count in Job Title is Cashier which has a total of 118 employees in Store Number 42. On the other hand, the least count in Job Title is Dairy person that only has 42 employees in that Job.

### 7.3.4 Conclusion for Question 3

According to the findings from the 3 analyses on question 3, a simple conclusion is made such that only a few attributes are affected to justify the highest demand job in a crowded city. It clearly shows that Vancouver has the most "ACTIVE" employees compare to other cities, as Vancouver is one of the largest cities in Canada. The company's Head Office is at Vancouver and has 11 employees, while the store at Vancouver has 1086 employees. by filtering into Store No., the data shown that a total of 370 employees are working in Store No.42, it could be a very large store as it is required for so many employees to work compared to the other stores. Lastly, the most popular job in the most crowded city would be Cashier as the job scope itself is relatively simple and easy to do, most workers will opt for Cashier. Human resource should consider allocating new employees to different job titles other than cashier because cashier position might be packed .

## 7.4 Question 4: Is there gender inequality in the organization?

This question is created to study the organization and whether or not there is there any gender inequality. It is important to Human Resources as having different ratios in gender might be leading to employee dissatisfaction, and leave the company, which can affect the growth of the organization too. Gender inequality has many reasons, such as unequal pay for different genders, unfavorable hiring strategy, and sexual harassment (BrightHR, 2021). Hence, Human Resources could take note of this analysis and make some improvements.

### 7.4.1 Analysis 1: Determine the relationship between Year and Gender

#### 7.4.1.1 Source Code

Text

Description automatically generated

Figure 68: Source code for Year and Gender

Explanation:

The source code in Figure 68 is showing that fct\_rev() function from forcats library is used to reverse the order for Year and fill by Gender. 2 geom\_bar() is used for creating population pyramid graph, with subset() function to select different Gender. For mapping, a negative count is assigned on y value to allow negative numbers. By using scale\_y\_continuous() function, labeling the sequence from -2500 to 2500 by each 500 counts. Then, a theme is applied from ggthemes to make the graph easier to visualize. Title, X-Axis labeling and Y-Axis labeling is then added into the graph. Lastly, coord\_flip() is used to flip the cartesian coordinates where horizontal becomes vertical, and vertical becomes horizontal.

#### 7.4.1.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure 69: Population Pyramid for Year and Gender

Explanation:

A population pyramid chart with Employee Count as X-Axis and Year as Y-Axis to determine the gender inequality throughout the years from 2006 to 2015. Red color indicates Female and blue color indicates Male.

#### 7.4.1.3 Outline of the findings based on results

From the population pyramid chart above, it is clearly seen that there are quite equal on gender throughout the years from 2006 until 2015. It is interesting to go deeper for analysis and find out if some variables define that is there any gender inequality in this organization.

### 7.4.2 Analysis 2: Determine the relationship between HireYear and Gender

#### 7.4.2.1 Source Code

Text

Description automatically generated

Figure 70: Source code for HireYear Gender

Explanation:

The source code above shows that X-Asix is HireYear and fill by Gender. 2 geom\_bar() is used for creating population pyramid graph, with subset() function to select different Gender. For mapping, a negative count is assigned on y value to allow negative numbers. By using scale\_y\_continuous() function, labeling the sequence from -1500 to 1500 by each 500 counts. Then, a theme\_gdocs() is applied from ggthemes for clearer visual of the graph. Title labeling, X-Axis labeling and Y-Axis labeling is then added into the graph. Lastly, coord\_flip() is used to flip the cartesian coordinates where horizontal becomes vertical, and vertical becomes horizontal.

#### 7.4.2.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure 71: Population Pyramid for Year and Gender

Explanation:

A population pyramid chart with Employee Count as X-Axis and Year as Y-Axis to determine the gender inequality throughout the years from 2006 to 2015. Red color indicates Female and blue color indicates Male.

#### 7.4.2.3 Outline of the findings based on results

From Figure 71 above, the population pyramid graph shows that female is slightly more employee count than male. This has proved that it is worth to look into more in depth analysis.

### 7.4.3 Analysis 3: Determine the relationship between Status and Gender within 2011-2015

#### 7.4.3.1 Source Code

Text

Description automatically generated

Figure 72: Source code for Status and Gender within 2011-2015

Explanation:

The source code is showing that the first filter() function is used to get “Resignation” and “Layoff”, notice that a %in% is used rather than == operator, for different approach, %in% is value matching and can be used to compare vectors. Then, second filter() is used to filter the "2015", "2014", "2013", "2011", "2012" from Year variable as a 5-year sample. Third filter() function is used to filter the “TERMINATED” from Status. For mappings, Status is X-Axis and Y-Axis will be the employee count, fill by Year. The ylim() is used to add limit from 0 to 300 counts for cleaner graph display. A theme\_economist() from ggthemes library is applied for clearer visual of the graph. Then, geom\_text() is used for displaying the label of the employee count. Lastly, facet\_wrap() is applied for separating Gender into 2 panels.

#### 7.4.3.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure 73: Bar Chart for Status and Gender within 2011-2015

Explanation:

A bar chart with X-Axis displays the Status of “TERMINATED” and the Y-Axis is the employee count. This chart is to determine the terminated employee and gender by the year. Red color is indicating year 2011, yellow color is showing year of 2012, green color is showing the year of 2013, blue color is indicating 2014, and pink color is indicating year of 2015.

#### 7.4.3.3 Outline of the findings based on results

The female termination has more count than male, but this is consider that gender is equal as the hire for female is slightly more than male in analysis 2, Figure 71.

### 7.4.4 Analysis 4: Determine the relationship between Department and Gender

#### 7.4.4.1 Source Code

Text

Description automatically generated

Figure 74: Source code for Department and Gender

Explanation:

A bar chart is chosen to show and compare the Department in which the column name “Department” and the numbers of employees. For mappings, Department will be on Y-Axis and X-Axis will be the employee count. The theme\_clean() function is used from ggthemes library for cleaner and tidier to read. The labs() function is then used for title, x-axis, y-axis, and fill labeling. Then. facet\_wrap() is applied for separating Gender into 2 panels.

#### 7.4.4.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure 75: Bar Chart for Department and Gender

Explanation:

A bar chart with X-Axis showing the Employee Count and the Y-Axis is the Department. There are 2 panels separated which is Female and Male. This chart is to determine the which Department that has gender inequality.

#### 7.4.4.3 Outline of the findings based on results

In this bar chart, it is shown that female has interest in the accounting field where the male has not much interest in this field. However, the male has interest in the information technology field where the female has no interest in it. Hence, it is also a 50:50 ratio in comparison to gender inequality.

### 7.4.5 Conclusion on Question 4

In conclusion for the 4 analyses from this question, throughout the year of 2006 to 2015, there has no gender inequality in the organization. However, by looking back to analysis 2 Figure 71, there has a slight increase for female than male in terms of hiring strategy by comparing from the year the organization started to hire employees. Then, the sample that was analyzed over a 5-year period from 2011 to 2015 shows that female that got terminated is also slightly higher than the male which got balanced from the hire count. Therefore, there is no gender inequality in the organization. Furthermore, the research shown that there are some fields in which each gender has a distinct interest that the opposite gender does not share in Figure 75. For instance, the male shows more interest in the information technology field than the female does. On the other hand, the female shows more interest in accounting field than the opposite gender. Human Resources did a great job in having gender equally and should always maintain a 50:50 ratio to avoid gender inequality in the company.

## 7.5 Question 5: What is the organization of the store?

This question is to investigate the organization structure of the store. Human Resources will be able to see the numbers of employees in each store and look at the structure of each store. Human Resources could take note of this analysis and make some improvements for the organization in the future.

### 7.5.1 Analysis 1: Determine the relationship between Store Number and Department

#### 7.5.1.1 Source Code

Text

Description automatically generated

Figure : Source code for Store Number and Department

Explanation:

The source code from Figure 76 explains that “2015” Year will be filtered as to find out the organization structure of the store, latest year is capable of showing the structure of each store that is still operating. Then, second filter is filtering the “ACTIVE” from Status for getting all the active employees from year of 2015. “StoreNo” as store number will be X-Axis and fill with “Department”. A Y-Axis limit is added to cap at 500 for tidier graph design. The theme\_economist() is applied from ggthemes to provide clearer visuals for the whole bar chart. Lastly, labs() function is use to label the title, X-Axis, Y-Axis.

#### 7.5.1.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Store Number and Department

Explanation:

A bar chart shows that X-Axis is the Store Number and Y-Axis is Employee Count. Legends are from Department which has Bakery, Customer Service, Dairy, Executive, Meats, Processed Foods, Produce, Recruitment, and Store Management. A total of 9 departments and 31 stores are involved in this bar chart.

#### 7.5.1.3 Outline of the findings based on results

From the bar chart in Figure 77, it shows that Store Number 46 has the most employees working but there is no store manager to manage the store. Furthermore, there are a few stores that has relatively less amount of employee working, which could be a rural area or they just startup a store to test the market on the area. Other than that, there were a few stores that has store management.

### 7.5.2 Analysis 2: Determine the Highest Employee Count in Store with Job Position

#### 7.5.2.1 Source Code

Text

Description automatically generated

Figure : Source code for Highest Employee Count in Store with Job Position

Explanation:

The source code above shows that the “2015” Year will be filtered to determine the store’s organization structure, the most recent year can be used to display each store that is still operating. Then, second filter is filtering the “ACTIVE” from Status for getting all the active employees from year of 2015. A third filter is applied to “46” from StoreNo as from Figure 77 shows that it has the most employees. “StoreNo” as store number will be X-Axis and fill with “JobPosition”. A Y-Axis limit is added to cap at 500 for tidier graph design. The theme\_economist() is applied from ggthemes to provide clearer visuals for the whole bar chart. Then, labs() function is used to label the title, X-Axis, Y-Axis. Labeling the employee count in the bar is used by geom\_text. Lastly, although facet\_grid() is used for multiple discrete variables, but it is applied here to just show the city name.

#### 7.5.2.2 Output of the graph

Chart

Description automatically generated

Figure : Bar Chart for Highest Employee Count in Store with Job Position

Explanation:

A bar chart is used to show the highest employee count in store with job position. The X-Axis is Store Number and Y-Axis is the Employee Count. In the JobPosition legends, red color indicates Baker, yellow color shows that it is Cashier, green color shows Dairy Person, light blue represents Meat Cutter, dark blue indicates Produce Clerk, and Shelf Stocker is pink color. City name can be seen on top of the bar.

#### 7.5.2.3 Outline of the findings based on results

According to the bar chart above from Figure 79, it has the most employee count among all the stores, a total of 473 employees. This chart also shows that the store is based on Victoria, which is the capital city of British Columbia and also the largest urban area on the island, approximately 100km to the province’s largest city, Vancouver (McGillivray, 2021). Since it is the capital city of British Columbia, Canada, the store requires a huge number of employees to work to grow the business. However, there has no store manager in this store to manage them. Therefore, this is worth to look into the matter.

### 7.5.3 Analysis 3: Determine the Lowest Employee Count in Store with Job Title

#### 7.5.3.1 Source Code

Text

Description automatically generated

Figure : Source code for Lowest Employee Count in Store with Job Position

Explanation:

The source code from Figure 78 explains that the “2015” Year will be filtered to determine the store’s organization structure, the most recent year can be used to display each store that is still operating. Then, second filter is filtering the “ACTIVE” from Status for getting all the active employees from year of 2015. A third filter is applied to a vector with “3”, "19", "35" from StoreNo as from Figure 77 shows that it has the least employees. Note that there is a change to %in% operator because it is much simpler to use vector for matching multiple values at once, “StoreNo” as store number will be X-Axis and fill with “JobPosition”. The theme\_economist() is applied from ggthemes to provide clearer visuals for the whole bar chart. Then, labs() function is used to label the title, X-Axis, Y-Axis. Labeling the employee count in the bar is used by geom\_text. Lastly, facet\_grid() is used on City for multiple discrete variables, scales are set to free to reduce the empty space from other stores.

#### 7.5.3.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Bar Chart for Lowest Employee Count in Store with Job Position

Explanation:

A bar chart is used to show the store with job position that has the lowest employee count. The X-Axis is Store Number and Y-Axis is the Employee Count. JobPosition legends are listed above all three bars, and city name has also labeled above the bars.

#### 7.5.3.3 Outline of the findings based on results

According to the bar chart above from Figure 81, these 3 cities are considered having the lowest employee count among all stores which is lower than 50 employees. Bella Bella is having only 13 employees working in the store, follow by Store 35 from Vancouver. Store Number 35 is the head office as all employees can be seen that their job scope is more suitable in office. Then, Store Number 19 from Nelson city has a total of 32 employees working. With the exception of the head office, it can be seen that these 2 cities that has low employee count also lack a store manager, perhaps it is a small store in the city, so store manager is unnecessary.

### 7.5.4 Analysis 4: Determine the store that has Store Manager

#### 7.5.4.1 Source Code

**Text

Description automatically generated**

Figure : Source code for finding store that has Store Manager

Explanation:

The first section of the source code from above shows that the “2015” Year will be filtered Then, second filter is filtering the “ACTIVE” from Status for getting all the active employees from year of 2015. A third filter is applied to a vector with “17”, “18”, “32”, “44” from StoreNo as from Figure 77 shows that it has Store Manager. The %in% operator is used for vector for matching multiple values at once, “StoreNo” as store number will be X-Axis and fill with “JobPosition”. A theme is applied from ggthemes to provide clearer visuals for the whole bar chart. Then, labs() function is used to label the title, X-Axis, Y-Axis. Labeling the employee count in the bar is used by geom\_text. Lastly, facet\_grid() is used on City for multiple discrete variables, scales are set to free to reduce the empty space from other stores.

The second section of the source code has the same filtering as it is to do a more in depth and clearer view of the graph. The “StoreNo” as store number will be X-Axis and fill with “JobPosition”. Then, ylim() is applied for Y-Axis from 0 to 3 to show a better visual on the bar chart. A theme is applied from ggthemes to provide clearer visuals for the whole bar chart. The labs() function is used for labeling the overall bar chart and facet\_grid() is used on City for multiple discrete variables.

#### 7.5.4.2 Output of the graph

Chart, bar chart

Description automatically generated

Figure : Result 1 for finding store that has Store Manager

A picture containing chart

Description automatically generated

Figure : Result 2 for finding store that has Store Manager

Explanation:

Figure 83 bar chart shows that the store that has Manager. The X-Axis is Store Number and Y-Axis is the Employee Count. . In the JobPosition legends, red color indicates Baker, yellow color shows that it is Cashier, green color shows Dairy Person, light blue represents Manager, dark blue indicates Meat Cutter, Produce Clerk is purple color and Shelf Stocker is pink color. City name has also labeled above the bars.

Then, in Figure 84 bar chart that is run by the second section code shows the Number of Store Manager in Store. Store Number is on the X-Axis where Y-Axis is Employee Count. The city name is labeled above the bars.

#### 7.5.4.3 Outline of the findings based on results

According to these 2 bar charts above, these 4 cities have store manager in the store. Each of the store which are store number 17 in Langley, store number 18 in Nanaimo, store number 32 in Terrace and store number 44 in Vancouver has 1 store manager to manage the store.

### 7.5.5 Conclusion for Question 5

In conclusion for the 4 analyses from this question, the organization structure of the store is inconsistent. It is clearly stated from analysis 1 that Store Number 46 has the most employees working and a few stores that have relatively a smaller number of employees. The Store Number 46 is a very large store filled with a lot of employees and it is on a capital city, Victoria, yet the store has no store manager. It is also reasonable that is has no store manager because the head office is 100km away from Victoria, which is in Vancouver so that the directors or CEO might come once in a while to survey the market, but it is recommended for human resource to allocate store manager in the store for much more organized store and assist in the growth of business.

In Figure 81, bar chart shows that there were 3 stores that has low employee count, it can be considered that if the store has low employee count, it is a small store, so it does not have a store manager to manage the store. Human resource could promote an employee as a supervisor for a smaller store which could take care stuff in all aspects or suggest a store manager to drop by every 2 weeks for an inspection of the store. By having these recommendations, not only the organization will grow effectively, but also can monitor every single store.

In Figure 83 and 84 bar chart shows that each store which the store number 17 in Langley, store number 18 in Nanaimo, store number 32 in Terrace and store number 44 in Vancouver has 1 store manager to manage the store. This proves that some stores have store manager and not all stores have store manager which show inconsistency in this organization.

# 8.0 Extra Features

## 8.1 fct\_reorder

**Before**Text

Description automatically generated

Figure : Source Code before fct\_reorder()

**After**

Text

Description automatically generated

Figure : Source Code after fct\_reorder()

Explanation:

The 2 source codes above shows before and after adding the fct\_reorder() function. The fct\_reorder() function allows to reorder the factor by sorting along another variable. By applying it in Question 2, Analysis 2, the count() function is used for city to count the variable rows, then mutate() it on city with fct\_reorder() function to reorder the factor levels by sorting along with n, which is the employee count.

**Before**

Chart, bar chart

Description automatically generated

Figure : Output before fct\_reorder()

**After**

Chart, bar chart

Description automatically generated

Figure : Output after fct\_reorder()

Explanation:

In comparison to Figure 87 and Figure 88, the Figure 88 which is after placing fct\_reorder() function definitely makes the highest value stands out a lot, it captures the readers focus point when it is displayed comparing to Figure 87 that has value all over the place. Therefore, this feature does improve the results from the analysis.

## 8.2 ggthemes

**Before**

Text

Description automatically generated

Figure : Source Code before ggthemes

**After**

Text

Description automatically generated

Figure : Source Code after ggthemes

Explanation:

The 2 source codes above shows before and after adding the ggthemes features. The ggthemes has a total of 14 unique themes that will allow clearer visual of the graph. In Question 1, Analysis 1, the theme\_stata() from ggthemes library is chosen and applied to ease the reader when reading the graph.

**Before**

Chart, histogram

Description automatically generated

Figure : Output before ggthemes

**After**

Chart, bar chart, histogram

Description automatically generated

Figure : Output after ggthemes

Explanation:

In comparison to Figure 91 and Figure 92, the graph in Figure 92 that after adding theme\_stata() from ggthemes library really makes the different where it is tidier and cleaner graph. Hence, it can improve the result by having clearer visual of the graph by applying this feature.

## 8.3 Population Pyramid Chart

**Before**

Text

Description automatically generated

Figure : Source Code before applying Population Pyramid Chart

**After**

Text

Description automatically generated

Figure : Source Code after applying Population Pyramid Chart

Explanation:

The 2 source codes above shows before and after adding the population pyramid chart. The population pyramid chart is very suitable when comparing 2 attributes. By applying it in Question 4, Analysis 2, 2 geom\_bar is required with subset() function to select different Gender. For mapping, a negative count is assigned on y value to allow negative numbers. By using scale\_y\_continuous() function, labeling the sequence from -1500 to 1500 by each 500 counts. Lastly, coord\_flip() is used to flip the cartesian coordinates.

**Before**

Chart, bar chart, histogram

Description automatically generated

Figure : Output before applying Population Pyramid Chart

**After**

Chart, bar chart

Description automatically generated

Figure : Output after applying Population Pyramid Chart

Explanation:

By using the same data, the graph in Figure 96 that after applying population pyramid chart clearly shows that the female got hire more than male compared to using a bar chart in Figure 95 that can hardly identify which gender has more count. Therefore, it can improve the result by having clearer visual of the graph by applying this chart.

# 9.0 Conclusion

First of all, I would like to express my gratitude to lecturer “Liew Yee Jing” for his guidance and advice on this assignment. I managed to complete the assignment on time and deliver the required knowledge from this module such as data exploration, data pre-processing, and data manipulation to apply in this assignment with explanations on source code and output, as well as express the outline of findings based on the results. In addition, I also managed to explore the internet and add in extra features to improve the result for better visualization of the graph for ease of reading.

This assignment has enhanced my knowledge and skills in R programming language on cleaning raw data and pre-process them into useful data in the correct format. Furthermore, it is also essential to transform data into useful and meaningful information before analyzing the data in order to give the human resource a precise information to do decision making. Then, I also learned about how to improve the appearance of the graph.

Aside from the data analysis, this dataset has allowed me to do research about British Columbia, Canada from the internet. I found this exploration of this area a real eye-opener by looking into landmarks of Victoria, the capital city, and Vancouver, the largest city in British Columbia. Other than that, it sparks my interest to research more about the population in British Columbia.

Lastly, the information and result that gather from all the analysis and findings of several relationships between the attributes and variables provided a meaningful and valuable insight for the human resource to make decisions after receiving the suggestions from data analyst.

# References

BrightHR. (2021, November 21). *Gender inequality at work*. Retrieved from BrightHR: https://www.brighthr.com/articles/equality-and-discrimination/gender-equality/

DigiMarConCanadaWest. (2021, November 18). *British Columbia Cities*. Retrieved from DigiMarConCanadaWest: https://digimarconcanadawest.ca/british-columbia-cities/

Government of Canada. (2021, November 19). *CPP retirement pension:Overview*. Retrieved from Government of Canada: https://www.canada.ca/en/services/benefits/publicpensions/cpp.html

macrotrends. (2021, November 19). *Canada Inflation Rate*. Retrieved from macrotrends: https://www.macrotrends.net/countries/CAN/canada/inflation-rate-cpi

McGillivray, B. (2021, November 21). *Victoria*. Retrieved from britannica.com: https://www.britannica.com/place/Victoria-British-Columbia

Oberholtzer, A. (2017, September 13). *5 Reasons Why Your Customer Service Reps are Leaving (& What to Do About it)*. Retrieved from Baudville: https://ideas.baudville.com/the-baudville-blog/5-reasons-why-your-customer-service-reps-are-leaving-what-to-do-about-it

Statistics Canada. (2021, November 22). *Census Profile,2016 Census*. Retrieved from Statistics Canada: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=POPC&Code1=0293&Geo2=PR&Code2=59&SearchText=Fort%20Nelson&SearchType=Begins&SearchPR=01&B1=All&TABID=1&type=0

WorldAtlas. (2021, November 19). *The Largest Cities In Canada*. Retrieved from WorldAtlas: https://www.worldatlas.com/articles/biggest-cities-in-canada.html