Abstract.
This study investigates the factors influencing employee turnover by analyzing a comprehensive institutional dataset. By examining variables such as age, tenure, department, and termination reasons, the study identifies patterns and correlations associated with employee departures. Key findings reveal distinct trends in turnover across various employee demographics and organizational units. The research underscores the importance of understanding these dynamics for the development of effective human resource strategies to mitigate turnover. I wish to convey my sincere thanks to the university for supplying the dataset and to express my gratitude to the academic community for their invaluable support.

Acknowledgement.

I wish to extend my heartfelt thanks to the organization for generously supplying its employee dataset. This crucial resource played a pivotal role in the successful completion of my research.

The unwavering support and guidance of the academic faculty have been invaluable throughout this project. Their expertise has significantly enriched my research.

I would also like to extend my heartfelt thanks to Isak Thapa, Sankalpa Ojha, Ankit Acharya, and Sajag Shrestha for their invaluable contributions to this research. Their assistance with documentation and coding was essential to its success.

Introduction.

Human resources are a critical function in organizational performance. This study analyzes a comprehensive dataset of 49,645 employees to identify potential challenges within the organization's human resources management. Under the guidance of the Human Resources Department manager, the research aims to uncover underlying issues, derive actionable insights, and inform strategic decision-making. By examining employee attributes, the study seeks to pinpoint minor but impactful problems that may be affecting overall organizational efficiency. This analysis will inform the development of targeted solutions and optimized HR strategies.

Data Import.

The dataset, formatted as a comma-separated values (CSV) file, will be imported as the initial step in our analysis. The CSV format is a widely accepted standard for data storage and exchange. To facilitate this process, we will leverage the read_csv() function from the readr package. This function is specifically designed for efficient CSV importation and will serve as the foundation for subsequent data exploration and modeling endeavors.

```r
# Get working directory
getwd()

#Set working directory
setwd("C:\Users\sangi\Downloads")

#install necessary packages
install.packages("ggplot2")
install.packages("dplyr")
install.packages("crayon")
install.packages("readr")
install.packages("na.tools")
library(na.tools)
library(crayon)
library(ggplot2)
library(dplyr)

#Read employee_attrition.csv and set as dataset variable
library(readr)
employee_details <- read.csv("employee_attrition.csv")

#view dataset
view(employee_details)
```

Fig: Establishing the project environment, acquiring necessary tools, and importing the data.

To commence the project, a dedicated working directory was established to optimize file organization and accessibility. Essential software packages were then procured and installed to equip the analytical environment with the necessary tools. The subsequent step involved importing the dataset into the R environment and assigning it to the variable "employee_details" for convenient reference. A meticulous visual inspection of the imported data was undertaken to ensure data integrity, accuracy, and completeness. This preliminary examination provided a foundational understanding of the dataset's structure and content, paving the way for subsequent exploratory data analysis and modeling efforts.

Data Cleaning.

Purging and refining data is a crucial preliminary phase in the data analysis process, indispensable for guaranteeing the integrity and dependability of the data. This meticulous process involves systematically identifying and rectifying various data anomalies that can compromise the integrity of the dataset. Missing values, outliers, duplicates, and inconsistencies are common culprits that can skew analytical results if left unaddressed. A clean and well-prepared dataset is instrumental in deriving meaningful insights and avoiding erroneous conclusions, ultimately contributing to the overall success of the analytical endeavor.

```
#column removed
##gender_short
employee_details$gender_short <- NULL
```

Fig: Eliminating superfluous columns.

To streamline the dataset for subsequent analytical endeavors, the "gender_short" column is removed from the "employee_details" dataset. This procedural action is carried out by allocating a null value to the specified column through the utilization of the assignment operator <-. By effectively eliminating this superfluous column, the dataset becomes more concise and efficient, thereby facilitating subsequent data manipulation, exploration, and modeling processes. The removal of redundant information, such as the "gender_short" column, is a standard practice in data preprocessing to improve data quality and facilitate efficient analysis.

Data Pre-Processing.

Data preprocessing represents a pivotal and frequently labor-intensive phase within the data analysis workflow. It involves a comprehensive array of precise methodologies aimed at converting raw, unprocessed data into a well-organized, pristine, and insightful format ready for analytical purposes. This vital procedure entails a comprehensive strategy to tackle prevalent data quality challenges, including missing entries, anomalies, inconsistencies, and inaccuracies, which can profoundly affect the precision and dependability of ensuing analyses. By adhering to best practices and leveraging appropriate tools and techniques, analysts can ensure that their data is adequately prepared for the challenges of modern data analysis.

```
#DataPre-Processing
#view dataset
View(employee_details)

#structure
str(employee_details)

#number of rows and columns
dim(employee_details)

#column Names
names(employee_details)

#summary of the data
summary(employee_details)
```

Fig: Preprocessing of data.

To gain initial insights into the dataset, several R functions are employed. The View() function provides a visual representation of the data, allowing for interactive exploration. To obtain a structured summary of the dataset, encompassing data types and a preview, the str() function is employed. The dimensions of the dataset, specifically the number of rows and columns, can be swiftly ascertained using the dim() function.

To identify the specific variables within the dataset, the names() function is helpful. Lastly, the summary() function generates descriptive statistics for each variable, offering valuable information about data distribution and central tendencies.

Data Transformation.

Data transformation is a pivotal phase in the data analysis process, involving the conversion of raw data into an appropriate format for analysis. This process includes altering data structures, patterns, and values to improve its usefulness. Common transformation techniques include normalization, aggregation, discretization, and feature engineering.

```
# Transformation: Convert "STATUS_YEAR" to a factor variable
data$STATUS_YEAR <- as.factor(data$STATUS_YEAR)

# Transformation: Create a new variable "service_length_group" based on length_of_service
data$service_length_group <- cut(data$length_of_service, breaks = c(0, 5, 10, 15, 20, Inf),
                                 labels = c("0-5", "6-10", "11-15", "16-20", "20+"))

# Transformation: Recode "STATUS" to a binary variable (Terminated or Not Terminated)
data$STATUS <- ifelse(data$STATUS == "TERMINATED", "Terminated", "Not Terminated")
```

Fig: Transforming data.

The code implements several data preprocessing techniques. Initially, the "STATUS_YEAR" variable undergoes transformation into a factor, enabling categorical data analysis. Subsequently, a new variable, "service_length_group", is created through categorization of the "length_of_service" variable into specified intervals. This step facilitates analysis by grouping employees based on tenure. Finally, the "STATUS" variable is recoded into a binary format, distinguishing between terminated and active employees, thereby simplifying subsequent analysis focused on employee attrition. These data manipulations establish a suitable foundation for in-depth data exploration, modeling, and visualization.

Data Analysis.

Analysis 1.1. Examining the Link Between Job Role and Employee Turnover

This study will perform an in-depth analysis of the dataset to investigate the complex relationship between job roles and employee turnover. By thoroughly examining the data, we aim to identify patterns and trends that reveal the connection between these variables. This investigation will go beyond simply finding a correlation; it will strive to uncover the underlying factors that drive these trends.

The findings from this analysis will be crucial for human resource managers in crafting targeted strategies to tackle the specific challenges associated with different job roles. By identifying the positions most prone to turnover, HR professionals can implement proactive measures to enhance employee retention and satisfaction. Additionally, this research will provide a deeper understanding of the organizational factors that influence turnover, enabling data-driven decisions to optimize workforce management and improve overall organizational performance.

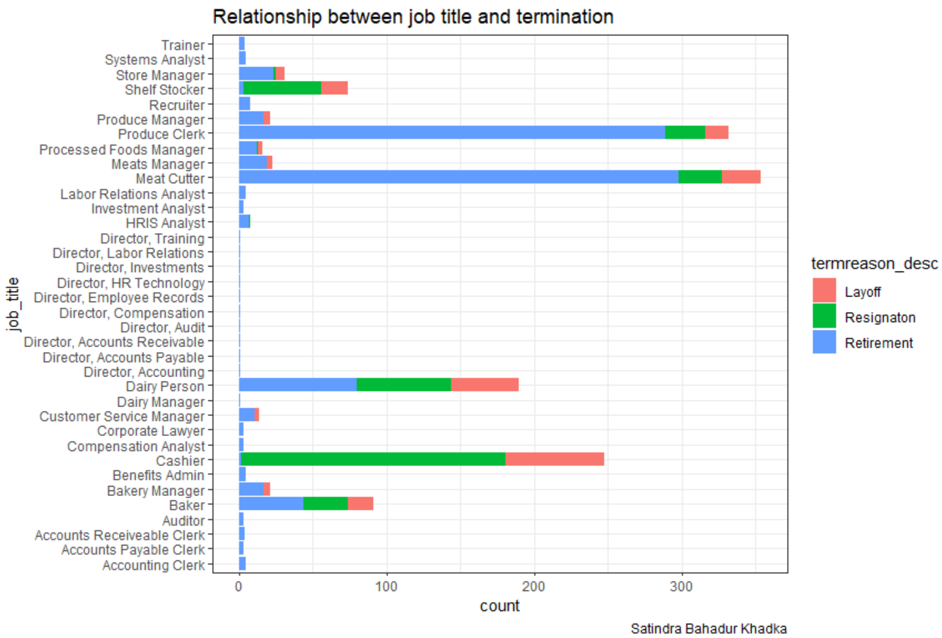
Code:

```
#Analysis 1.1 Determine the relationship between a work position and termination
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot(fill = "#FFC0CB")+
  aes(x = job_title, fill = termreason_desc)+ geom_bar()+coord_flip()+
  labs(title = "Relationship between job title and termination",
       caption ="Satindra Bahadur Khadka")+
  theme_bw()
```

A visual examination was performed to investigate the correlation between job titles and termination causes within the dataset using R. The dplyr and ggplot2 packages were utilized for data manipulation and visualization, respectively. To ensure data accuracy, entries with a termination reason of “Not Applicable” were excluded from the analysis.

A horizontal bar chart was generated, displaying job titles along the x-axis and using color-coded bars to differentiate termination reasons. Titles and captions were added to enhance clarity and interpretation. This visualization provides valuable insights into termination patterns across various job roles, helping to identify trends and areas for targeted human resource interventions.

Output:



Key Findings:

The analysis reveals that retirement was the leading cause of employee turnover across multiple departments, with higher retirement rates observed in produce and meat cutting roles. Resignation was identified as the most common reason for termination, significantly affecting employees in cashier, dairy, and shelf stocking positions. Although layoffs accounted for a smaller portion of total terminations, they disproportionately affected employees in cashier and dairy roles.

Analysis 1.2. Exploring Gender Disparities in Terminations.

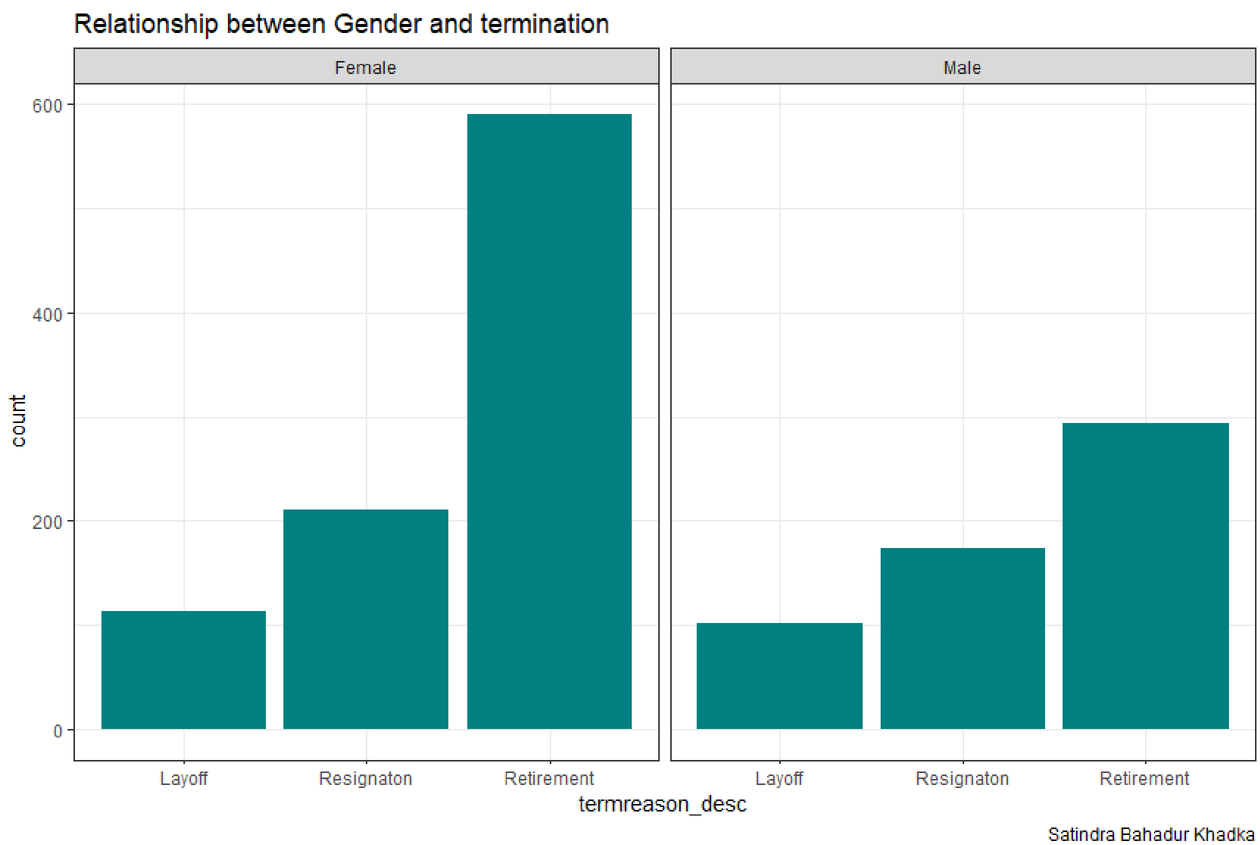
A thorough analysis will be undertaken to investigate potential gender differences in termination rates. Using statistical methods and visual tools, the study will identify patterns or inconsistencies in termination outcomes based on gender. The goal is to provide actionable insights for human resource management by determining if gender-based disparities exist in termination decisions. Addressing any identified inequities will help the organization promote a more equitable and inclusive workplace.

Code:

```
#Analysis 1.2 Investigate the relationship between gender and termination.
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = termreason_desc)+ geom_bar(fill = "#008080")+
  labs(title = "Relationship between Gender and termination",
       caption ="Satindra Bahadur Khadka")+
  theme_bw()+ facet_wrap(vars(gender_full))
```

A bar chart was generated to visualize the frequency of different termination reasons across gender categories. This visual representation effectively compares the distribution of termination types between male and female employees. The chart is designed with a specific color scheme and includes titles and captions for better clarity. To facilitate comparison, the chart is split into separate panels for each gender, enabling a direct visual comparison. This visualization helps identify potential gender-related trends in termination outcomes.

Output:



Key Findings:

The bar graph reveals clear differences in termination rates between male and female employees. Women had significantly higher retirement rates, with about twice as many cases as men. Voluntary resignations were also more common among women. Although layoffs affected both genders, women faced a disproportionately higher rate. These results suggest a potential gender disparity in termination outcomes within the organization.

Analysis 1.3. Analyzing Business Unit Impact on Employee Turnover.

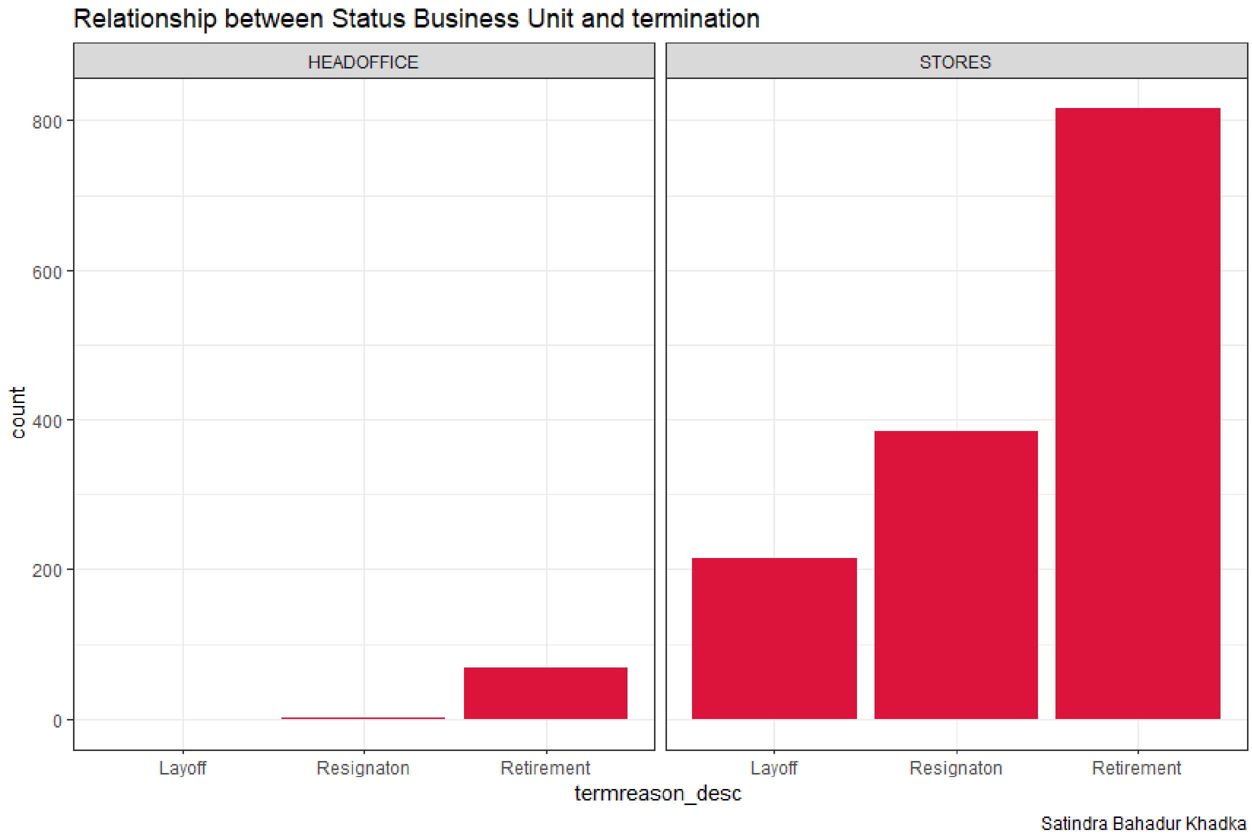
A thorough analysis will be performed to investigate the link between business units and employee turnover. This study seeks to identify trends and patterns in termination rates across diverse organizational units through the application of data manipulation and visualization techniques. Key factors, including business unit, termination reason, and other relevant variables, will be examined to uncover insights that can guide strategic human resource management decisions.

Code:

```
#Analysis 1.3 Establish a connection between the business unit and the termination.
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = termreason_desc)+ geom_bar(fill = "#DC143C")+
  labs(title = "Relationship between Status Business Unit and termination",
       caption ="Satindra Bahadur Khadka")+
  theme_bw()+ facet_wrap(vars(BUSINESS_UNIT))
```

A bar chart was meticulously constructed to visually depict the frequency distribution of termination reasons across distinct business units within the organization. The chart uses a crimson color scheme (#DC143C) for visual consistency. This visualization helps identify potential patterns or anomalies in termination rates among the various business units.

Output:



Key Findings:

Employee turnover is significantly higher in store locations compared to the head office. The data reveals that retirement is the primary reason for termination in both areas, followed by resignation as the second most common cause. Interestingly, layoffs were only observed in store operations.

Analysis. 1.4. Analyze the connection between age and employee turnover.

A thorough analysis will be undertaken to explore the relationship between age and employee termination. Utilizing statistical methods such as logistic regression, the study will examine potential correlations between age and the likelihood of termination. The results will offer valuable insights into age-related termination trends, aiding in the creation of strategic initiatives to enhance workforce management, succession planning, and retention strategies.

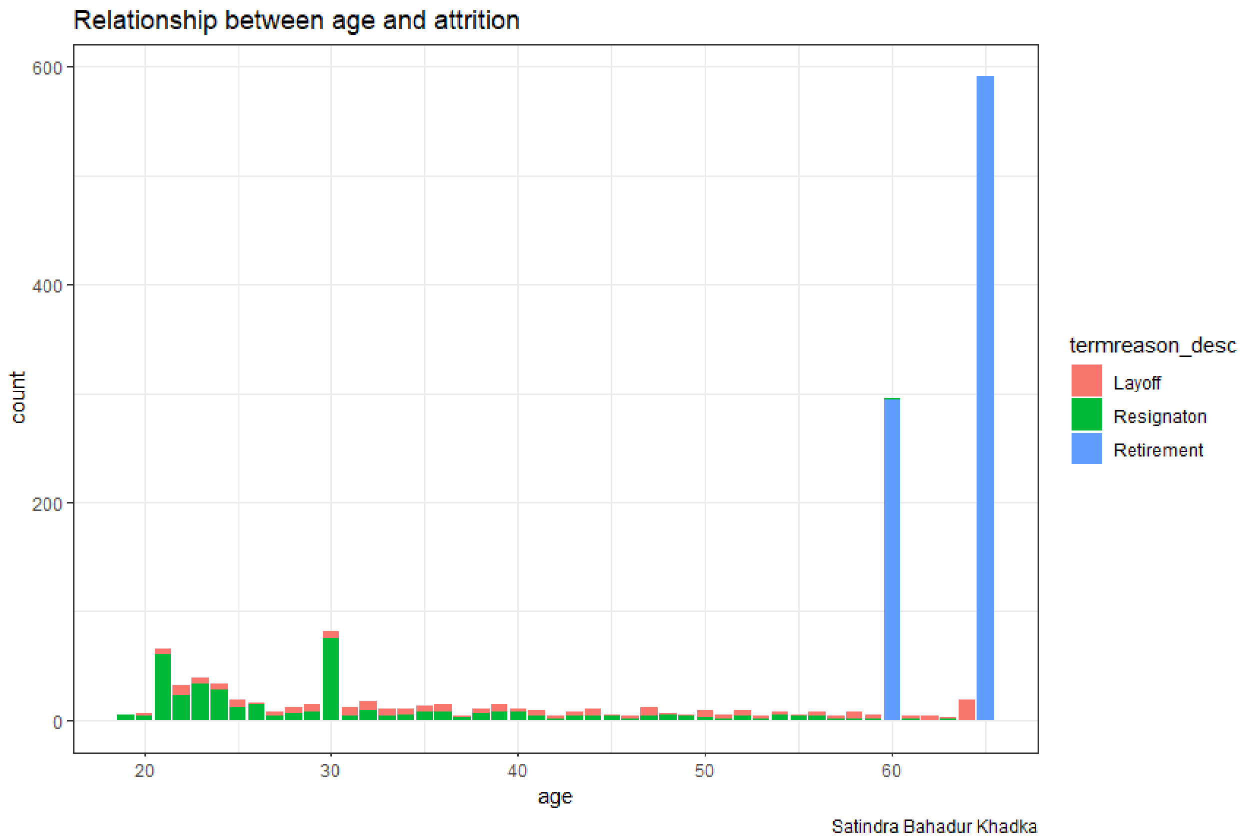
Code:

```
#Analysis 1.4 Determine the relationship between age and termination.

data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = age, fill = termreason_desc)+ geom_bar()+
  labs(title = "Relationship between age and attrition", caption ="Satindra Bahadur Khadka")+
  theme_bw()
```

Observations were excluded from the analysis where the termination reason was designated as "Not Applicable." A bar chart was subsequently generated to illustrate the age distribution of employees, categorized by termination reason. The chart, titled “Relationship between Age and Attrition,” was presented in a monochromatic style for better visual clarity.

Output:



The analysis reveals a significant link between age and termination reasons. Retirement is the main reason for employees aged 60 and above, whereas resignation rates are considerably higher among those in the 20-30 age group. This trend likely mirrors the career stages and life cycle considerations of different age groups.

Analysis: 1.5. Analyze the impact of tenure on employee termination across different age groups.

A comprehensive analysis will be conducted to examine the interplay between employee tenure, age, and termination. Through the application of statistical and visual methodologies, this study seeks to uncover patterns and trends in workforce dynamics. The findings will inform the development of strategic initiatives to optimize employee retention, facilitate career progression, and cultivate an age-inclusive organizational culture.

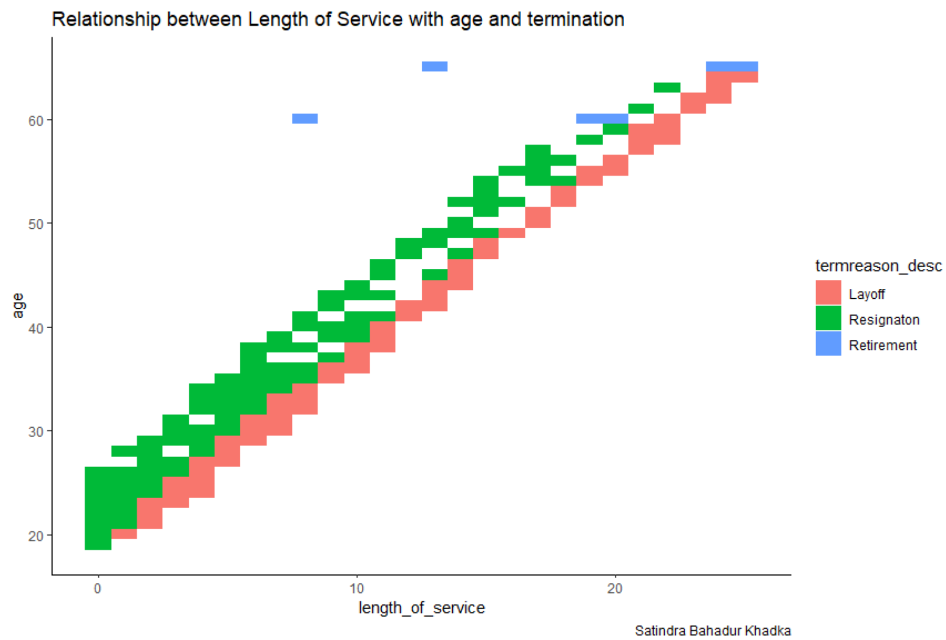
Code:

```
#Analysis 1.5 Examine the relationship between duration of service and termination with respect to age.

data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = length_of_service, y= age , fill = termreason_desc)+ geom_tile()+
  labs(title = "Relationship between Length of Service with age and termination",
       caption ="Satindra Bahadur Khadka")+
  theme_classic()
```

To investigate the interplay between tenure, age, and termination reasons, a heatmap was constructed using the ggplot2 package in R. Data points where the termination reason was categorized as "not applicable" were excluded from the analysis. The heatmap, employing the geom_tile() function, visually represents the frequency of terminations across varying age and tenure cohorts through color intensity. Enhanced with a title, subtitle, and a classic theme, the visualization effectively communicates patterns and trends in employee turnover across different demographic segments.

Output:



Key Findings:

The analysis shows that employees with shorter tenures, especially those in the preliminary stages of their careers (20-30 years old), have higher turnover rates. This tendency decreases with longer organizational tenure and advancing age. Layoffs occur consistently across all tenure and age groups, while retirement is mainly seen among employees with significant tenure and in older age brackets.

Analysis: 1.6. Examine the correlation between departmental affiliation and employee turnover.

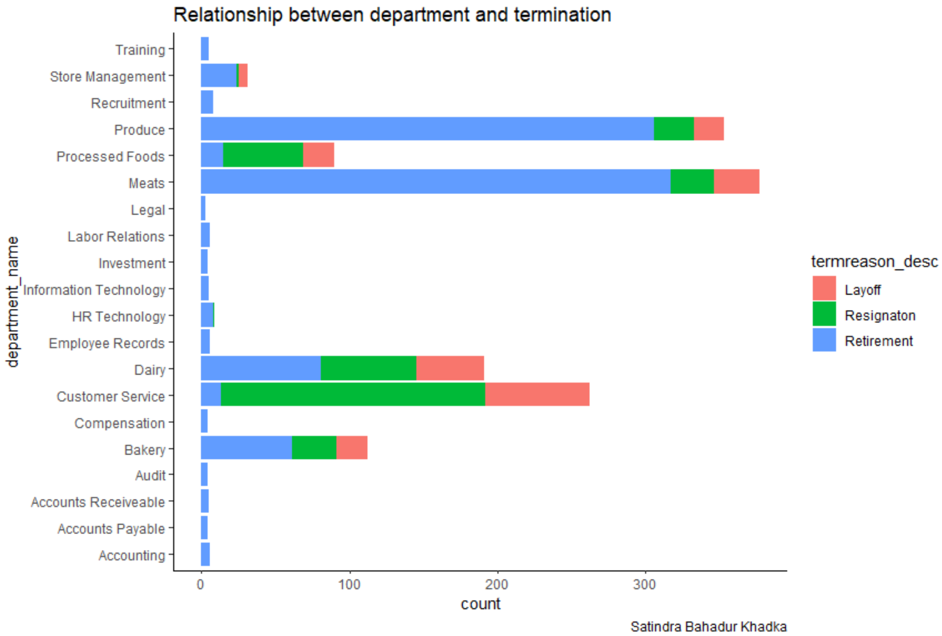
This analysis primarily examines the link between departmental affiliation and employee termination. The study's findings will guide the creation of targeted HR strategies to address potential disparities in termination rates across different organizational units.

Code:

```
#Analysis 1.6 Identify the relationship  between department and termination
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = department_name, fill = termreason_desc)+ geom_bar()+
  labs(title = "Relationship between department and termination", caption ="Satindra Bahadur Khadka")+
  coord_flip()+
  theme_classic()
```

The provided R script utilizes the dplyr and ggplot2 packages to analyze the correlation between departmental affiliation and termination status within the dataset. Observations where the termination reason is labeled as "Not Applicable" have been omitted from the analysis. A horizontal bar chart was created to visually display the frequency of terminations across different departments. The chart's design, including its title, caption, and thematic elements, enhances clarity and interpretability. This visual representation offers insights into departmental turnover patterns.

Output:



Key Findings:

The analysis reveals that employees in lower-level, operational roles have higher termination rates. In contrast, those in higher-level positions show significantly lower turnover. Retirement is the main reason for termination among managerial and executive roles.

Analysis: 2.1. Analyze the temporal distribution of employee status changes.

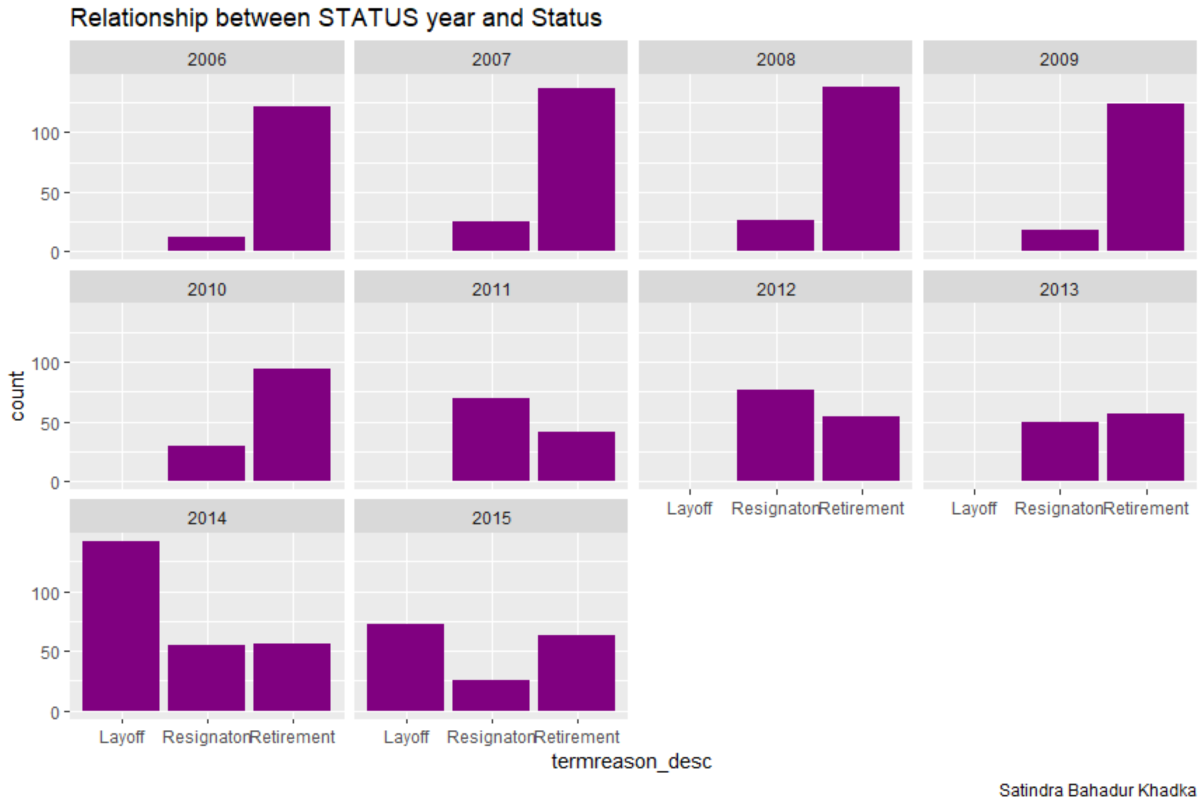
The primary objective of this research is to uncover recurring patterns in the temporal sequence of organizational changes. By meticulously examining historical data and employing advanced analytical techniques, the study seeks to identify consistent trends and cycles in the timing of key organizational events. These insights will serve as a foundational basis for developing predictive models and forecasting future organizational shifts.

Code:

```
#Analysis 2.1 Determine the relationship between status and year of status.
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = termreason_desc, fill = STATUS )+ geom_bar(fill = "#800080")+
  theme_gray()+
  labs(title = "Relationship between STATUS year and Status",
       caption ="Satindra Bahadur Khadka")+
  facet_wrap(vars(STATUS_YEAR))
```

Data points with the termination reason marked as “Not Applicable” were excluded from the analysis. Using ggplot2, a bar chart was created to visualize the distribution of termination reasons by status. Faceting by status year enabled a comparative analysis of termination trends over time.

Output:



Between 2006 and 2008, employee turnover remained relatively stable, mainly due to resignations and retirements. However, in 2010, retirement rates began to decline, while resignations increased. This trend continued for two years, until 2013, when resignations and retirements reached a similar level. In 2014, there was a significant increase in layoffs, which was a departure from the previous pattern. The following year, 2015, saw a decline in both layoffs and resignations. Throughout this period, retirement rates remained consistent.

Analysis: 2.2. Analyze the relationship between employment status and termination rates..

This study seeks to elucidate the intricate relationship between employee status and termination rates. By delving into the nuances of this association, the research aspires to uncover patterns and trends that can guide the development of effective employee retention strategies and workforce management practices. This study will shed light on the factors driving employee turnover, allowing organizations to anticipate and address potential issues, and make better decisions about how to allocate their human resources. Ultimately, the results of this research will provide valuable insights for decision-makers seeking to improve their organization's performance and long-term viability by optimizing their workforce planning and management.

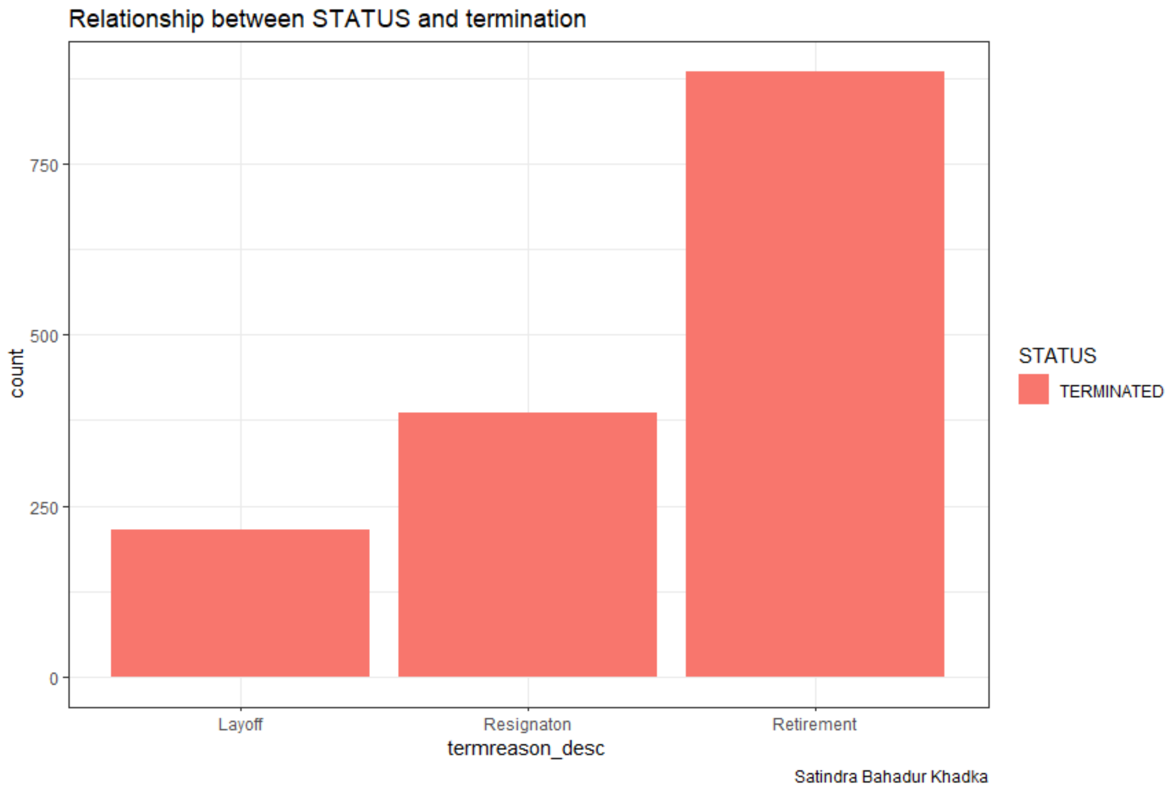
Code:

```
#Analysis 2.2 Determine the relationship between status and termination.

data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = termreason_desc, fill = STATUS )+ geom_bar()+
  labs(title = "Relationship between STATUS and termination",
       caption ="Satindra Bahadur Khadka")+
  theme_bw()
```

A bar chart was created to show the connection between employee status and reasons for termination. The chart has two main axes: the x-axis lists the reasons for termination, and the y-axis shows how often each reason occurs. Employee status is highlighted using different colors. The chart, titled 'Relationship Between Employee Status and Termination Reasons', includes a brief description and a single-color scheme to make it easy to read. This visual representation successfully conveys the spread of termination reasons across various employee groups.

Output:



Key Findings:

Retirement was identified as the leading cause of employee termination, exceeding both resignations and layoffs in frequency. This indicates that a significant portion of the workforce reached the end of their employment lifecycle and opted to retire. Resignations were the second most common reason for termination, suggesting that many employees voluntarily left the organization to pursue other opportunities or for personal reasons. Layoffs, while present, were less frequent compared to retirements and resignations, indicating that workforce reductions through termination were relatively limited during the analyzed period.

Analysis 3.1. Analyze the relationship between employee status and geographic location.

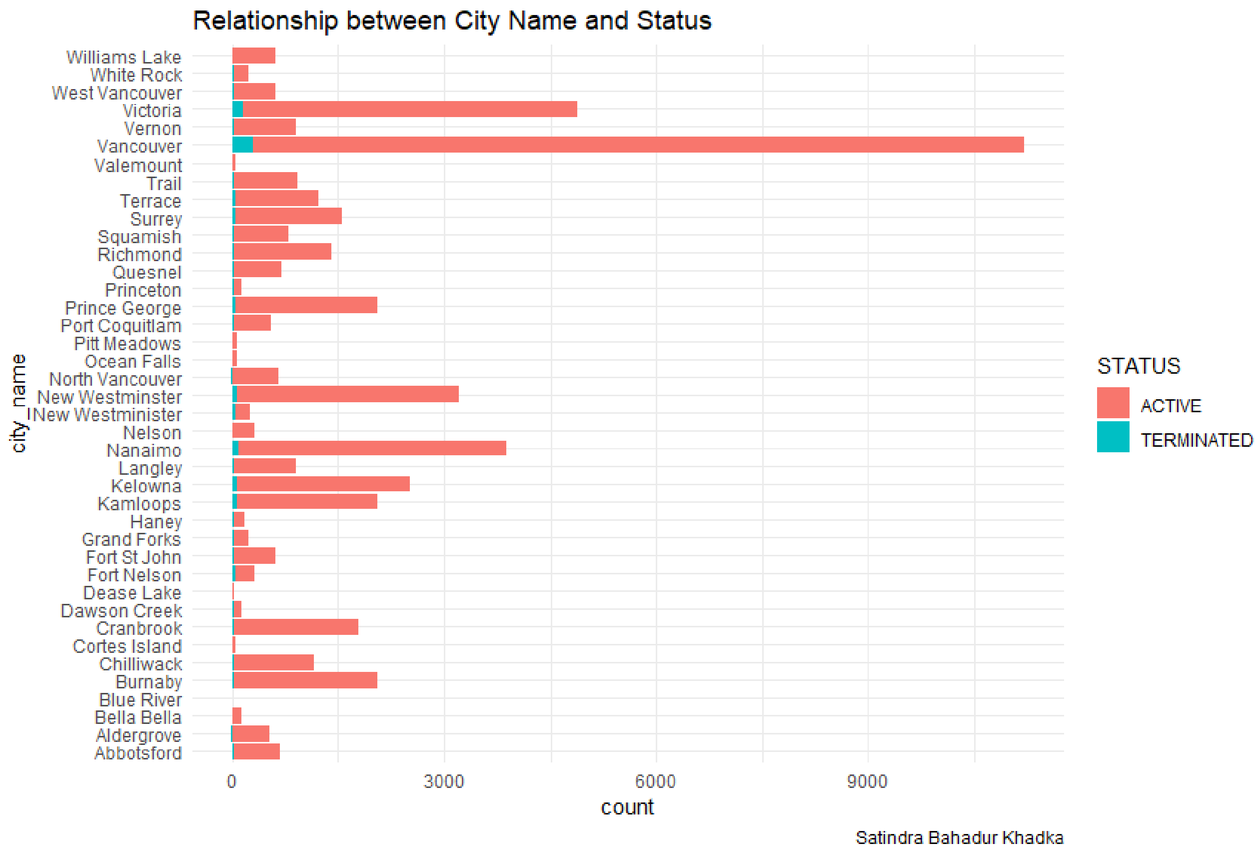
This analysis seeks to identify and quantify regional disparities in employee status. By understanding these variations, strategic decisions can be made regarding resource allocation, service delivery, and targeted interventions tailored to specific geographic contexts.

Code:

```
#Analysis 3.1 Examine the link between a status and city name.
data %>%
  ggplot()+
  aes(x = city_name, fill = STATUS )+ geom_bar()+
  labs(title = "Relationship between City Name and Status",
       caption ="Satindra Bahadur Khadka")+
  coord_flip()+
  theme_minimal()
```

A horizontal bar chart was created to visualize the distribution of employee statuses across various geographic locations. The chart features city names along the y-axis and uses color-coded bars to represent the frequency of each status within each city.

Output:



Key Findings:

Metropolitan areas like Vancouver and Victoria typically have a higher density of active employees than smaller cities. These urban centers also experience higher termination rates. As a result, the combination of a larger workforce and increased turnover leads to a higher number of active employees in these regions.

Analysis 3.2. Analyzing The Impact of Geography on Organizational Structure.

This research aims to explore the spatial distribution of organizational divisions by analyzing the relationship between cities and divisions. The study seeks to identify potential patterns, disparities, or concentrations of specific divisions across various geographic regions. This analysis will offer valuable insights into the organization's geographic footprint, operational structure, and potential areas for optimization or expansion.

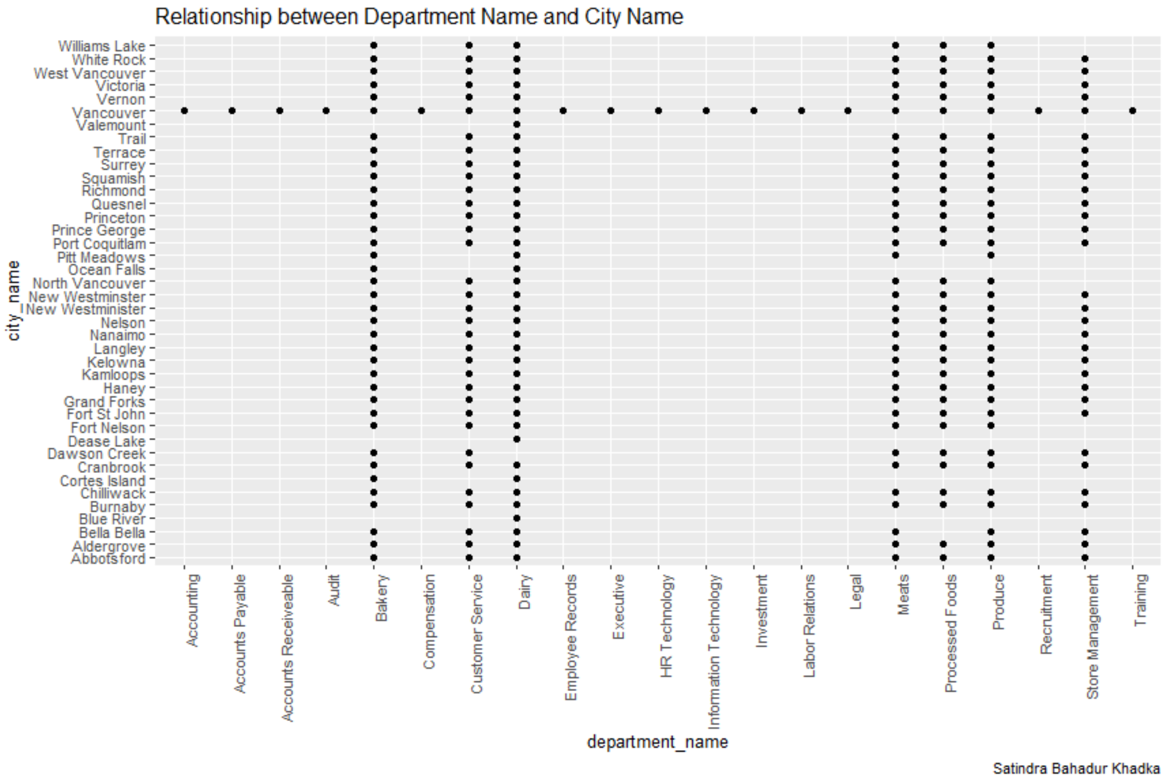
Code:

```
#Analysis 3.2 Unearth the relationship between a city's name and its division organization.

data %>%
  ggplot()+
  aes(x = department_name, y = city_name )+ geom_point()+
  labs(title = "Relationship between Department Name and City Name",
       caption ="Satindra Bahadur Khadka")+
  theme(text=element_text(size=10),axis.text.x = element_text(angle = 90,hjust = 1))
```

To better understand the connection between a company's organizational structure and its geographical presence, a statistical analysis was conducted using R programming. A scatter plot was created to visualize the relationship between the company's departments and the cities where they are located. Each data point on the plot illustrates a unique pairing of department and city, thereby offering a visual depiction of the organization's structural distribution across various geographical locations. This graph can help identify patterns, concentrations of specific departments, or geographical disparities in the organization's distribution. By analyzing the graph, experts can gain valuable insights into the company's global footprint and how its different departments are distributed across cities.

Output:



Key Findings:

The visualization highlights a significant disparity in departmental distribution across the analyzed cities. Vancouver stands out as a comprehensive hub, encompassing all departmental functions. In contrast, many cities show a notable absence of specific departments, particularly accounting, human resources, and information technology. This pattern indicates potential imbalances in organizational structure and resource allocation across regions. Such disparities could impact operational efficiency, service delivery, and overall organizational performance. Further investigation into the factors driving these departmental variations is necessary to inform strategic decision-making and resource optimization.

Analysis 3.3. Analyse The Relationship Between Store Name and City.

This study investigates the spatial distribution of retail outlets to uncover the underlying patterns and relationships between store locations and urban environments. By analyzing the connection between store names and city locations, we aim to identify geographic trends in retail placement and explore how different store types interact with urban characteristics. A thorough examination of these factors will provide a detailed understanding of market penetration, consumer behavior, and the competitive landscape within the retail industry.

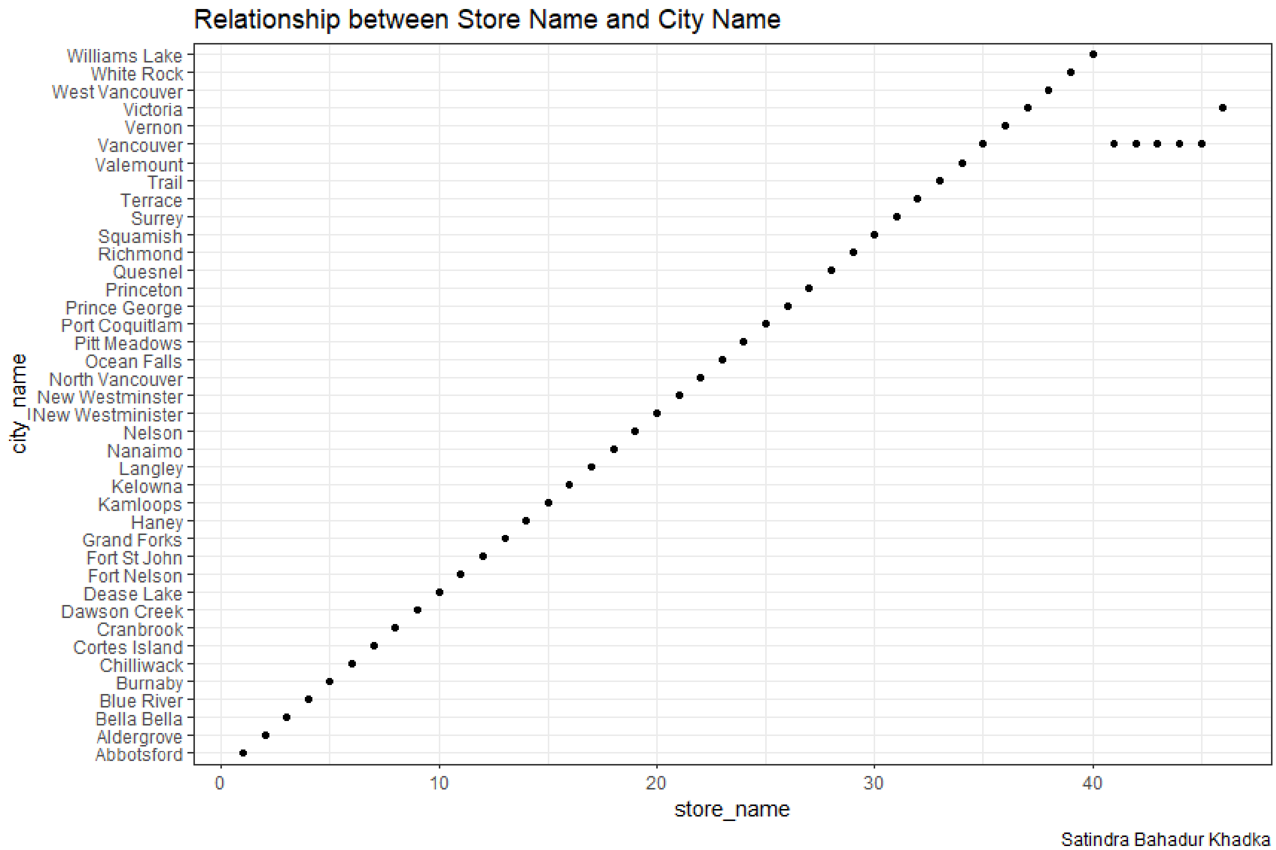
Code:

```
#Analysis 3.3 Determine the link between the store's name and the city name.

data %>%
  ggplot()+
  aes(x = store_name, y = city_name )+ geom_point(fill = "#FF0000")+
  theme_gray()+
  labs(title = "Relationship between Store Name and City Name",
       caption ="Satindra Bahadur Khadka")+
  theme_bw()
```

A geographic analysis was performed to explore the connection between store names and their respective cities. Using the ggplot2 package, a scatter plot was created to display the locations of retail stores across different cities. This visual representation helped to identify patterns such as clustering, dispersion, and outliers in the distribution of retail stores. These findings provide a basis for understanding market coverage, competition, and potential areas for expansion or contraction in the retail industry.

Output:


Relationship between Store Name and City Name
Williams Lake
White Rock
West Vancouver
Victoria
Vernon
Vancouver
Valemount
Trail
Terrace
Surrey
Squamish
Richmond
Quesnel
Princeton
Prince George
Port Coquitlam
Pitt Meadows
Ocean Falls
North Vancouver
New Westminster
New Westminster
Nelson
Nanaimo
Langley
Kelowna
Kamloops
Haney
Grand Forks
Fort St John
Fort Nelson
Dease Lake
Dawson Creek
Cranbrook
Cortes Island
Chilliwack
Burnaby
Blue River
Bella Bella
Aldergrove
Abbotsford
city_name
0
10
20
30
40
store_name
Satindra Bahadur Khadka

Analysis 4.1. Analyze the relationship between employee age and termination reasons.

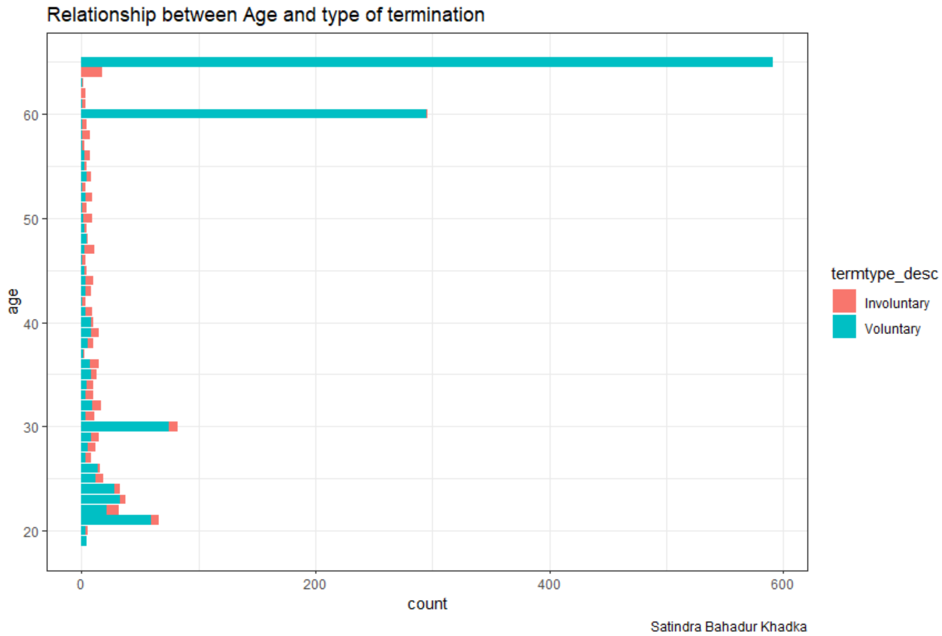
This research endeavors to elucidate the intricate relationship between employee age and termination rates. By analyzing termination patterns across various age groups, the study aims to identify correlations, trends, and disparities that may exist. The results will provide valuable insights into the factors driving employee turnover among different age cohorts, enabling organizations to develop targeted strategies and interventions to improve employee retention.

Code:

```
#Analysis 4.1 Discover the correlation between age and the type of termination.
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(y = age,fill = termtype_desc )+ geom_bar()+
  scale_fill_hue(direction = 1)+
  labs(title = "Relationship between Age and type of termination",
       caption ="Satindra Bahadur Khadka")+ theme_bw()
```

The R script performs a thorough analysis to identify the connection between employee age and termination reasons. To visually represent this relationship, a bar chart is constructed using the ggplot2 package. In this chart, age groups are plotted on the y-axis, while termination types are displayed on the x-axis. To enhance visual clarity, different colors are employed to differentiate between the various termination categories. The color palette is carefully selected using the scale_fill_hue function to ensure that each termination type is represented by a distinct and easily recognizable color, making it easier to interpret the results.

Output:



Key Findings:

Employees aged 60 and older frequently experience voluntary termination, mainly due to retirement. In contrast, those in the 20-30 age range tend to leave voluntarily, often motivated by career growth and exploration. Meanwhile, individuals aged 30 to 60 face a higher incidence of involuntary termination, such as layoffs, indicating a possible organizational strategy to open positions for younger employees.

Analysis 4.2. Examine how the duration of employment influences the type of termination.

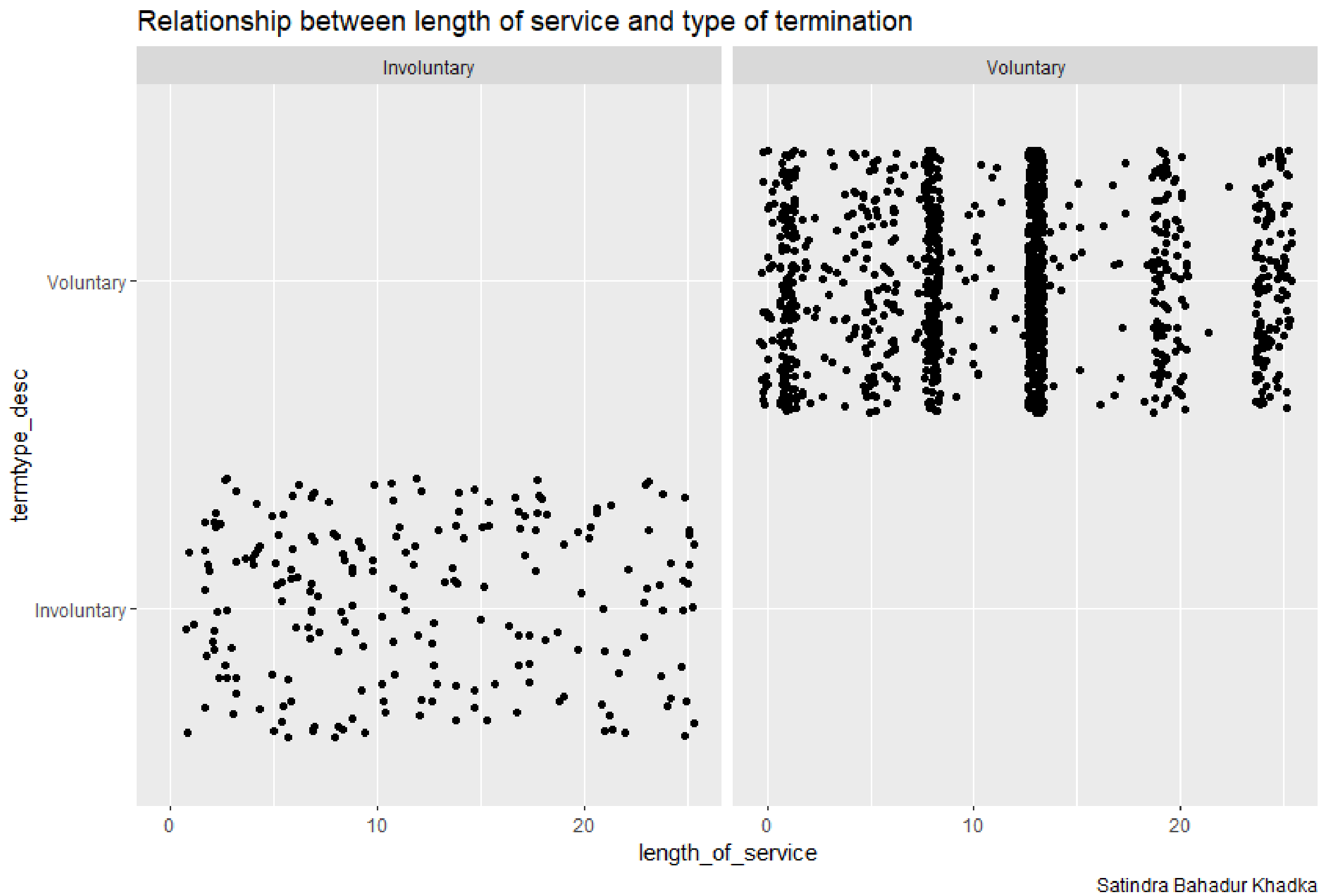
This research investigates how an employee's service within a company correlates with how their employment is ended. By analyzing various withdrawal plans across different time frames of employment, the study aims to uncover potential correlations or recurring themes. The findings could provide valuable insights into how organizations handle departures based on an employee's longevity, potentially influencing HR policies and targeted interventions for staff at different career stages within the company.

Code:

```
#Analysis 4.2 Understand the relationship between the length of service and the method of termination.
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(x = length_of_service, y = termtype_desc )+ geom_jitter(fill = "#00FFFF")+
  labs(title = "Relationship between length of service and type of termination",
       caption ="Satindra Bahadur Khadka")+ theme_bw()+
  facet_wrap(vars(termtype_desc))
```

This script utilizes R to examine how an employee's tenure relates to their departure method from a company. The resulting visualization plots each worker's service duration against their exit type, with slight randomization added to prevent overlap. The code employs a paneling technique to create distinct sections for each termination category, enabling easy comparison of service length patterns across different exit scenarios.

Output:



Key Findings:

The data suggests that forced departures occur at a similar rate regardless of how long an employee has been with the company. Newer hires tend to leave of their own accord more frequently. Interestingly, the peak in voluntary exits is observed among those who have worked for the organization between a decade and a decade and a half.

Analysis 4.3. Investigate the relationship between department and termination type.

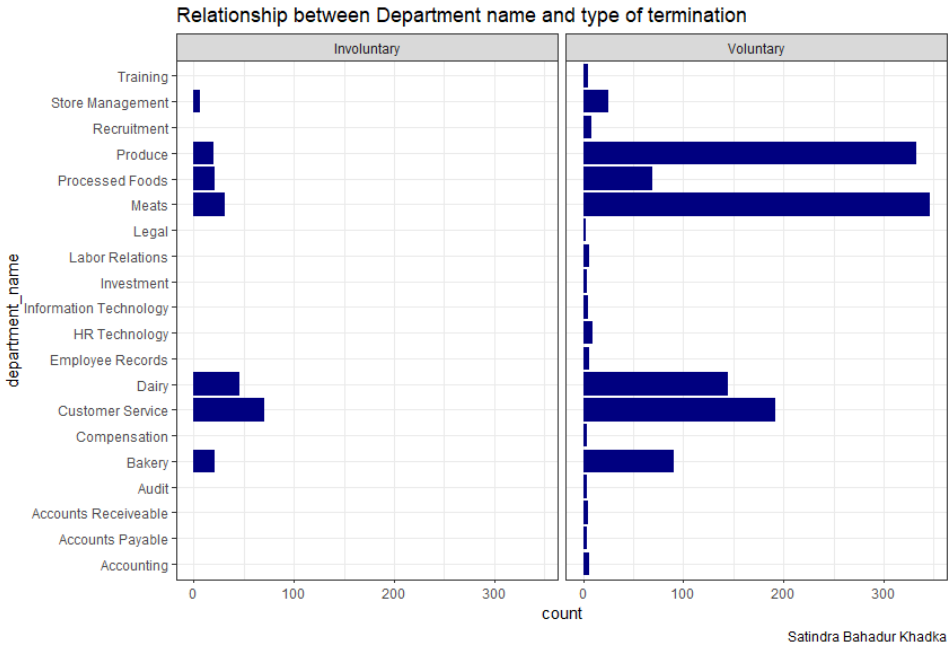
This study explores how an employee's workplace division might relate to the way their job ends. By analyzing termination patterns across various company sectors, the research aims to uncover any notable connections or recurring themes in how different departments handle staff departures.

Code:

```
#Analysis 4.3 Discover the correlation between department and the type of termination.
data %>%
  filter(!(termreason_desc %in% "Not Applicable")) %>%
  ggplot()+
  aes(y = department_name )+ geom_bar(fill = "#000080")+
  labs(title = "Relationship between Department name and type of termination",
       caption ="Satindra Bahadur Khadka")+ theme_bw()+
  facet_wrap(vars(termtype_desc))
```

This script employs R to examine how employee departures vary across different company divisions. It produces a horizontal bar graph using the ggplot2 library, with company sectors listed vertically. Each bar illustrates the number of job endings in a particular department, colored in a deep blue shade to enhance readability.

Output:



Key Findings:

The data shows that employees in sectors like fresh produce, packaged goods, meats, dairy products, customer support, and baked goods tend to leave their jobs voluntarily more often than others. These same areas also see a substantial number of staff retiring. Additionally, these departments experience forced departures, typically due to workforce reductions.

Conclusion.

A thorough examination of employee termination data has uncovered a complex web of interconnected factors. Our analysis reveals that age, tenure, department, and termination type are all intertwined, with distinct patterns emerging across different age groups. For instance, we have identified specific termination reasons associated with certain age ranges and observed correlations between termination type and age. Furthermore, our findings suggest that employee tenure and department also play a significant role in shaping termination patterns.

Understanding the reasons behind employee turnover is crucial for effective HR management. By analyzing employee termination data, organizations can identify trends and patterns, allowing them to develop targeted strategies to improve job satisfaction and retention. By recognizing specific employee groups that are more likely to leave, organizations can create customized programs to address their needs, such as mentorship or career development initiatives. This data-driven approach enables HR departments to anticipate and address employee needs, ultimately leading to a more stable workforce.

Ultimately, our findings suggest that a data-driven approach to employee retention can be a powerful tool for driving business success. By recognizing the complex patterns and trends underlying employee turnover, HR teams can develop targeted strategies that drive retention, improve engagement, and enhance overall business performance.

References.

R Data Visualization - javatpoint. (n.d.). www.javatpoint.com. https://www.javatpoint.com/r-data-visualization

Kabacoff, R. (n.d.). *Modern Data Visualization with R*. https://rkabacoff.github.io/datavis/

N. (2023, December 4). *Data Visualization in R*. Intellipaat. https://intellipaat.com/blog/tutorial/r-programming/data-visualization-in-r/

R - CSV Files. (n.d.). https://www.tutorialspoint.com/r/r_csv_files.htm

N. (2024, January 8). *How to Read CSV File into DataFrame in R*. Spark by {Examples}. https://sparkbyexamples.com/r-programming/r-read-csv-file-with-examples/

Quick-R: Scatterplots. (n.d.). https://www.statmethods.net/graphs/scatterplot.html

R Graphics - Scatter Plot. (n.d.). https://www.w3schools.com/r/r_graph_scatterplot.asp

R. (2024, January 2). *Scatter plot in R*. RCODER. https://r-coder.com/scatter-plot-r/

Scatter plots in R Language. (2023, May 10). GeeksforGeeks. https://www.geeksforgeeks.org/scatter-plots-in-r-language/

R Graphics - Plotting. (n.d.). https://www.w3schools.com/r/r_graph_plot.asp

Pedamkar, P. (2023, March 22). *Graphs in R*. EDUCBA. https://www.educba.com/graphs-in-r/#:~:text=The%20most%20commonly%20used%20graphs,plots%20for%20exploratory%20data%20analysis.

Producing Simple Graphs with R. (n.d.). https://sites.harding.edu/fmccown/r/

R Functions. (n.d.). https://www.w3schools.com/r/r_functions.asp

Functions in R Programming. (2023, November 25). GeeksforGeeks. https://www.geeksforgeeks.org/functions-in-r-programming/

Kosourova, E. (2023, March 6). *How to Write Functions in R (with 18 Code Examples)*. Dataquest. https://www.dataquest.io/blog/write-functions-in-r/#:~:text=A%20function%20in%20R%20is%20an%20object%20containing%20multiple%20interrelated,user%20(user%2Ddefined).