

I would say "improve". You have no proof/claim of optimality.

---

? **!!!!** Utilizing a Data-Driven Approach to Optimize  
Transportation Warehouse Recruitment

---

**Echo Luan**

zixial@andrew.cmu.edu

**Malik Khan**

@andrew.cmu.edu

**Yanxi Zhou**

yanxiz@andrew.cmu.edu

**Xiaofan Zhu**

@andrew.cmu.edu

**Lanyi Xu**

@andrew.cmu.edu

Please see rubric for grading this paper ("Final project/materials" under Assignments in Canvas) to make sure you haven't missed anything.

## Executive Summary

good draft summary, despite all the red ink!

This sounds weird and I'm not sure what you mean.

you can delete this without loss

The foodservice distribution industry is essential to the U.S. economy and serves to connect food producers and food operators such as restaurants and foodservice outlets. Although the foodservice distributors sustain the need to consume food outside the home, there have been growing recruiting challenges in developing a sufficient pipeline of qualified candidates to keep up with the workforce demands in warehouse and truck driving job families. This problem is further complicated by the COVID-19 pandemic as jobs in the foodservice distribution industry have been affected by government economic measures and identified as high-risk in the social perception. Therefore, this project aims to design a data-driven recruitment improvement plan for members of the International Foodservice Distributors Association (IFDA) by providing summary findings on available data as well as a roadmap that offers detailed recommendations on future approach.

delete

cite SHRM study, etc., here, using ASA format

cite BLS publication(s) here, using ASA format.

Delete, so that parag starts with stuff you did \*for IFDA\*

To achieve this goal, we started with reading IFDA studies on the foodservice distribution industry, recruitment research by the Society for Human Resource Management (SHRM), and publications about truck drivers and warehouse workers from the Bureau of Labor Statistics (BLS). After gaining a better understanding of the recruitment challenge from the industry perspective, we conducted interviews with five IFDA members and collected their qualitative responses. We then analyzed recruitment-related data provided by the participating companies and aggregated and merged anonymized records through a normalization process in order to conduct data analysis on the information provided in this project. Because the data collection across participating IFDA members is uneven, we manually mapped different data sources into uniformed values and merged all private mappable data received.

To gather data on the recruitment problem, we first

not meaningful here; can delete

A better phrase: "a common set of"

Our exploratory data analysis of the merged data surveyed employee dispositions, employee tenure by job role, seasonality in hiring, and turnover vs. hiring rate over time. We found that most applicants for warehouse and transportation jobs left their applications incomplete, followed by withdrawals during screenings or interviews as the second most common disposition. Our results also show that the truck driving role has a higher proportion of employees who stayed in the company for more than six months compared to warehouse workers. The detailed analysis results of private data are included in separate appendices and will be delivered to participating companies individually.

confidential company

we propose a "roadmap" for future data collection and analysis, to better understand IFDA member companies' recruitment challenges.

Because of the varying formats of the data received and the absence of some important information such as applicant sourcing data, we provided suggestions on how to further our understanding of the findings above and put statistical rigor behind validating anecdotal hypotheses in the proposed roadmap. Our roadmap identifies questions of interest, shows the importance of these questions, lists out variables that we recommend tracking, and explains methods of analysis to answer the proposed questions. We arranged the roadmap following the order of the recruitment pipeline that will be further discussed in the Introduction section. An overview of the questions that the roadmap identifies is provided below:

- Which sourcing channel provides candidates that are most likely to get hired?
- What proportion of visitors access the company website careers page using mobile, desktop, and tablet, respectively?
- What are the sources of the visitors on the website careers page?

- What is the conversion rate of each web sourcing channel?
- What types of applicants tend to apply early / late?
- Do shorter interview cycles increase the likelihood of hires, among those who were offered a position?
- What is the average time to hire for each job role?
- What is the predicted time to fill for job openings in different seasons, positions, and locations?
- What are the common traits of hired applicants?
- What types of employees are more likely to stay at the company for more than 6 months?
- Which sourcing channel provides the candidates that are most likely to stay for more than a year once hired?
- What is the number of qualified applicants per hire?

By following this recommended roadmap, IFDA members can extract more data-driven insights based on the anecdotal findings about recruitment challenges and localize the problem through detailed examination of the recruitment pipeline. After foodservice distributors track the recommended variables over a period of time, they can benchmark Key Performance Indicators in recruitment analytics and detect unusual patterns in the recruitment process. **and address**

**and/or unfavorable trends**

very useful!

## Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Data</b>	<b>5</b>
<b>3</b>	<b>Method</b>	<b>8</b>
<b>4</b>	<b>Results</b>	<b>8</b>
4.1	Interview Summary . . . . .	8
4.2	EDA . . . . .	9
4.3	Roadmap . . . . .	15
4.3.1	Awareness & Interest . . . . .	15
4.3.2	Application & Interviewing: . . . . .	15
4.3.3	Hiring: . . . . .	17
<b>5</b>	<b>Discussion</b>	<b>18</b>
<b>6</b>	<b>Reference</b>	<b>20</b>
<b>7</b>	<b>Appendix</b>	<b>21</b>
7.1	Glossary . . . . .	21
7.2	Interview Questions Used . . . . .	22
7.3	Detailed Roadmap . . . . .	22
7.3.1	Awareness Interest . . . . .	26
7.3.2	Application Interviewing: . . . . .	28
7.3.3	Hiring: . . . . .	29
<b>8</b>	<b>Individual Appendix</b>	<b>40</b>

delete colons

Use ASA style for the citations highlighted yellow here, and throughout paper.

# 1 Introduction

The foodservice distribution industry accounts for over \$250 billion in annual sales that affect millions of restaurants and foodservice outlets everyday (IFDA, A Comprehensive Economic Impact Study on Foodservice Distributors). However, the transportation and warehouse labor market is facing a scarce pool of talent that creates recruitment challenges for food distributors across the United States; the International Foodservice Distributors Association (IFDA), in collaboration with the American Trucking Association, estimates that the industry will need to hire about 1.1 million drivers over the next 10 years to keep pace with turnover and retirements. In particular, there exists a gap in filling both driving and warehouse roles with qualified candidates, as the industry faces competition from business-to-consumer and retail grocery companies (Kickham, Demand for Drivers on the Rise in Foodservice Industry). How can we improve the existing talent acquisition process through a data-driven approach? We propose a design for a better and hiring data collection recruitment process that helps food distributor services identify high-quality candidates by recording data from the following stages of the recruitment pipeline: Awareness Interest, Application Interviewing, and Hiring for a given job requisition (SmartDreamers Recruitment Marketing Automation).

start new paragraph here

To raise awareness about their brand, companies market through advertisements, direct sourcing, networking events, and other efforts that help passive and active applicants identify the company. The interest from a candidate is created once the candidate identifies the food distributor company as an organization they can work in. After, it is imperative that food distributors are able to select candidates who they deem most qualified for the job role, and interview them accordingly. At the hiring stage, candidates have the offer in hand and it is time to observe whether the investment into the selected candidate pays off. By identifying consistent and inconsistent data being tracked from food distributor services, we aim to address the following question:

This is not working for me, and I think it could all be deleted without loss.

1. How can we develop a data-driven framework to further our understanding of the candidate pool and improve recruitment efforts?

# 2 Data I like this section!

Data used for this report consists of three parts: public data obtained from Bureau of Labor Statistics (BLS)<sup>1 2</sup>, confidential company private data obtained from five IFDA member companies, and qualitative data collected during our interviews with executives and human resources managers from the five companies. Client data will be used to conduct statistical analysis while public data was used as a reference to the market. For the yearly public data, we mainly focus on three sectors: Wholesale trade (Non-durable goods), Food Beverage stores, and Transportation Warehousing. For these sectors, we are interested in the following statistics from 2017 to 2019: Employment, Hourly Wage, Annual Wage, Inter-quartile Ranges for Hourly Annual Wage. Client data are categorized to four sections: Employee Data, Job Data, Applicant Data and Requisition Data. Details of the interview process are discussed in the Methods section.

Because there are no results with BLS data in the paper, I suggest omitting this.

I don't see BLS data used anywhere in the results section or in the roadmap.

To explore a holistic picture of the recruitment problem environment faced by the IFDA members, we merged data from three client companies of the five provided to us, after data normalization. We mainly focused on mapping job titles and

I suggest a colon (":") here to signal that the definition of data normalization is coming next.

<sup>1</sup><https://data.bls.gov/oes//home>

<sup>2</sup>[https://www.bls.gov/news.release/jolts.t02.htmjolts\\_table2.f.1](https://www.bls.gov/news.release/jolts.t02.htmjolts_table2.f.1)

Add a footnote saying that two of the companies did not provide data amenable to statistical analysis.

applicant disposition (recruitment status) from different data sources in the normalization process. For job titles, we examined all job titles from the three IFDA members and categorized them into five groups: CDL Delivery Drivers, Non-CDL Drivers, Warehouse Selectors, Other Warehouse Workers, and Other Workers. Because each company names job titles and divides job responsibilities differently, we manually identified the job titles of each company and grouped them based on the responsibilities suggested. For example, the "Warehouse Order Selector" role from one client and the "1st Shift Selector" role from another client are categorized under "Warehouse Selectors" in our merged dataset. The group "Other Warehouse Workers" contains non-selector warehouse positions, such as replenishers and warehouse leads, while the group "Other Workers" includes all other job titles that are not identified by the former three groups, including supervisor and managerial positions.

Add a footnote defining CDL

The same procedure <sup>was</sup> also applied to the normalization of applicant disposition. We used five <sup>normalization</sup> categories to map the variables: Withdrew - Screening/Interviews, Unsatisfied - Experience/History/Skills, No Show - Pre-screen/Post-hiring/Interviews, Incomplete - Applications/Portal, and Hired. The first group (Withdrew - Screening/Interviews) <sup>includes</sup> applicants where the candidate voluntarily withdrew from the recruitment process in either screening or interviews. The second group (Unsatisfied - Experience/History/Skills) categorizes applicants where the company is not satisfied with the candidates due to their lack of experience, work history, or skills needed for the position. The third group (No Show - Pre-screen/Post-hiring/Interviews) aggregates applicants for which the candidate did not show up for pre-screens, interviews, or post-hiring. The fourth group (Incomplete - Applications/Portal) contains all other recruitment processes that are not completed in the application portal or in the application process. Finally, the last group includes all the hired candidates.

Again, a colon here will help reader to see you will define "normalization" in the next lines.

After mapping the shared variables of the <sup>three</sup> IFDA members, we <sup>created</sup> two separate data files that contain the normalized employee ~~data~~ and applicant data across the three companies. The detailed descriptions of variables contained in each dataset can be found in Table 1 and Table 2.

Table 1: Description of Merged Applicant Data

Variable Name	Description
Position	Original position title that the applicant applied to
Merged Position	Merged job title group that the position was mapped to
Date Applied	Application date
Merged Disposition	Merged disposition (recruitment status) of the applicant

Based on the data that we received, most organizations collected warehouse and transportation employee data <sup>as well as</sup> ~~and~~ <sup>KPI</sup> key performance indicators <sup>(KPIs)</sup> for the company. For example, turnover rate is tracked by most client companies. This KPI summarizes the company's retention status and facilitates long-term planning for recruitment and employee development. However, we also observed absences of important variables and inconsistencies of granularity in the data provided by different organizations. Some variables

Table 2: Description of Merged Employee Data

Variable Name	Description
Merged Job Title	Merged job title group
Department Name	Name of the department the job title belongs to
Date of Hire	Date when the employee was hired
Term Date	The date when the employee was terminated
Hourly Rate	Hourly rate in dollars
Race Code	Race of the employee
Gender	Gender of the employee
Full/Part Time	Full time or part time
Shift	Night shift or day shift
Rehire	Whether the employee was rehired
Tenure to Today	Tenure in number of days
Tenure Range	The range that tenure falls in: "<45 days", "45-89 days", "90-119 days", and "120+ days"
Termed?	Whether the employee was terminated
Tenure to Termed	Tenure range before termination

not tracked or not accessible <sup>for</sup> by all clients include applicant sourcing data, applicant work history, requisition close date. There is also data <sup>and</sup> not provided by most clients: applicant-level data and requisition-level data. There are mainly two reasons that these variables were not provided to the team. One reason is that the data was tracked by the company, but they are difficult to access or tracked inconsistently across time and regions. This ~~challenge of data extraction~~ <sup>within and across companies</sup> makes it difficult to transform data into a usable format for recruitment analytics. On the other hand, some variables were not collected by the company because of technical challenges or difficulties to analyze them effectively.

If tracked, these important variables can be used to calculate ~~Key Performance Indicators (KPIs)~~ <sup>from which</sup> where we can extrapolate more insights ~~from~~. For example, with requisition open date and fill date, we can calculate the average time to fill for each position throughout the year. By tracking time to fill, the company can plan for recruitment more accurately and understand the amount of time to obtain a qualified candidate for different seasons, positions, and locations. The company can also establish a benchmark for this metric and track any noticeable changes to detect problems early on. <sup>Another example would be the sourcing</sup>

new paragraph

channel. If we have access to the sourcing channel data and applicant-level data, we can calculate the source of hire, which is defined as the percentage of hires entering the recruitment pipeline from each sourcing channel. This KPI shows the effectiveness of each recruiting source and helps the human resources management evaluate different recruitment strategies. We will provide further recommendations for data collection and analysis in the Roadmap section. Our recommendations require the IT department to assist in establishing data markers through the application process and post-hiring in order to thoroughly evaluate any patterns or trends that arise over time. It is crucial that these metrics are tracked consistently over multiple time periods so that a baseline comparison of these metrics can be established, and so that the company can identify trends that are inconsistent with the company's standard.

new parag

### 3 Method

This brief methods section seems just about right.

To understand the business problems, we read multiple resources provided by our client and generated interview questions regarding the recruitment issues of warehouse and driving labor forces that IFDA members are facing. The interview questions can be found in Appendix page 21. We had five interview sessions with participating IFDA members, respectively, to gain the qualitative and quantitative information from them.

, as well as discussions with IFDA representative Annika .... , we

Based on

To analyze datasets from five of our clients, we compared the public data from the Bureau of Labor Statistics with data from five client companies, variables compared including turnover rate, hire rate, etc. We also merged client data based on variable mapping. After that, we conducted statistical analysis on individual company data and the merged dataset, plotted various graphs to visualize the data and gathered insight. After gaining a better understanding of the data through visualizations, we built decision trees and multiple regression models to predict employee tenure. Add a description of what the models do here

just one or two, right?

I don't see this in the results.

In addition to this data, we also provided a roadmap to improve recruitment in warehouse and transportation for the foodservice distributors. We found that many important variables are not tracked or only tracked by part of the companies, and in the roadmap, we will identify those important variables we suggest tracking, including granularity of variables, and give recommendations on methods of analyzing those variables. A short version roadmap is provided in the Results section, and a more detailed roadmap can be found in the Appendix page 21.

ok good idea

## 4 Results

### 4.1 Interview Summary

Nice summary.

We interviewed five participating client companies with regards to their recruiting challenges in warehouse and driver positions, influence of pandemic on recruiting, preferred way of dealing with driver shortage, measures taken to keep employees, and their perspectives on high technology as an alternative to human labor. For these companies, recruiting challenges has been a growing problem for at least the past five years. Some companies think the recruiting challenges are not seriously affected by the pandemic but will continue post-pandemic. Other companies discussed the impact of the pandemic such as paycheck offered by the government, safety concerns, and COVID-19 restrictions on warehouses. Some companies



also mentioned a staff lay-off during the pandemic and rehiring when business picked up. Nevertheless, all companies mentioned that they are experiencing recruiting challenges such as suffering from high turnover far before the pandemic started, so the influence of pandemic on recruiting is limited.

Generally speaking, warehouse positions have applications with “not high quality but high flow” while driver positions are the opposite. Less driver applications are partly due to its more rigorous requirement such as customer service, professionalism and physical fitness. Some companies stated that the difficulty level in the recruitment of warehouse and driver positions varies, and they are more concerned about recruiting truck drivers compared to warehouse workers. In order to deal with driver shortages, companies take measures such as moving drivers from one place to another to cover the routes, investing more money in recruiting advertisements such as TV commercials in addition to basic advertisements for driver positions, and raising payment for drivers. Some companies suggested they prefer to offer higher salaries to keep drivers in position given the money and energy they ~~spent~~ <sup>spend</sup> on training drivers. If higher payment could potentially solve the difficulty in recruiting, some companies are willing to do so.

Given the difficulties in recruiting warehouse and driver positions, these companies sometimes extend offers to candidates who are satisfactory but not ideal. For example, one company mentioned in the interview that they “are satisfied with these employees enough to extend the offer but wish to have more experienced workers”. Keeping a steady retention rate is also a main interest for these companies. To increase retention, they take measures such as ensuring a pleasing workforce environment, proposing promising career paths, and offering rewards and recognition. Among the client companies, all of them mentioned that warehouse and driver positions are stable, pay well and do not have seasonal layoffs. They also expressed the possibility for warehouse workers and drivers to rise to higher management positions. Some companies list this promotion possibility in their hiring ads. These promotional opportunities are further emphasized in some companies through programs such as leadership <sup>development</sup> programs and mentorship programs that are designed to help employees with building a career path at the company. Each company also has its own rewards and recognition program such as a benefit program with gift cards and drawings, education reimbursement program and safety incentives program. If recruiting difficulties cannot be relieved in the near future, these companies also expressed interests in utilizing high technology. All companies are considering automation, though they admit that automation is in its infancy and is unlikely to be vastly carried out in the near future.

only one space

#### (Exploratory Data Analysis)

### 4.2 EDA

Nice subsection

After combining the data from the participants of the project, our team answered meaningful questions we derived from the merged dataset. All the analyses in the results section are for the merged data sets where no individual company information can be identified. Some examples of individual company analyses will be in the appendices. **(provided separately to each participating company)**

The first question we addressed is: **What is the main disposition for different job positions?** We used the merged applicant data set and drew a stacked bar plot (An extension of standard bar plots with additional segmentation of the bars. <sup>(and add space)</sup> While standard bar plots only show the relationship between a categorical variable and a quantitative variable, stacked bar plots also divide each bar into stacked sub-bars based

italics

(space)

on another categorical variable) that displays the disposition of applicants by job position. From the graph (figure 1. ~~Disposition for different positions~~), we observe that the majority of the applicants leave applications incomplete, as depicted by the green bar, followed by withdrawals in interviews and screening (as depicted by the pink bar). ~~The reasons for understanding this insight is unknown, and it could be due to various reasons such as candidate interest in the role, inconsistencies in the portal system, or other parts of the application process.~~ **We are not sure why this is so, but it could be for various reasons** Since there is no available sourcing data provided from our clients, we will address how we can enhance this finding in our roadmap section.

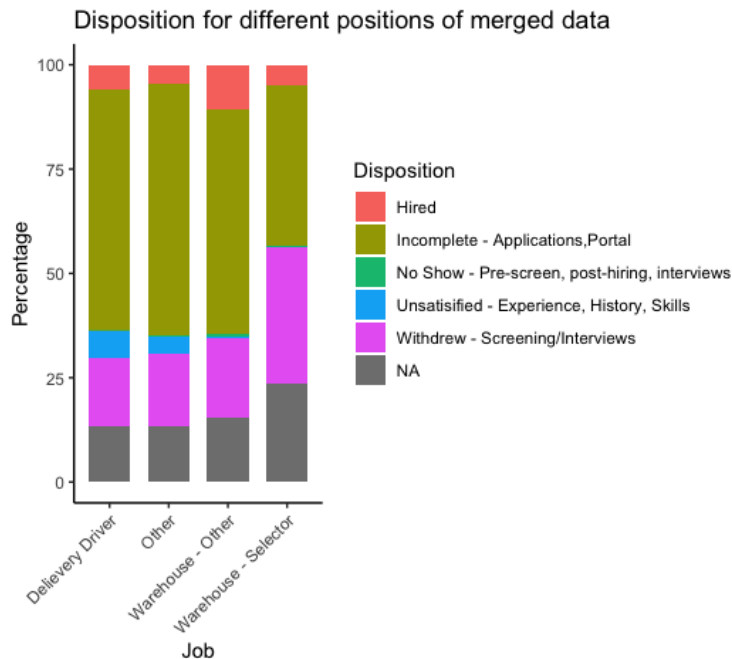


Figure 1: Disposition for different positions

The second question we were able to answer from the merged dataset is: **What types of employees are more likely to stay at the company for more than 6 months?** For this question, we define the types of employees as day shift vs night shift, and full-time vs part-time. Then, we visualized the number of months spent at the company, labeled as 'duration', for all employees in different job roles in figure 4 and then segmented the plot by shift and by full/part time status in figure 2 and 3.

In figure 2 (~~Duration of different Jobs subset by Shift~~), we can see that for day shift jobs, there are more people in 'delivery drivers' and 'other (positions such as supervisor, i.e not in warehouse or transportation department)' job positions that stay more than 6 months. For night shift jobs, there are more people in delivery drivers and warehouse-other positions that stay more than 6 months. Comparatively, we see more employees in warehouse and transportation positions that stay more than 6 months in the night shifts but not for other job positions. Based on the current data sets, we think the higher pay rate of night shifts might be the reason which is drawn from the interview sessions and hourly pay rate. In order to further validate our assumption, we suggest conducting surveys on reasons for termination, which is not provided for now. Furthermore, if we could locate the sourcing channel for the employees, we can analyze from which

source/channel the employees tend to stay more than 6 months. Based on that result, the company can decide different recruiting strategies ~~on~~ <sup>for</sup> different sources.

In figure 3 (~~Duration of different Jobs subset by Full/Part Time~~), we could see that for full-time positions, delivery driver positions have the highest rate of employees who stayed for more than 6 months, even higher than "other" positions. For warehouse positions, we could see that they usually stay more than 6 months as a part-time job. Comparatively, we see that people in warehouse positions would stay long as a part-time job but not as a full-time job. The reasons for termination will help us better analyze why such a difference exists.

In figure 4 (~~Duration of different Jobs (Aggregated data)~~), we see that indeed, transportation positions have a higher <sup>(space)</sup> ~~(similar)~~ percentage of staying more than six months than "other" jobs, so warehouse department seems to have a more severe challenge on retention.

I don't get "similar"...

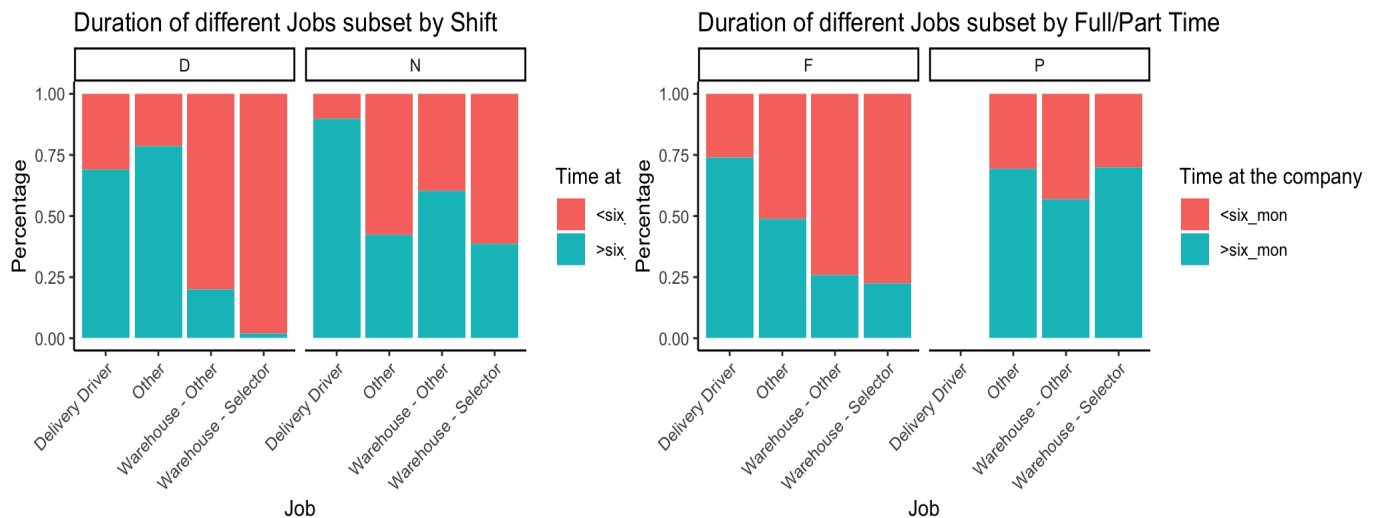


Figure 2: Duration of Jobs subset by Shift

Figure 3: Duration of Jobs subset by Full/Part Time

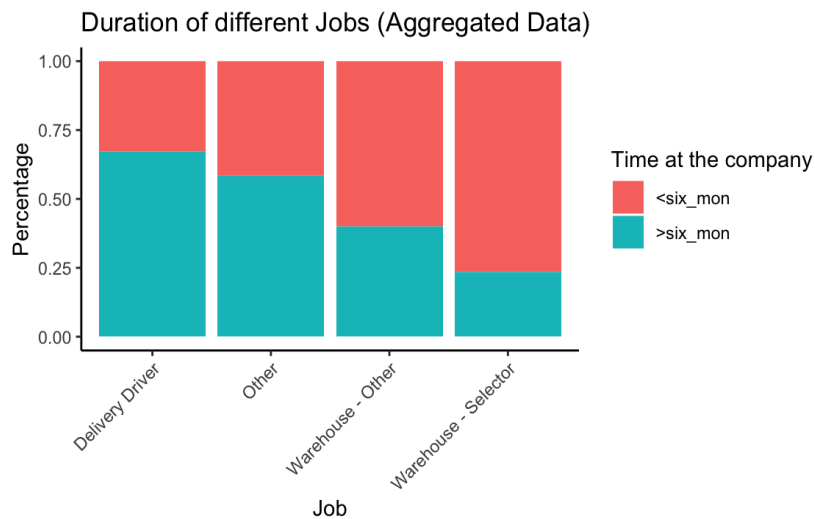


Figure 4: Duration of Jobs (Aggregated data)

(space)  
The third question we are interested in is: Is there seasonality in hiring and termination? Identifying the seasonality in hiring can help us analyze whether there is a certain time in a year a foodservice distributor should ramp up recruiting efforts in anticipation of a busier hiring period. To answer this question, we drew two histograms for both hiring and termination.

*italics*

In ~~Figure 5 (Seasonality in Hiring)~~, we observe that for delivery drivers, there is no obvious seasonality in hiring and each month is at a similar level (the count number is about 20 for all twelve months). For warehouse positions, we could see that there is a high number of hires in November, December and January (Winter time) probably because of the holidays.

In ~~Figure 6 (Seasonality in Termination)~~, we could see that delivery drivers usually left the company in the first two quarters (higher bars in the first few months) and after that the termination is relatively low. There would be more insights into the detailed reason of why people left in the first quarters if the company could conduct surveys on reasons ~~of~~ <sup>for</sup> termination. For warehouse selector positions, we see that high termination happens from December to March. Compared to the number of hires, we see that there is a lag in termination. From our previous analysis, we see warehouse selectors are more likely to stay more than 6 months as a part time job. Therefore, for warehouse selectors, the problem is not just about hiring, but more about retention. For "warehouse other" positions, we see a similar lag as warehouse selectors and the previous analysis suggest that people who work part-time are more likely to stay for more than 6 months. Therefore, for warehouse positions, the problem is more about retention.

**Not sure how this follows from the previous sentences. Make that clearer.**

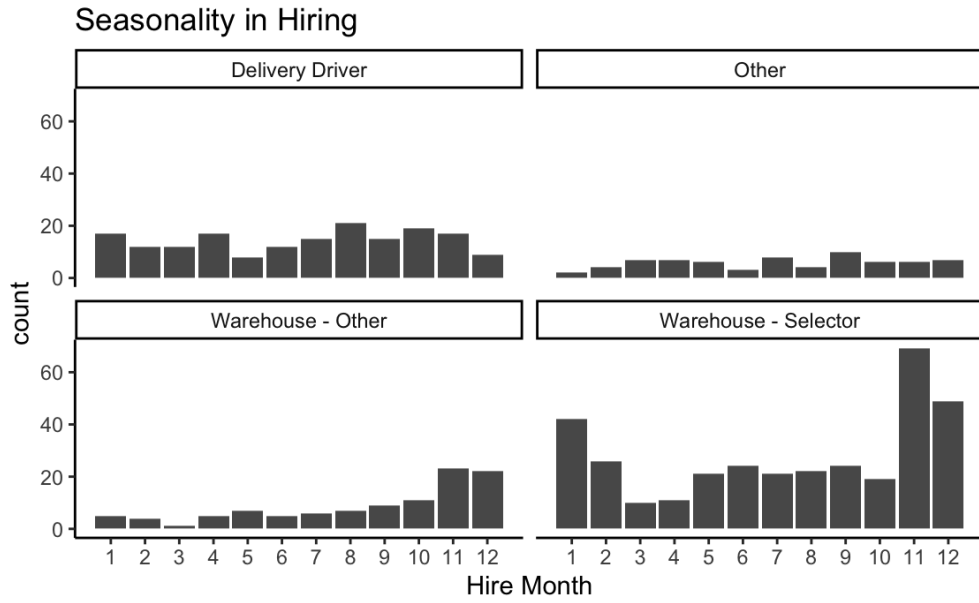


Figure 5: Seasonality in Hiring

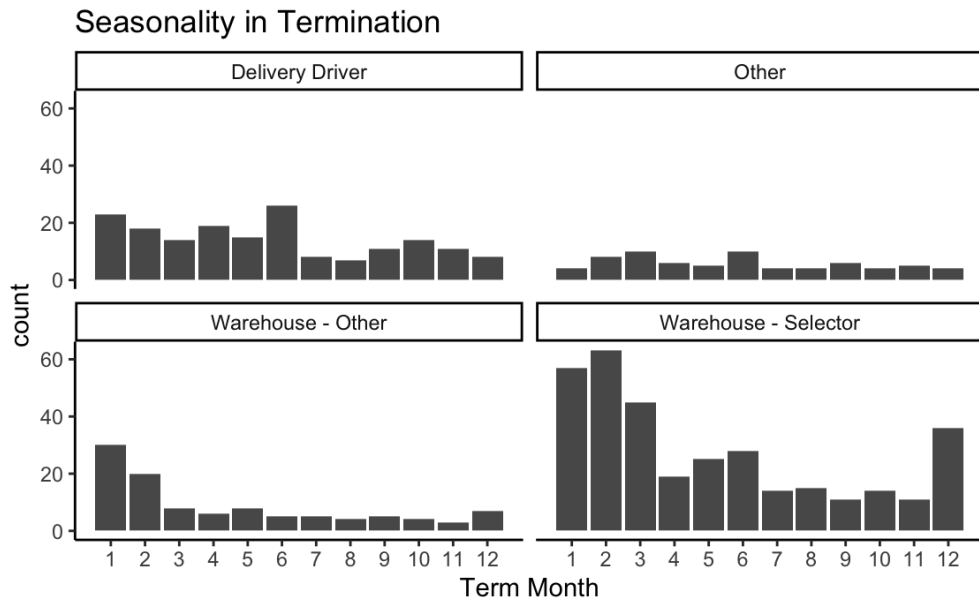


Figure 6: Seasonality in Termination

The last question we were able to answer from the merged dataset is: *How does hire rate and turnover rate change over time?* We would like to investigate if the recruiting challenge has been consistent and remains severe. To answer this question, we visualized the turnover rate and hire rate for warehouse and transportation departments over the years.

In figure 7 (Hire and Turnover rate over time), the hire rate is denoted by the solid lines and the turnover rate is denoted by the dashed lines; the red lines are for warehouse positions, and the black lines are for drivers positions. We observe that the hire rate for warehouse positions increases after 2018 and exceeds the turnover rate in 2019. However, for transportation positions, the hire rate is higher than the turnover rate in 2018 and 2019 with a slightly decreasing trend. The hire rate drops drastically in 2020 while the turnover rate surges in 2020 *perhaps* due to COVID-19.

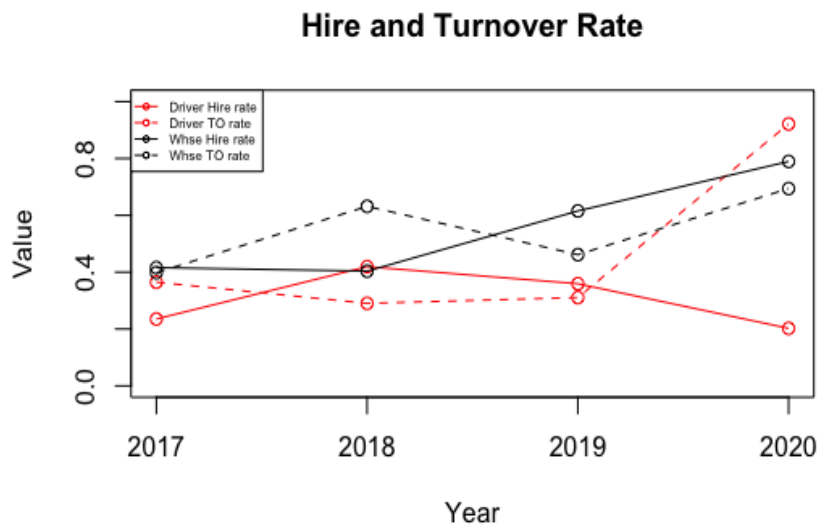


Figure 7: Hire and Turnover rate over time

These exploratory analyses show some ideas of how to use current data sets to generate insightful information about the employees and applicants in warehouse and transportation departments. A more detailed roadmap with questions, key performance indicators, and methods will be discussed in the next section. *great*

### 4.3 Roadmap

1. During the process of reviewing IFDA members' datasets, we found that <sup>several</sup> a lot of important variables which can help us analyze the recruitment and retention problem, are only tracked by some of the companies or missed by all the companies. Thus, we introduce a roadmap to suggest the variables for the companies to track based on the questions listed, which related to recruitment and retention. We arrange the roadmap according to the stages of the recruitment pipeline: Awareness Interest, Application, <sup>&</sup> Interviewing, and Hiring for a given job requisition, which is also mentioned in the introduction section. <sup>For</sup> Under each stage of the pipeline, we listed the questions and the variables we think are important for us to analyze the recruitment and retention problem. <sup>&</sup> This short version of the roadmap is a very brief version of the complete roadmap, which includes a few examples of questions, variables, <sup>some illustrative</sup> and simplified analysis methods. ~~This short version of the roadmap aims to make it easy to read for the audience and show how we analyze variables in the complete roadmap.~~ In the Appendix, we include the complete roadmap, which shows the complete list of questions, variables that we recommend tracking to address the questions, and how to use data to optimize the talent acquisition process.

comma belongs with previous word

comma before space

new paragraph

#### 4.3.1 Awareness & Interest

1. Which sourcing channel provides the candidates that are most likely to get hired?

italics

With this information, one can choose to spend more effort on the most efficient source of hiring to speed up the recruiting process.

Table 3: Variable table for Awareness & Interest

Variable	Definition
Applicant - Date applied	The date that applicant submitted the application
Applicant - Open Date	The date that the requisition was posted and open
Requisition - Division	The division of the company that the requisition refers to
Requisition - Status	The status of requisition, whether it is filled or not
Applicant - Referring source	The source where the applicant applied for the job
<b>Visualizations we can complete with this information</b>	<b>Data</b>
Sourcing Channel Comparison	Visualization of Distinct Total Applicants from each Sourcing Channel

#### 4.3.2 Application & Interviewing:

1. Do shorter interview cycles increase the likelihood of hires, among those who were offered a position?

italics

We think shorter interview cycles save costs in revenue by reducing the time to search for quality hires, while also increasing their chances of getting a qualified candidate in this market where the demand is high.

**Too much whitespace. Can you get rid of it?**



Table 4: Variable table for Application & Interviewing

Variable	Definition
Applicant- Interview Bin of candidate	The progress including all stages of the interview
Applicant - Interview Date	The date that the applicant was interviewed
Applicant - Status of interview	Whether the applicant was rejected or hired
<b>Visualizations we can complete with this information</b>	<b>Data</b>
calculate the offer acceptance rate and plot a bar graph displaying the offer acceptance rate for each bin	Interview Bin of Candidate

### 4.3.3 Hiring:

1. What is the average time to hire for each job role?

~~Importance:~~ It's informative to understand the time it takes in hiring for specific job titles, and relate that to what the client would expect it would take to hire someone for a particular role.

*italics*

How is this different from question posed in 4.3.2? Can you differentiate it better, or ask a different illustrative question?

Table 5: Variable table for Hiring

Variable	Definition
Requisition - Job Title	The job title of then requisition
Requisition - Open Date	The date that the requisition was posted and open
Employee - Hire Date	The date when the employee was hired
Requisition - Division	The division of the company that the requisition refers to
<b>Calculations we can complete with this information</b>	<b>Formula</b>
average time to hire for each job/department/location	average (Hire Date - Date applied) By job title/ department/location


In addition, we interviewed the companies to find out more about their recruitment and hiring challenges.

## 5 Discussion

The ~~participating~~ <sup>five</sup> IFDA members <sup>participating in</sup> of this project have a variety of information <sup>tracked</sup> tracked to address the recruitment problem in warehouse and transportation job families. ~~In section 4.1 of our Results, we found that the foodservice distribution industry is being impacted from the COVID-19 pandemic in multiple ways, including early-pandemic layoffs and a decreasing supply of applicants which could be because of the unemployment checks that individuals are receiving. In section 4.2 of our Results, we displayed the dispositions across driver and warehouse roles, and we saw a consistent pattern of incomplete applications across the varying roles. The next leading status of disposition were withdrawn applications. In our extended roadmap, we suggest tracking data that could explain an applicant's decision to withdraw or leave incomplete, including the time taken to conduct interviews, specific requisitions that experience a higher applicant incompleteness rate or applicant withdrawal rate, and other metrics surrounding a company's web-trafficking data for these requisitions. By recording this information, foodservice distributors lay the groundwork for analyzing which job requisitions are highly affected by the recruitment problem, while supplementing the analysis with descriptive statistics on the requisition and the candidates.~~ <sup>based on interview data</sup> <sup>new paragraph</sup> <sup>based on data aggregated and merged from the participating companies</sup>

We also visualized the proportion of driver and warehouse associates who spent less than six months versus those who spent more than six months at their company, which is detailed in section 4.2 <sup>above</sup> of ~~Results~~. The warehouse job role consistently has a higher proportion of employees who stay for less than six months at their company. Foodservice distributor companies can enhance their understanding behind this termination statistic by conducting candidate feedback surveys during an associate's tenure so there is greater understanding of the associate experience while working in the transportation and warehouse roles. We can further our understanding by keeping track of what company the employee is leaving from and the reasoning behind the departure, to better align the associate's needs with their current employer. Anecdotally, the foodservice distribution industry has been facing competition from other industries. Though this sentiment was not conveyed in our interviews, the threat of increasing competition from other industries was discussed in an article written by DCVelocity <sup>above</sup> about the possible reintroduction of the DRIVE-Safe Act, including a conversation with Mark Allen, CEO of IFDA and Paul Saval, CEO of Saval Foods. Allen states, "the rapid acceleration of e-commerce and increasing volume of products flowing through the supply chain are creating a more competitive driver landscape that is likely here to stay." We think this is an important <sup>delete comma</sup> change in the labor market to consider, and we introduce variables to track about the candidate experience and their previous employment experience in the extended Roadmap located in our Appendix (REFER IT). <sup>put an ASA style citation to the article here</sup>

<sup>I don't get how this is functioning in the sentence</sup> ~~Given the qualified hires within a company, we can improve the candidate selection process by identifying several pieces of information about the qualified hire: the source of hire, feedback from the candidate experience surveys, time spent to hire, requisition location, and more. These metrics provide companies supplemental information on how qualified candidates were hired, how quickly they were interviewed, how these metrics vary by location and seasons, and how turnover can be reduced by targeting the segment of qualified candidates that exhibit longer tenure than expected. We provide further explanations on these metrics in our extended Roadmap in the Appendix. We also recommend impactful web-analytics variables that can optimize the web traffic to a company's requisition posting. By identifying questions that are beneficial to answer about the recruitment process, we derived a set of variables that may help answer~~ <sup>ok. don't forget in final draft!</sup> <sup>(pp. xx-yy)</sup> <sup>new parag</sup>

these questions and ~~to~~ motivate foodservice distributor companies to leverage their recruiting data to their ~~potential~~. <sup>advantage</sup> 

## 6 Reference

IFDA. *A Comprehensive Economic Impact Study of the U.S. Foodservice Distribution Industry*. International Foodservice Distributors Association, 2017.

Kickham, Victoria. *Demand for Drivers on the Rise in Foodservice Industry*. DC Velocity RSS, DC Velocity, 18 Feb. 2021, [www.dcvelocity.com/articles/49591-driver-shortage-revisits-foodservice-industry](http://www.dcvelocity.com/articles/49591-driver-shortage-revisits-foodservice-industry).

Team, SmartDreamers. *A Recruitment Marketer's Guide to the Candidate Journey*. SmartDreamers Recruitment Marketing Automation, 2019, [www.smartdreamers.com/blog/candidate-journey-guide](http://www.smartdreamers.com/blog/candidate-journey-guide).

**Reformat to ASA style. Add BLS references and any other references you have cited in the main body of the paper.**

## 7 Appendix

### 7.1 Glossary

This is fine, but put the terms in alphabetical order (I cannot tell what order they are in now, and alphabetical will help readers find things).

1. **EDA (Exploratory Data Analysis):** The process of performing initial explorations on data to summarize the main characteristics of the data and to discover underlying patterns with numerical summaries and data visualizations.
2. **Categorical Variable:** A variable that consists of different categories which each observation can be grouped into. Some examples of categorical variables are gender, race, and department.
3. **Quantitative Variable:** A variable that take numerical values such as counts, percents, and numbers. Some examples of quantitative variables are wage, number of applicants, and turnover rate.
4. **Stacked Bar Plot:** An extension of standard bar plots with additional segmentation of the bars. While standard bar plots only show the relationship between a categorical variable and a quantitative variable, stacked bar plots also divide each bar into stacked sub-bars based on another categorical variable.  
**with vertical bars**
5. **Histogram:** A plot showing the frequency of a quantitative variable in different ranges. For example, we can show the number of people in each age group (15-25, 25-35, etc.) with a histogram.
6. **Mosaic Plot:** A plot that displays the relationship between two categorical variables with blocks representing the number of observations in the intersection of two categories from the variables. For example, we can construct a mosaic plot of location and reason for termination and examine whether the proportion of different reasons for termination differs for locations.
7. **Standard Deviation:** A statistic that measures how spread out a quantitative variable is. A variable with similar values in the data will have a small standard deviation, while a variable that adopts a variation of values in the data will have a large standard deviation.
8. **Logistic Regression Model:** A model that predicts the probability of a certain category or event existing for each observation. For example, the logistic regression model can model the probability of an employee terminating in 6 months based on other variables such as sourcing channel, demographics information, and annual performance review.
9. **Classification Decision Tree Model:** A model that classifies a observation as a certain category by subsetting data based on values of selected variables. This subsetting process constructs a set of rules that divide data in the form of a tree structure. An example of a simple decision tree with only one subsetting step would be to predict if a patient is sick or not by the number of times the patient coughs during the session. If a patient coughs more than five times, then the model predicts that the patient is sick. Otherwise, the model predicts that the patient is healthy.
10. **Confusion Matrix:** A table that describes the performance of a model that classifies a observation as a certain category. It shows the number of records that the model successfully predict for each category, as well as the number of records that the model fails to predict correctly for each category.

I didn't see any mosaic plots in the main paper...

nice succinct description

11. **Statistically Significant:** A claim that it is unlikely to observe the data collected if the ~~claimed~~ <sup>hypothesized</sup> relationship between variables or the ~~claimed~~ <sup>hypothesized</sup> effect does not exist. It adds statistical evidence to the model that we build or the hypothesis that we make.
12. **Model Summary Statistics:** A summary of model statistics that help interpret the model and evaluate how well the model fits the data.

## 7.2 Interview Questions Used

1. Restate Our Understanding of the Problem
  - Can you tell me about your perspective on the challenge and how long it's been a concern?
  - Can you tell me about your perspective on the competition for workers?
  - How do you think this problem has been affected by the pandemic? Do you see this problem continuing post-pandemic?
2. Problem Diagnostics
  - Are you content with the candidates that you do accept into the jobs?
  - What are some growth opportunities for these jobs? What are their career paths forward?
    - Are potential candidates and new employees made aware of career paths, for example in job descriptions?
  - What are some rewards and recognition programs for these jobs, if any?
  - Has your company considered or implemented technology solutions to relieve job shortages, such as autonomous labor? If autonomous labor has been employed, what stage is it at?
3. Problem Measurement - What Data Did They Track

I don't think this is the exact text of your questionnaire, is it? If not, please either (a) replace with the exact text of your questionnaire, or (b) add the exact questionnaire after these three general points.

## 7.3 Detailed Roadmap

The aim of the roadmap is to answer the question ~~raised in the introduction paragraph:~~ <sup>(space)</sup> How can we develop a data-driven framework to further our understanding of the candidate pool and improve recruitment efforts? ~~We construct the following roadmap to suggest to our clients what variables we recommend tracking,~~ <sup>In the roadmap below, we list</sup> why these variables are crucial in recruitment analytics, and how to use them to ~~optimize~~ <sup>improve</sup> the talent acquisition process after tracking them for a while. We organize the roadmap according to the recruitment pipeline, Awareness & Interest, Application & Interviewing, and Hiring for a given job requisition. For every variable ~~we suggest our clients to track,~~ <sup>in the roadmap</sup> we list the questions we want to answer with the variable, the importance of the questions, and how to use the variables to ~~optimize~~ <sup>improve</sup> the recruitment pipeline. We also include a variable table ~~. Roadmap recommendation variable table,~~ <sup>provide</sup> at the beginning of the roadmap section, to give our clients a clear overview of which variables we are recommending to track for future analysis to improve employee recruitment.

Table 6: Roadmap recommendation variable table

Variable	Tracked by Participating Companies	Explanation and Granularity	Issue revelant to
Requisition - Job description	yes	Include all the information in the job posting	Recruitment
Applicant - Referring source	no	Accompanied with applicant information, Include the website name or other way of source, such as LinkedIn, referral	Recruitment
The number of hit on the company website job page	yes	Including the time and number of hit, such as Date: 04/21/2021, Hit:870	Recruitment
Applicant- Date applied	yes	The date that applicant applied for the job, such as 04/21/2021	Recruitment
Requisition- Open Date	yes	The date that the job application channel open, such as 04/21/2021	Recruitment
Employee- Disposition	yes	As detailed as possible, such as "rejected in the phone interview stage due to short work history"	Recruitment
Employee - Tenure	yes	Can be recorded as years and months or can calculated by hire date and termination date if the employee has been terminated	Recruitment
Employee - Performance metric	no	As detailed as possible, such as 5 out of 5 for high efficiency	Recruitment
Applicant Unique ID (Connection between applicant and employee data)	no	Assign an unique applicant ID for each applicant and include that in the employee table, such as a 4 digit ID	Recruitment
Requisition- Division	yes	The division the employees are in or the applicants are applying for	Recruitment
number of openings	yes	The number of job openings	Recruitment
Requisition - Location	yes	The location of the posted job, such as Pittsburgh, PA	Recruitment
Applicant - demographics (Gender, Age, etc)	yes	The base of the applicants, such as Pittsburgh, PA	Recruitment

<b>Variable</b>	<b>Tracked by Companies</b>	<b>Explanation and Granularity</b>	<b>Issue relevant to</b>
Requisition- Close	no	Such as 04/21/2021	Recruitment
Applicant- Work history	no	As detailed as possible, include job title, company, and duration.	Recruitment
Requisition Unique ID (The connection between requisition and applicant data)	no	Include an unique requisition ID for every job posted, and include that in the applicant data	Recruitment
Requisition - Full/part-time	yes	Include this in requisition data and employee data	Recruitment/ retention
Requisition(Day/night shift)	yes	Include this in requisition data and employee data	Recruitment/ retention
Employee - demographics (Gender, Age,etc)	yes	the base of the employees, such as Pittsburgh, PA	Recruitment/ retention
Termination - Reason for termination	yes	As detailed as possible, indicate it is voluntary or involuntary. Such as voluntary termination due to transfer to another job	Recruitment/ retention
Offer acceptance rate	no	Should be tracked or calculated by each requisition	Recruitment
Employee - satisfaction	no	Descriptive record, can be done either through surveys or websites like Glassdoor.com	Recruitment/ retention
Requisition - status	yes	Such as closed and filled/ closed and not filled	Recruitment
Requisition - wage budget	no	Can be separately recorded as hourly wage and headcount, such as 19\$/hr and 17 headcount	Recruitment
Requisition - sourcing channels	no	The website or other channel that the job is been posted, such as LinkedIn or career fair	Recruitment
Applicant - Interview Date	no	Such as "interview stage:phone interview, Date: 04/21/2021"	Recruitment



<b>Variable</b>	<b>Tracked by Companies</b>	<b>Explanation and Granularity</b>	<b>Issue relevant to</b>
Applicant - Status of Interview	no	As detailed as possible, such as phone interview	Recruitment
Employee - Hire Date	yes	Such as 04/21/2021	Recruitment/ retention
Termination- Termination Date	yes	Such as 04/21/2021	Recruitment/ retention
Employee - Hourly rate	yes	Such as 19\$/ hr, and if there is extra payment for certain situations, please include.	Recruitment/ retention
Applicant - The time the employee took the offer	no	Can be tracked as days or calculated by date of accepting the offer minus date of receiving the offer	Recruitment/ retention
Applicant- Reason for leaving the last job	no	As detailed as possible, and can such as low payment, etc and it should be categorized to voluntary and involuntary.	Recruitment/ retention
Interest conversion rate	no	It is defined as $\frac{\text{of visitors who clicked 'Apply'}}{\text{of unique visitors to Job Posting}}$	Recruitment
Applicant conversion rate	no	It is defined as $\frac{\text{of applicants who completed application}}{\text{of applicants who clicked the 'Apply' button}}$	Recruitment

justified text,  
not centered  
text

Under each pipeline below, we list questions with four sections: Importance describes why is the question important, Variables Tracked are variables that the companies already record, Variables not Tracked are variables that the companies have not recorded, and Method of analysis shows how to use the data to gain information.

gives examples of

### 7.3.1 Awareness Interest

#### 1 Which sourcing channel provides the candidates that are most likely to get hired?

- (a) Importance: With this information, one can choose to spend more effort on the most efficient source of hiring to speed up the recruiting process.
- (b) Variables tracked:
  - Applicant - Date applied
  - Requisition - Open Date
  - Requisition - Division
  - Requisition - Status
- (c) Variables not tracked:
  - Referring source
- (d) Methods of analysis:
  - i. For each source of hiring, calculate the hire rate for each period of time.

use bullets instead of roman numerals to be consistent with other examples below.

#### 2 What proportion of visitors access the company website careers page using each device (mobile / desktop tablet)?

where are (b), (c) & (d) ???

- (a) Importance: User experience from different devices usually vary a lot. If there is a considerable amount of users accessing the company recruitment page using mobile devices, the company can consider developing a mobile-friendly layout.

#### 3 By tracking the number of visitors from different sources, the company can have a more holistic view of the effectiveness of sourcing channels.

- (a) Importance:<sup>(space)</sup>By tracking the number of visitors from different sources, the company can have a more holistic view of the effectiveness of sourcing channels.
- (b) Traffic sources examples:
  - Social media
  - Job boards
  - Paid advertisement
  - Direct traffic (website visits that arrived on your site either by typing your website URL into a browser or through browser bookmarks)
  - Organic traffic (visits from search engines)

Not consistent with what you promised at top of p 26

where are (c) & (d) ???

#### 4 What is the conversion rate of each web sourcing channel?

- (a) Variable definition:<sup>(space)</sup> 1. Interest conversion rate = (number of visitors who clicked the "apply" button) / (number of unique visitors): This is essentially the apply button "click-through rate".

where are (b),  
(c) & (d) ???

(space)

2. Application conversion rate = (number of applicants who completed application) / (number of visitors who clicked the "apply" button)

### 5 What proportion of visitors access the company website careers page using each device (mobile / desktop tablet)?

- (a) Importance: User experience from different devices usually vary a lot. If there is a considerable amount of users accessing the company recruitment page using mobile devices, the company can consider developing a mobile-friendly layout.

where are (b),  
(c) & (d) ???

### 6 What types of applicants tend to apply early/late?

- (a) Importance:<sup>(space)</sup> Although the company does not have access to potential candidates who did not apply, we can investigate whether applicants who apply early have different characteristics compared to applicants who apply late.

- (b) Variables tracked::<sup>(space)</sup> only one colon

- Requisition - Open date
- Requisition - Job title
- Requisition - location
- Applicant- Date applied
- Applicant - Demographics

- (c) Variables not tracked::<sup>(space)</sup> only one colon

- Requisition - Close date
- Applicant - Referring Source
- Applicant - work history (years of experience)
- Requisition Unique ID (The connection between requisition and applicant data)

- (d) Methods of analysis:

- Use applicant application date - requisition open date to obtain the response variable "application days"
- Use box plots and line plots to compare the distribution of "application days" for different applicant sourcing channels and for categorical demographic variables such as gender, race, education
- Analyze the plots and understand whether the differences result from different posting dates in sourcing channels / different time for information in each sourcing channel to reach potential applicants
- Use a scatter plot to explore the relationship between "application days" and applicant age
- Segment data by job positions, department, location, and repeat the process to examine whether early applicants and late applicants behave differently for these groupings.
- Use requisition job title, requisition location, applicant application date, applicant sourcing channel, applicant demographics, and applicant work history (years of experience) to predict "application days".

## 7 How can we attract more candidates to apply?

- (a) Importance:<sup>(space)</sup> We can see that for driver and warehouse positions, especially driver positions, our clients are facing recruitment issues, hard to fill the expected headcount. If we can attract more candidates to apply, at least a part of the recruitment issue should be addressed.
- (b) Variables tracked::<sup>only one colon</sup>
- Job descriptions
- (c) Variables not tracked::<sup>only one colon</sup>
- Applicant - Referring Source
  - The number of hit on the company website job page, including the hit on different job positions
- (d) Example of analysis:<sup>Example or method? You have used "method" elsewhere. It doesn't matter which word you choose, as long as you always use the SAME word.</sup>
- We can build a pie chart and a bar chart to visualize the proportion of candidates from each source. And also we can build a time series plot for each source to see if there is any change by time. Thus, we can determine which source to market and to focus on so that we can attract candidates. Say, one of the clients said that they used to attend some career fairs in person. However, only a few people came to the career fair nowadays, and they decided not to attend those events now.

### 7.3.2 Application Interviewing:

#### 1 Do shorter interview cycles increase the likelihood of hires, among those who were offered a position?

- (a) Importance:<sup>(space)</sup> We think shorter interview cycles save costs in revenue by reducing the time to search for quality hires, while also increasing their chances of getting a qualified candidate in this market where the demand is high.
- (b) Variables tracked::<sup>only one colon</sup>
- Applicant - Talent Acquisition Interview Bin
- (c) Variables not tracked::<sup>only one colon</sup>
- Applicant - Interview Bin of Candidate
  - Applicant - Interview Date
  - Applicant - Status of Interview
- (d) Method of analysis:
- With each column representing the different stages of the interview process, including hiring, obtain the dates of each stage for a given requisition.
  - Calculate the difference between the offer date and the application date for a candidate, for a given requisition. Keep count of whether the candidate accepted the offer or not.
  - Aggregate the differences between the offer date and application date (in days) for candidates who accepted the offer.

- Split the length of the interview cycles (the difference between the application date and offer date) into the number of bins you see fit: 1 month, 2 months, 3 months, . . . , etc. These bins indicate the duration of the interview process in order for us to observe the offer acceptance rate.
- For each bin, calculate the offer acceptance rate and plot a bar graph displaying the offer acceptance rate for each bin, which indicates how long the interview process was for them. You can also observe the most common time it takes to complete the interview cycle and the average time for the interview cycle.

### 7.3.3 Hiring:

#### 1 What is the average time to hire for each job role?

- (a) Importance:<sup>(space)</sup> It's informative to understand the time it takes in hiring for specific job titles, and relate that to what the client would expect it'd take to hire someone for a particular role
- (b) Variables tracked::<sup>only one colon</sup>
- Requisition - Job Title
  - Requisition - Open Date
  - Employee - Hire Date
  - Requisition - Division
- (c) Method of analysis:
- Locate the Application Date for each applicant who was hired.
  - Locate the Hire date for each applicant who was hired.
  - Compute the difference between Hire Date and Application Date for each applicant. This can be done in days as the unit of time.
  - Aggregate all of the differences between Hire Date and Application Date and compute the average difference, which is the average time to hire for a particular job role.

where is variables not tracked? If there are none, still include the section but say "none" or something like that

#### 2 Is there seasonality in hiring and termination?

- (a) Importance:<sup>(space)</sup> Identifying the seasonality can help us analyze whether there is a certain time in a year we need to pay more attention to. If the seasonality (histogram) of hiring and termination is different, then we need to figure out ways to fill the gap between high termination and low hiring.
- (b) Variables tracked::<sup>only one colon</sup>
- Termination - Termination Date
  - Employee - Hire Date
- (c) Variables not tracked::<sup>only one colon</sup>
- Requisition - Open Date
  - Termination - Reasons for Termination
- (d) Example of analysis:

- For delivery drivers, there is no obvious seasonality in hiring and each month is at a similar level. For warehouse positions, we could see that there is a high hiring number in Nov, Dec and Jan (Winter time), it is probably because of the holidays.

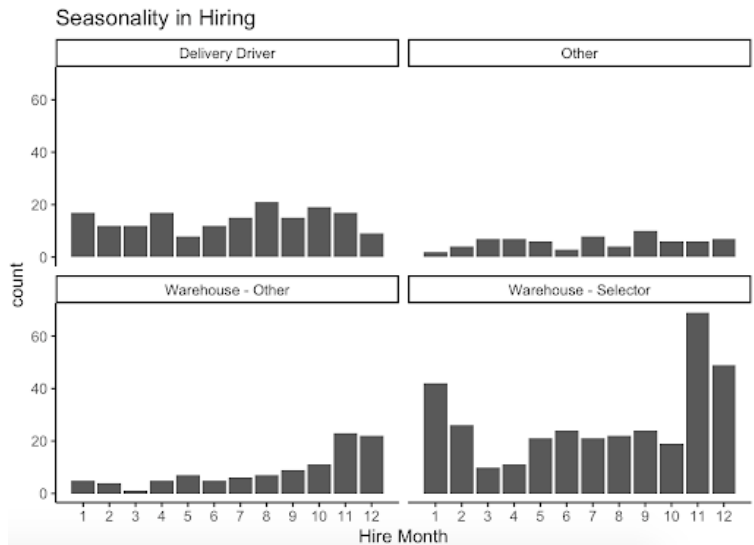


Figure 8: Seasonality of Hiring

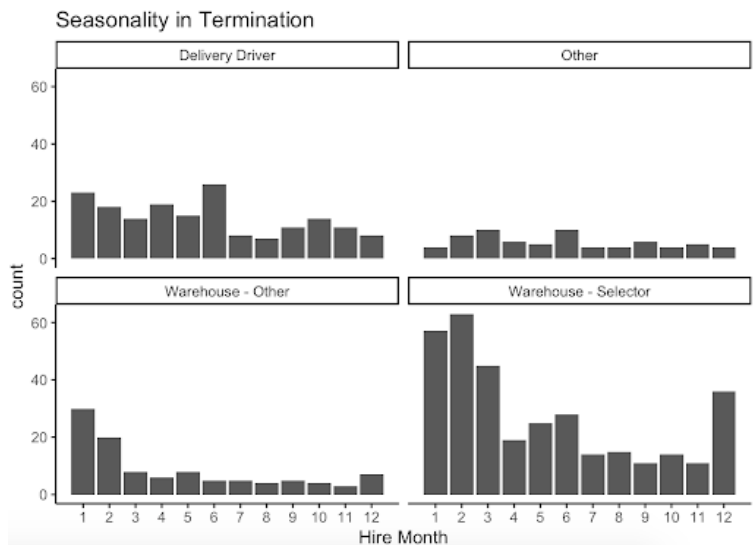


Figure 9: Seasonality of Termination

- For termination, we could see that delivery drivers usually left the company in the first two quarters and after that the termination is relatively low. The detailed reason of why people left in the first quarters will be done if the company could conduct surveys on reasons of termination. For warehouse selector positions, we see that high termination happens from December to March. Compared to the hiring number, we see that there is a lag in termination. From our previous analysis, we see warehouse selectors are more likely to stay more than 6

months as a part time job. Therefore, for warehouse selectors, the problem is not just about hiring, but more about retention. For “warehouse other” positions, we see a similar lag as warehouse selectors and the previous analysis also suggests that people who work as a part time are more likely to stay more than 6 months. Therefore, for warehouse positions, we should focus more on retention.

### 3 Are there difficulties in filling requisitions (does the status change from time to time, region to region)?

(a) Importance: With this information, one can have a general idea on hiring that may “verify/disprove” the conception (Of course, difficulty is defined).

(b) Variables tracked::: only one colon

- Application - Date applied
- Requisition - Division
- Requisition - Status
- Requisition - Open Date
- of Openings
- Employee - Hire Date
- Requisition- ID

where is the variables not tracked section??

(c) Method of analysis:

- For the calculations below, one can either assume simultaneous recruiting hiring or lag by the average time to hire throughout (needs extra calculation).
- Calculate the ratio of applicants to requisitions for each time period.
- Aggregate total of openings and number hired for requisitions.
- Calculate the ratio of number hired to of openings for each time period.
- Calculate the hire rate (hired/applicants) and compare to fill rate (hired/ of openings) for each time period.
- Calculate the average time to hire for each time period.
- Calculate the average time to fill (if requisition ever gets filled) for each time period.

### 4 What is the time to fill for job openings in different seasons, positions, and locations?

(a) Importance: With an estimate of time to fill a position, the company can plan for recruitment more accurately and understand the amount of time to obtain a qualified candidate for different seasons, positions, and locations. The company can also establish a benchmark for this metric and track any noticeable changes to detect problems early on.

(b) Variables tracked::: only one colon

- Requisition - open date
- Requisition - status (closed and filled / closed and not filled)
- Requisition - job title
- Requisition - location

(c) Variables not tracked:: only one colon

- Requisition - close date
- Requisition - wage budget
- Requisition - sourcing channels

(d) Method of analysis:

- Calculate time to fill for all closed requisitions in at least a one-year period
- Use mosaic plots to compare requisitions that were filled and not filled for different positions and locations
- Use line plots to visualize the number of filled requisitions and the number of unfilled requisitions over requisition open time
- Filter out requisitions that were not filled, use boxplots to visualize time to fill for each job position and location
- Filter out requisitions that were not filled, use line plots to visualize average time to fill over requisition open time.
- Filter out requisitions that were not filled, use boxplots to visualize time to fill for each month.
- Segment data by job position, department, location, and repeat the process to examine possible effect of seasonality for different job positions/departments/locations
- Filter out requisitions that were not filled, use scatter plot to visualize the relationship between requisition time to fill and requisition budget wage
- Linear regression model, tree-based models (decision tree, random forest, XGBoost): Use requisition job title, requisition location, applicant application date, applicant sourcing channel, applicant demographics, and applicant work history (years of experience) to predict “application days”.

## 5 What are the common traits of hired applicants?

(a) Importance: Traits exhibited before the screening process such as time takes to apply for the position, application source, and traits exhibited during the screening process such as time to get hired may have a correlation with how applicants will react to the job once they get hired. With this information, one can have a general idea of the likelihood of the applicant accepting the offer or performing satisfactorily in the job. Assumption: “Successfully hired applicants” is defined as hired applicants that accept the offer, stay in the company for more than one month, and with satisfactory performance.

(b) Variables tracked:: only one colon

- Applicant - Date applied
- Requisition - Division
- Requisition - Open Date
- of Openings
- Requisition - Status

(c) Variables not tracked:: only one colon



- Applicant - Referring Source
- Applicant - Unique ID(Connection between applicant and employee data)
- Employee - Performance metric

(d) Method of analysis:

- Group hired applicants based on time to apply for the position, hire source and time to get hired. For each group of applicants, calculate their rate of acceptance. For each group of applicants, calculate their distribution of performance ratings.

## 6 What types of employees are more likely to stay at the company for more than 6 months?

(a) Importance: Understanding the difference of people staying at their jobs will help the company design specific hiring strategies for those departments. If “other” jobs (non-warehouse/ drivers ) have a higher leaving rate, then we know that the problem is not unique to the warehouse and transportation department.

(b) Variables tracked:: only one colon

- Requisition - Job title
- Requisition - full/ part-time
- Requisition - Day/Night Shift
- Termination - Reason for termination

(c) Variables not tracked:: only one colon

- Applicant - Referring Source
- Applicant - demographics
- Applicant - Job history
- Offer acceptance rate
- Employee - satisfaction ( either through surveys or websites like Glassdoor.com)

(d) Method of analysis:

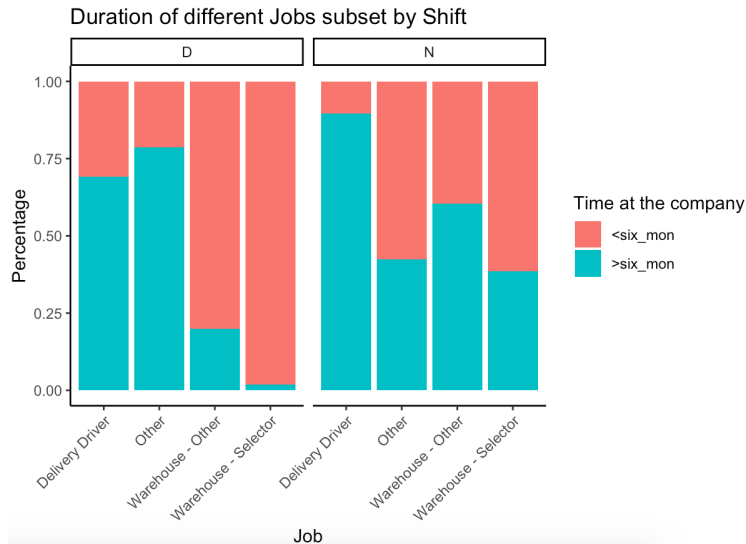


Figure 10: Duration of different Job subset by Shift

- From this plot, we can see that for Day shift jobs, there are more people in delivery drivers and other job positions stay > 6 months. For night shift jobs, there are more people in delivery drivers and warehouse other positions that stay > 6 months. Comparatively, we see that the warehouse and transportation positions usually stay more than 6 months in the night shifts but not for other job positions. Our guess is that the pay rate is usually high for Night shifts and in order to further analyze our assumption, we need to do a survey on reasons for termination.

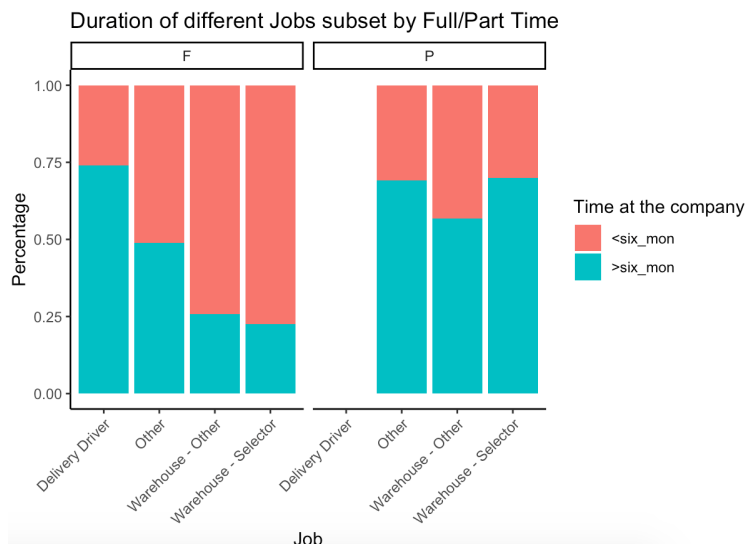


Figure 11: Duration of different Job subset by Full/part time

- Another factor we could analyze based on our current data sets is full and part-time. From this plot, we could see that for full-time positions, people in delivery driver positions have

the highest rate of staying > 6 months, even higher than “other” positions. For warehouse positions, we could see that they usually stay more than 6 months as a part-time job. Comparatively, we see that people in warehouse positions would stay long as a part-time job but not as a full-time job. The “reasons of termination ” will help us better analyze why such a difference exists.

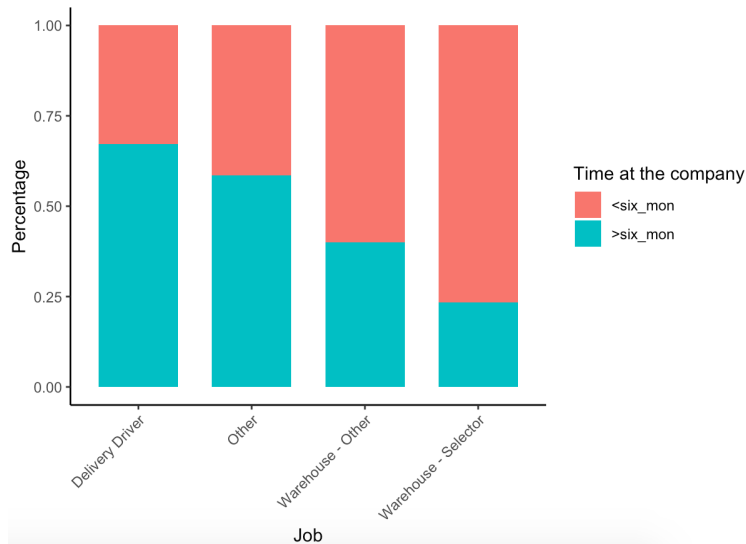


Figure 12: Percentage of people who stay more than 6 months

- This is the overall percentage of people (Aggregated data ) who stay more than 6 months in our merged data set. We see that indeed, transportation positions have a higher rate(similar rate as) than “other” jobs, so we should focus more on the warehouse selector jobs

## 7 Which sourcing channel provides the candidates that are most likely to stay for more than a year once hired?

- (a) Importance:<sup>(space)</sup> Answering this question helps the company understand which sourcing channel provides candidates with the best quality. The company can put more resources into effective sourcing channels and perhaps drop some lagging recruiting strategies.
- (b) Variables tracked::: <sup>only one colon</sup>
- Applicant- Disposition
  - Employee - tenure
- (c) Variables not tracked::: <sup>only one colon</sup>
- Applicant - Referring Source
  - Applicant - Unique ID(Connection between applicant and employee data)
  - Employee - Performance metric
- (d) Method of analysis:
- Calculate source of hire (the percentage of hires entering the pipeline from each sourcing channel)

- Use mosaic plots to compare the percentage of dispositions for each sourcing channel
- Analyze the plots and understand whether the differences reflect biases in the recruitment process
- Filter out rejected/withdrawn candidates, use boxplots to visualize the performance metric for each sourcing channel
- Filter out rejected/withdrawn candidates, use boxplots to visualize tenure for each sourcing channel (Be mindful of newly employed sourcing channels)
- Segment data by job position, department, location, and repeat the process to examine whether the effectiveness of sourcing channels differ for job positions/departments/locations
- Consider grouping sourcing channels or ungrouping sourcing channels if differences/similarities are observed
- Linear regression model, tree-based models (decision tree, random forest, XGBoost): Use only sourcing channels to predict employee tenure and employee performance

## 8 How can I predict who will terminate among current employees?

(a) Importance: Answering this question will help us analyze the difference between different jobs/departments and come up with specific retention plans for those different jobs

(b) Variables tracked:

- Hire Date
- Term Date
- Job
- Hourly Rate
- Full/Part-time
- Shift
- Reason for termination

(c) Variables not tracked:

- Source of hiring
- The time the employee took the offer
- Work history
- Reason for leaving the last job

(d) Methods of analysis:

i. Tree-based Model:

```
library(MASS)
library(party)
library(rpart.plot)
# READ, CLEAN DATA
cd = openxlsx::read.xlsx("Merged_Employee.xlsx", sheet=1, detectDates = TRUE)
cd$`Termed?` = as.factor(cd$`Termed?`)
cd$Company = as.factor(cd$Company)
```

I do not think you should put raw R code here!!!

If you want to, make me a **\*\*separate\*\*** technical appendix with raw R code and English narrative so I can see what you did. **But not for IFDA.**

```

cd$Department.Name = as.factor(cd$Department.Name)
cd$Merged.Job = as.factor(cd$Merged.Job)
cd$Race.Code = as.factor(cd$Race.Code)
cd$Gender = as.factor(cd$Gender)
cd$Full.Part.Time = as.factor(cd$Full.Part.Time)
cd$Shift = as.factor(cd$Shift)
cd$Tenure.Range = as.factor(cd$Tenure.Range)
cd_terminated = cd%>% dplyr::select(-Is.Rehire,-Tenure.to.Today,
-Tenure.to.Termed.Date.Range,-Department.Name,-Term.Date,-Tenure.Range,
-Race.Code,-Gender,-X1,-Job.Title)
cd_tc = na.omit(cd_terminated)
summary(cd_tc)
levels(cd_tc$'Termed?') = c("No","Yes")

# TREE MODEL --> Termination as Y
seat_idx = sample(1:nrow(cd_tc), 583)
trn = cd_tc[seat_idx,]
tst = cd_tc[-seat_idx,]
output.tree <- rpart(
  'Termed?' ~.,
  data = trn,method = "class")
rpart.plot(output.tree)
plotcp(output.tree)
min_cp = output.tree$cptable[which.min(output.tree$cptable[, "xerror"]), "CP"]
min_cp
rpart_prune = prune(output.tree, cp = min_cp)
rpart.plot(rpart_prune)
predict(rpart_prune, tst, type = "prob")
confusionMatrix(predict(rpart_prune, tst, type = "class"), tst$'Termed?')

```

No raw R code please!

Unreadable by mere mortals....

Give some results and pictures instead

- A. From the decision tree model results, people who are more likely to stay (termination = No) at the company are those who have an hourly rate greater than 20; for hourly rate between 16 and 20, people in Other and Warehouse-other positions are more likely to stay (termination = No).
- B. From the confusion matrix, our model has an accuracy of 78% . For the 146 observations (people) used in the test set, the model correctly predicted whether or not somebody terminated 78% of the time.

ii. Logistic Regression Model:

```

# LOGISTIC REGRESSION MODEL --> Termination as Y
full.model <- glm('Termed?' ~Merged.Job +Hourly.Rate+Full.Part.Time+Shift,
data = trn, family = "binomial")
summary(full.model)
pred = as.factor(ifelse(predict(full.model,tst, type = "response")>=0.6,

```

no raw R code!

there needs to be a discussion of this for the readers. IFDA folks cannot possibly understand this without your copious help!

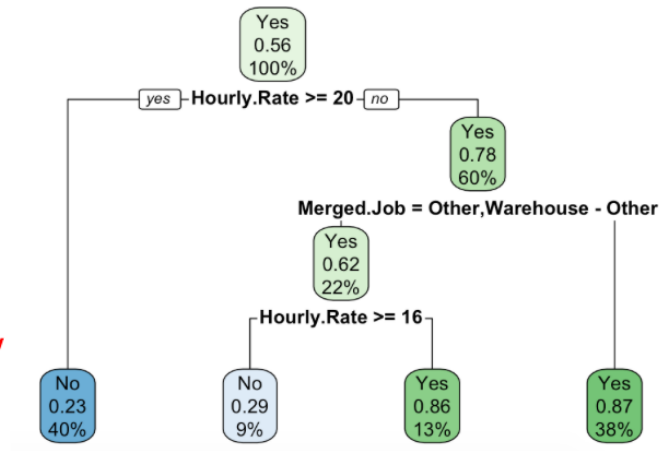


Figure 13: Tree Model Results

### Confusion Matrix and Statistics

also need discussion of this.

	Reference	
Prediction	No	Yes
No	56	17
Yes	14	59

Accuracy : 0.7877

Figure 14: Confusion Matrix using Decision Tree Model

```
"Yes", "No"))
levels(pred) = levels( tst$'Termed?')
levels( tst$'Termed?')
confusionMatrix(pred, tst$'Termed?')
```

no raw R code!

discuss this for readers you have to teach readers how to understand this stuff!

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	11.64754	1.23238	9.451	< 2e-16 ***
Merged.JobOther	-2.96472	0.46419	-6.387	1.69e-10 ***
Merged.JobWarehouse - Other	-2.31818	0.48877	-4.743	2.11e-06 ***
Merged.JobWarehouse - Selector	-0.82588	0.38418	-2.150	0.0316 *
Hourly.Rate	-0.55628	0.05623	-9.893	< 2e-16 ***
Full.Part.TimeP	-0.57837	0.53930	-1.072	0.2835
ShiftN	0.50806	0.29529	1.721	0.0853 .
---				

Figure 15: Logistic Regression Model Summary

- A. From the logistic regression model results, people with higher hourly rate, part time, day shift jobs are more likely to stay at the company (Termination = No). As for positions, people in Other and Warehouse -other positions are more likely to stay.

**explain for reader**

Prediction	Reference	
	No	Yes
No	60	16
Yes	10	60

Accuracy : 0.8219

Figure 16: Confusion Matrix using Logistic Regression Model

- B. With a 0.6 threshold(After estimating the probability for a certain class, the logistic regression model uses threshold to finalize its prediction of category for a observation. The model will classify a observation as a category when the estimated probability for that category passes the specified threshold), our model has an 82% accuracy from the confusion matrix. For the 146 observations (people) used in the test set, the model correctly predicted whether or not somebody terminated 82% of the time.
- C. Combining the results in those two models, we conclude that people in Other and warehouse positions, with higher hourly rate and work during day shifts are more likely to stay at the company.

**9 What is the number of qualified applicants per hire?**

- (a) Importance:<sup>(space)</sup> The company can establish a benchmark with this metric and tracks fluctuations effectively.
- (b) Variables tracked:: **only one colon**
- Requisition - Job title
  - Requisition - Location - Disposition
- (c) Variables not tracked:: **only one colon**
- Applicant - work history (years of experience)
  - Applicant - Unique ID(Connection between applicant and employee data)
- (d) Method of analysis:
- Average (number of applicants who passed an initial interview or phone screen for each requisition) for each job title / department / location

## 8 Individual Appendix

### Figures and Tables

Use the table and tabular commands for basic tables — see Table 7, for example. You can upload a figure (JPEG, PNG or PDF) using the project menu. To include it in your document, use the include graphics command as in the code for Figure 17 below.

Figure 17: An example image of a frog.

Item	Quantity
Candles	4
Fork handles	?

Table 7: An example table.

### Citations

LaTeX formats citations and references automatically using the bibliography records in your .bib file, which you can edit via the project menu. Use the cite command for an inline citation, like [? ], and the citep command for a citation in parentheses [? ].

### Mathematics

LaTeX is great at typesetting mathematics. Let  $X_1, X_2, \dots, X_n$  be a sequence of independent and identically distributed random variables with  $E[X_i] = \mu$  and  $\text{Var}[X_i] = \sigma^2 < \infty$ , and let

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_i^n X_i$$

denote their mean. Then as  $n$  approaches infinity, the random variables  $\sqrt{n}(S_n - \mu)$  converge in distribution to a normal  $\mathcal{N}(0, \sigma^2)$ .

### Lists

You can make lists with automatic numbering ...

1. Like this,
2. and like this.

... or bullet points ...

- Like this,
- and like this.

... or with words and descriptions ...

**Word** Definition

**Concept** Explanation

**Idea** Text

Obviously delete all this.