PAGE NUMBERS!!

CMU - IFDA Practicum: Preliminary Draft

Utilizing a Data-Driven Approach to Optimize Transportation & Warehouse Recruitment

nice title

Student Team: Echo Luan, Malik Khan, Yanxi Zhou, Xiaofan Zhu, Lanyi Xu

Faculty Advisors: Jamie McGovern, Brian Junker

IFDA Client: Annika Stennson

Date: 04/23/2021

Abstract or Exec Summary will be needed in final draft

I've written a lot below, but that's because this is a big, complex project that requires really clear descriptions and writing, and I want the final report to the client to be really useful for them. You have all done great work so far!

Introduction

The transportation and warehouse labor market is facing a scarce pool of talent that creates recruitment challenges for food distributors across the United States. There exists a gap in filling the demand for these positions with qualified candidates, which affects the distribution chain from the processing plant, the warehouses, and to the stores food distributors have invested a relationship with. How can we improve the existing talent acquisition process through a data-driven approach? Designing an efficient recruitment process can help food distributor services identify high-quality candidates while mitigating the number of unqualified candidates. This includes recording data from the following stages of the recruitment pipeline: Awareness & Interest, Selecting & Interviewing, and Hiring for a given job requisition. Companies can raise awareness about their brand through advertisements, direct sourcing, networking events, and other efforts that help passive and active applicants identify the company. Once the candidate identifies the food distributor company an organization they can work in, that is when the interest is developed. After, it is imperative that food distributors are able to select candidates most relevant 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 qualified candidate pays off. By identifying the data being tracked and not tracked from food distributor services, we aim to answer the main question posed above, and we will address the following questions:

1. Can we validate the hypothesis that the food distributor industry is faced with increasing competition for qualified candidates from other industries (i.e. eCommerce)?
2. How can we develop a data-driven framework to further our understanding of the candidate pool and improve recruitment efforts?

Citations for some of these assertions would really improve your credibility here.

I don't get this sentence.

as

Check this sequence of events with Annika or some-one else at IFDA

is this how the process actually works?

the work you highlight in this draft addresses this question. It might be useful to break this down into several sub-questions This might help you describe your methods and results better "for subquestion A we used logistic regression blah blah", "for subquestion B we used tree models blah blah", "for subquestions C and D we used EDA blah blah".... But don't say "subquestion A". Instead, repeat what the subquestion actually is (and similarly for other subquestions).

you seem to have forgotten this question, and/or looking at Bureau of Labor data, comparing it with IFDA data, etc.

We have to decide:

(a) include analysis and results for this question; or

(b) mention the question in the introduction but not as a main research question, and then return to it as "future work" in the discussion section

start the data section by introducing the 5 companies, anc generally how you got data from them.

per our most recent meeting, you will describe the process of collecting qualitative data from the companies in the Methods section, and then one part of the Results section will be a summary of the qualitative data; the detailed answers to the questionnaire questions will be in a separate appendix for each company

Data

you can omit this if you are not going to have the first research question above.

Data used for this report consists of two parts: public data obtained from Bureau of Labor Statistics (BLS) and private data obtained from five IFDA member companies. Client data will be used to conduct statistical analysis while public data will be used as a reference to the market. Yearly public data are retrieved from url (**attach as footnote)**: https://data.bls.gov/oes/#/home and monthly public data on economics condition are retrieved from url (**attach as footnote)**: https://www.bls.gov/news.release/jolts.t02.htm#jolts_table2.f.1. For 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, Employment RSE, Hourly Wage, Annual Wage, Interquartile Ranges for Hourly & Annual Wage. Client data are categorized to four sections: Employee Data, Job Data, Applicant Data and Requisition Data. **Table 1** shows the availability of these four types of data for each of the client companies. In addition to the four above-mentioned categories, we are also interested in Candidate Sourcing Data, which none of the companies were able to provide. More discussion on Candidate Sourcing Data will be given in the roadmap section. Please also note that companies are not anonymized at this stage. Special digits will be given to these companies to de-identify them in the final report.

new parag here

explain what this is

Generally, make your tables look more like the examples here:
https://texblog.org/2017/02/06/proper-tables-with-latex/
(these were done in LaTeX but you can do the same thing in msword!)

Also, single-space in tables generally looks better, and try not to break tables across pages.

Table 1: Client Data Collection on Important Variables

Client	Employee Data?	Applicant Data?	Job Data?	Requisition Data?
Saladino's	Y	Y (7/20/2019 - 1/10/2020	Y	Derived from the applicant data, not sure if there is just a set of ongoing open requisitions
Gordon Food Service	Y	Y (After 10/19/18)	N	Y

Good idea to include dates of dat
But make "Date Range Covered" be a separate column, and make sure the dates fit on a single line, for each company.

Martin Bros.	Y	N	Y- annual	N - need to see job description from the website
PFG	Y - annual	N	N	N
Ben E Keith	Y - only turnover data	N	Y	Y

Talk about the categories here and list all the possible categories, so the reader will know what they are looking at in Tables 2, 3, etc.

Table 2 shows variables that are common in all client companies and we will omit these variables when listing important variables for each client company.

Point out that there are only 3 variables common across all 5 companies...

Table 2: Share variables

Variable Name	Category	Description
Job Title	Employee Data	Job title
Department Name	Employee Data	Department job belongs to
Division Name	Employee Data	Division job is in (if any)

in Tables 3, 4, 5, 6, and 7,

Next, for each client data, we list important variables and their categories and descriptions in tables.

talk about why these are important and maybe give an example of an unimportant variable that you have omitted.

Table 3: Saladino's Data Description

Variable Name	Category	Description
Race	Employee Data	The race of the employee
Gender	Employee Data	The gender of the employee
Current Salary	Employee Data	Current hourly rate of pay
Hire Date	Employee Data	The date employee gets hired
Term Date	Employee Data	The date employee departs
Category	Employee Data	Voluntary/Involuntary
Reason for Termination	Employee Data	Reason of termination
Position Applied For	Applicant Data	The position applicant

It would be nice if you have some-thing to say about each of tables 3, 4, 5, 6 and 7...

		applied
Date Applied	Applicant Data	The date application filed
Disposition	Applicant Data	The recruitment status
Average Weight per Case	Job Data	Average weight in lb per case
Average Hours	Job Data	Average hours per week
Gross Number of Stops	Job Data	Average number of stops

Table 4: Gordon Food Service Data Description

Variable Name	**Category**	**Description**
Time in Position	Employee Data	Tenure in number of days
Is Rehire	Employee Data	Whether employee is rehired
Work Shift	Employee Data	Type I/II/III shift
Employee Type (Full/Part time)	Employee Data	Full time or part time
Hired Date	Employee Data	The date employee gets hired
Tenure Range	Employee Data	The time range tenure belongs
Termed Date	Employee Data	The date employee departs
Status	Applicant Data	Status of Application
Application	Applicant Data	The date application filed
X* Bin	Applicant Data	The date application reaches stage X
First Open	Req Data	The date requisition first open
# of Openings	Req Data	# of openings per requisition

# of Openings Remaining	Req Data	# of openings remaining
Part-time/Full-time	Req Data	Full time or part time
Status Hired	Req Data	Number of applicants hired

Table 5: Martin Bros. Data Summary

Variable Name	**Category**	**Description**
Date of Hire	Employee Data	The date employee gets hired
Term Date	Employee Data	The date employee departs
Hourly Rate	Employee Data	The hourly rate in dollars
Race Code	Employee Data	The race of the employee
Gender	Employee Data	The gender of the employee
Full/Part Time	Employee Data	Full time or part time
Shift	Employee Data	Night shift or day shift
Average Weight per Case	Job Data	Average weight in lb per case
Average Hours	Job Data	Average hours per week
Gross Average Pay	Job Data	Average annual payment

Table 6: Performance Food Group Data Summary

Variable Name	**Category**	**Description**
Job descriptions	Employee Data	Job descriptions
Insurance	Employee Data	Costs & Insurance Rate
Turnover	Employee Data	Turnover rate

Table 7: Ben E. Keith Data Summary

Variable Name	Category	Description
Total Termination	Employee Data	Distribution of termination
Employment Status	Employee Data	Status of employment
Termination Reason	Employee Data	Reason of termination
Turnover Rate	Employee Data	Turnover rate

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

The specific code and so forth that you used for the merging should be in a 'technical appendix'

The same procedure is also applied to the normalization of applicant disposition. We used five 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) groups 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)

Again, code for normalization should be in the same tech appx

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.

After mapping the shared variables of the IFDA members, we have 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 8: Description of Merged Applicant Data

Variable Name	**Description**
Company	The company of the data source
Position	The original position title that the applicant applied to
Merged Position	The merged job title group that the position was mapped to
Date Applied	Application date
Merged Disposition	The merged disposition (recruitment status+subject) of the applicant

Table 9: Description of Merged Employee Data

Variable Name	**Description**
Company	The company of the data source
Merged Job Title	The merged job title group
Department Name	The name of the department the job title belongs to
Date of Hire	The date when the employee was hired
Term Date	The date when the employee was terminated
Hourly Rate	The hourly rate in dollars

Race Code	The race of the employee
Gender	The gender of the employee
Full/Part Time	Full time or part time
Shift	Night shift or day shift
Is Rehire	Whether the employee was rehired
Tenure to Today	Tenure in number of days (only available for current employees)
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 (only available for terminated employees)

Based on the data that we received, most organizations collected warehouse and transportation employee data and key performance indicators 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 not tracked or not accessible by all clients include applicant sourcing data, applicant work history, requisition close date. There is also data that is not provided by most clients: applicant-level data and requisition-level data. (**-- further explained in the final draft WIP** after we receive responses from our clients about whether some data is not tracked or difficult to access)

spell out and indicate the abbreviation on first use

If the unavailable variables were tracked, we can use them to calculate KPIs 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

This is not quite the same term that you used before. Change this, or change your earlier usage, so that it is very consistent throughout the document

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. Another example would be the sourcing 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.

Method

Put qualitative methods first

To analyze datasets from five of our clients, we compared the public data 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 client data and merged the dataset, plotted various graphs to visualize the data and gathered insight. Aside from that, we also built some models to explore the relationships such as building decision trees and multiple regression models to predict employee tenure.

tell us what methods you used (logistic regression, trees, EDA, etc.) for which subques-tions.

Aside from analyzing the dataset, we also provided needs assessment and roadmap for our clients. We found out 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.

great

Results

State upfront that all of the analyses here are for the merged data, and that some examples of indivdual company analyses will be in the appendices.

Don't give your team history, but rather get right to the point:

The first question we tried to address was ... For this, we started with EDA (show EDA results) and refined our results with ... (whatever you did)...

Before constructing the roadmap, our team first conducted initial exploratory data analyses that sheds light on the recruitment challenge given the data we received from IFDA members. This helped us construct our Roadmap in which we provide supplemental information that can help companies take a deep dive into the initial plots we have created below.

Using the merged applicant data set, we drew a stacked bar plot that displays the disposition of applicants by the varying job positions.. From the graph (figure1.), we observe that the majority of the applicants leave applications incomplete, followed by withdrawals in interviews and screening. This might illustrate some problems in the portal system instead of the qualification of applicants. Since there is no available sourcing data provided from our clients, we will address how to better analyze this problem in our roadmap section.

It's fine (and good!) to talk about gaps in data and how filling in those gaps would improve the analysis.

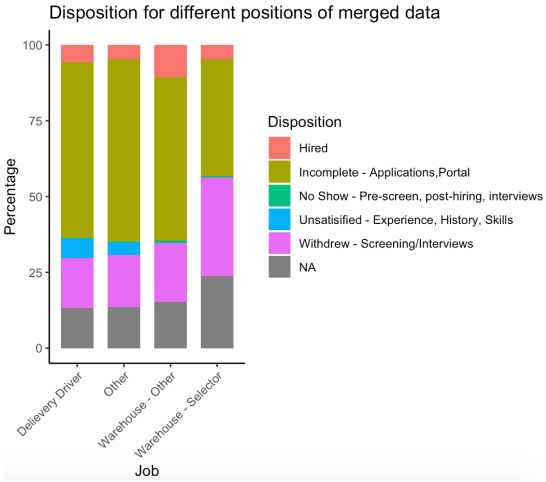


Figure 1: Disposition for different positions

We also web scraped employee reviews of order selector from Amazon and Sysco from Indeed and created a word cloud of frequently mentioned words. The larger the word size is, the more number of times it has been mentioned on the website. From Figure 2 & 3, we see that the reviews on Indeed often mentioned managers or management, feedback, time or hours and place or environment. Even though those words may not reflect the reviewers opinions exactly, it is apparent that they care more about the work environment, timing and management.

say up front why you did this. Say it here, and also include it as a subquestion above. Otherwise reader will not know why he/she should care about your word clouds.

what are you trying to say here? Be more clear/direct. It's fine to mention uncertainly, but be clear in what conclusion you are trying to draw.

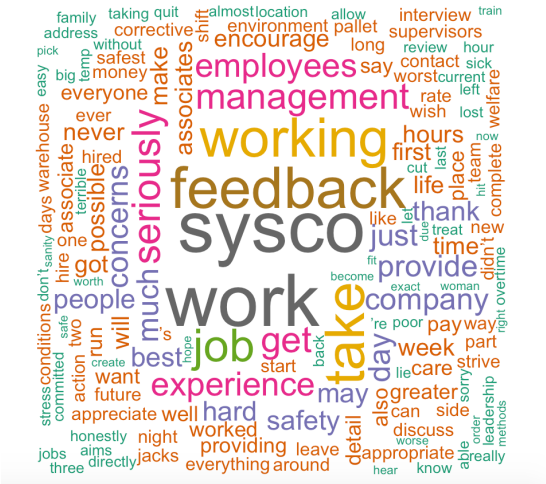


Figure 2: Sysco Wordcloud

Figure 3: Amazon Wordcloud

Since different companies used different values and metrics to keep track of employees and applicants information, there are more informative EDAs from each individual company instead of the merged data. We will show some examples of how analysis can be done within those specific datasets.

What specific question are you trying to address here?

State the question before diving into the analysis

Even though we did not have much available sourcing information, we did have termination and hiring data that is helpful to understand current and former employees in warehouse and transportation departments. Given reasons for job terminations (Figure 4 below -- **add footnote to refer to tables**), we observe that the dominant reasons for shipping departments were

terminations in

terminations for

Performance, No Call No Show, No Reason or Notice and the dominant reasons for drivers are Policy Violations, Disability or Death. Since the no call no show and no reason or notice are the major termination reasons, it would be beneficial if the company could investigate more in details about why employees did not show up.

capitalization and punctuation!

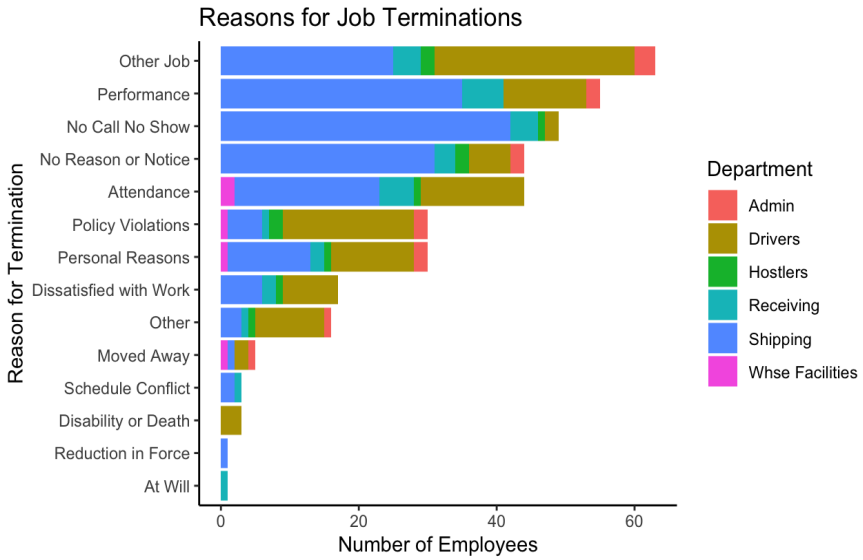


Figure 4: Reasons for Termination by Department

it will be obvious that it is useful, if it sheds light on a question of interest. you do not have to tell the reader it is useful.

say upfront what question the exploratory analysis addresses.

Another useful exploratory analysis we did was to visualize the turnover rate and hire rate for warehouse and transportation departments over the years to investigate if the problem has been consistent and how severe. From the hire and turnover rate plot (Figure 5 below), we observed that in general the hire rate is lower than the turnover rate over the years for both of these departments. Specifically, in 2018, the turnover rate reached its highest for the warehouse department but for transportation, its hire rate is actually slightly higher than the turnover rate in that year. However, since the company did not specify whether the employee would get retired, we include people who retired as termination as well. Therefore, we suggest companies indicate retired employees by adding a column or creating another sheet since termination is not the same as separation and combining those two information would make our results less accurate.

Figure cut off.

(TO)

Figure 5: Hire and Turnover rate over time

Those 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.

Roadmap

This sentence not needed. Instead you could simply restate the main research question and state that you are building a roadmap for future data collection to better answer this question in the future.

There is opportunity to recommend important variables that can help food distributors address the recruitment problem in the transportation and warehouse departments. We construct the following roadmap to suggest to our clients what variables we recommend tracking, why these variables are crucial in recruitment analytics, and how to use them to optimize the talent acquisition process after tracking them for a while. For every variable we suggest our clients to track, we list the questions we want to answer with the variable, the importance of the questions, and how to use the variables to optimize the recruitment pipeline. We also include a variable table at the beginning of the roadmap section, to give our clients a clear overview.

refer to "part 1" and "part 2" below and let reader know what to expect specifically in each part.

in Table X below

of which variables we are recommending to track for future analyses to improve employee recruitment.

Give this table a table number and a caption, like "Roadmap: Recommendations of variables to track" or something like that.

Variable	Tracked by companies	Explanation and Granularity	Issue relevant to
Job description	Yes	Include all the information in the job posting	Recruitment
Source of applications of all candidates	No	Accompanied with applicant information, Include the website name or other way of source, such as LinkedIn, referral	Recruitment

Add a column indicating what units the data are collected from, e.g.: applicants, job requisitions, employees, job categories, etc.

Different variables are collected on different units (eg the first row is collected on "job postings" and the second row is collected on "applicants"). Add a 5th column that indicates the units on which each variable is to be collected. This will help companies implement the recommendations.

The number of hits on the company website job page	No	Including the time and number of hit, such as Date: 04/21/2021, Hit:870	Recruitment
Application Date	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
Source of hiring	No	The source of all employees, such as LinkedIn, referral and etc	Recruitment
Application 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
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
Division	Yes	The division the employees are in or the applicants are applying for	Recruitment

# 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	Yes	The base of the applicants, such as Pittsburgh, PA list some other demographics that you'd like to see here, too	Recruitment
Requisition close date	No	Such as 04/21/2021	Recruitment
Applicant work history	No	As detailed as possible, include job title, company, and duration.	Recruitment
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
full/part-time	Yes	Include this in requisition data and employee data	Recruitment/retention
day/night shift	Yes	Include this in requisition data and employee data	Recruitment/retention
Employee demographics	No	the base of the employees, such as Pittsburgh, PA	Recruitment/retention
Reason for termination	Yes	As detailed as possible, indicate it is voluntary or involuntary. Such as voluntary termination due to transfer to another job maybe suggest a set of categories	Recruitment/retention
Offer acceptance rate	No	Should be tracked or calculated by the applicant data	Recruitment

collected on job posting, not applicant.

Satisfaction data from individual employees (surveys) vs data on companies or job categories from glassdoor.com will have rather different uses. Could stand to elaborate on which uses each type of data would be good for.

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 rage and headcount, Such as 19$/hr and 17 headcount	Recruitment
Requisition sourcing channels	No	The website or other channel that the job is been posted, such as LinkedIn or career fair	Recruitment
Interview Date	No	Such as “interview stage:phone interview, Date: 04/21/2021”	Recruitment
Status of Interview	No	As detailed as possible, such as phone interview -	recruitment
Hire Date	Yes	Such as 04/21/2021	Recruitment/retention
Termination Date	Yes	Such as 04/21/2021	Recruitment/retention
Hourly rate	Yes	Such as 19$/ hr, and if there is extra payment for certain situations, please include.	Recruitment/retention
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

In the first paragraph, describe the 4 parts in each item below:
a. Importance: why is the question improtant
b. variable tracked - variables that the companies already record
c. variables not tracked - variables that we recommend recording, that are not yet tracked by all companies
d. Method of analysis: some illustrated

Reason for leaving the last job	No	As detailed as possible, and can such as low payment, etc	Recruitment/retention
Interest conversion rate	No	It is defined as $\frac{\#\ of\ visitors\ who\ clicked\ 'Apply'}{\#\ of\ unique\ visitors\ to\ Job\ Posting}$ shouldn't this be inverted?	Recruitment
Applicant conversion rate	No	It is defined as $\frac{\#\ of\ applicans\ who\ completed\ applicatin}{\#\ of\ visitors\ who\ clicked\ the\ "apply"\ button}$	Recruitment

Part I : Action Questions

The action questions aim to provide guidance for company policies. Once the companies track the suggested data and analyze it through the method we suggested, they can change their approach in hiring and management, mitigating recruitment and retention issues.

1. How can we attract more candidates to apply?

- **a.** Importance: 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:
    1. Job descriptions (this can help us determine whether clear information is presented. If not, some candidates that could apply may miss the chance. Or, some candidates who apply will finally find out he or she does not fit this position, and it will cause a waste of time for both the company and the candidate)
- **c.** Variables Not tracked:
    1. Source of application of all candidates (if we can know the source of the application, we can see the proportion of candidates of each source, and determine which source to focus more on. Say, if most of the candidates apply for the job

through Linkedin, the company may post and update the job description more frequently. Also, HR can reach potential candidates through Linkedin.

2. The number of hit on the company website job page, including the hit on different job positions (This is also a source that we can track)

d. Method of analysis

I suggest you change the part d name to "examples of analyses"

1. 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.

2. 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. Assumption: Hiring status happens within the same period as requisition opening/application Although we have the date of hire, it is more reasonable to use requisition opening/application date to link back as "for requisitions opening at a certain period, what are their hiring rate". Assumption: Each position is subject to equal opportunity of the source of hiring.

b. Variables Tracked:

Make capitalization consistent throughout report. E.g. this is not consistent with item 1 above, nor with item 3 below, etc.

1. Application Date
2. Division, Requisition Open Date
3. Status

c. Variables Not Tracked:

1. Source of Hiring

d. Method of Analysis:

1. For each source of hiring, calculate the hire rate for each period of time.

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

I did not read the rest of the road map in detail. There are just one or two comments on the roadmap below.

a. Importance: 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:

1. Applicant disposition
2. Employee tenure

c. Variables not tracked:

1. Source of all applicants
2. Employee performance metric
3. Connection between applicant data and employee data

d. Methods of analysis

1. Exploratory data analysis
    a. Calculate source of hire (the percentage of hires entering the pipeline from each sourcing channel)
    b. Use mosaic plots to compare the percentage of dispositions for each sourcing channel
    c. Analyze the plots and understand whether the differences reflect biases in the recruitment process
    d. Filter out rejected/withdrawn candidates, use boxplots to visualize the performance metric for each sourcing channel
    e. Filter out rejected/withdrawn candidates, use boxplots to visualize tenure for each sourcing channel (Be mindful of newly employed sourcing channels)
    f. 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
    g. Consider grouping sourcing channels or ungrouping sourcing channels if differences/similarities are observed
2 Linear regression model, tree-based models (decision tree, random forest, XGBoost) **-- further explained in the final draft WIP**

a. Use only sourcing channels to predict employee tenure and employee performance

4. 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:

1. Application Date
2. Division
3. Requisition Open Date
4. # of Openings
5. Status

c. Variables Not Tracked:

1. Source of Hiring
2. Connection between applicant and employee data
3. Employee Performance metric

d. Method of Analysis

1. Group hired applicants based on time to apply 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.

5. What types of applicants tend to apply early/late?

a. Importance: 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:

1. Requisition open date
2. Job title
3. Requisition location
4. Applicant application date
5. Applicant demographics

c. Variables not tracked:

1. Requisition close date
2. Source of all applicants
3. Applicant work history (#years of experience)
4. The connection between requisition and applicant data

d. Methods of analysis

1. Exploratory data analysis
    a. Use applicant application date - requisition open date to obtain the response variable “application days”
    b. 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
    c. 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
    d. Use a scatter plot to explore the relationship between “application days” and applicant age
    e. Segment data by job positions, department, location, and repeat the process to examine whether early applicants and late applicants behave differently for these groupings.
    f. 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”.
2. Linear regression model, tree-based models (decision tree, random forest, XGBoost) **-- further explained in the final draft WIP**

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.

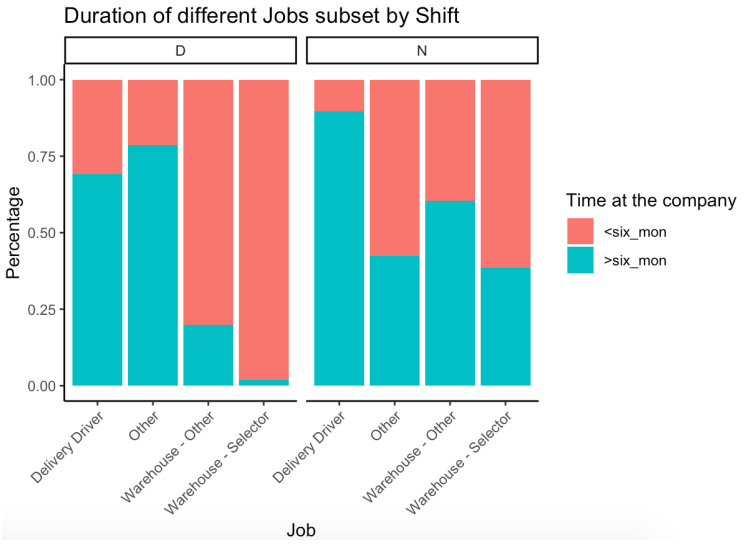
b. Variables tracked:

1. Job title
2. full/ part-time
3. Day/Night Shift
4. Reason for termination

c. Variables not tracked:

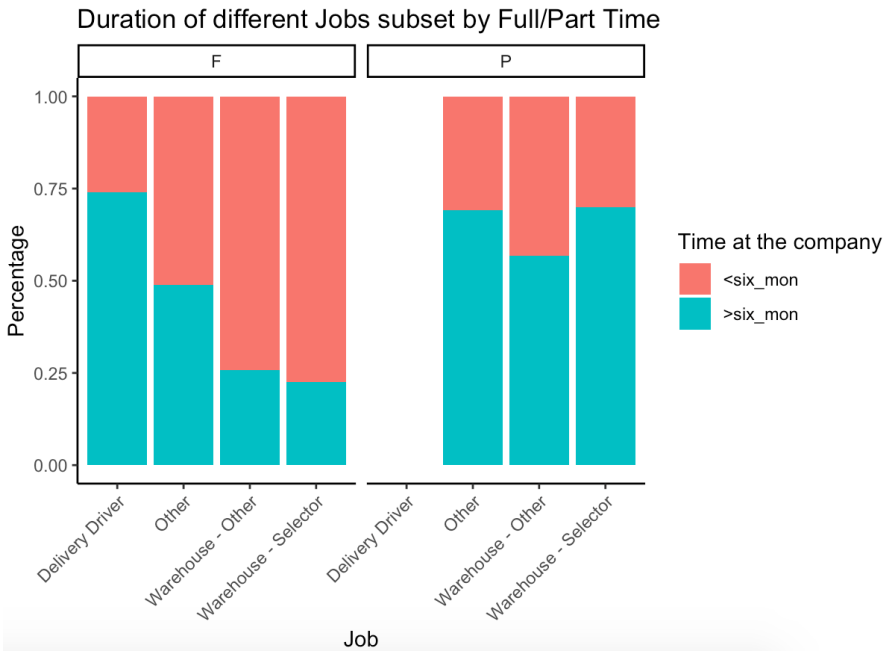
1. Source of hiring
2. Employee demographics
3. Job history
4. Offer acceptance rate
5. Employee satisfaction (either through surveys or websites like Glassdoor.com)

d. Methods of analysis:

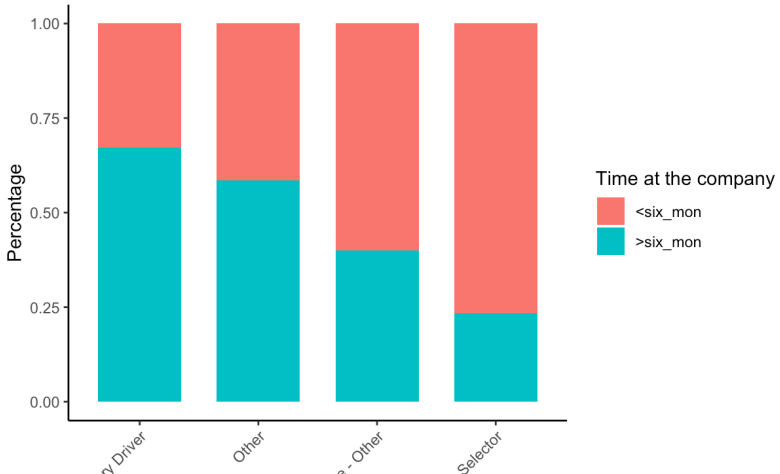


1. From this plot, we can see that for Day shift jobs, there are more people in delivery drivers and other job positions stay > 6 months. For night shift jobs, there are more people in delivery drivers and warehouse other positions that stay

> 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.



2. 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.

3. 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. What is the predicted time to fill for job openings in different seasons, positions, and locations?

7a. How is it likely to fill a position 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.
- **a.** Variables tracked:
    1. Requisition open date
    2. Requisition status (closed and filled / closed and not filled)
    3. Requisition job title
    4. Requisition location
- **b.** Variables not tracked:
    1. Requisition close date
    2. Requisition wage budget
    3. Requisition sourcing channels
- **c.** Methods of analysis
    1. Exploratory data analysis
        a. Calculate time to fill for all closed requisitions in at least a one-year period
        b. Use mosaic plots to compare requisitions that were filled and not filled for different positions and locations
        c. Use line plots to visualize the number of filled requisitions and the number of unfilled requisitions over requisition open time

d. Filter out requisitions that were not filled, use boxplots to visualize time to fill for each job position and location
e. Filter out requisitions that were not filled, use line plots to visualize average time to fill over requisition open time.
f. Filter out requisitions that were not filled, use boxplots to visualize time to fill for each month.
g. Segment data by job position, department, location, and repeat the process to examine possible effect of seasonality for different job positions/departments/locations
h. Filter out requisitions that were not filled, use scatter plot to visualize the relationship between requisition time to fill and requisition budget wage

2. Linear regression model, tree-based models (decision tree, random forest, XGBoost) **-- further explained in the final draft WIP**
    a. 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".

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

a. Importance: 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:
1. Talent Acquisition Interview Bin

c. Variables Not Tracked:
1. Interview Bin of Candidate
2. Interview Date
3. Status of Interview

d. Methods of analysis
1. With each column representing the different stages of the interview process, including hiring, obtain the dates of each stage for a given requisition.

2. 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.
3. Aggregate the differences between the offer date and application date (in days) for candidates who accepted the offer.
4. Aggregate the differences between the offer date and application date (in days) for candidates who did not accept the offer.
5. 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.
6. 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.

9. 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:

1. Hire Date
2. Term Date
3. Job
4. Hourly Rate
5. Full/Part-time
6. Shift
7. Reason for termination

c. Variables not tracked:

1. Source of hiring
2. The time the employee took the offer

3. Work history
4. Reason for leaving the last job

d. Methods of analysis

1. Logistic regression: Termed?` ~ Merged.Job + Hourly.Rate + Full.Part.Time + Shift **-- further explained in the final draft WIP**

```
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 .
---
```

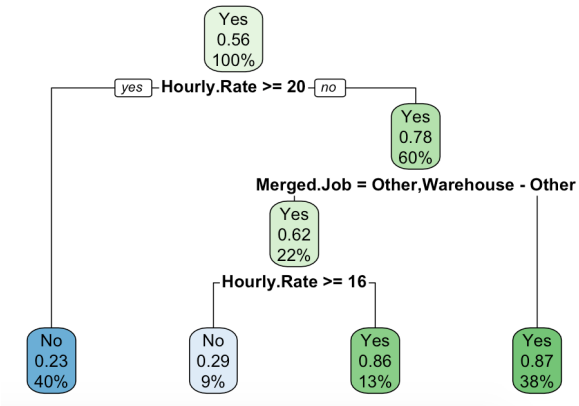
a. Higher Hourly rate, Part time job, day shift jobs are more likely to stay at the company (Termination = No). People in Other and Warehouse -other

```
          Reference
Prediction No Yes
       No  60  16
       Yes 10  60

               Accuracy : 0.8219
```

positions are more likely to stay.

b. With a 0.6 threshold, our model has an 82% accuracy. For the 146 observations (people) used in the test set, the model correctly predicted whether or not somebody terminated 82% of the time



2. Tree model **-- further explained in the final draft WIP**:
    a. People who are more likely to stay at the company are those who have an hourly rate > =20, and people in Other and Warehouse-other departments with an hourly rate between 16 and 20.
    b. Our model has a 78% accuracy. For the 146 observations (people) used in the test set, the model correctly predicted whether or not somebody terminated 78% of the time.

Part II: Web Analytics Questions

This part seems really incomplete. Can it be folded into the descriptive part below?

Web analytics are helpful to understand the conversion from awareness to interest, as well as from interest to application. By analyzing web-specific activity data, the recruitment team can have more perspectives on the digital recruiting landscape and increase the number of applicants. We address the concept of Awareness in the recruiting pipeline to quantify how many viewers are looking at the job postings online through the various sourcing channels, and mainly the clients' careers page. By observing how many views these job postings are receiving, we can derive Interest in these roles as viewers who start the application process, whether it is complete or incomplete. We can identify gaps between Interested candidates and completed applications by observing the pageviews or web traffic within each page of the application process online.

1. 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.

2. What is the conversion rate of each web sourcing channel?

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

3. What are the sources of the visitors on the website careers page?

a. Importance: 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
    1. Social media
    2. Job boards
    3. Paid advertisement
    4. Direct traffic (website visits that arrived on your site either by typing your website URL into a browser or through browser bookmarks)
    5. Organic traffic (visits from search engines)

Part III: Descriptive Questions

Descriptive questions can help us better understand the different aspects of recruitment and retention situations in the companies, and we may propose action questions and conduct actions based on the answers.

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

a. 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'd take to hire someone for a particular role

b. Variables Tracked:

1. Job Title
2. Req Open Date
3. Hire Date
4. Division

c. Method:

1. Filter out the applicants who were not hired. For each job role / job family, you can do the following:
    a. Locate the Application Date for each applicant who was hired.
    b. Locate the Hire date for each applicant who was hired.
    c. Compute the difference between Hire Date and Application Date for each applicant. This can be done in days as the unit of time.
    d. 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.

2. Is there seasonality in hiring and termination?

a. Importance: 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:

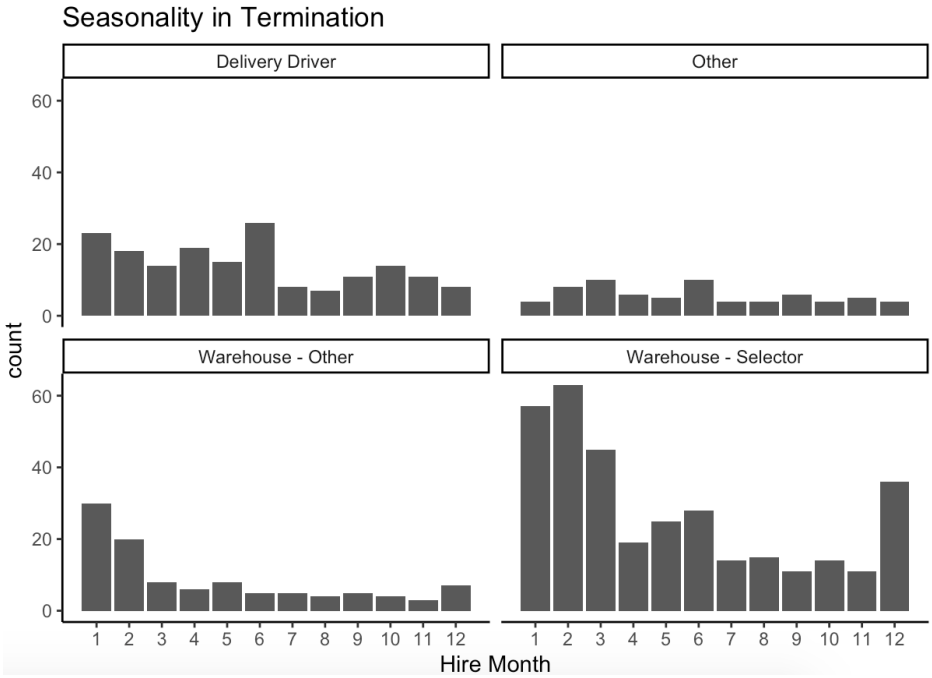
1. Termination Date
2. Hire Date

c. Variables not tracked:

1. Requisition Date
2. Reasons for Termination

d. Example: 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.



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 by the companies). We can segment our analysis into divisions for our client.
- **b.** Variables Tracked:
    1. Application Date
    2. Division
    3. Status
    4. Requisition Open Date
    5. # of Openings
    6. Hire Date
    7. Req ID
- **c.** Method:
    1. For the calculations below, one can either assume simultaneous recruiting & hiring or lag by the average time to hire throughout (needs extra calculation).
    2. Calculate the ratio of applicants to requisitions for each time period.
    3. Aggregate total # of openings and number hired for requisitions.

4. Calculate the ratio of number hired to # of openings for each time period.
5. Calculate the hire rate (hired/applicants) and compare to fill rate (hired/# of openings) for each time period.
6. Calculate the average time to hire for each time period.
7. Calculate the average time to fill (if requisition ever gets filled) for each time period.

4. Are requisitions that take longer to apply harder to fill (if so, by what percentage)?

- **a.** Importance: By examining the lagged time between requisition opening and application, one can have a general idea on the success of hiring before reaching the end of the screening process. With this information, measures can be taken at an earlier stage to avoid under filling positions.
- **b.** Variables Tracked:
    1. Application Date
    2. Division
    3. Applicant Status
    4. Req Open Date
    5. Job of Openings
    6. Hire Date

5. What is the number of qualified applicants per hire?

- **a.** Importance: The company can establish a benchmark with this metric and tracks fluctuations effectively.
- **b.** Variables tracked:
    1. Requisition job title
    2. Requisition location
    3. Applicant disposition
- **c.** Variables not tracked:
    1. Applicant work history (#years of experience)
    2. Connection between requisition and applicant data
- **d.** Methods of analysis

1. Average (number of applicants who passed an initial interview or phone screen for each requisition) for each job title / department / location

Discussion

Good start!

The participating IFDA members of this project have a variety of information tracked to address the recruitment problem. ~~In the Results section~~, Using data merged from all 5 companies, we saw that the majority of applicants leave their application incomplete, or they withdraw their candidacy. However, the reasons for understanding these decisions are unknown. In our Roadmap Section____, we suggest tracking data that could explain this decision, including the time taken to conduct interviews, specific requisitions that experience a higher applicant incompletion 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.

Briefly summarize the process of interviewing companies, collecting and merging data, and developing a data collection roadmap for the future (all in one paragraph).

(new paragraph)

Do you mean section, or part, or a specific question in your roadmap?

We also visualized the reasons for termination across departments for our clients using the merged data. The leading reasons for termination include 'Other Job', 'Performance', 'No Call - No Show', and 'No Reason - No Notice'. Food distributor companies can enhance their understanding behind this summary of termination reasons by conducting candidate feedback surveys during an associate's tenure so that 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 food distributor employer. Conversely, food distributor companies can learn more about what attracts a qualified associate to their company by learning about their previous experiences and consistently tracking the associate's experience during their tenure. More details on how to track these data points are described in the Roadmap Section ____.

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, the variance in these metrics by location and seasons, and how turnover can be reduced by targeting the segment of qualified candidates that exhibit longer tenure than expected. We suggest recommendations on the types of variables to track and how to analyze them to fill requisitions with qualified candidates in a timely manner in our Roadmap, Section____ .

The discussion is a work in progress. In the final report:

OK, good to know.

- We will further discuss how we might incorporate descriptive questions.
- We will determine the balance between the limitations of the project and what is included in the roadmap.
-We will wrap up our findings, generalizations, and conclusions about the project.

* References section missing. Did you read internal docs from IFDA about their work? Did you find some external references about the problem of hiring in this sector? What about general information on the Bureau of Labor website, or similar?

* Tech Appx missing. I'll give you credit for this for the moment, but we should talk about what the tech appx should look like for your final paper.