Company name

Description automatically generated

**INDIVIDUAL ASSIGNMENT**

**022023-HKB**

**Programming for Data Analysis**

**HAND IN DATE: 29h May 2023**

**Online Submission Time before = 09: 00 PM Malaysia Time**

**Student Name:** WONG KANG SHIN

**Student ID:** TP068522

**INSTRUCTIONS TO CANDIDATES:**

1. Assignment is to be submitted through online submission (Moodle).
2. Students are advised to underpin their answers with the use of references (cited using the APA name system of Referencing).
3. Late submission will be awarded zero (0) unless Extenuating Circumstances (EC) are upheld.
4. Cases of plagiarism will be penalized.
5. You must obtain 50% overall to pass this module.

Table Of Contents

[1.0 Introduction and Assumptions 4](#_Toc136292591)

[2.0 Data Import / Cleaning / Pre-processing / Transformation 5](#_Toc136292592)

[3.0 Question and analysis 7](#_Toc136292593)

[3.1 Questions 1: What is the attrition for different aspects of the company? 7](#_Toc136292594)

[3.1.1 Analysis 1.1 What is attrition based on year? 7](#_Toc136292595)

[3.1.2 Analysis 1.2 What is attrition based on departments. 8](#_Toc136292596)

[3.1.3 Analysis 1.3 What is the attrition based on departments with year. 9](#_Toc136292597)

[3.1.4 Analysis 1.4 What is the attrition based on Job Title of all years. 10](#_Toc136292598)

[3.1.5 Analysis 1.5 What is the attrition based on Stores of all years. 11](#_Toc136292599)

[3.1.6 Analysis 1.6 Which city or stall has closed? 12](#_Toc136292600)

[3.1.7 Conclusion 13](#_Toc136292601)

[3.2 Questions 2: What happened to the company in 2009, 2014, and 2015? 14](#_Toc136292602)

[3.2.1 Analysis 2.1 What is the attrition for job titles in the year 2009 Specifically? 14](#_Toc136292603)

[3.2.2 Analysis 2.2 What is attrition for job titles in the year 2014 Specifically? 15](#_Toc136292604)

[3.2.3 Analysis 2.3 What is the attrition for job titles in the year 2015 Specifically? 16](#_Toc136292605)

[3.2.4 Analysis 2.4 What is the reason for Job titles termination for 2009? 17](#_Toc136292606)

[3.2.5 Analysis 2.5 What is the reason for Job titles termination for 2014? 18](#_Toc136292607)

[3.2.6 Analysis 2.6 What is the reason for Job titles termination for 2015? 19](#_Toc136292608)

[3.2.7 Analysis 2.7 What is the age of people who retired? 20](#_Toc136292609)

[3.2.8 Analysis 2.8 Which department is in danger of retirement? 21](#_Toc136292610)

[3.2.9 Analysis 2.9 Which store is in danger of retirement? 22](#_Toc136292611)

[3.2.10 1Analysis 2.10 How long did employees who are laid off work for? 23](#_Toc136292612)

[3.2.11 Analysis 2.11 Are employees Laid off in 2015 and 2014 related to Age. 24](#_Toc136292613)

[3.2.12 Analysis 2.12 What is the of age managers who Remained? 25](#_Toc136292614)

[3.2.13 Analysis 2.13 What is the of age managers who Remained excluding high-volume jobs? 26](#_Toc136292615)

[3.2.14 Analysis 2.14 Is Employees Laid off in 2015 and 2014 related to Location Demographic? 27](#_Toc136292616)

[3.2.15 Analysis 2.15 Is high Laid Off due to generally having high employee count in that city or otherwise. 28](#_Toc136292617)

[3.2.16 Analysis 2.16 Is Laid off related to gender. 29](#_Toc136292618)

[3.2.17 Conclusion 30](#_Toc136292619)

[3.3 Questions 3: Is there demographics imbalance in the company? 31](#_Toc136292620)

[3.3.1 Analysis 3.1 What is the gender distribution across the whole company? 31](#_Toc136292621)

[3.3.2 Analysis 3.2 What is gender distribution across departments? 32](#_Toc136292622)

[3.3.3 Analysis 3.3 Is gender distribution across job title balanced? 33](#_Toc136292623)

[3.3.4 Analysis 3.4 Is the gender distribution across Job title excluding the high-volume job balanced? 34](#_Toc136292624)

[3.3.5 Analysis 3.5 What is the gender distribution across stores? 35](#_Toc136292625)

[3.3.6 Conclusion 36](#_Toc136292626)

[3.4 Questions 4: What is the relationship between Age, Job title and years worked? 37](#_Toc136292627)

[3.4.1 Analysis 4.1 Which year contained employees who worked the longest? (To find best example for answering question) 37](#_Toc136292628)

[3.4.2 Analysis 4.2 What is the age of the employee when they're first hired. 38](#_Toc136292629)

[3.4.3 Analysis 4.3 What is the relationship between hired age and years worked? 39](#_Toc136292630)

[3.4.4 Analysis 4.4 What is the correlation and regression age and years worked? 40](#_Toc136292631)

[3.4.5 Analysis 4.5 What is the relationship between Job Title and Years worked? 41](#_Toc136292632)

[3.4.6 Conclusion 42](#_Toc136292633)

[3.5 Questions 5: What is the year’s prediction of how long more employee’s world work before leaving? 43](#_Toc136292634)

[3.5.1 Analysis 5.1 What is the prediction of How long each employee would stay based on Age 43](#_Toc136292635)

[3.5.2 Analysis 5.2 What is the prediction of How long each employee would stay based on Job Title 44](#_Toc136292636)

[3.5.3 Analysis 5.3 What are the weights of Job Title and Age's predictions? 45](#_Toc136292637)

[3.5.4 Analysis 5.4 What is the prediction of length of work based on both Age and Job Title 46](#_Toc136292638)

[3.5.5 Analysis 5.5 What is the difference between predicted length of service vs actual length in produce department. 47](#_Toc136292639)

[3.5.6 Analysis 5.6 What is the predicted remining life span of each Job Title 48](#_Toc136292640)

[3.5.7 Analysis 5.7 How many employees are predicted to lose for the following years in each Job Title? 49](#_Toc136292641)

[3.5.8 Conclusion 50](#_Toc136292642)

[3.0 extra features 50](#_Toc136292643)

[3.1 Regression summary 50](#_Toc136292644)

[3.2 Prediction 51](#_Toc136292645)

[3.3 theme\_minimal() 52](#_Toc136292646)

[3.4 Position stack 53](#_Toc136292647)

[3.5 Scale\_y\_discrete() 54](#_Toc136292648)

[3.6 Violine graph 55](#_Toc136292649)

[3.7 Pie chart 56](#_Toc136292650)

[3.8 Vjust and Hjust 57](#_Toc136292652)

[Bibliography 58](#_Toc136292653)

# 1.0 Introduction and Assumptions

Data analysis involves processes with the goal of identifying trends and creating reasoning and predictions. The purpose of this analysis was to find hidden problems and issues in a company.

The dataset provided includes the company’s employee details from the year 2006 to 2015. The dataset provided variables Employee ID, Record Date, Birthdate, Hired Date, Termination Date, Age, Years worked, City, Department, Job Title, Store Name, Gender Short, Gender Full, Termination Reason, Termination Type, Year Status, Status, Head Office.

There were limitations to the dataset as data variables such as salary, fringe benefits, working hours and leaves were not available. These variables could have a significant impact on the companies performance and attrition rates.

R version 4.2.2 was used to analyze the data provided. The package broom, dplyr, lubridate, readxl, ggplot2 and datasets were used to assist analyzation. These tools provided adequate functions for data cleaning, manipulating, and visualizing.

The techniques used for analyzation were data cleaning, manipulating, and visualizing. Data cleaning involves deleted duplicate data, reformatting and renaming columns. Data manipulation allows us to create new variables, use calculation, filter wanted data and make conclusions out of new ones. Visualization displays graphs that help identify trends and patterns in the data set.

Trends found in dataset were reasoned and given hypothesis. Predictions were done using past data for better decision making and preparation for identified possible threats.

Lastly, dataset does not include employee’s personal identity, contacts, or background. Analysis was done for improvement of business operation and our conduct was done within the rights of employee’s privacy.

# 2.0 Data Import / Cleaning / Pre-processing / Transformation

A screenshot of a computer program

Description automatically generated with low confidence

*Figure 1.1 Data import and cleaning code*

The excel containing employee data was imported to R studio. The columns were renamed first for easier access and reference to in throughout the program. Miss spellings such as “Resignaton” to “Resignation” was addressed for professional labeling. The data contains duplicated data for several years record of the same employee. Therefore, the data was filtered to only having 2015, the latest year records. This is stored in a new variable just in case the whole data needs to be analyzed. The format of columns with dates was then standardized.

# 3.0 Question and analysis

## 3.1 Questions 1: What is the attrition for different aspects of the company?

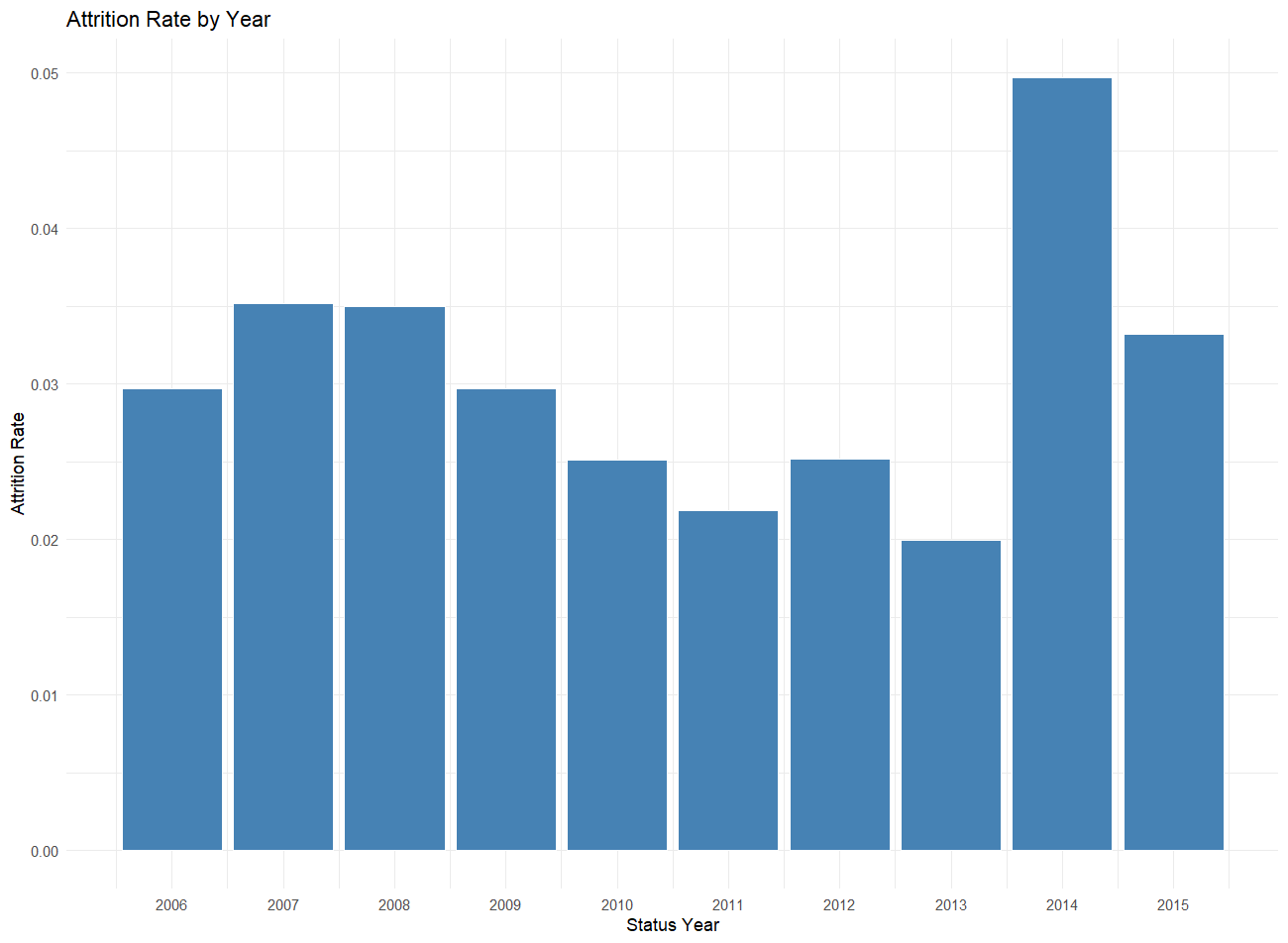
### 3.1.1 Analysis 1.1 What is attrition based on year?

A picture containing text, font, screenshot

Description automatically generated

*Figure 1.2 Attrition of year code*

The purpose of this analysis was to calculate the attrition rate of the company for the past 9 years. This will provide a general view of the companies’ current state.



*Figure 1.3 Attrition of year graph*

The results we got were pleasing. An online article on Chron, suggest that an attrition rate of below 10% is healthy for the company(Gartenstein, 2019). The attrition rate according to the results ranges from 2% to 5%, which provides adequate room for improvements of staff in the company.

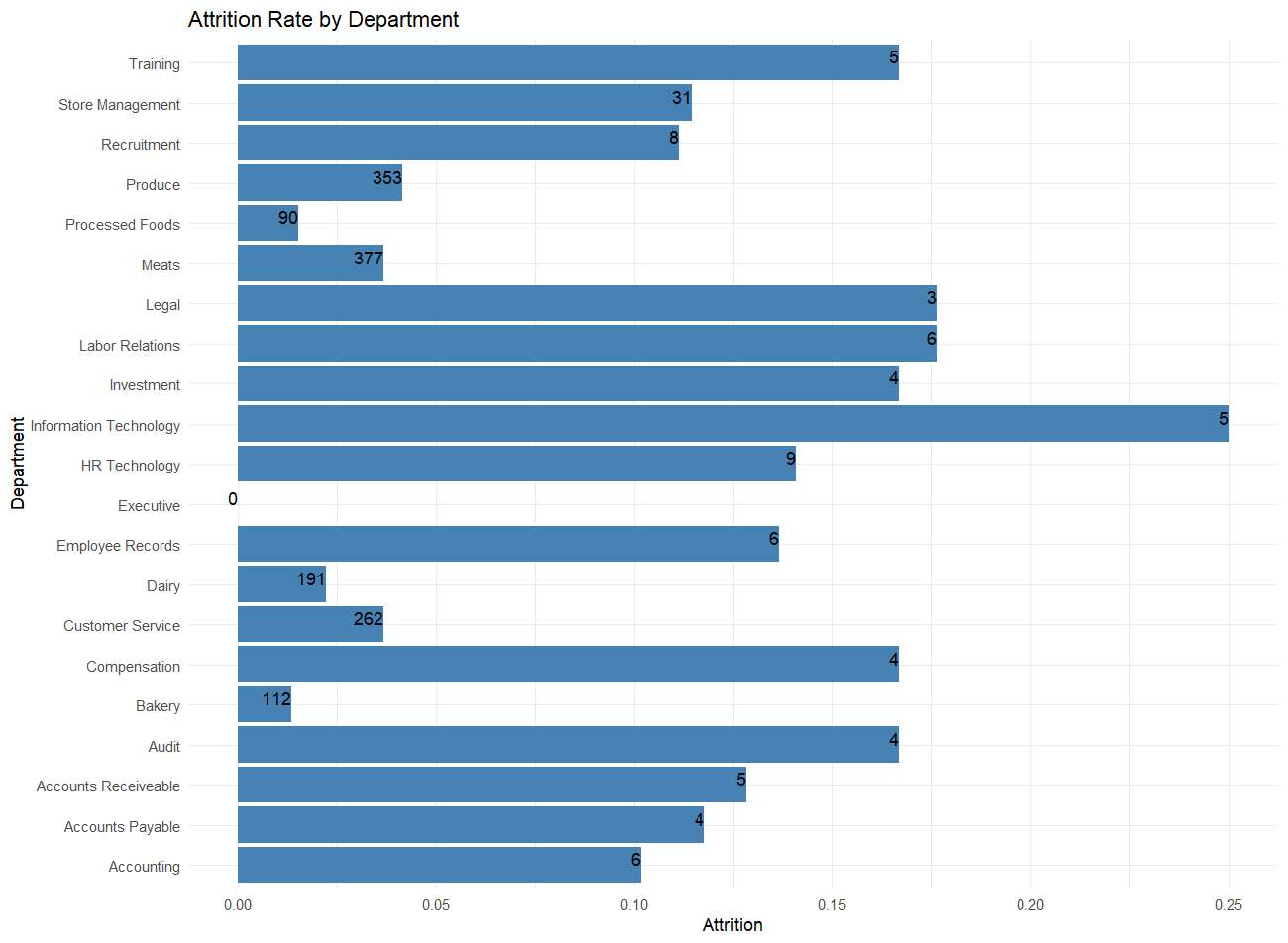
### 3.1.2 Analysis 1.2 What is attrition based on departments.

A screenshot of a computer

Description automatically generated with low confidence

*Figure 1.4 Attrition department code*

From the previous analysis, we studies attrition of the entire company, this analysis focuses on attrition of each department in the company. The labels show the termination count of each department, while the bar displays attrition rate



*Figure 1.5: Attrition department graph*

The results displayed relatively high attrition for Information technology, followed by training, legal, labor relation, investment, compensation, and audit. This department has less employee counts, therefore losing a few employees can have a significant impact on the attrition rate.

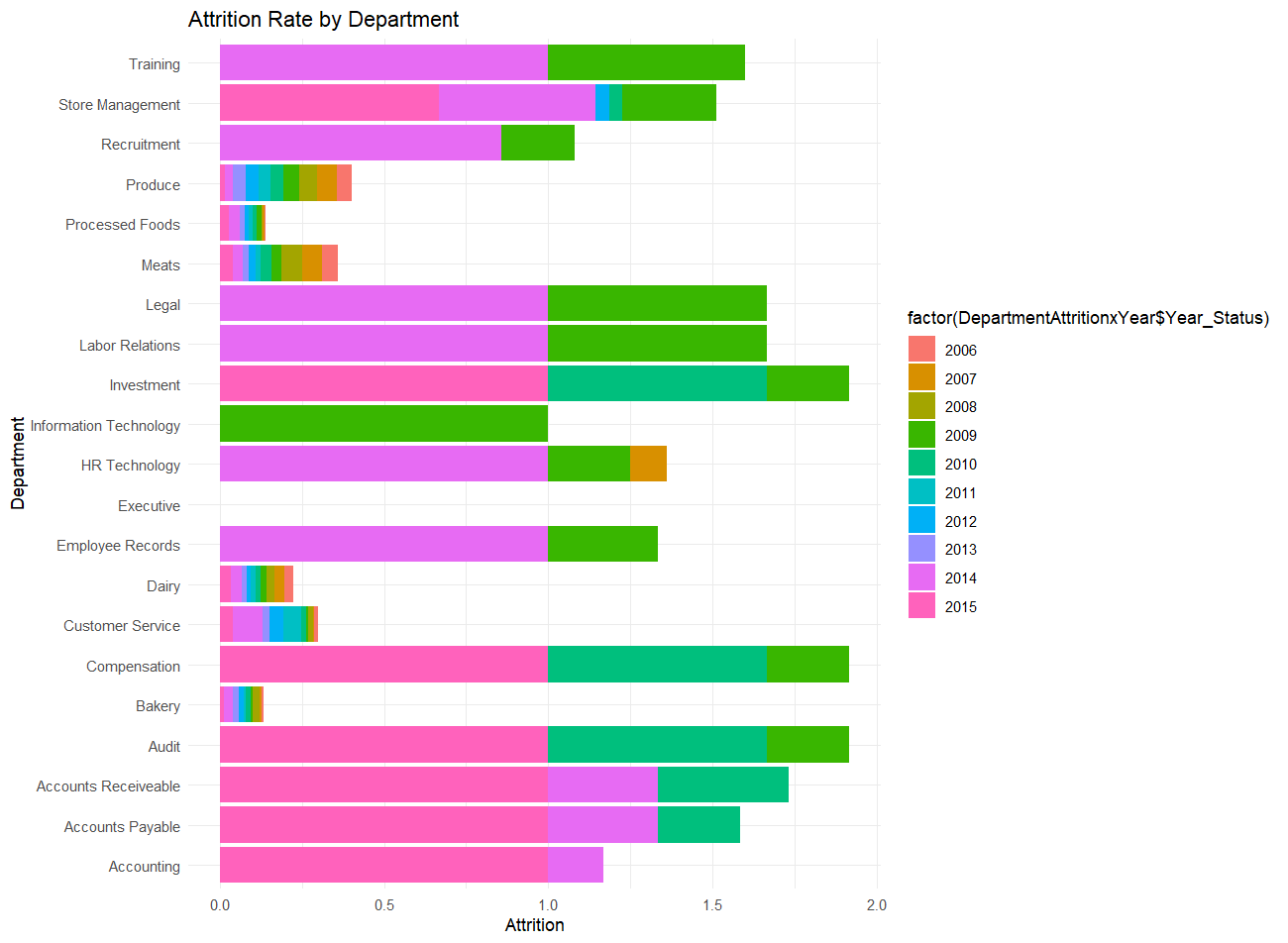
### 3.1.3 Analysis 1.3 What is the attrition based on departments with year.

A picture containing text, font, line, screenshot

Description automatically generated

*Figure 1.6: Attrition department with year code*

This analysis calculates the attrition rate in departments, for every single year. This will show us the state of the company’s department for each year.



*Figure 1.7: Attrition department with year graph*

The results from this analysis are detrimental. In the year 2015, Account, Accounts Payable, Audit and Compensation department has lost all its employees. In the year 2014, Training, Legal, Labor Relations, HR Technology and Employe Records Department has lost its employees. While 2009 displayed very high attrition rates as well. The department losing employees tends to be management departments. A positive reason could be the outsourcing of accounting management to a third-party company, which allows the company to focus on production. A negative reason could be poor planning done by the recruitment team, who failed to recruit sufficient employees to compensate for the termination of employees.

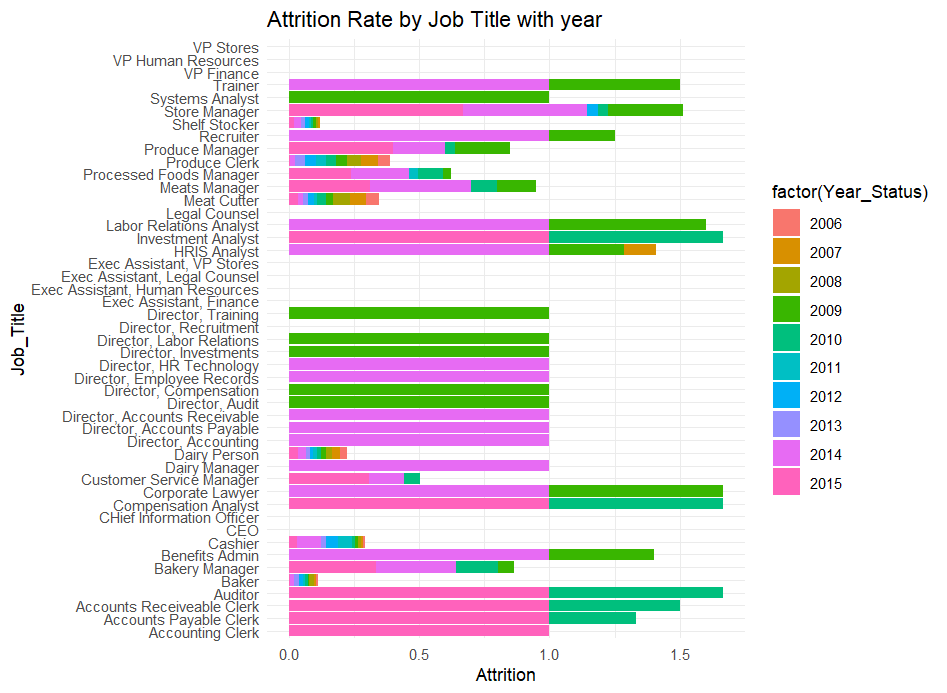
### 3.1.4 Analysis 1.4 What is the attrition based on Job Title of all years.

A screenshot of a computer

Description automatically generated with low confidence

*Figure 1.8 Attrition job title with year code*

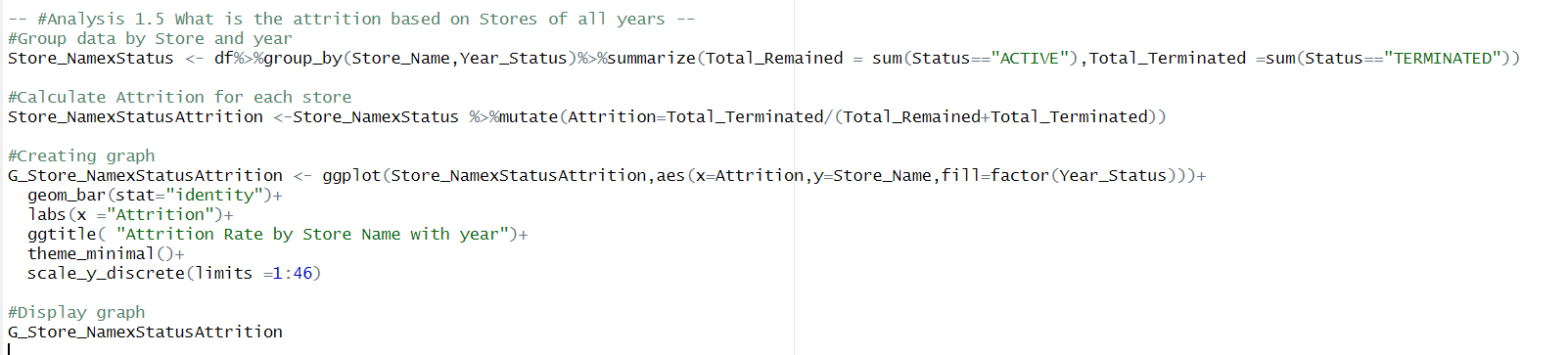
This analyses the loss of employees throughout the years for each Job Title.



*Figure 1.9 Attrition job title with year graph*

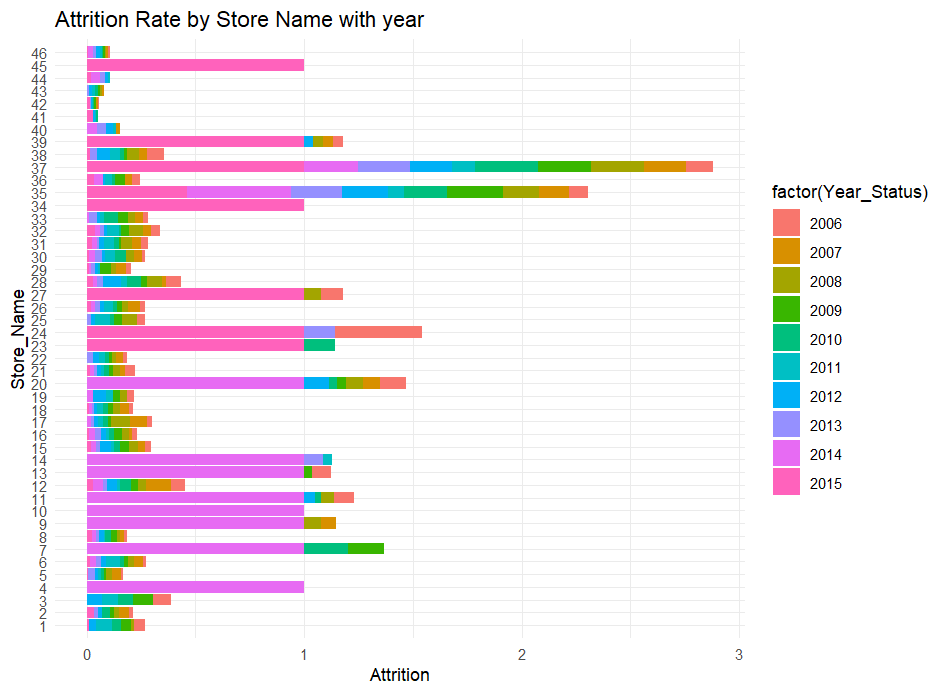
The company lost almost all its directors in the years 2009 and 2014. It loses most of the accounts department crew in year 2015. The reason for such high attrition in the year 2009,2014,2015 are unknown and must be analyzed in the following analyses.

### 3.1.5 Analysis 1.5 What is the attrition based on Stores of all years.



*Figure 1.10 Attrition stores title with year code*

The purpose of this analysis is to find out the attrition rate of each store.



*Figure 1.11 Attrition stores title with year graph*

The results indicate that store 4,7,9,10,11,13,14,20,23,24,25,27,34,37,39,45. Has lost all it’s employees by 2015. There are many possible reasons for the closure of these stores. Management might have decided to reduce store count for quality control. Or these stores are closed due to non-profitability. The reason for these attrition rates must be further analyzed.

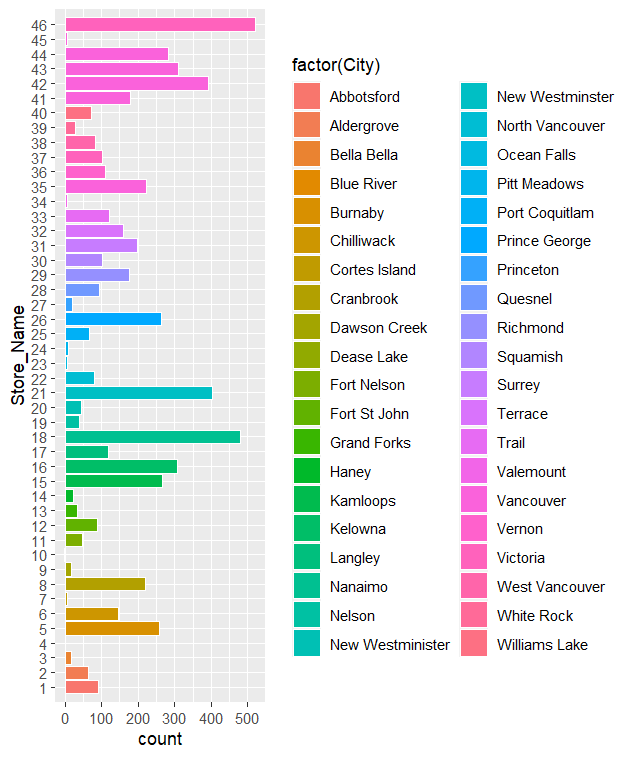
### 3.1.6 Analysis 1.6 Which city or stall has closed?

A picture containing text, font, screenshot, line

Description automatically generated

*Figure 1.12 Attrition stores title with city code*

These analyses tell us the location of each stall, and the employee count of each of them.



*Figure 1.13 Attrition stores title with city graph*

The findings reveal that the store location and its location of employees are directly related. Therefore, the loss of some employees may be due to the closing down of stores in that state.

### 3.1.7 Conclusion

In question 1, we discovered that the company generally has a relatively low attrition. However, once analysis was narrowed down into specific departments and job titles, the problems and detrimental attrition rates were discovered. Most departments and job titles have lost all their employees in the years 2009,2014 and 2015. The cause behind these high attrition rates will be further analyzed in the next question.

## 3.2 Questions 2: What happened to the company in 2009, 2014, and 2015?

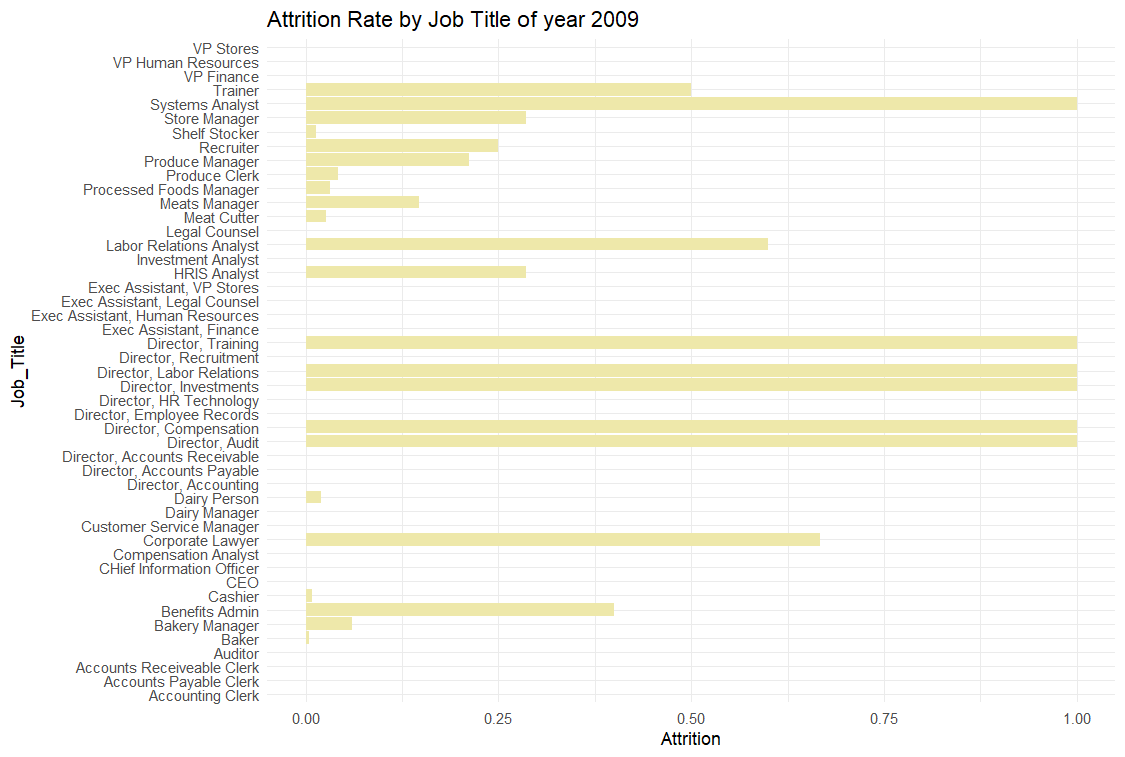
### 3.2.1 Analysis 2.1 What is the attrition for job titles in the year 2009 Specifically?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.1 Attrition job title of year 2009 code*

The analysis focuses on the great attrition rates found in year 2009 for Job titles.



*Figure 2.2 Attrition job title of year 2009 graph*

The findings provide insights into the root issue of the high attrition rates. In the year 2009, all the Director of recruitment, training, compensation, audit, has been terminated. The recruitment and training team are crucial for rehiring and the retraining of employees. Without these departments, there will be no renewal of employees for the years to come.

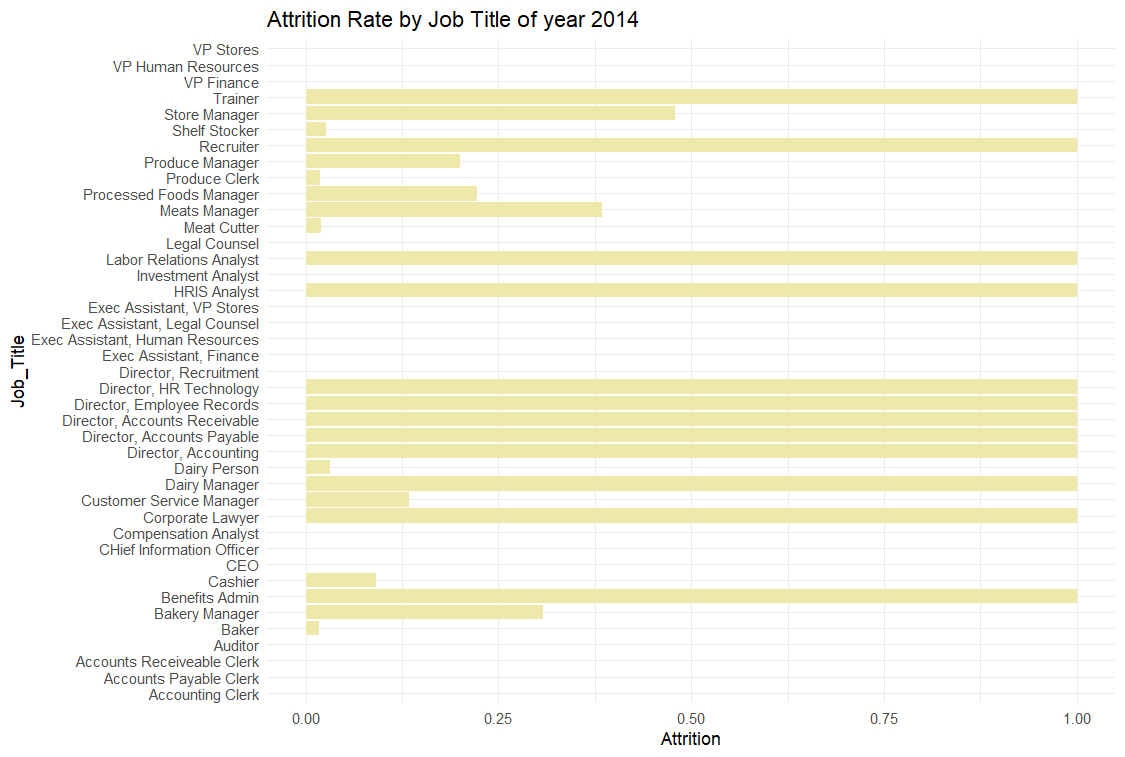
### 3.2.2 Analysis 2.2 What is attrition for job titles in the year 2014 Specifically?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.3 Attrition job title of year 2014 code*

The analysis focuses on the great attrition rates found in year 2014 for Job titles.



*Figure 2.4 Attrition job title of year 2014 graph*

It is no surprise that the data not long after the year 2009 displays an even detrimental stat of the company. In the year 2014, the company lost all its Benefits Admin, Corporate Lawyer, dairy manager, HRIS analyst, Labor relation Analyst, Recruiter, Trainer, Director of Accounting, Accounts Payable, Accounts Receivable, Employee Records, HR Technology.

The lack of management, without managers or directors in the accounting department and dairy department. The company may soon lose control and organization of that department. This could lead to further loss of employees in that department due to low working morale with no direction. Or inefficiency in the department as there are no supervisors to be looked after.

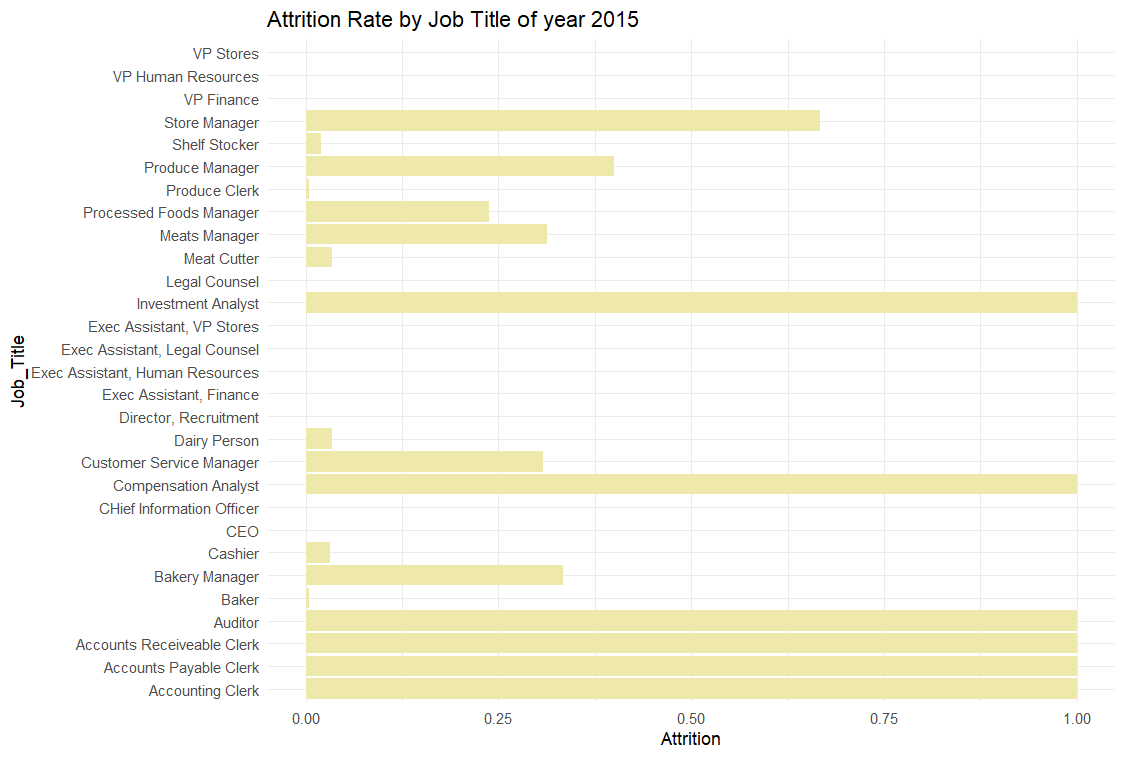
### 3.2.3 Analysis 2.3 What is the attrition for job titles in the year 2015 Specifically?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.5 Attrition job title of year 2015 code*

The analysis focuses on the great attrition rates found in year 2014 for Job titles.



*Figure 2.6 Attrition job title of year 2015 graph*

In the year 2015, all employees in the accounting department have been terminated. This might be the product of having no directors and managers across the accounting department. All investment analysts, compensation analyst and auditor have been terminated.

There is a spike in store manager and produce manager attrition. This might be due to the closure of stores in many states.

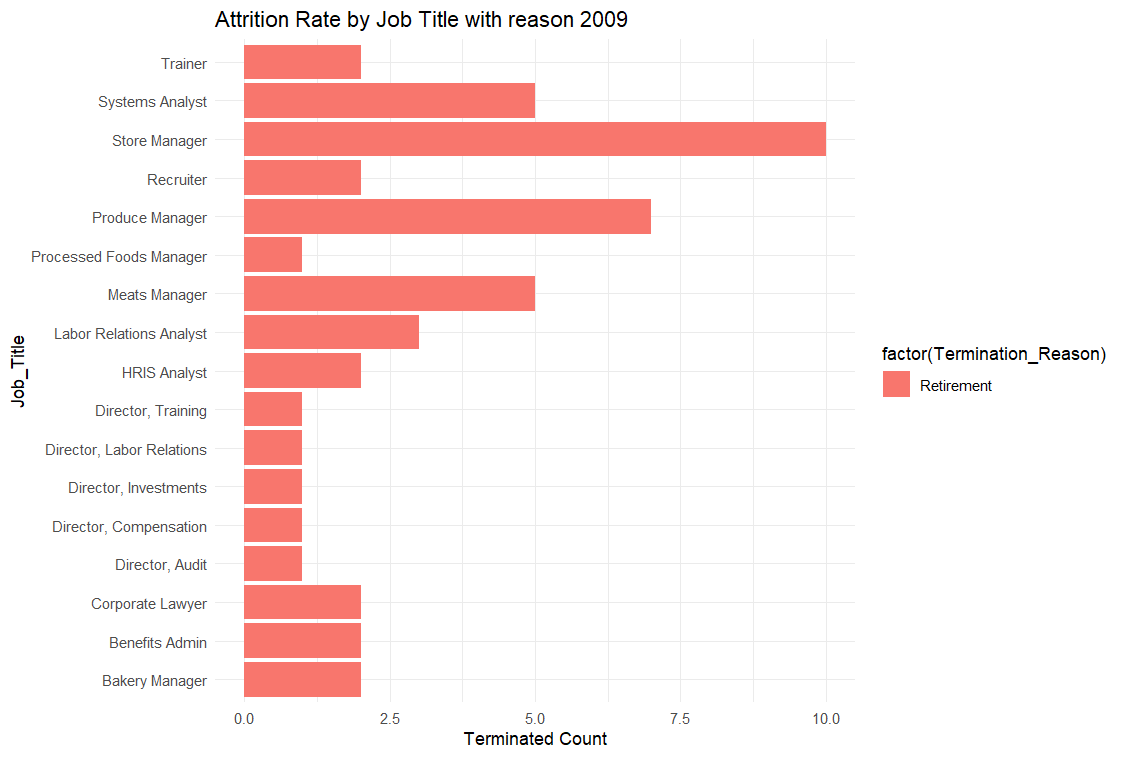
### 3.2.4 Analysis 2.4 What is the reason for Job titles termination for 2009?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.7 termination reason title of year 2009 graph*

This analysis focuses on the reason behind the spike in termination for 2009.

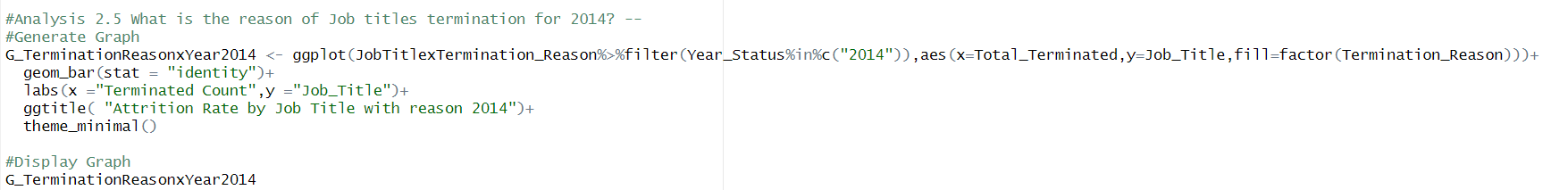


*Figure 2.8 termination reason title of year 2009 graph*

Before the analysis was done, it was suspected that high termination was due to the great recession from deregulation of financial institution with consumer’s mortgage overextension. (Nora Dunn, 2022) However,

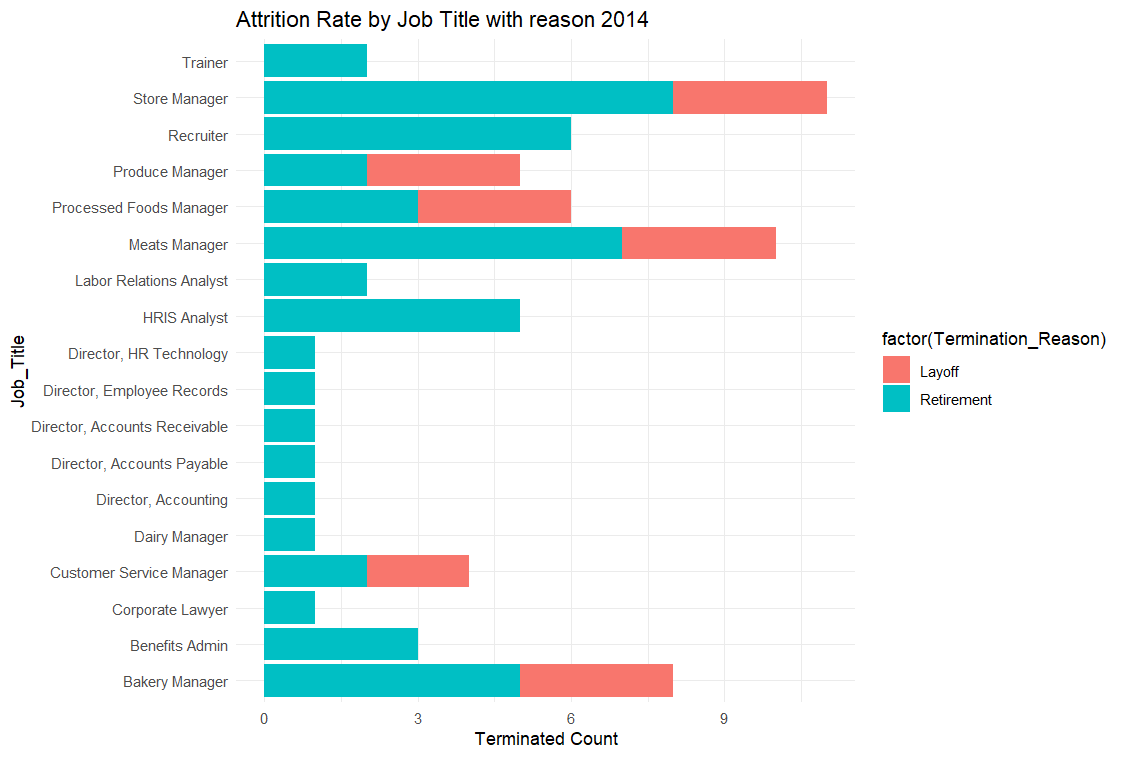
This result suggests that all the termination occurred in year 2009 was due to retirement. Therefore, the age of retirement must be analyzed to confirm this data and finding.

### 3.2.5 Analysis 2.5 What is the reason for Job titles termination for 2014?



*Figure 2.9 termination reason title of year 2014 code*

This analysis focuses on the reason behind the spike in termination for 2014.



*Figure 2.10 termination reason title of year 2014 graph*

The analysis result indicates that the termination in 2014 occurred due to employees being laid off and retirement. The possible reasons for being laid of may be the lack of security in the company after witnessing the lost of branches and managers. Further analyses are done in the next question to reason for this finding.

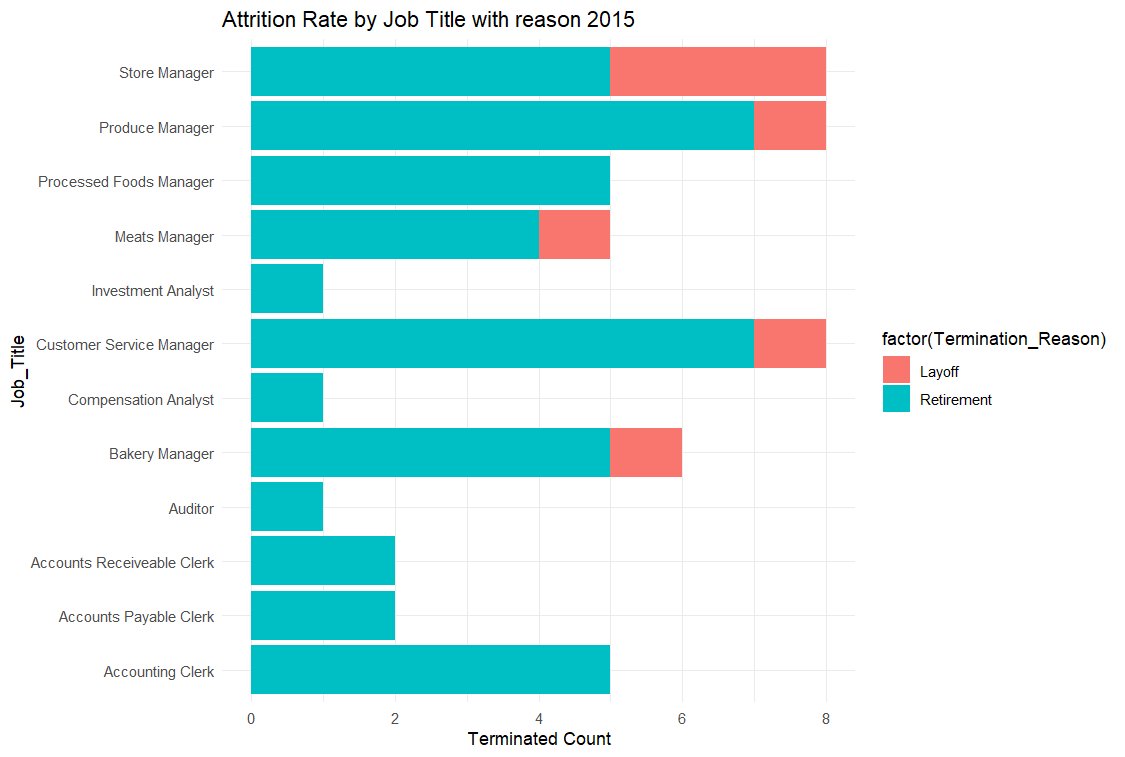
### 3.2.6 Analysis 2.6 What is the reason for Job titles termination for 2015?

A picture containing text, font, line, screenshot

Description automatically generated

*Figure 2.11 termination reason title of year 2015 code*

This analysis focuses on the reason behind the spike in termination for 2014.



*Figure 2.12 termination reason title of year 2015 graph*

Similarly, to 2014, the result indicates a mixed reason of termination between retirement and layoff.

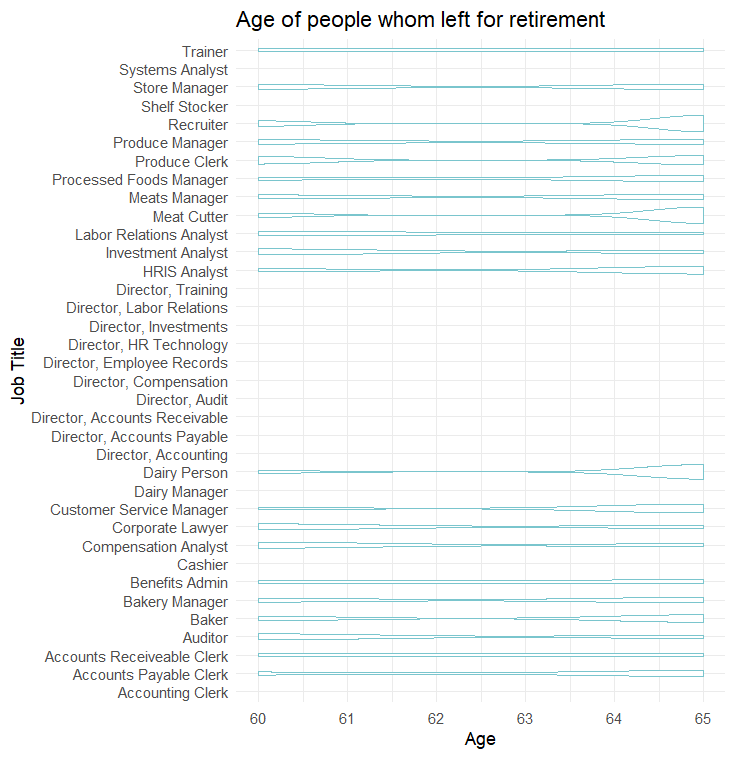
### 3.2.7 Analysis 2.7 What is the age of people who retired?

A picture containing text, screenshot, font, line

Description automatically generated

*Figure 2.13 Age retired code*

This analysis was done to reconfirm the termination reason “Retirement”, as it was hard to believe all termination occurred in a year was due to retirement. We verify the data by finding the age of all employees who are terminated due to retirement. An elder range of ages would reconfirm and strengthen the validity of the data.



*Figure 2.14 Age retired graph*

The violine graph suggests that all employees retire after the age of 60 and before the age of 65. According to Ramsey Solution. Most people retire at the age of 61-65.(Solutions, 2023) Therefore we can conclude the dataset was indeed accurate.

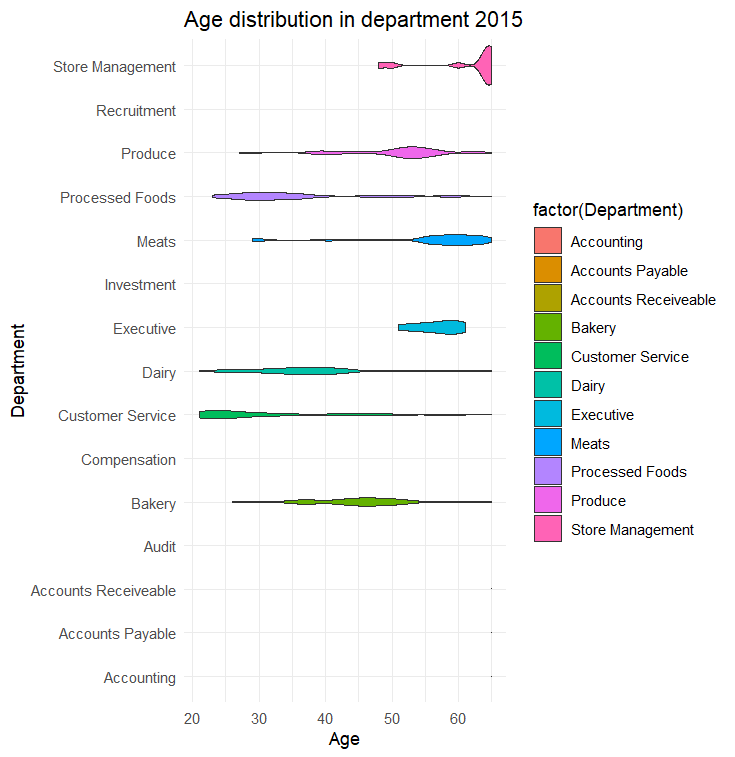
### 3.2.8 Analysis 2.8 Which department is in danger of retirement?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.15 Department in danger code*

After knowing the major reason for termination was retirement. We analyzed the possible departments that were at stake of being hit by another wave of generational retirement.



*Figure 2.16 Department in danger graph*

The department of Store Management and Meats displayed a very high risk of spiking attrition due to retirement. Therefore, the company was suggested to find an alternative solution, by recruiting younger staff. Or capitalizing and automatic the department.

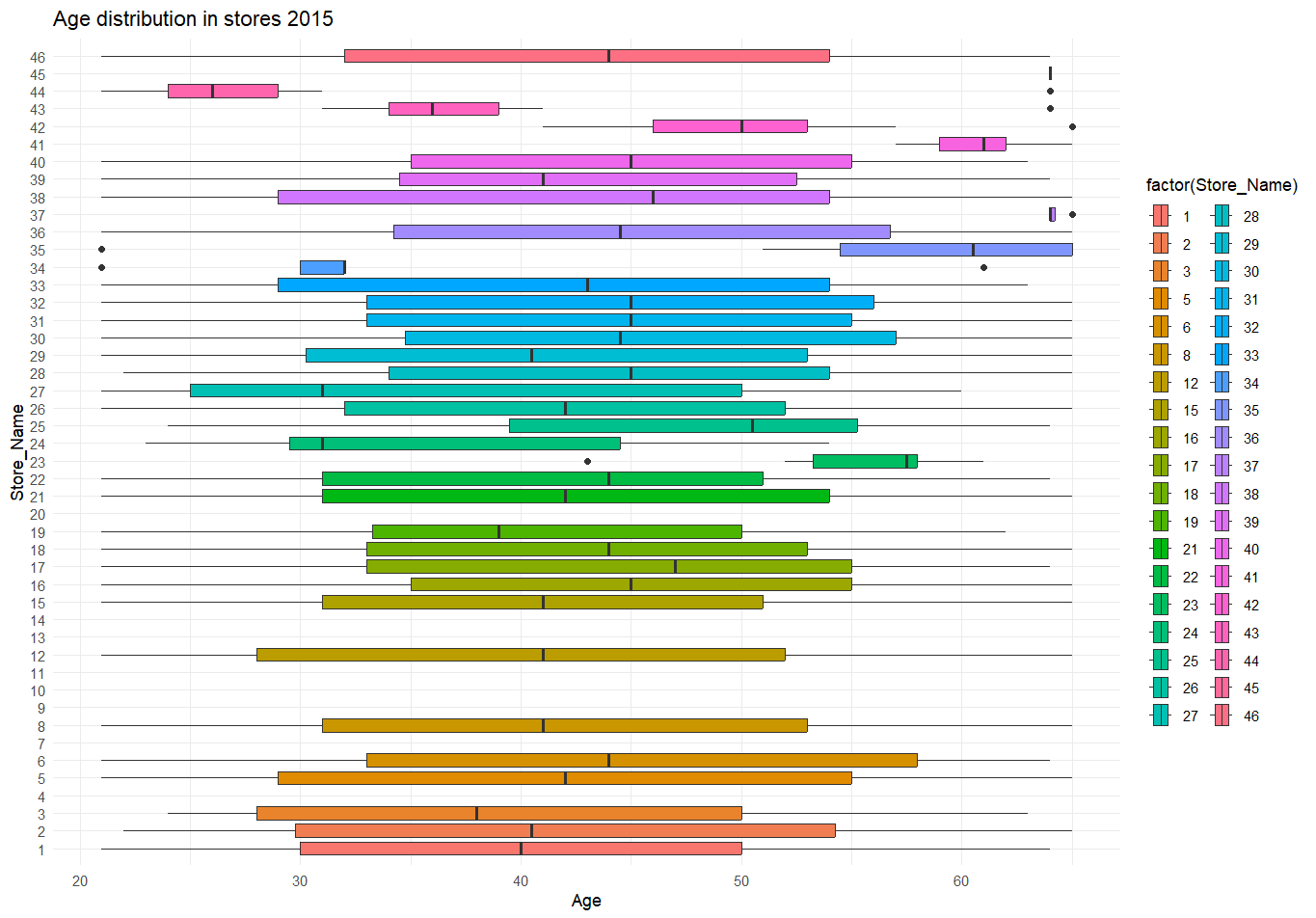
### 3.2.9 Analysis 2.9 Which store is in danger of retirement?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.17 store in danger code*

The analysis was done to search for stores with a high risk of losing employees from retirement soon.



*Figure 2.18 store in danger graph*

The findings suggest that stores 35 and 41 have a mean age of above 60. Therefore, it is recommended to recruit more staff in that city and store urgently.

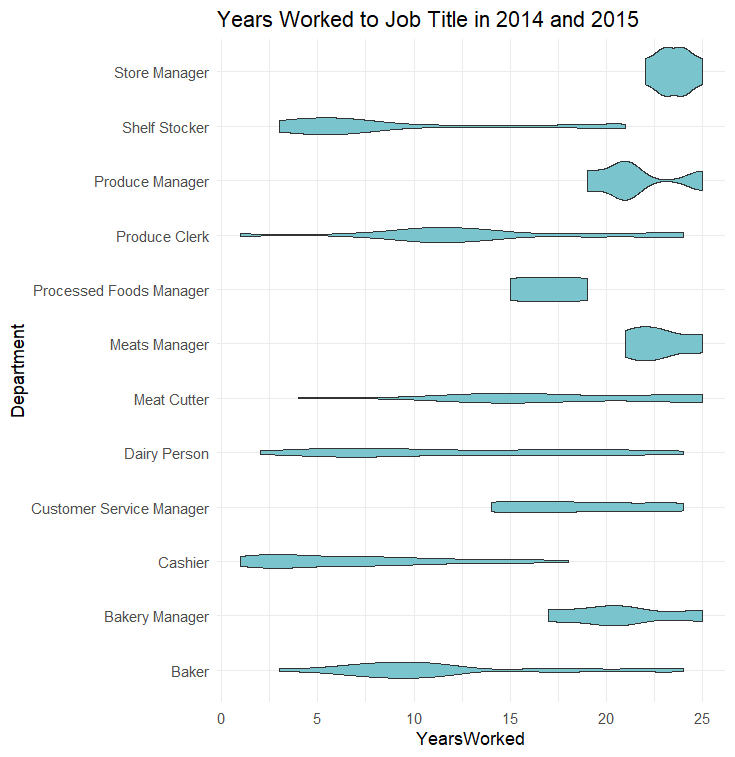
### 3.2.10 1Analysis 2.10 How long did employees who are laid off work for?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.19 Length worked before laid off code.*

This analysis was done to find out why employees are terminated due to being laid off. A short working year may suggest the incapability of new recruits to adapt to a company’s environment and a long working year may suggest skills being out of date, or the companies’ incapability to sustain a high volume of employees.



*Figure 2.20 Length worked before laid off graph*

The findings suggest that the Bakery Manager, Meat Manager, produce manager, and store managers have worked long duration before being laid off. This may be due to skills being out of date, digitalization of stores management, or companies’ incapability of upkeeping stores.

Bakery, cashier, and shelf stockers have less years worked. This might be due to the low skill requirements needed for the position. Therefore, having high turnover rain for this department is less harmful to the company as employees are easily replaceable.

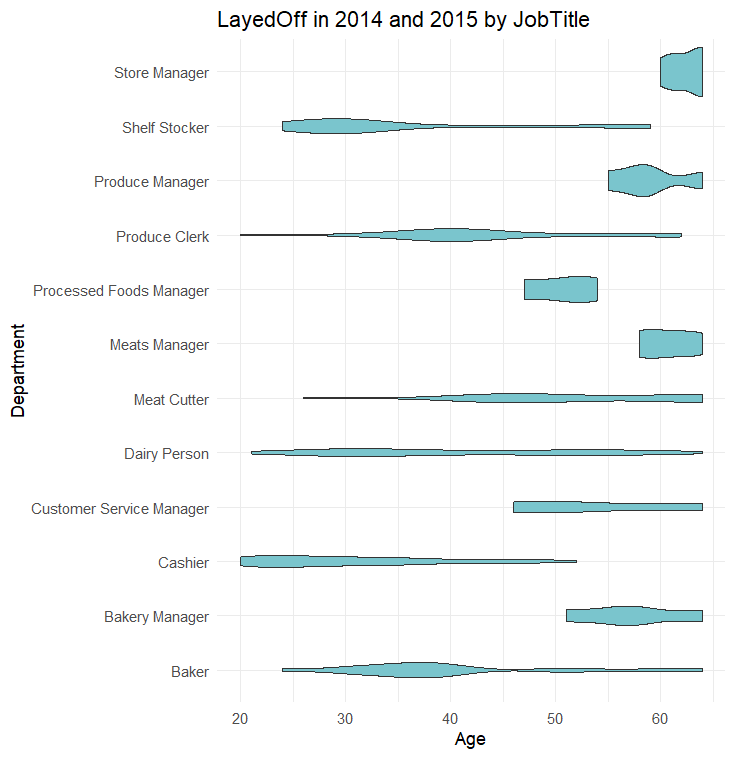
### 3.2.11 Analysis 2.11 Are employees Laid off in 2015 and 2014 related to Age.

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.21 Age of laid off in 2015,2014 code*

The previous analysis made a comparison of years worked with laid off termination. This analysis focuses of age to further confirm and strengthen the finding.



*Figure 2.22 Age of laid off in 2015,2014 graphs.*

Similarly, to the previous analysis, positions such as Meats Manger, Bakery Manager, Store Manager, produce manager requires time for skills to be developed to be qualified for these job titles. However, further findings need to be made to reason termination of loyal employees.

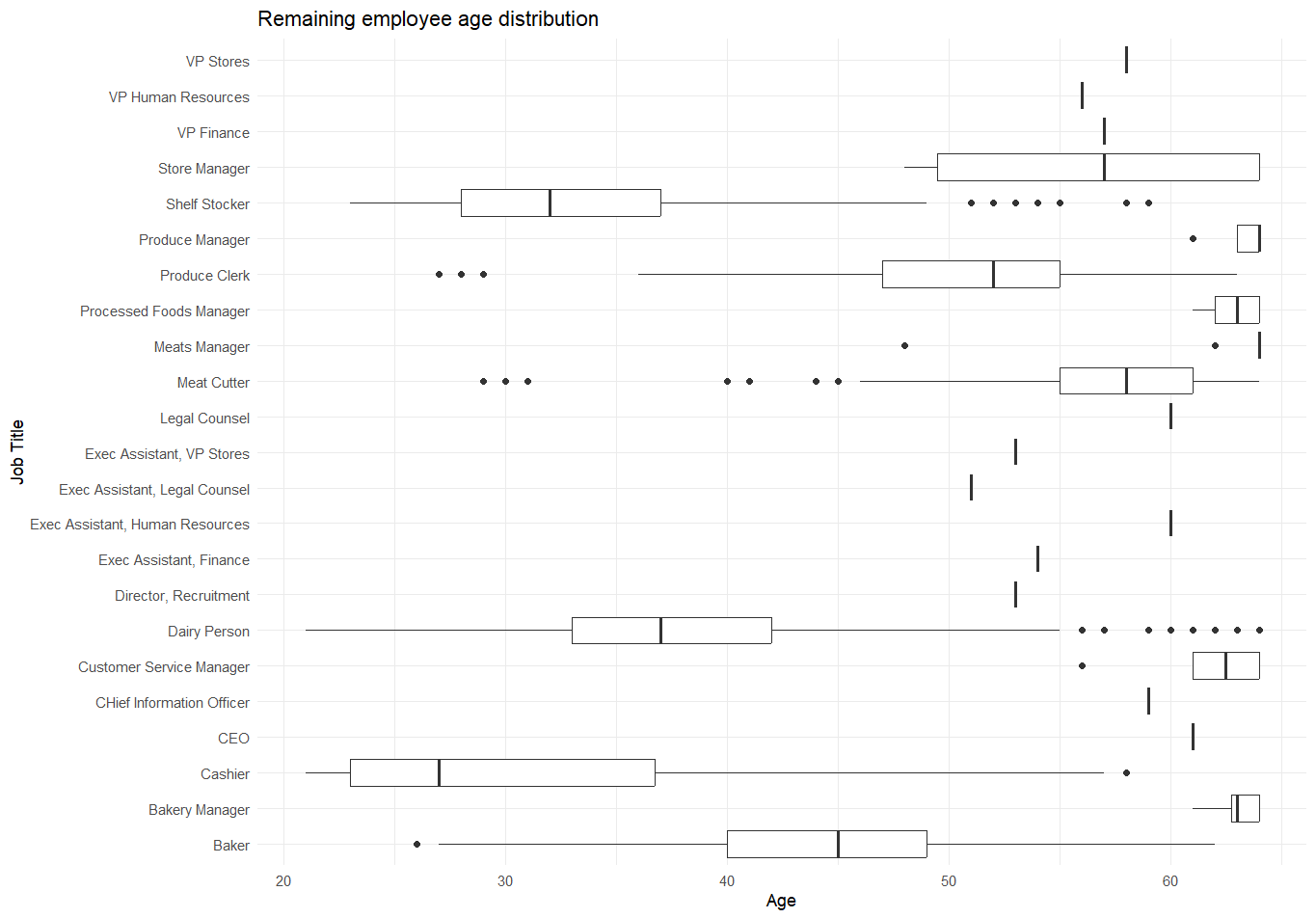
### 3.2.12 Analysis 2.12 What is the of age managers who Remained?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.23 Age of remaining employees in 2015,2014 codes*

This analysis was done to find out if the termination of employees who worked more than 10 years are terminated due to out of date skills or incapability to fund employees.



*Figure 2.24 Age of remaining employees in 2015,2014 codes*

The findings indicate that the age of employees remaining in managerial positions was above the age of 50. Therefore, it is unlikely that the termination of managers shown in the previous analysis was done due to out-of-date skills. But the closing of stores and the company’s incapability of funding these high position employees.

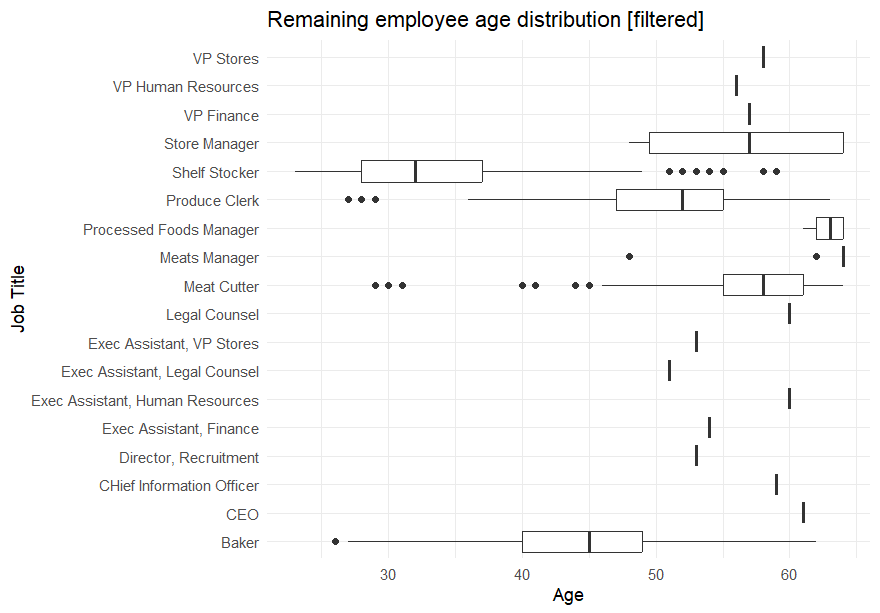
### 3.2.13 Analysis 2.13 What is the of age managers who Remained excluding high-volume jobs?

A picture containing text, font, screenshot

Description automatically generated

*Figure 2.25 Age of remaining employees in 2015,2014 (Low volume job titles) codes*

This analysis was done to display a clearer view of employees remaining in managerial positions.



*Figure 2.26 Age of remaining employees in 2015,2014 (Low volume job titles) graphs*

Likewise, the average age shown in most job titles and departments was above the age of 50. This is detrimental to these departments as most employees are likely to retire within the next 10 years if precautions are not taken.

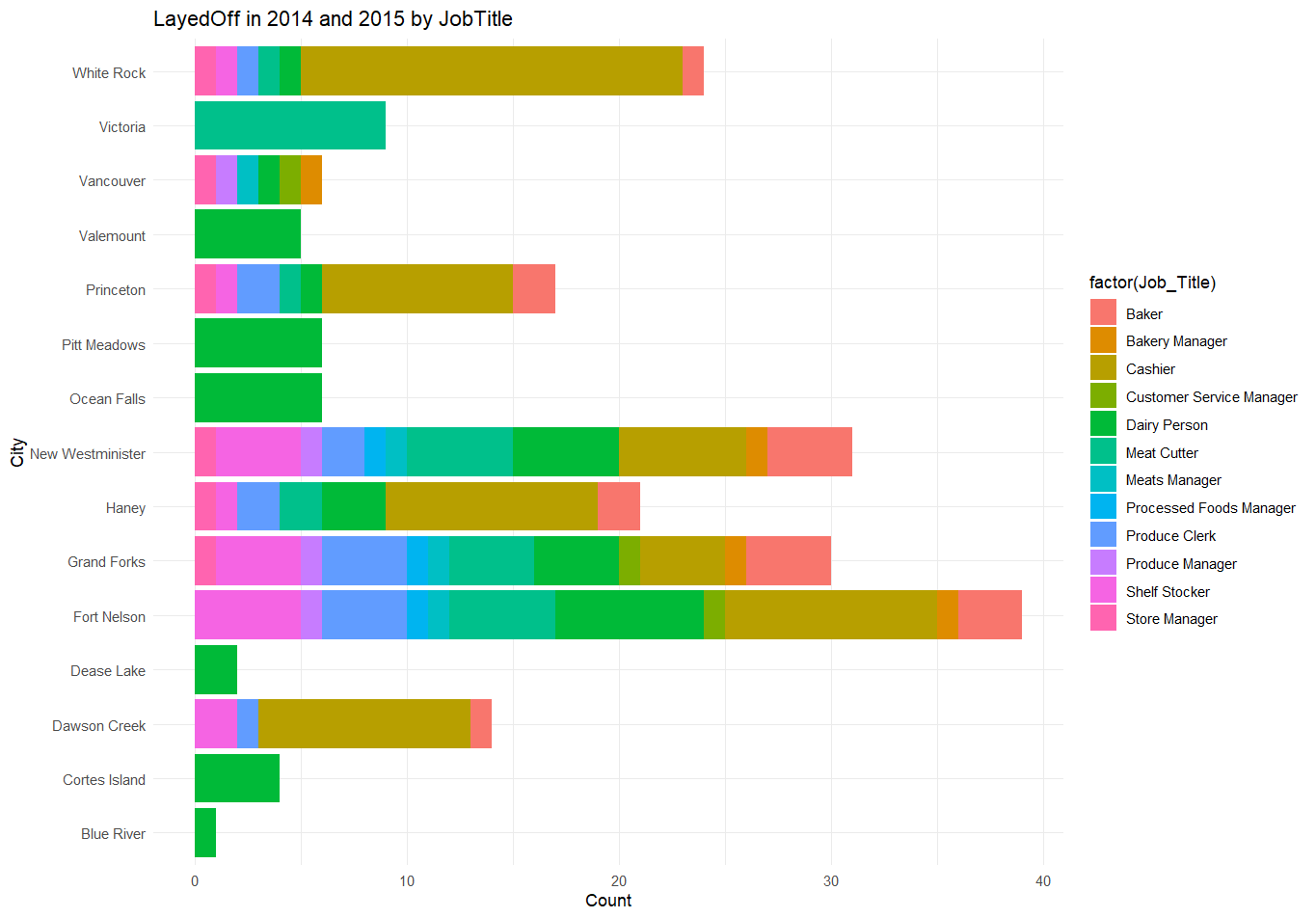
### 3.2.14 Analysis 2.14 Is Employees Laid off in 2015 and 2014 related to Location Demographic?

A picture containing text, font, line, screenshot

Description automatically generated

*Figure 2.27 Demographics of employee laid off in 2015 and 2014 codes*

As location and city was directly related to the stores. We analyzed the relationship between employees laid off with location demographics to find out, whether employees being laid of was due to the closure of stores.



*Figure 2.28 Demographics of employee laid off in 2015 and 2014 graph*

The finding suggests that New Westminster, Haney, Grand Forks, Fort Nelson, White Rock, and Princeton lose a high number of employees with a wide variety of employees. These are departments that have lost almost all their employees as shown in analysis 1.6. This further confirms that the laid off was done due to closure of shops.

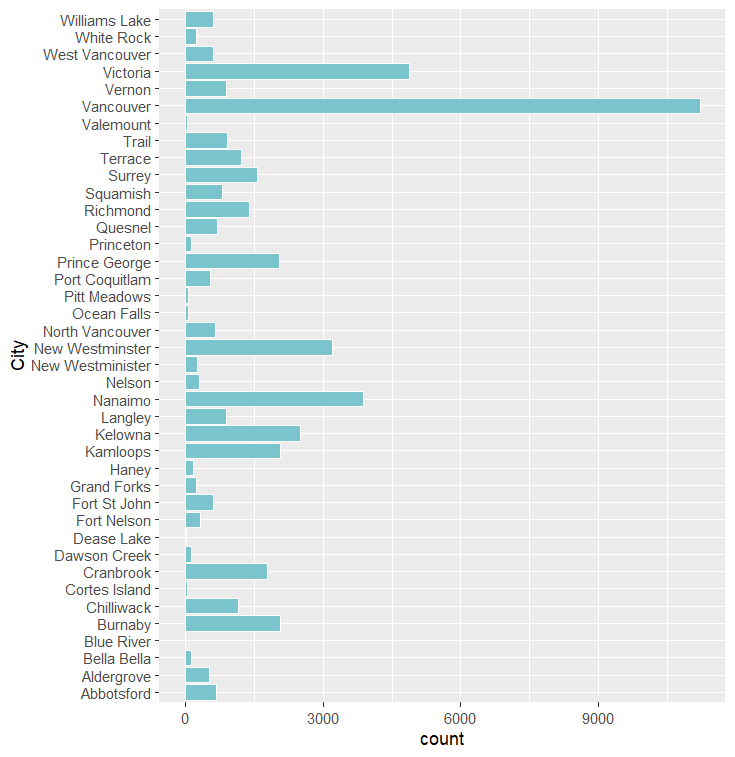
### 3.2.15 Analysis 2.15 Is high Laid Off due to generally having high employee count in that city or otherwise.

A picture containing text, screenshot, font, line

Description automatically generated

*Figure 2.29 Frequency of employees in cities codes*

In the previous analysis New Westminster, Haney, Grand Forks, Fort Nelson, White Rock, and Princeton displayed high termination. This analysis focuses on finding the employee count in each department to further strengthen the hypothesis of the previous analysis.



*Figure 2.30 Frequency of employees in cities graph*

The results display low employee count in Haney, Grand Forks, Fort Nelson, White Rock, and Princeton except New Westminster. This trend might spiral down as the remaining employees may get over the burden by workload increase done due to the termination of employees.

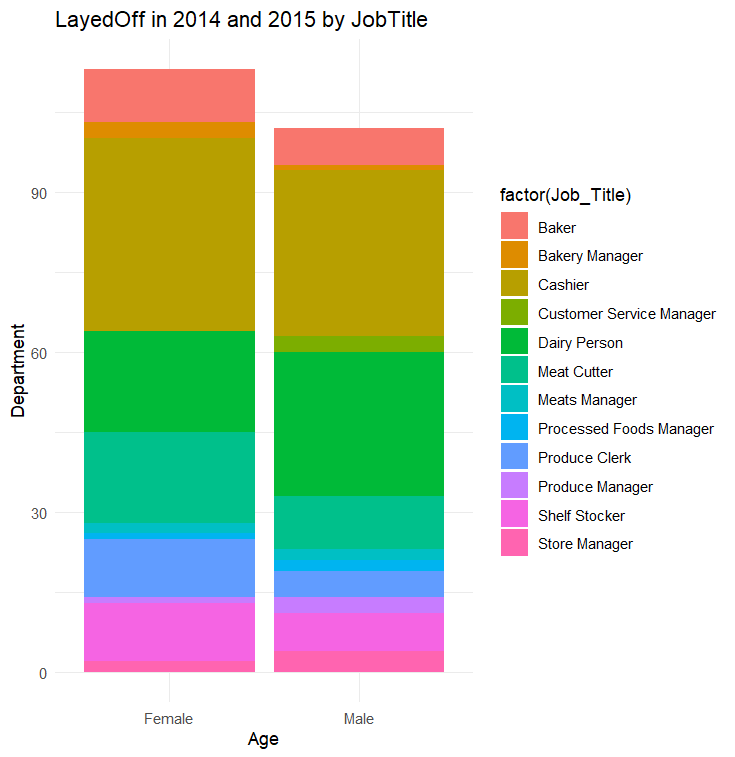
### 3.2.16 Analysis 2.16 Is Laid off related to gender.

A screenshot of a computer

Description automatically generated with low confidence

*Figure 2.31 Laid off employee’s gender codes.*

The analysis was done to figure out if being laid off was due to biases of employers.



*Figure 2.32 Laid off employee’s gender graph.*

The result displayed relatively balanced termination across gender of all departments. Therefore, termination was unlikely due to gender biases.

### 3.2.17 Conclusion

In question 2, we discovered that most terminations are due to retirement. This led to further analysis of which department would be greatly affected by termination in the coming years. Store Management and Meats are generally populated with elderly employees, which could be dangerous for the company. Several lay offs were made throughout the years, analysis suggest that it was neither due to employees having out of date skills, gender bias or digitalization of company, but the closure of stores in several cities and incapability to upkeep high employee count.

## 3.3 Questions 3: Is there demographics imbalance in the company?

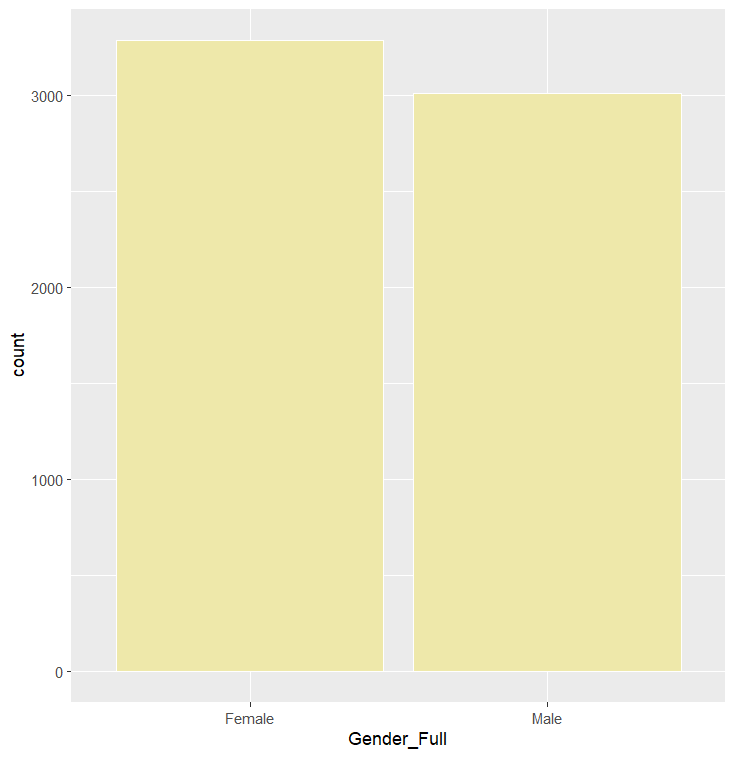
### 3.3.1 Analysis 3.1 What is the gender distribution across the whole company?

A picture containing text, screenshot, font, line

Description automatically generated

*Figure 3.1 Gender distribution of company codes*

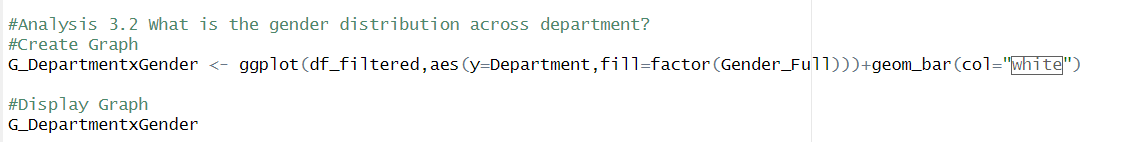
This analysis focuses on the general distribution of gender to spot biases in gender.



*Figure 3.2 Gender distribution of company graph*

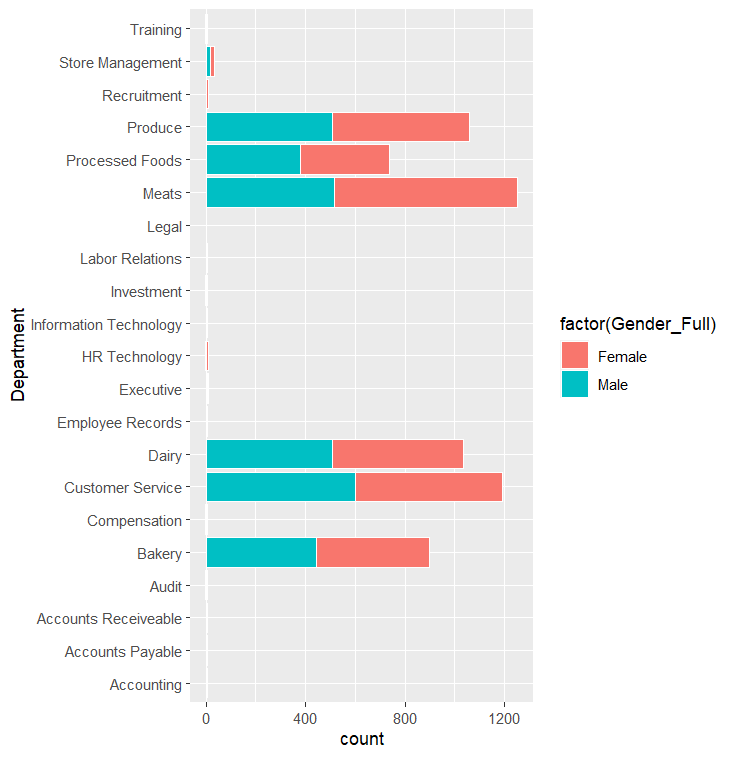
The results displayed a relatively balanced distribution. However, the target analysis must be narrowed down to further strength this hypothesis.

### 3.3.2 Analysis 3.2 What is gender distribution across departments?



*Figure 3.3 Gender distribution of departments codes*

This analysis aims to display gender distribution across all departments.



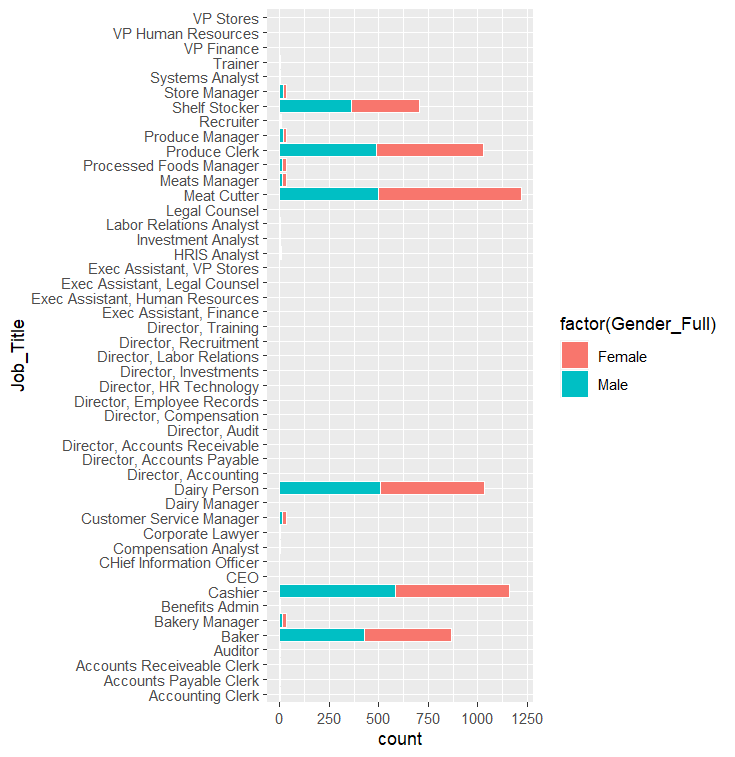
*Figure 3.4Gender distribution of departments graph*

The result indicates that the company has a very balanced gender across departments.

### 3.3.3 Analysis 3.3 Is gender distribution across job title balanced?

*Figure 3.5 Gender distribution of job titles code*

The analysis focuses on analyzing the distribution of employees across job titles.



*Figure 3.6 Gender distribution of job titles graph*

The results conclude that the distribution of employees is balanced across job titles with high volume. Further analysis must be done to assess the gender distribution in the department with low employee count.

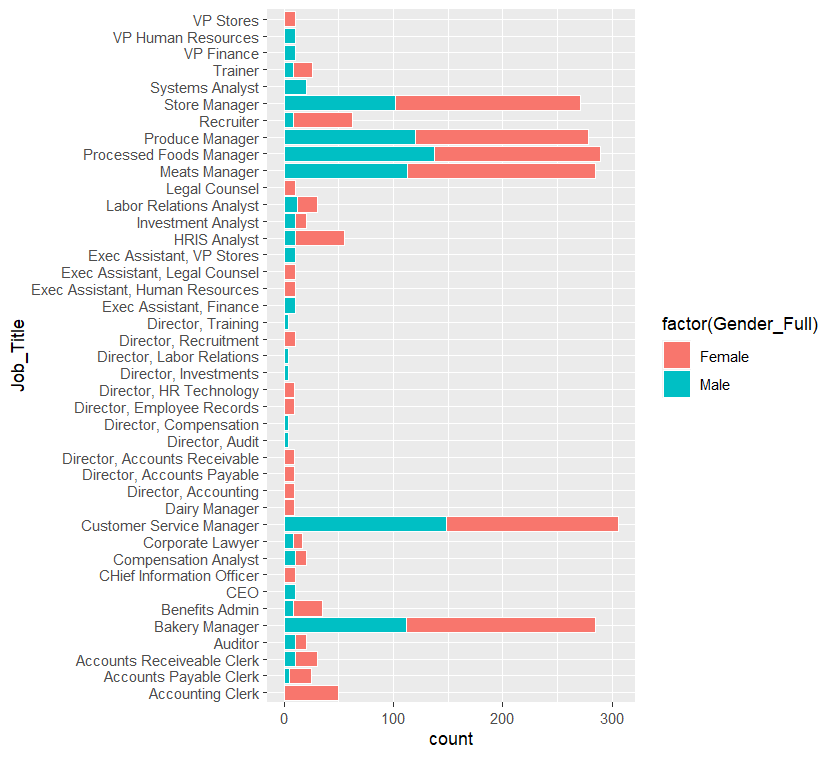
### 3.3.4 Analysis 3.4 Is the gender distribution across Job title excluding the high-volume job balanced?

A picture containing text, font, screenshot

Description automatically generated

*Figure 3.7 Gender distribution of job titles (In low volume department) code*

This analysis was done to search for gender equality within departments with low employee count and managerial positions.



*Figure 3.8 Gender distribution of job titles (In low volume department) graph*

The results indicates that Job title related to accounting are more dominated by female. This is common as female are usually more detailed oriented and suited for accounting roles. Job titles such as labor, system analyst, investment, compensation is more dominated by male. This could lead to gender biases in the further as most roles related to human management are managed by male.

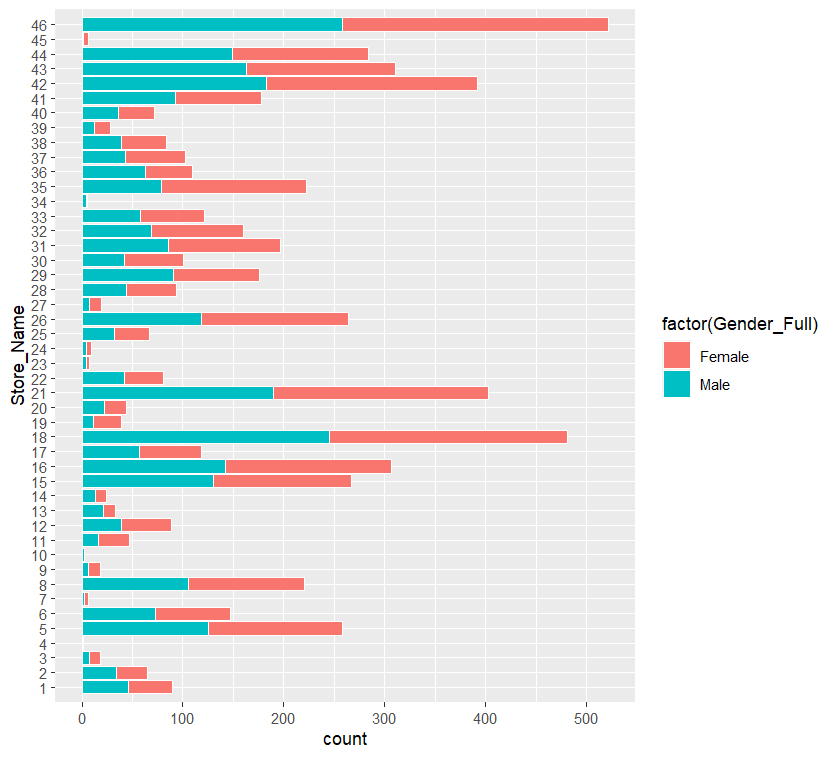
### 3.3.5 Analysis 3.5 What is the gender distribution across stores?

A close-up of a computer screen

Description automatically generated with low confidence

*Figure 3.9 Gender distribution across stores codes*

This analysis is done to ensure gender equality across stores of the company.



*Figure 3.10 Gender distribution across stores graph*

The results indicate a very positive distribution of gender across the store is very balanced.

### 3.3.6 Conclusion

In question 3, we search for any demographic biases in all aspects of the company to predict possible inequality. The gender distribution in aspect of department, high volume job titles and stores of the company is very positive and balanced. However, low volume job titles such as labor, compensation system analyst and investment are most dominated by male. This could lead to management biases in the future if not cautious.

## 3.4 Questions 4: What is the relationship between Age, Job title and years worked?

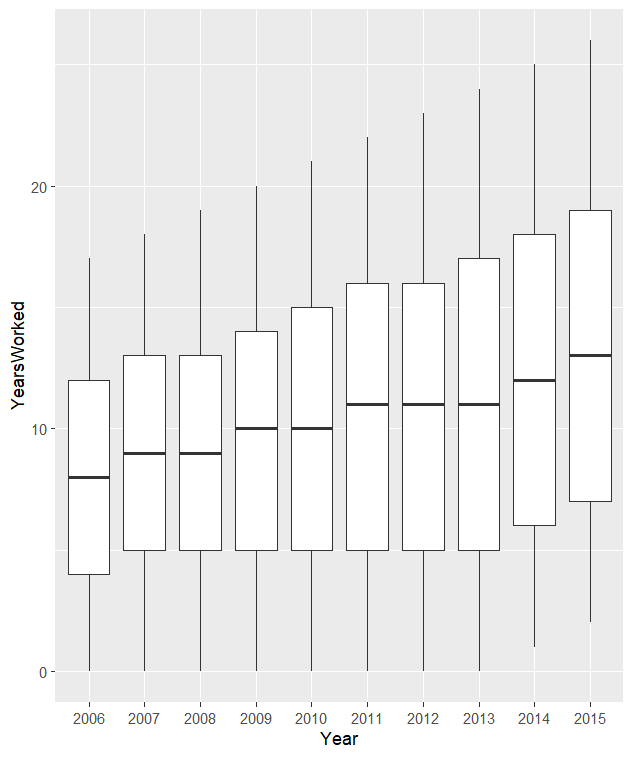
### 3.4.1 Analysis 4.1 Which year contained employees who worked the longest? (To find best example for answering question)

A picture containing text, font, screenshot

Description automatically generated

*Figure 4.1 Longest length of work year codes*

This analysis is done to find the year with the highest number of years the average employee in the company has worked.



*Figure 4.2 Longest length of work year codes*

The year 2015 displays the longest length employees worked. The linear trend indicates that most of the employees in the company are loyal. The year 2015 will be used for prediction and calculation samples.

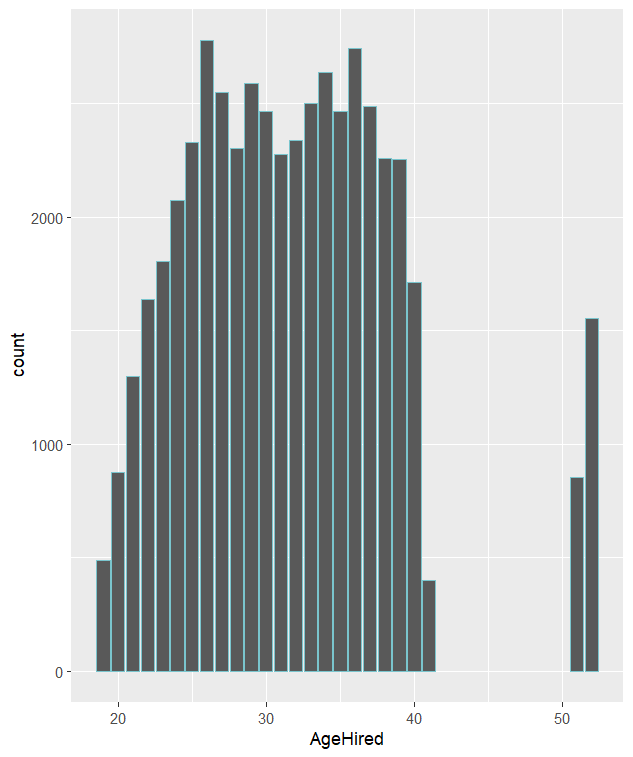
### 3.4.2 Analysis 4.2 What is the age of the employee when they're first hired.

A picture containing text, font, line, screenshot

Description automatically generated

*Figure 4.3 Age of employee hired codes.*

This analysis is done to figure out what age is every employee in the company hired.



*Figure 4.4 Age of employee hired graph.*

The age of the employees hired ranges from 20 to 40. Most of the hired employees were aged 27 to 38. This is positive as employees from age 27 to 38 has a good balance between experience and longevity in the company.

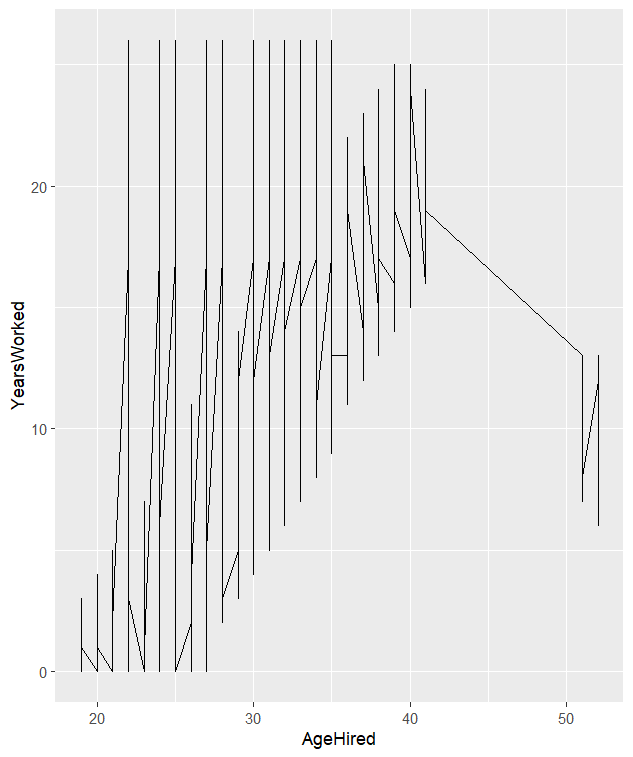
### 3.4.3 Analysis 4.3 What is the relationship between hired age and years worked?

A close-up of a computer screen

Description automatically generated with low confidence

*Figure 4.5 Age hired and years worked codes*

This analysis displays a graph of the relationship between years worked and hired age for employees who have been terminated.



*Figure 4.6 Age hired, and years worked graph*

The findings displayed a liner trend between age hired and years worked. This might be due to employees having less commitment at a young age to risk leaving the company and finding a new one. While employees at elder age are more concerned about job security. Another possible reason may be the jobs title young employees are employed for generally has less room for career advancement. Job title such as cashier are more dominated by young employees which may be the reason for low years worked in young employees.

### 3.4.4 Analysis 4.4 What is the correlation and regression age and years worked?

A picture containing text, screenshot, font

Description automatically generated

*Figure 4.7 Correlation and Regression of age and years worked codes*

This analysis is done to find out what percentage of years worked is related to age for employees who have been terminated.

A picture containing text, screenshot, font, number

Description automatically generated

*Figure 4.8 Correlation and Regression of age and years worked codes.*

The findings suggest that 48% of the length of employees working in the company is related to age of employee.

### 3.4.5 Analysis 4.5 What is the relationship between Job Title and Years worked?

A picture containing text, screenshot, font

Description automatically generated

*Figure 4.9 Relationship between Job title and years worked codes*

This analysis finds the general relationship between job titles and years worked for employees who have been terminated

A picture containing text, screenshot, font, diagram

Description automatically generated

*Figure 4.10 Relationship between Job title and years worked codes.*

The finding suggests that job titles that require more skills have higher years worked. This may be due to higher salary, opportunity for career advancement and much more.

A black text on a white background

Description automatically generated with low confidence

*Figure 4.10: Regression of job title and years worked result.*

The findings suggest that years worked is 64% effected and related to job title.

### 3.4.6 Conclusion

In question 4, we discovered that 2015 has the highest mean years worked in the company. Relationship between age, job title and years worked was analyzed and reasoned. Calculations were done to identified trends in data where employees were terminated. These trends can later be used to predict life span of remaining employees in further analysis.

## 3.5 Questions 5: What is the year’s prediction of how long more employee’s world work before leaving?

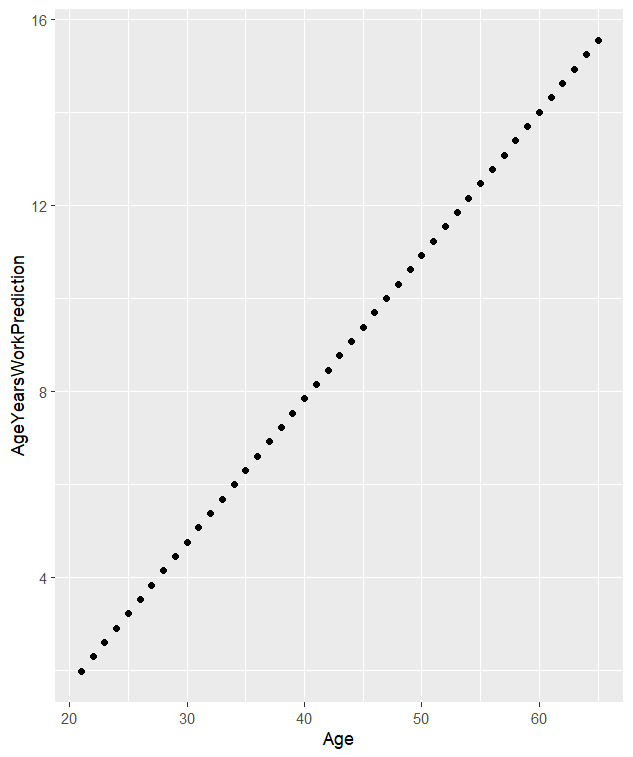
### 3.5.1 Analysis 5.1 What is the prediction of How long each employee would stay based on Age

A picture containing text, font, screenshot

Description automatically generated

*Figure 5.1: Employee life span prediction based on age codes.*

This analysis predicts the years of remaining employees depending on age.



*Figure 5.2: Employee life span prediction based on age graph*

The findings displayed a linear trend for employees who remain in the company. This prediction can be used to predict the remaining life span of employees in the company currently

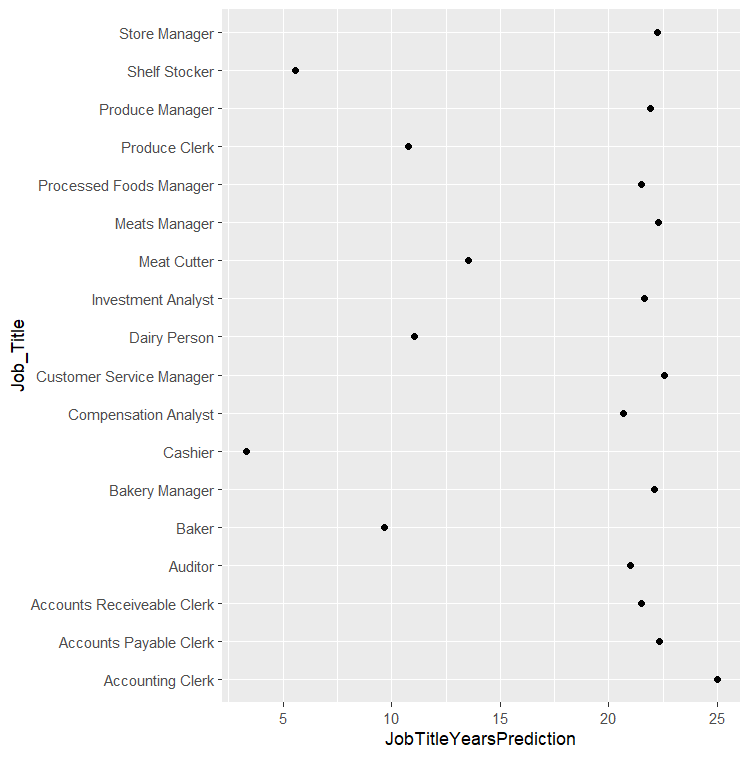
### 3.5.2 Analysis 5.2 What is the prediction of How long each employee would stay based on Job Title

A picture containing text, screenshot, font, algebra

Description automatically generated

*Figure 5.3: Employee life span prediction based on job title codes*

This analysis predicts the years of remaining employees depending on job titles



*Figure 5.4: Employee life span prediction based on job title graph.*

Just like analysis 4.5. Jobs with higher skill ceiling and career advancement are expected to have longer working life span of employees. This prediction can be used to predict the remaining life span of employees in the company currently.

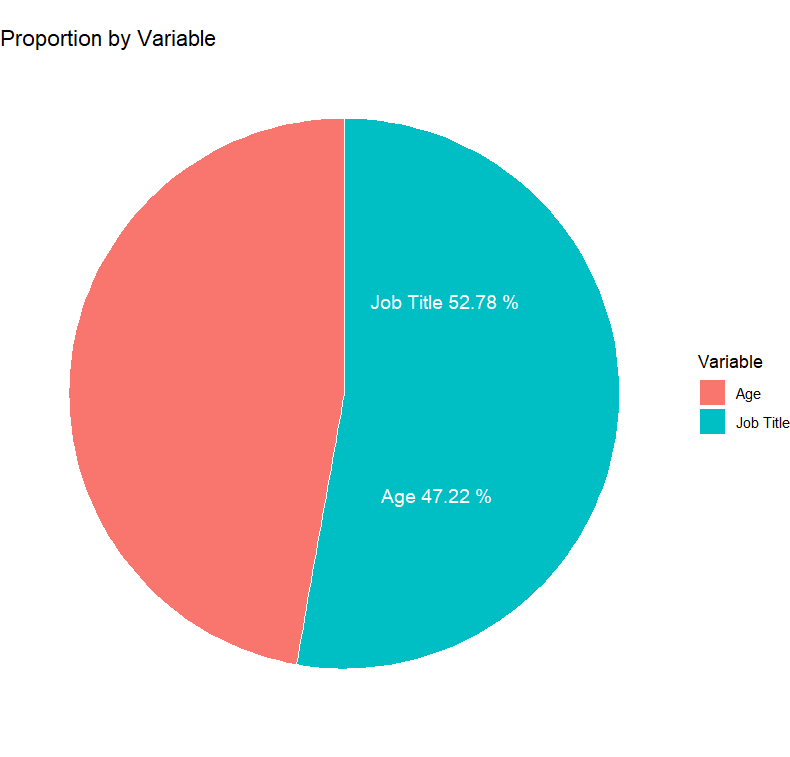
### 3.5.3 Analysis 5.3 What are the weights of Job Title and Age's predictions?

A picture containing text, font, screenshot

Description automatically generated

*Figure 5.5: Weights of prediction codes*

These calculations and analysis are done to find the weights of Job title and Age for more accurate predictions.



*Figure 5.6: Weights of prediction pie chart*

The findings suggest that Job title is 5.56%. A job title brings in factor of career advancement, salary, personal growth while age includes factor such as room for risk, skills and experience, and commit and responsibility. These two weights would give a balance prediction for upcoming analysis.

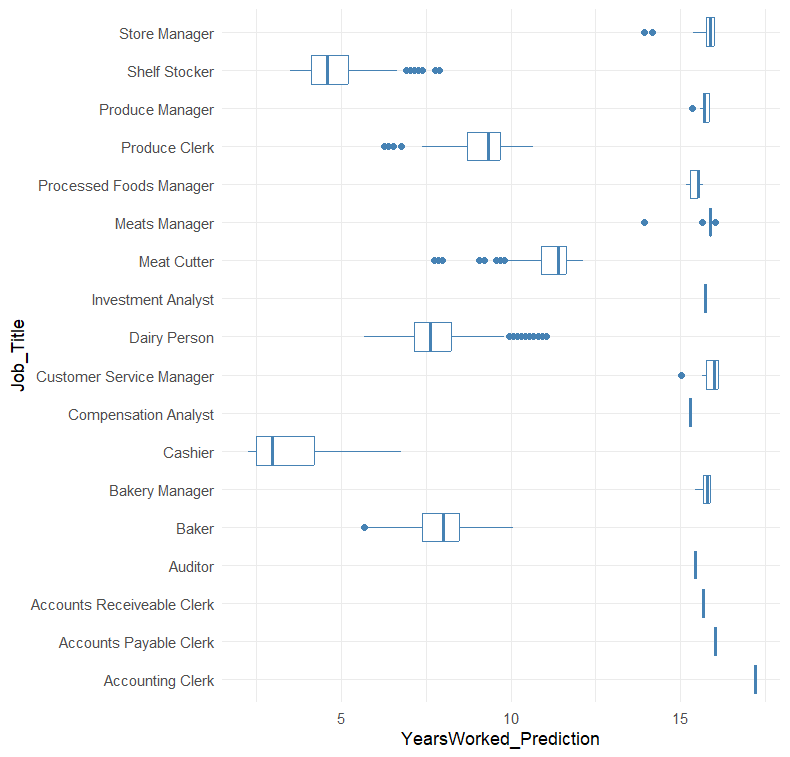
### 3.5.4 Analysis 5.4 What is the prediction of length of work based on both Age and Job Title

A picture containing text, font, screenshot, line

Description automatically generated

*Figure 5.7: Predicted life span of job title with age codes*

This analysis combines both factor age and job title to create the prediction of employees working life span.



*Figure 5.8: Predicted life span of job title with age graph*

This analysis suggests that the remaining employees in managerial positions with the factor age included are expected to work for a little over 15 years. While jobs that requires skills and experiences such as dairy, bakery and meat cutting are expected to work for 7-10 years. While basic skills requirement roles such as cashier are expected to only work for 1-4 years.

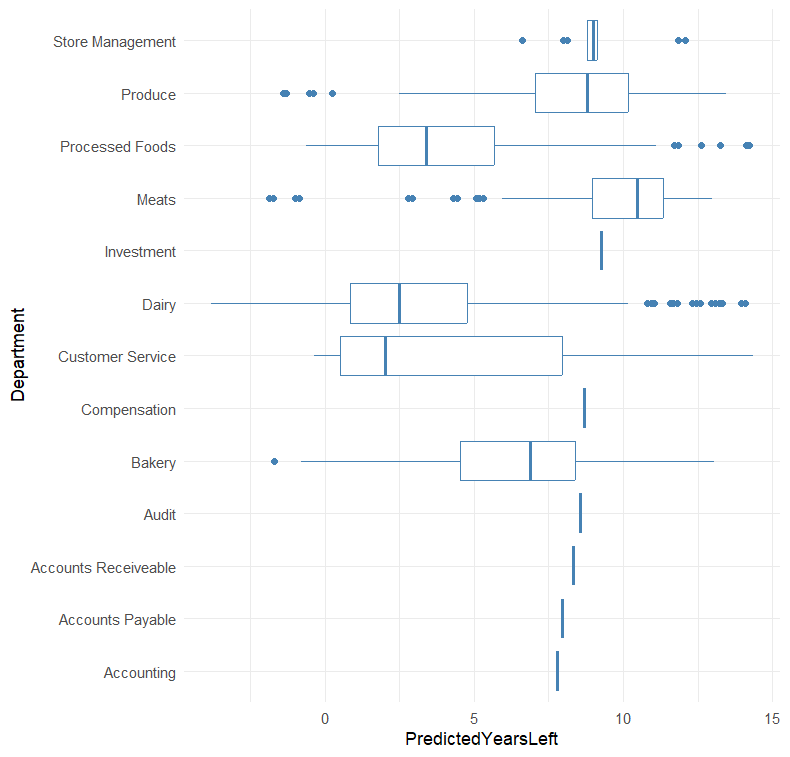
### 3.5.5 Analysis 5.5 What is the difference between predicted length of service vs actual length in produce department.

A picture containing text, font, screenshot

Description automatically generated

*Figure 5.9: Predicted life span of department with age codes*

This graph shows the general life span of departments across the company. Which could assist decision making of employer of identifying which department requires more attention and aid.



*Figure 5.10: Predicted life span of department with age graph*

The prediction suggests that Dairy, Customer service and processed food have life span of below 5 years. While Accounting and bakery departments have lifespan of about 7 years, and meat cutting, and produce have about 10 years.

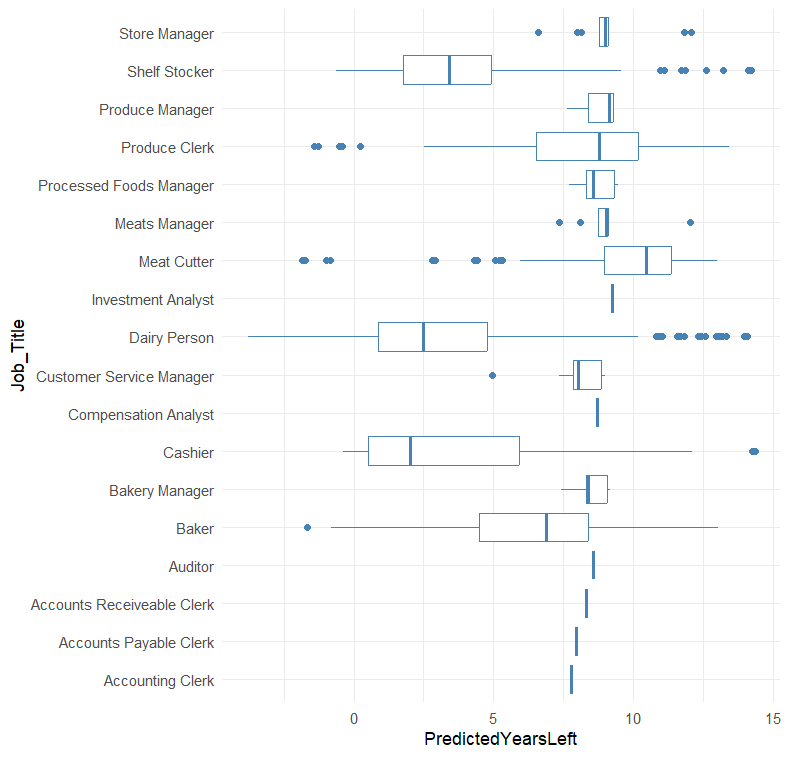
### 3.5.6 Analysis 5.6 What is the predicted remining life span of each Job Title

A picture containing text, font, screenshot

Description automatically generated

*Figure 5.11: Predicted remaining life span of job title with age codes*

This graph shows the general life span of specific job titles across the company. Which could assist employers in identifying which job titles are in need.



*Figure 5.12: Predicted remaining life span of job title with age graph*

The prediction suggests that job titles relating to Dairy, Customer service and processed food have life span of below 5 years. While Accounting and bakery roles have lifespan of about 7 years, and most clerks in the company have life span of about 10 years.

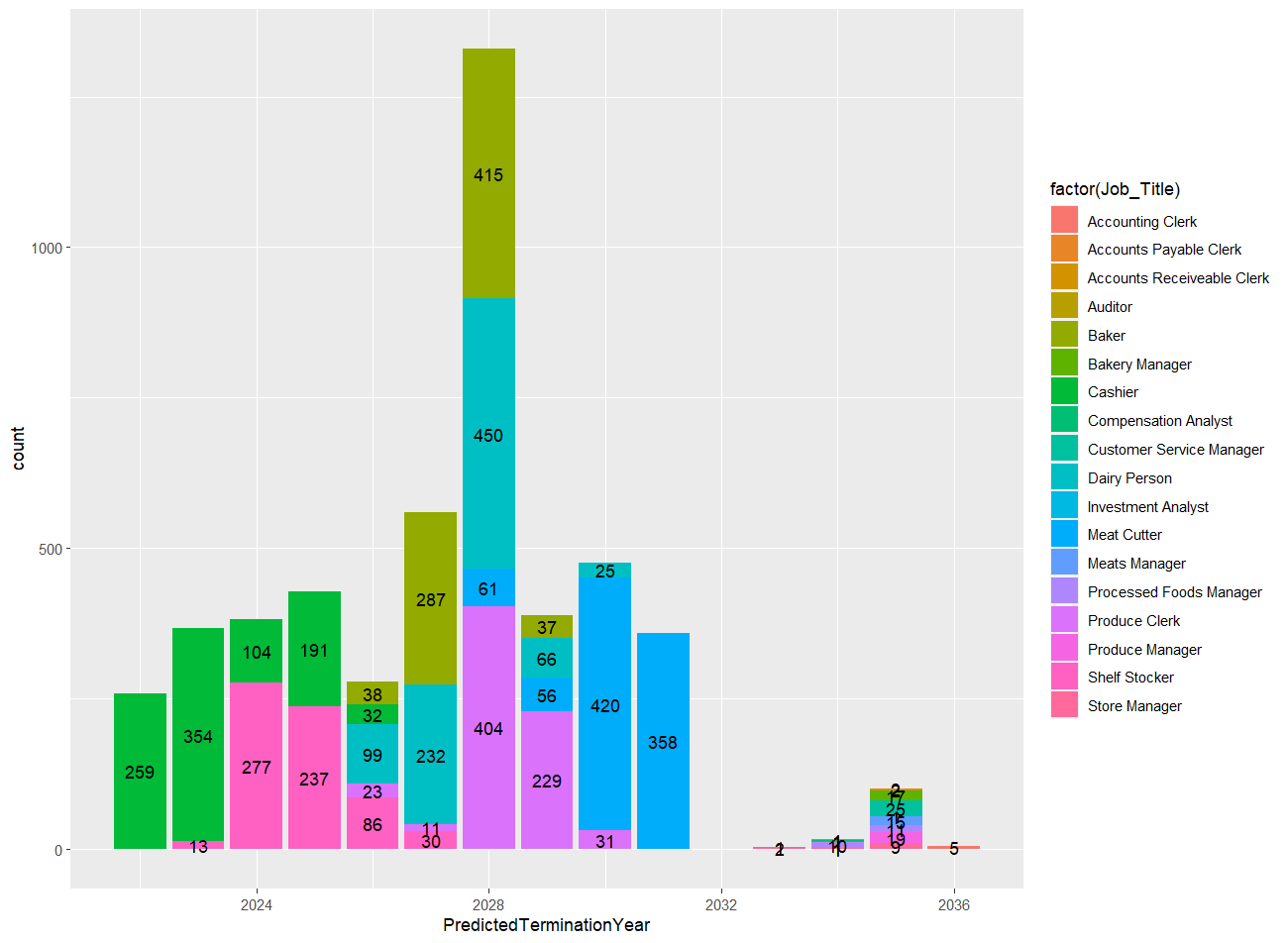
### 3.5.7 Analysis 5.7 How many employees are predicted to lose for the following years in each Job Title?

A picture containing text, font, line, screenshot

Description automatically generated

*Figure 5.13: Predicted employee termination for each year codes*

This analysis focuses on counting the expected termination in each year for each job title. This could assist employer in preparation for the lost of employees in the years to come.

A 

*Figure 5.14: Predicted employee termination for each year graph*

Most cashiers and self-stockers are expected to be terminated in the next 5 years. These job titles are easily replaceable and are of less concern. A spike in employee termination is expected in the year 2028. Employers should calculate the time of training of employees and number of employees to employ for each job title to counter this wave.

### 3.5.8 Conclusion

In conclusion, the trends and predictions of terminated data were used to predict the remaining life span of employees in the company. A spike in employee termination in the year 2028 was forecasted. The count of employees in each job title needed to be re-recruited for each year was calculated in hopes of assisting employers in better decision making and resources allocation.

# 3.0 extra features

## 3.1 Regression summary

A picture containing text, screenshot, font

Description automatically generated

*Figure 6.1: Predicted employee termination for each year codes*

A black text on a white background

Description automatically generated with low confidence

*Figure 6.2: Predicted employee termination for each year codes*

Regressions calculation and summaries allows us to find the percentage of how much a dependent variable is dependent on another independent variable. In this case, we used it to find the degree to which years worked was related age and job title.

## 3.2 Prediction

A picture containing text, font, screenshot

Description automatically generated

*Figure 6.3: Predicted employee termination for each year graph*

Predict function uses the old data of RegressionOfProduceAge dataframe to calculate predictions for dfRemainProduce which is the data of active employees.

A picture containing text, screenshot, diagram, colorfulness

Description automatically generated

*Figure 6.4: Predicted employee termination for each year graph*

Predictions allow us to foresee the future based on pass patterns and dataset. In this case, prediction allows the company to plan for future termination waves. We used prediction function to calculate the number of employees expected to leave for each year.

## 3.3 theme\_minimal()

A picture containing text, font, screenshot

Description automatically generated

*Figure 6.5: Predicted life span of job title with age graph*

A picture containing text, screenshot, number, font

Description automatically generated

*Figure 6.6: Predicted life span of job title with age graph*

Theme\_minimal() turns graph into clean and clear designs for better reading and analysis.

## 3.4 Position stack

A picture containing text, font, line, screenshot

Description automatically generated

*Figure 6.7: Predicted employee termination for each year graph*

A picture containing text, screenshot, diagram, colorfulness

Description automatically generated

*Figure 6.8: Predicted employee termination for each year graph*

Position stacks tell the labels that the bar graph of the same categories is stacked on top of each other. This allows clear labels as shown in this figure.

## 3.5 Scale\_y\_discrete()

A picture containing text, font, screenshot

Description automatically generated

*Figure 6.9: store in danger graph*

A picture containing text, screenshot, colorfulness, parallel

Description automatically generated

*Figure 6.10: store in danger graph*

Scale\_y\_discrete allows us to specify the y axis accordingly. This is useful when precise readings must be shown.

## 3.6 Violine graph

A picture containing text, font, screenshot

Description automatically generated

*Figure 6.11: Age of laid off in 2015,2014 graph.*

A picture containing text, screenshot, diagram, number

Description automatically generated

*Figure 6.12 Age of laid off in 2015,2014 graphs.*

Violine graphs allow us to visualize the frequency of the data as continuous variable changes. In this case, violine graph displayed the distribution of age in department across a wide range of ages.

## 3.7 Pie chart

## A picture containing text, font, screenshot Description automatically generated

*Figure 6.13: Weights of prediction pie chart*

A picture containing text, screenshot, diagram, font

Description automatically generated

*Figure 6.14: Weights of prediction pie chart*

Pie chart helps sectors variables, proportions, and percentage into clean visualizable diagrams for easier presentation and communication. In this case, we could visualize the proportion of job and age effecting predictions of employee life span.

## 3.8 Vjust and Hjust

A screenshot of a computer

Description automatically generated with low confidence

*Figure 6.15: Attrition department graph*

Vjust and Hjust indicate the location where labels are shown. 0 representing the center, -1 to the left or below and 1 top and right.

A picture containing text, screenshot, diagram, parallel

Description automatically generated

*Figure 6.16: Attrition department graph*

In this graph, vjust and hjust displayed the termination count of employee at the left side and center of the bar graph as shown in Figure 6.16.

# Bibliography

Gartenstein, D. (2019, February 5). *What is a Healthy Employee Turnover Rate?* Retrieved from Chron: https://smallbusiness.chron.com/healthy-employee-turnover-rate-12145.html

Nora Dunn, S. T. (2022, December 12). *What Is a Recession in Canada?* Retrieved from nerdwallet: https://www.nerdwallet.com/ca/personal-finance/what-is-a-recession#:~:text=The%202008%2D2009%20%E2%80%9CGreat%20Recession,in%20the%20U.S.%20to%20collapse.

Solutions, R. (2023, May 1). *What Is the Average Retirement Age?* Retrieved from Ramsey: https://www.ramseysolutions.com/retirement/average-retirement-age#:~:text=Okay%E2%80%94let's%20get%20the%20age,keep%20working%20until%20age%2066.&text=Also%2C%20many%20retirees%20go%20back,again%20in%20a%20few%20years.

Team, I. E. (2023, March 11). *How Long Should You Stay at a Job? 6 Questions To Ask*. Retrieved from indeed: https://www.indeed.com/career-advice/career-development/how-long-should-you-stay-at-a-job#:~:text=The%20typical%20employee%20stays%20at,company%20than%20their%20younger%20counterparts.