Table of Contents

EXECUTIVE SUMMARY

The purpose of this report is to help <REDACTED> Recruiting understand what factors are most relevant in filling job requisitions from companies that approach <REDACTED> for that purpose. By identifying the most relevant factors, <REDACTED> Recruiting can focus its resources on those factors, leading to a more optimal allocation of resources.

The report is broken down into five independent but complimentary analyses that contribute to the big question. Namely:

1. **Exploratory Data Analysis**, where the properties of individual variables are summarized without further use of formal modelling or testing techniques
2. **Correlation of Time Variables,** where the factors that affect time variables the most are explored
3. **Outcome Analysis on Requisitions,** where the factors that affect the outcome of a requisition the most are explored
4. **Survival Analysis,** where factors that affect the chances of a requisition being filled in different stages is explored

Recommendations for <REDACTED> are summarized at the end of the report, but they will be brought up at various parts of the report whenever it fits the flow.

Overall, we believe that <REDACTED>'s allocation resources on pipeline management and obtaining requisitions to its strength is near optimal, but <REDACTED> can improve upon its clients and candidates selection to obtain more optimal results.

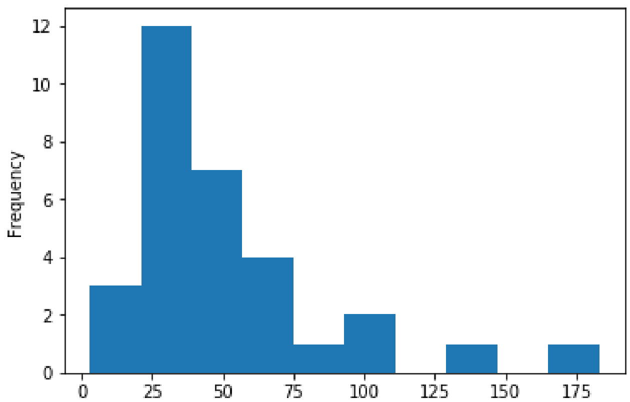
PART 1: EXPLORATORY DATA ANALYSIS

Exploratory Analysis of Time Variables

In this section, the following variables are considered:

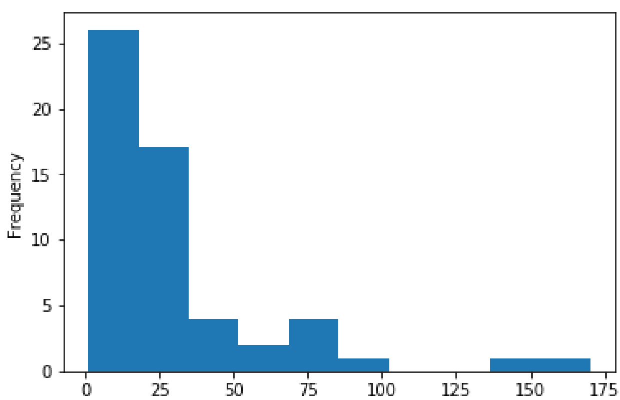
1. ***Time to Fill:*** The number of days it takes to fulfill a successful requisition
2. ***Time to Interview****:* The number of days it takes until the first interview
3. ***Decision Time:*** The number of days it takes for the first offer to be field after interview

The outputs below show the distribution and descriptive statistics of the three time variables, ***Time to Fill, Time to Interview*** and ***Decision Time.*** The unit of measurement is days.



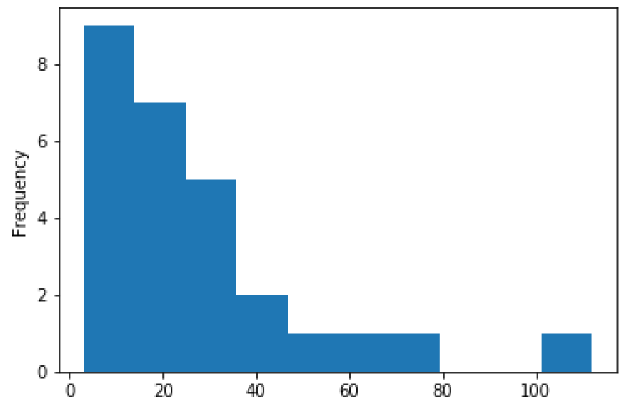
count	31.000000
mean	50.870968
std	38.099643
min	3.000000
25%	31.000000
50%	40.000000
75%	60.000000
max	183.000000

Exhibit 1: Distribution and Descriptive Statistics of ***Time to Fill***



count	56.000000
mean	30.071429
std	32.637322
min	1.000000
25%	9.750000
50%	18.000000
75%	32.250000
max	170.000000

Exhibit 2: Distribution and Descriptive Statistics of ***Time to Interview***



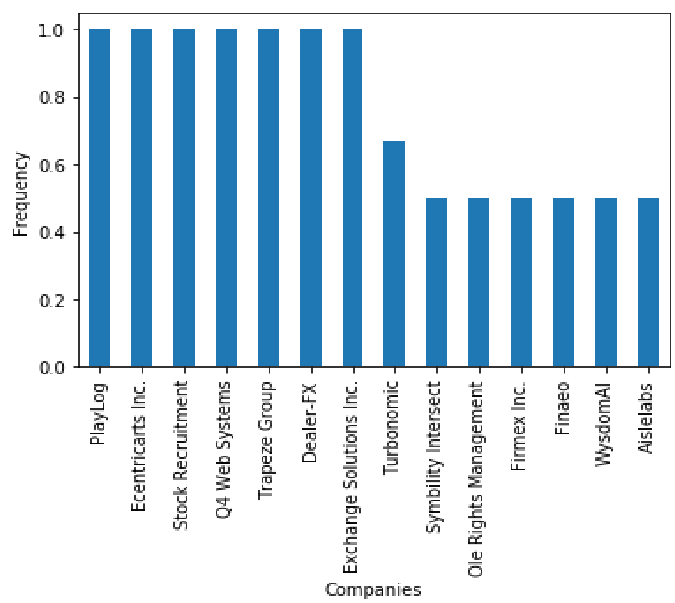

```
count      27.000000
mean       26.740741
std        24.588291
min         3.000000
25%        11.500000
50%        20.000000
75%        33.000000
max       112.000000
```

Exhibit 3: Distribution and Descriptive Statistics of ***Decision Time***

On average, a company that manages to fulfill a position with <REDACTED> takes **51** days from starting its engagement with <REDACTED> to do so. They take on average **30** days to select candidates for an interview and on average **27** days to field a job offer to one or more of the candidates.

Not all companies that interview candidates identified by <REDACTED> select <REDACTED> candidates. Of the 56 requisitions that had an interview, only 27 of them eventually selected a candidate whom they interviewed from <REDACTED> for various reasons, a rate of 48%.

What companies tend to hire candidates pushed by <REDACTED>? The table below shows all the companies in the dataset that have at least a 50% chance of hiring a <REDACTED> candidate if they interview at least one. The sample size is included for context.

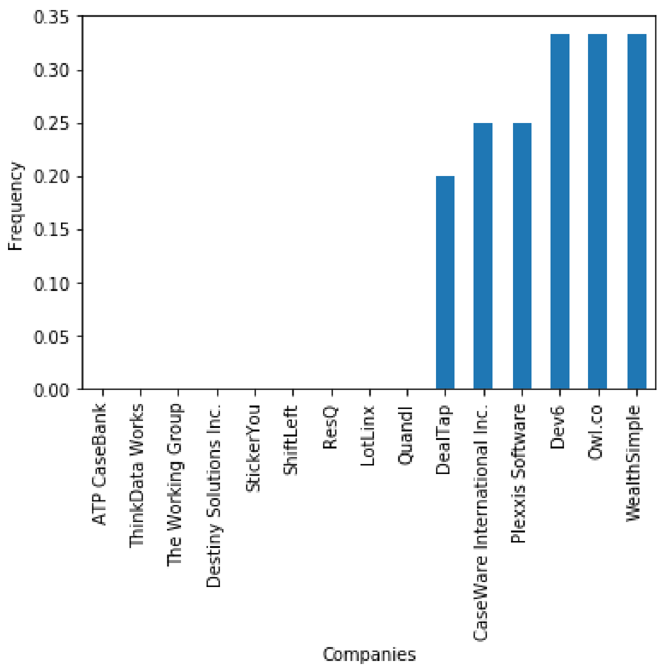


Company	% Hired	Sample
Ecentricarts	100%	2
Q4	100%	2
Stock Recruitment	100%	1
PlayLog	100%	1
Trapeze	100%	1
Dealer-FX	100%	1
Exchange Solutions	100%	1
Turbonomic	67%	3
Firmex	50%	10
Ole Rights	50%	2
Symbility	50%	2
Finadeo	50%	2
WysdomAI	50%	2
AisleLabs	50%	2

Exhibit 4: Companies with more than 50% hiring rate after interviews

Among the companies with 100% hiring rates following an interview with a <REDACTED>-backed candidate, most of them only have 1 engagement with <REDACTED>, and the results produced may not representative of their actual strike rate with <REDACTED>. Companies like **Firmex Inc.**, which has entrusted <REDACTED> repeatedly with job requisitions with an above-average hiring rate of 50%, have provided reliable matches with <REDACTED> candidates. Not accounting for other contexts, the evaluation methods that <REDACTED> used to select interview candidates for these companies appear to be working.

Conversely, what companies tend to reject candidates pushed by <REDACTED>? The table below shows all the companies in the dataset that have less than a 50% chance of hiring a <REDACTED> candidate if they interview at least one. The sample size is included for context.



Company	% Hired	Sample
ResQ	0%	6
Destiny Solutions	0%	4
ATP CaseBank	0%	3
Quandl	0%	2
StickerYou	0%	1
LotLinx	0%	1
The Working Group	0%	1
ThinkData Works	0%	1
DealTap	20%	5
CaseWare	25%	12
Plexxis	25%	4
WealthSimple	33%	6
Owl.co	33%	3
Dev6	33%	3

Exhibit 5: Companies with less than 50% hiring rate after interviews

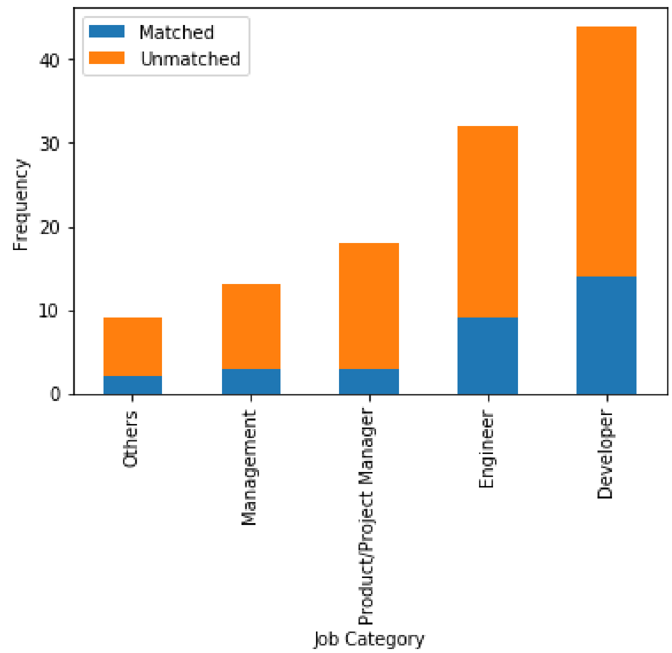
Companies such as **ResQ, Destiny Solutions** and **CaseWare International Inc.** have repeatedly approached <REDACTED> for interview candidates, but seldom if never select a candidate recommended by <REDACTED>. For these companies, <REDACTED>’s time and monetary investment in selecting candidates for their requisitions have often failed to pay off. Possible explanations are that these companies might have approached multiple recruitment firms or that <REDACTED>’s matching methods are not effective in identifying suitable candidates.

Exploratory Analysis on Job Variables

In this section, the main analysis will be focused on the variable **Job Category,** which is a custom variable that classifies the job titles into one of the following:

1. Project/Product Manager and Sales
2. Developer
3. Engineer
4. Management (Senior) roles
5. Others

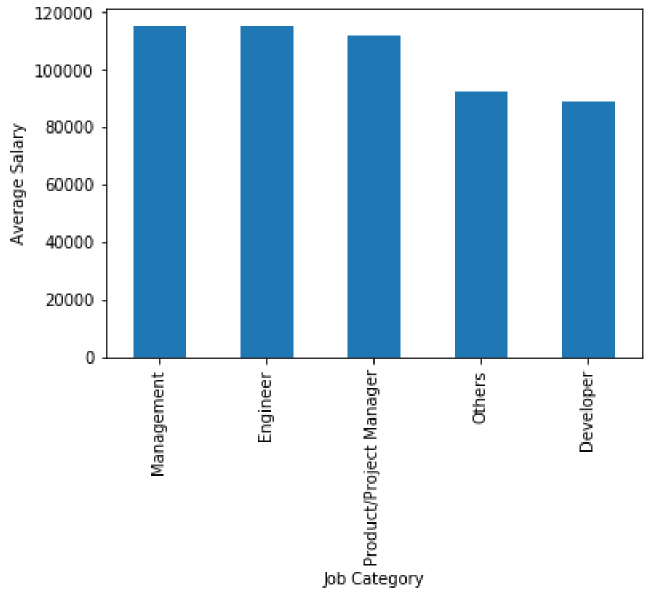
The distribution of the jobs is shown in **Exhibit 6**. Developer and Engineer roles are the most common types of requisitions as well as the most common types of requisitions that are filled.



Category	Matched	Unmatched
Developer	14	30
Engineer	9	23
Product Manager	3	15
Management	3	10
Others	2	7

Exhibit 6: Number of requisitions by job category

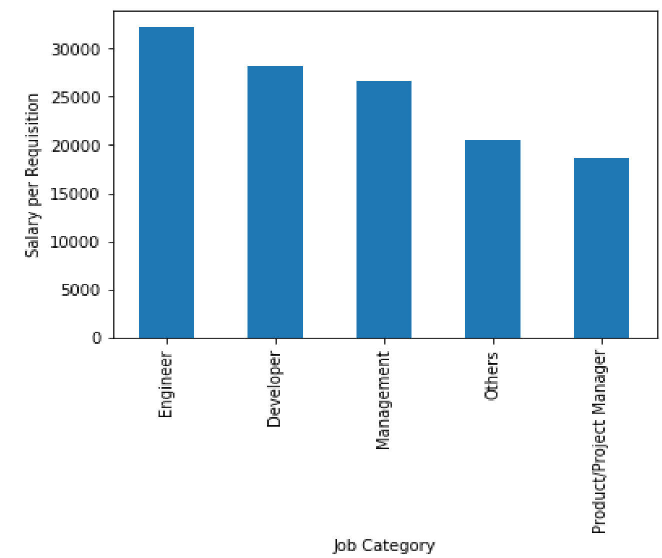
Which job category brings in the most revenue? As <REDACTED> charges a flat percentage on the annual salary of a requisition but the percentage value is not constant throughout the range of the dataset, the analysis would be more standardized by looking at the **Salary** instead of **<REDACTED> fee**. **Exhibit 7** shows that on average, a successful **Management** placement would generate the highest amount of revenue for <REDACTED> with **Engineer** positions not very far behind. This result assumes that the input needed to fulfill the position is constant among all categories, and that there is an equal probability of filling the position.



Category	Average Salary
Management	$115,333
Engineer	$114,888
Others	$111,666
Product Manager	$92,500
Developer	$88,792

Exhibit 7: Average Salary of Successful Requisitions

Another way to conduct this analysis is to look at the **Average Salary Per Requisition Request**. The key assumption required for using this measure over **Average Salary** is that regardless of whether a requisition is successful, <REDACTED> utilized similar amount of inputs. In other words, the results of this analysis consider the probability that a requisition in each category is successfully fulfilled. **Exhibit 8** shows that like before, **Engineer** jobs yield the highest salary (fee) per requisition. As the probability of successfully fulfilling a Management requisition is lower, the salary per requisition for Management jobs fall below those for Developer jobs. <REDACTED> management should compare the two analyses in this section, and take necessary actions based on whichever best reflects reality.



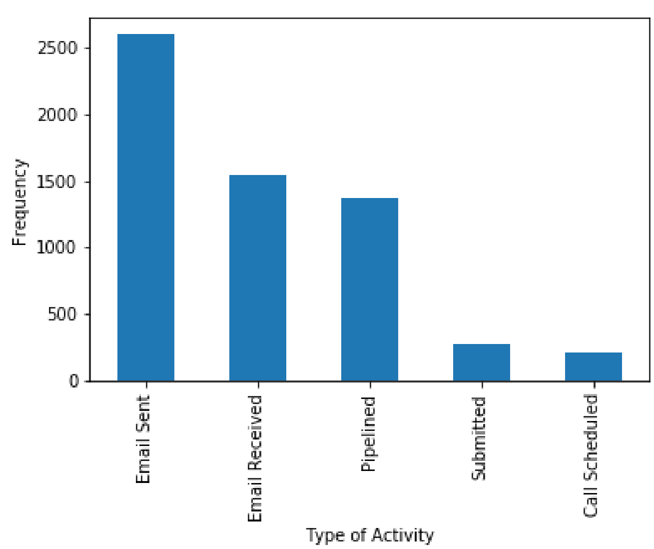
Category	Average Salary
Engineer	$32,312
Developer	$28,252
Management	$26,615
Others	$20,555
Product Manager	$18,611

Exhibit 8: Salary per Requisition by Category

Exploratory Analysis on Candidates

In this section, the analysis will focus on the individual candidates and their interaction with <REDACTED>. Data from the files “candidateprogress.csv” and “activities.csv” are the main sources.

Overall, only **30** candidates were matched with a job through <REDACTED> and **1208** candidates who had interactions with <REDACTED> did not end up with a job.



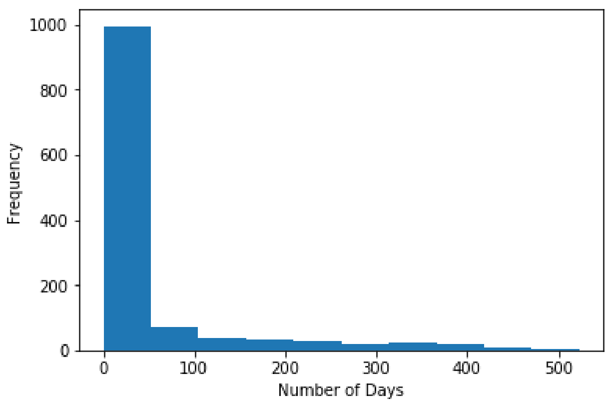
Activity	Unmatched	Matched
Email Sent	2277	323
Email Received	1234	309
Pipelined	1518	37
Submitted	253	48
Call Scheduled	191	20

Exhibit 8: Type of activities engaged by candidates with <REDACTED>

Exhibit 8 shows that the most prevalent type of activity engaged by candidates with <REDACTED> is when <REDACTED> sends an email to the candidates. Only slightly more than half of those emails are replied to. However, among candidates who have been matched, almost all emails sent are replied to. This may be an indication that **<REDACTED> may want to identify candidates who are more committed to the platform** before allocating more resources to them. The steep difference between the number under pipelined and submitted for the unmatched group does imply that only a small proportion of candidates who are pipelined for a job actually end up completing and submitting their application.

Additionally, among matched candidates, **the number of submitted applications are higher than the number of requisitions they are pipelined for.** <REDACTED> may want to inquire why this is the case. Is it a lack of clerical discipline, or did the candidates find the jobs through other methods and skipped the pipeline process?

Exhibit 9 in the next page shows the distribution of the number of days the candidates have been engaging with <REDACTED>.



Descriptive Stat	Unmatched	Matched
Count	1208	30
Mean	43.03	82.47
Std	92.51	110.36
Min	0	3
Max	523	435

Exhibit 9: Distribution of the days of engagement by candidates

On average, unmatched candidates engage with <REDACTED> for 43 days, while for the matched candidates it increases to 82.47 days. This result likely shows that many candidates do not rely upon <REDACTED> to find a job for them, and are using other sources in addition to <REDACTED>.

Most engagements with <REDACTED> last less than 50 days, however those who are matched with <REDACTED> takes a lot more than 50 days. Whenever <REDACTED> source a candidate, they should be budgeting more than 80 days from the start of their engagement in order for them to be placed.

PART 2: CORRELATION AND PREDICTION ON TIME VARIABLES

Connection to the Big Picture

The big question facing <REDACTED> is one of cost reduction. That is, how can <REDACTED> fulfill requisition requests from companies using the least amount of resources as possible. Under this context, **time** becomes a variable of interest. As the sayings go, “Time is money”. The longer the time spent working on a single requisition, the more resources <REDACTED> has to put in to fulfill a position and the lower the amount of time available to <REDACTED> to attend to other opportunities. Additionally, as organizations who approach <REDACTED> often seek candidates through other avenues, the longer it takes for <REDACTED> to fulfill a requisition, the higher the risk that the organization fulfills the position through an alternative method. As <REDACTED> only receives income if a requisition is fulfilled through them, the longer a requisition is active the higher the risk of <REDACTED> not receiving any income for their work. This analysis hopes to identify the biggest factors that affects fulfillment time. The idea is that if <REDACTED> focuses on those factors, they should be able to more efficiently distribute their resources.

Variables Definition

1. ***Time to Fill:*** The number of days it takes to fulfill a successful requisition
2. ***Time to Interview****:* The number of days it takes until the first interview
3. ***Decision Time:*** The number of days it takes for the first offer to be field after interview
4. ***Total Unique Candidates:*** The total number of candidates that was part of recruitment process. This is different from the number of candidates who are pipelined because there are candidates in the dataset who applied through the web and are not taken into the pipelined amount
5. ***Applicants:*** Number of candidates who filed an application for the position
6. ***Hires:*** Number of openings for that requisition
7. ***HasPhoneInterview:*** Whether or not a phone interview is part of the process
8. ***HasInPersonInterview:*** Whether or not an in-person interview is part of the process
9. ***HasVideoInterview:*** Whether or not a video interview is part of the process
10. ***Salary:*** The salary offered to the successful candidate of the requisition

Correlation Analysis

An analysis was conducted to identify the variables that are most correlated with **Time to Fill*.*** Correlation is represented on a scale of -1 to 1. If two variables have a correlation of -1, it means that if the value of one variable increases, the other decreases all the time. Conversely, a correlation of 1 implies that the two variables move in the same direction all the time. A correlation of 0 implies that the movement of variable has absolutely no effect on the other variable. **Exhibit 10** shows the variables of interest and their correlation coefficients with regards to **Time to Fill.**

Variable	Correlation Coefficient
Decision Time	0.5370
HasPhoneInterview	0.5215
Time to Interview	0.4156
Total Unique Candidates	0.2604
Applicants	0.2433
HasInPersonInterview	0.2144
Hires	0.0667
Salary	0.0513
HasVideoInterview	-0.1562

Exhibit 10: Correlation Coefficients of Explanatory Variables with Time to Fill

Relationship Between Time Variables

The following exhibits are graphs showing the graphical relationship between each of the variables listed in **Exhibit 11** with **Time to Fill**, as well as comments about the graphs. Several graphs are grouped together in the same exhibit to tell a more coherent story.

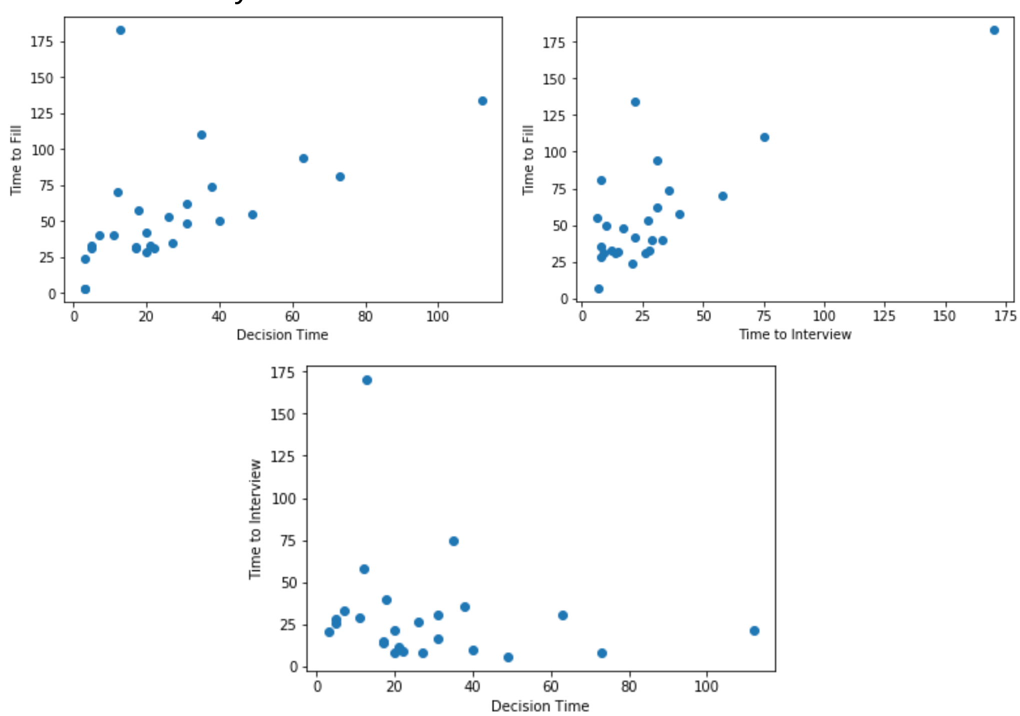


Exhibit 11: (Upper left, clockwise direction) Relationships between **Decision Time** and **Time to Fill**, **Time to Interview** and **Time to Fill,** and **Decision Time** and **Time to Interview**

As expected from the correlation coefficients, the **longer** it takes for a company to decide on an offer following an interview, the **longer** it takes to fill the requisition.

Equally unexpected, the **longer** it takes for a company to select candidates for an interview, the **longer** it takes to fill the requisition.

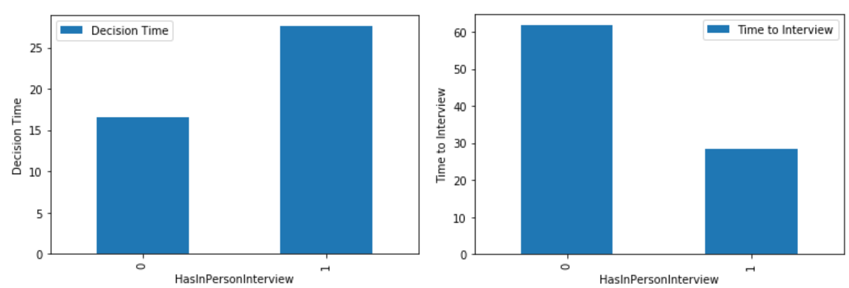
The third scatter plot produces an unexpected result: there is **no relationship** between Decision Time and Time to Interview. In other words, there are no companies that move inherently “slow” throughout the entire process. A company that takes longer to select candidates for interview may or may not take longer to select interviewed candidates to offer a job to.

The important takeaway from **Exhibit 11** is that <REDACTED> should not try to identify and give preference to companies who appear to select candidates for interview quickly, as it does not necessarily mean that the company in question will complete the requisition more quickly.

Relationship of Interview Types on Time Variables

Broadly speaking, companies that entrust <REDACTED> with requisition conduct three types of interviews: **In-person interviews, phone interviews** and **video interviews.** They are represented by the binary variables **HasInPersonInterview, HasPhoneInterview and HasVideoInterview,** which takes the value of 1 if that interview type is used for the requisition, and a value of 0 if not.

Exhibit 12 shows the effect of **in-person interviews** on the time variables.



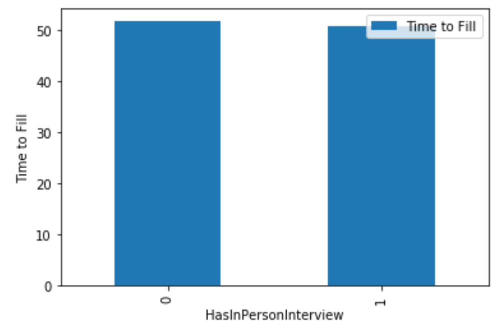


Exhibit 12: Effect of In-person interviews on time variables

Overall, having in-person interviews have a negligible effect on time to fill a requisition. However, requisitions that had an in-person interview took a shorter amount of time to select candidates to interview but a longer amount of time to make a decision as to which candidates to hire.

Exhibit 13 shows the effect of **phone interviews** on the time variables.

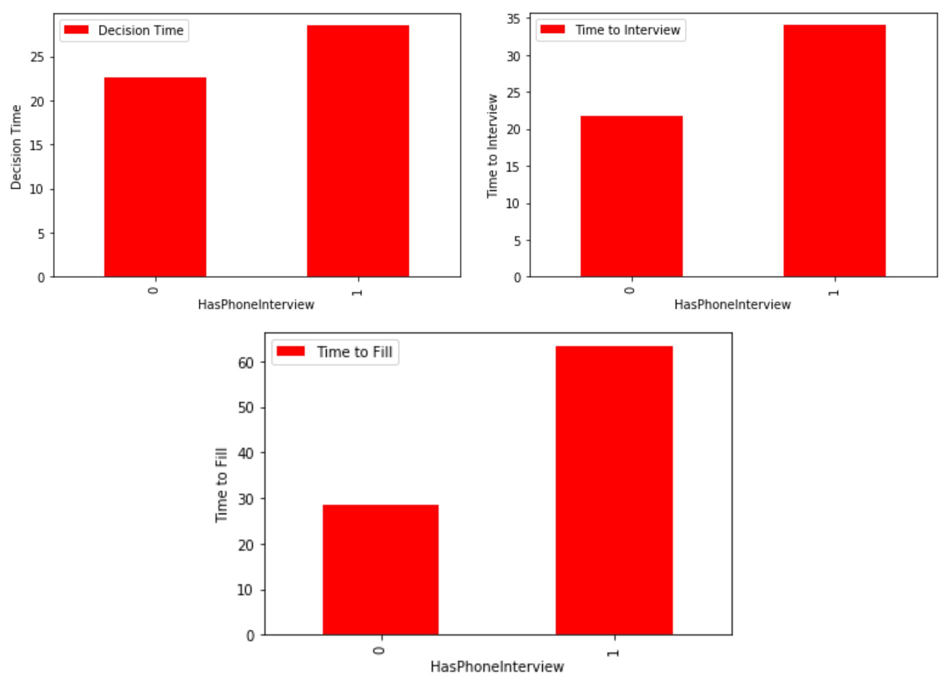


Exhibit 13: Effect of phone interviews on time variables

Overall, having phone interviews increases the time required to fill a requisition. Requisitions with a phone interview are associated with both longer times to select candidates for an interview and a longer time to make a decision following the interview. The cumulative effect results that on average, **companies take twice as long to fill a position if they use phone interviews.**

Exhibit 14 shows the effect of **video interviews** on the time variables. The exhibits are presented in the next page for better clarity and formatting.

Overall, having video interviews increases the time required to fill a requisition. Requisitions with a video interviews are associated with a much shorter decision time and a slightly longer time to interview.

As the most profound effect is found on requisitions with **phone interviews,** <REDACTED> should investigate why exactly do companies take longer to fill positions when they conduct a phone interview. One hypothesis could be that companies who feel the need to conduct a phone interview may have multiple rounds of selection or are more thorough with their process. To reduce the time cost incurred, <REDACTED> could look into what ways they can take over some of the processes so that their clients do not need to do it themselves.

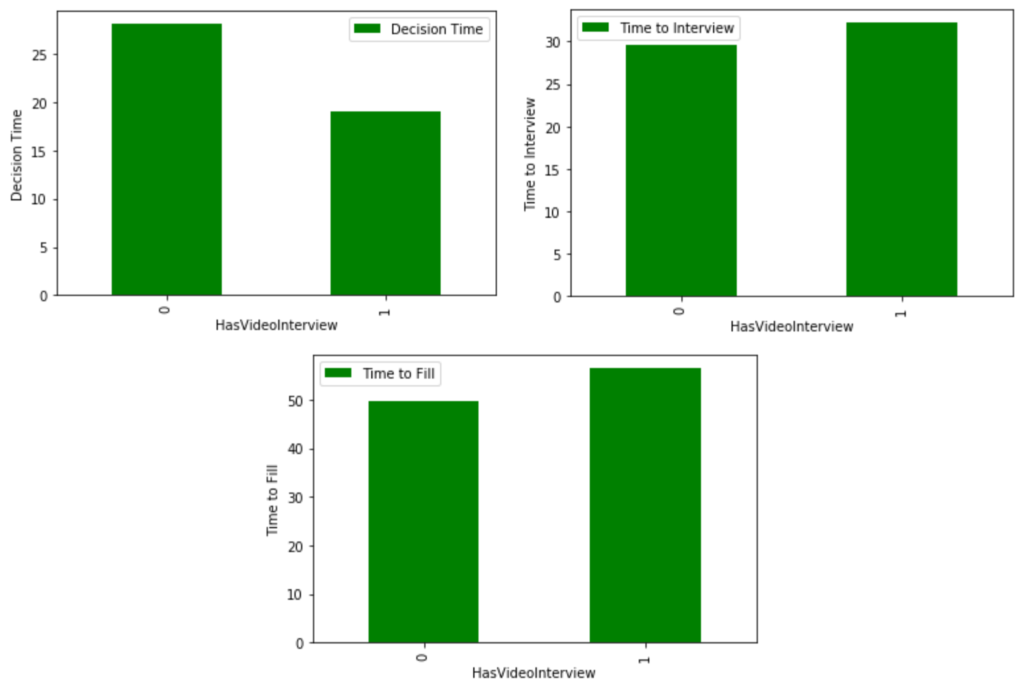


Exhibit 14: Effect of video interviews on time variables

Relationship of Quantity Variables with Time Variables

It is straightforward to assume that if the companies have to review more candidates, they will need more time to fulfill the requisitions. The exhibits in this section test those assumptions, and examine if the number of **unique candidates** (i.e. number of candidates involved at some stage during the process) and **applicants** have any effect on the time variables.

Exhibits 15, 16 and 17 on the next page shows the relationship between the quantity variables and the time variables. **None of the variables show a significant relationship with Decision Time and Time to Interview**, **but the number of unique candidates and applicants increases Time to Fill slightly.**

The takeaway from this exercise shouldn't be that <REDACTED> should reduce the number of candidates that they recommend to their client companies, but to understand that there is a trade-off between the number of candidates that they send to the companies and the time it takes for the requisition to be filled. In other words, it is important for <REDACTED> to determine the **right number of candidates** for each open position so that they don't end up delaying the process due to giving their clients too many choices to go through.

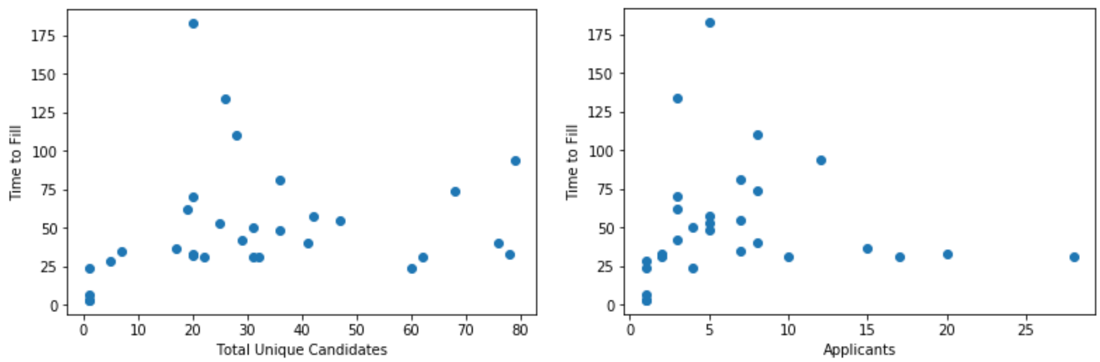


Exhibit 16: Relationship between quantity variables and Time to Fill

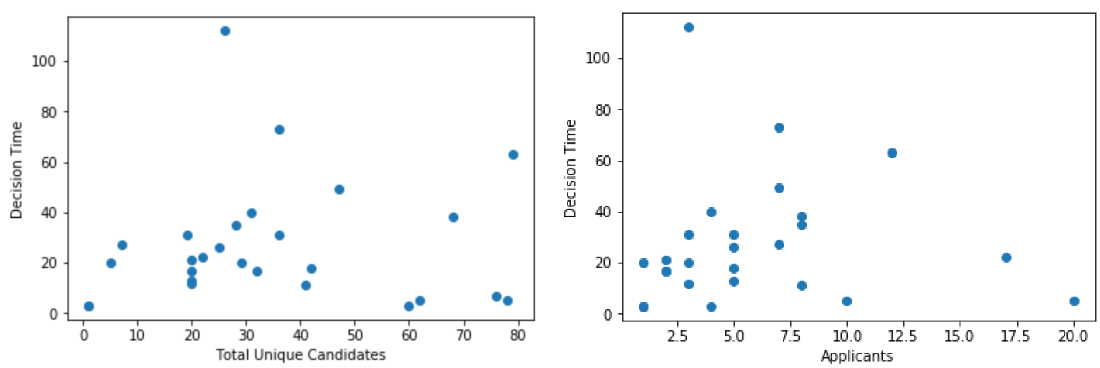


Exhibit 17: Relationship between quantity variables and Decision Time

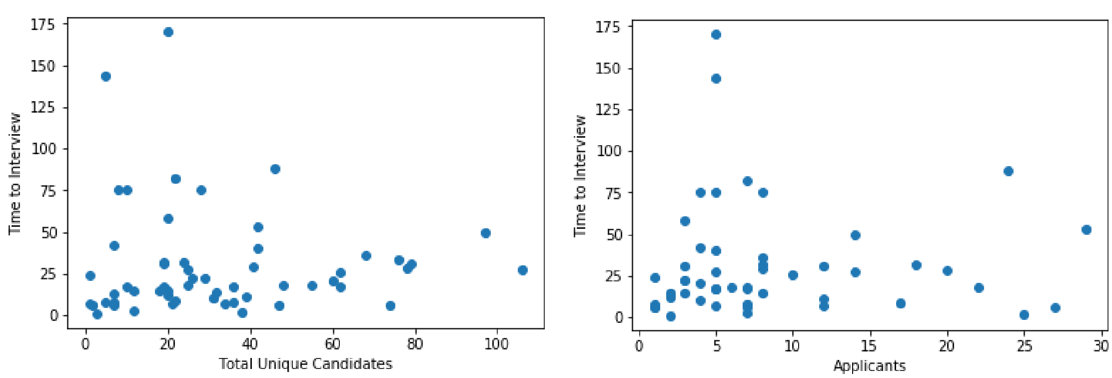


Exhibit 18: Relationship between quantity variables and Time to Interview

Prediction of Time Variables

Attempts were made to construct **linear models** that would predict the time variables. However, none of the models constructed using the variables stated in this section yielded results beyond what was presented above during the **correlation analysis.** Using those variables, all the models constructed shows that **most of the prediction on time variables are explained through randomness, or variables that are not available in the dataset.** Therefore, further details regarding the models are not included in this report, but the code used to construct those models are attached for reference.

Nevertheless, all the models constructed have one similarity in that **different interview types are by far the biggest determinants of time variables.**

Summary

- **Interview types** are the biggest determinants of time variables
    1. If a company does **in-person interviews,** they are more likely to **take longer to make a hiring decision** but **take less time to select candidates to interview**
    2. If a company does **phone interviews,** they are more likely to **take more time in general**
    3. If a company does **video interviews**, **they are more likely to take less time to make a hiring decision but more time overall**
- “Naturally slow” companies do not appear to exist
- The higher the **number of applicants**, the **longer it takes overall** to fill a requisition

PART 3: OUTCOME ANALYSIS ON REQUISITIONS

Connection to the Big Picture

The analysis in this part focuses on finding that **variables that most affect whether or not a requisition is filled.** If a variable is found to positively affect the chances of a requisition being filled, <REDACTED> should focus their efforts on that variable. If a variable is found to negatively affect the odds, <REDACTED> should try and avoid adding to that variable. Having a better understanding on what to focus could lead to reduction in efforts used to improve variables that do not contribute to better odds of fulfilling the requisition. This analysis focuses on three general variable types: **interview type, job category and quantity variables (e.g. number of applicants)**

Variables Definition

1. ***Total Unique Candidates:*** The total number of candidates that was part of recruitment process. This is different from the number of candidates who are pipelined because there are candidates in the dataset who applied through the web and are not taken into the pipelined amount
2. ***Applicants:*** Number of candidates who filed an application for the position
3. ***HasPhoneInterview:*** Whether or not a phone interview is part of the process
4. ***HasInPersonInterview:*** Whether or not an in-person interview is part of the process
5. ***HasVideoInterview:*** Whether or not a video interview is part of the process
6. ***IsDeveloper:*** Whether or not the position is a developer position
7. ***IsEngineer:*** Whether or not the position is an engineer position
8. ***IsProjectManager:*** Whether or not the position is a project manager position
9. ***IsManagement:*** Whether or not the position is a management position
10. ***IsFilled:*** Whether or not the requisition is filled

Correlation Analysis

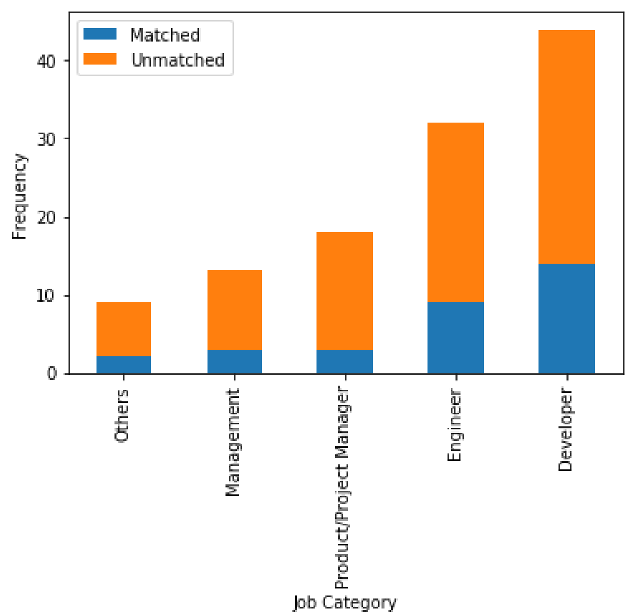
As with Part 2, an analysis was conducted to identify the variables that are most correlated with **IsFilled.** The correlation coefficients computed would present a basic intuition on what the relationship of each of the variables are like in relation to **IsFilled. Exhibit 19** shows the variables of interest and their correlation coefficients with regards to **IsFilled.** Overall, there were 116 requisitions, with 31 filled and 85 not filled.

Variable	Correlation Coefficient
HasInPersonInterview	0.4019
Total Unique Candidates	0.2204
HasPhoneInterview	0.2175
HasVideoInterview	0.1616
Applicants	0.1551
IsDeveloper	0.0900
IsEngineer	0.0195
IsManagement	-0.0293
IsProjectManager	-0.0974

Exhibit 19: Correlation Coefficients of Explanatory Variables with IsFilled

Relationship of Job Categories with IsFilled

Exhibit 20 shows the relationship between the 5 job categories: Developer, Engineers, Management, Project Manager and Others with the chances of the jobs being filled. Note that the charts below are exactly the same as **Exhibit 6** in Part 1.



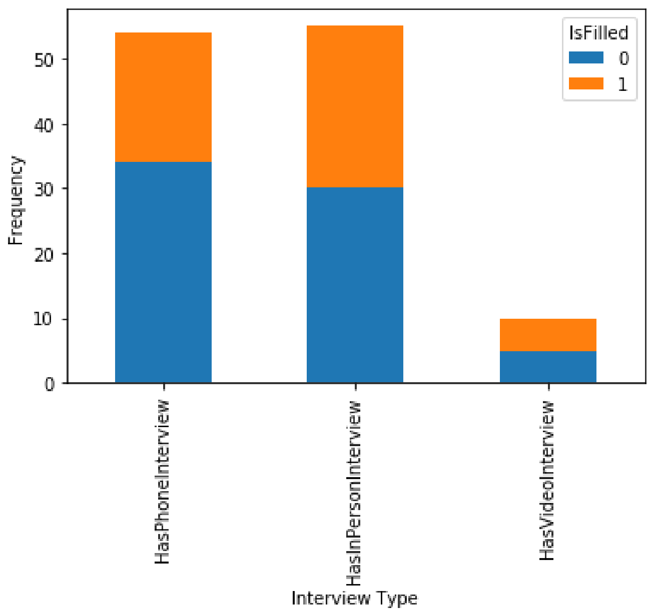
Category	Matched	Unmatched
Developer	14	30
Engineer	9	23
Product Manager	3	15
Management	3	10
Others	2	7

Exhibit 20: Number of requisitions by job category

Not only are **developer and engineer jobs most likely to be filled by <REDACTED>,** they are also make up the majority of all requisitions that <REDACTED> receives. Further specialization or activeness in sourcing these job categories over the others could help <REDACTED> utilize its resources more efficiently.

Relationship of Interview Types with IsFilled

Exhibit 21 shows the relationship between the 3 interview types: **Phone Interviews, In-Person Interviews and Video Interviews** with the chances of the jobs being filled.



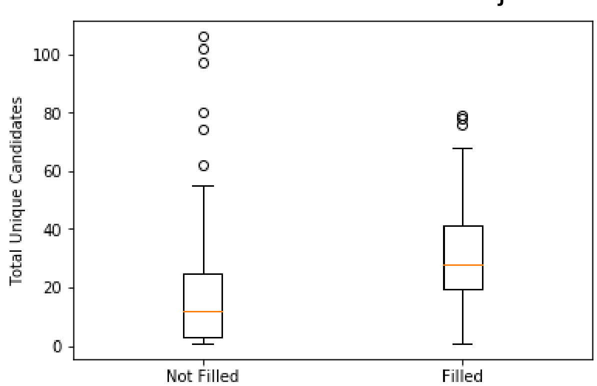
Interview Type	Filled	Not Filled
Phone	20	34
In Person	25	30
Video	5	5

Exhibit 21: Number of requisitions by interview type

Requisitions with video interviews are more likely to be filled, followed by in-person interviews and then phone interviews. Compared to the other two interview types, video interviews are less common, so the predictive capability in this analysis is less than the other two. Also note that it is possible for a requisition to have more than one type of interview.

Relationship of Quantity Variables with IsFilled

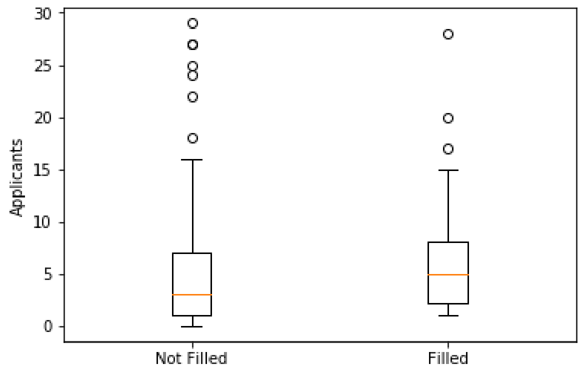
Exhibit 22 shows the relationship between the **total number of unique candidates** and the chances of the jobs being filled.



Descriptive Statistic	Not Filled	Filled
Count	85	31
Mean	19.84	31.65
Std	23.47	23.13
Min	1	1
25%	3	19.5
50%	12	28
75%	25	41.5
Max	106	79

Exhibit 22: Total Unique Candidates and IsFilled

Exhibit 23 shows the relationship between the **total number of applicants** and the chances of the jobs being filled. An outlier was removed to show that there is indeed a difference between both groups.



Descriptive Statistic	Not Filled	Filled
Count	85	31
Mean	5.69	8.77
Std	7	13.66
Min	0	1
25%	1	2.5
50%	3	5
75%	7	8
Max	29	74

Exhibit 23: Applicants and IsFilled

The higher the number of candidates that are involved with the requisition (at any stage), **the higher the chances of the requisition being filled.** The same positive correlation is seen in the **total number of applicants of a requisition.**

A possible hypothesis as to why this may be happening is that many of <REDACTED>'s lost requisitions could be early in the process, where <REDACTED> has not yet even sending the matching candidates to the company clients. It would be valuable to conduct further analysis on this hypothesis to test its validity.

Cluster Analysis on Quantity Variables

Can <REDACTED>'s clients be split into groups based on how many candidates are sourced in total, and how many applied to the role? If so, how does the groups affect the outcome of the requisition?

To answer the questions above, a cluster analysis was conducted to determine the groups (clusters). The clusters were determined based on just two features: **Total Unique Candidates** and **Applicants.** Requisitions that show similar features belong in the same cluster.

As per **Exhibit 24** in the next page**,** 4 clusters were constructed (the 1 outlier with high candidates count and high applicants count is in a cluster of its own as cluster 2).

Cluster	Total Unique Candidates	Applicants	% Candidates Applied	Colour
0	30.090909	8.090909	26.88%	Blue
1	7.523077	2.015385	26.60%	Purple
2	76.000000	74.000000	97.37%	Red
3	79.400000	10.200000	12.85%	Brown-Green
4	45.000000	26.000000	57.78%	Green

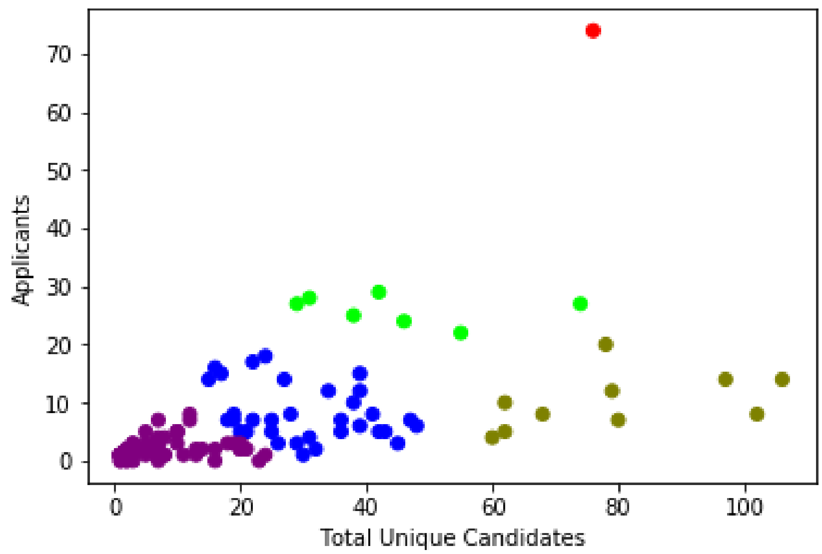


Exhibit 24: Clusters of requisitions based on Applicants and Total Unique Candidates

The clusters are:

0. Moderate candidates count, moderate number of applicants, moderate application percentage
1. Low candidates count, low number of applicants, moderate application percentage
2. Not a cluster as there is only one outlier
3. High candidates count, moderate number of applicants, low application percentage
4. Moderate candidates count, high number of applicants, high application percentage

Using the above clusters, an analysis was conducted to find out which of the clusters are associated with a higher chance of filling the requisition. The results are presented graphically in **Exhibit 25.**

Roughly half of the requisitions in **clusters 0 and 3** are filled, which is greater than the proportion in **clusters 1 and 4.** The results suggest that there may be an optimal number of <REDACTED> applicants **between 8 and 10** that maximizes the chances that the requisition will be filled by <REDACTED>. **Total number of candidates do not seem to matter (i.e. pipelined candidates).** As such, <REDACTED> may want to be **more selective** in the candidates that they pipeline, so that they don't waste resources recommending too many candidates that are either rejected by the client companies or have no interest in the job that they're pipelined for.

Requisitions with less than the optimal number of applicants (cluster 1) could be an indication that the companies do not have enough candidate choices, and more than that (cluster 4) could mean that the client companies received too many choices from <REDACTED>. Cluster 4 requisitions may also be the more competitive ones that attracted more applicants.

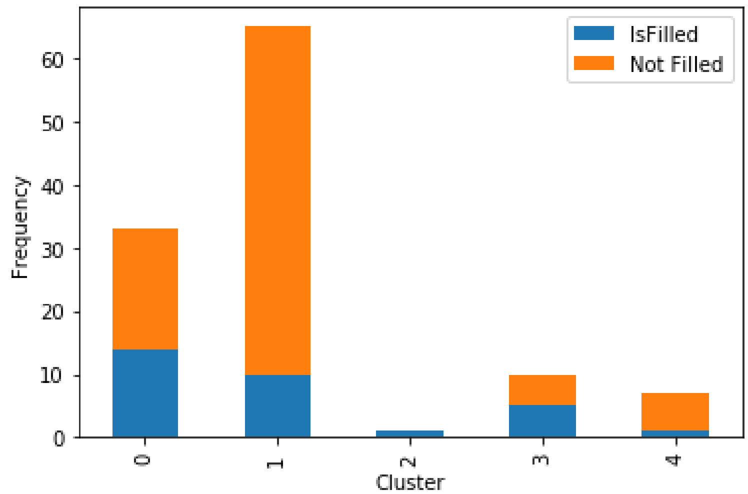


Exhibit 25: Result of requisitions by cluster

Logistic Regression Model

Using the **variables mentioned in Part 3 only,** we attempted to fit a logistic regression model that if successful, would be able to accurately predict the odds of success for any given requisition given the same variables. The results of the model are presented in the classification report in **Exhibit 26** below.

	precision	recall	f1-score	support
0	0.79	0.86	0.83	36
1	0.38	0.27	0.32	11
micro avg	0.72	0.72	0.72	47
macro avg	0.58	0.57	0.57	47
weighted avg	0.70	0.72	0.71	47

Exhibit 26: Classification of Logistic Regression

Overall, the model has an accuracy score of 0.72, meaning that the prediction given by the model is accurate 72% of the time, which looks good on surface. However, this is misleading. The model is able to **correctly classify unfilled**

requisitions 79% of the time, but only able to predict the requisitions that will be filled 38% of the time. Considering that 73% of requisitions actually go unfilled compared to 27% that fills, this regression model is only **slightly better** than randomly assigning the outcome to a requisition based on their probabilities.

In order to accurately predict the outcome of a given requisition, <REDACTED> may need to conduct a similar analysis as the one in this section, but with different variables. For example, the location of the company making the requisition, how <REDACTED> acquired those requisition, and more detailed exchanges between the company and <REDACTED> could all help in generating a working predictive model. **The present model has little to no predictive value.**

The simple reason why this predictive method is not useful is that among the data that was given, most of the requisitions are not filled. The method is most accurate when there is about a 50-50 split.

Summary

- **Developer** jobs are more likely to be filled than other job types
- Companies that conduct **phone interviews** are less likely to fill their requisition with <REDACTED>-sponsored candidates than those who don’t conduct phone interviews.
- Requisitions with **about 8 to 10 applicants** are more likely to be filled.

PART 4: SURVIVAL ANALYSIS ON CANDIDATES

Survival analysis is a statistical technique that measures the probability of a certain event occurring. Mostly commonly used for medical studies, survival analysis can be used to model the **odds of a candidate not being matched at a particular time** (e.g. staying in contact with <REDACTED> after 60 days of establishing first contact). Using those odds, the effect of each of the variables on the survival odds were examined.

Analysis Design

The data from ‘activities.csv’ and ‘candidateprogress.csv’ were rearranged and combined into one file that mainly has the following types of variables:

1. Number of **interviews** of each type
2. Number of **interaction** (e.g. Email sent, email received) with <REDACTED> of each type
3. Number of **requisitions** pipelined, submitted, and interviewed for.
4. Days between first interaction with <REDACTED> and last interaction with <REDACTED> (“Time”)
5. Whether or not the candidate is placed.

All the variables were run in **R** using the **Cox proportional hazards procedure.** Variables were gradually removed from the model until 4 of the ones that best model survival odds remain. The resulting **survival plot** using the variables **Emails Sent to <REDACTED>, Number of In-Person Interviews, Number of Video Interviews** and **Number of Pipelined Requisitions** is shown as **Exhibit 27** in the next page.

Survival Analysis Results

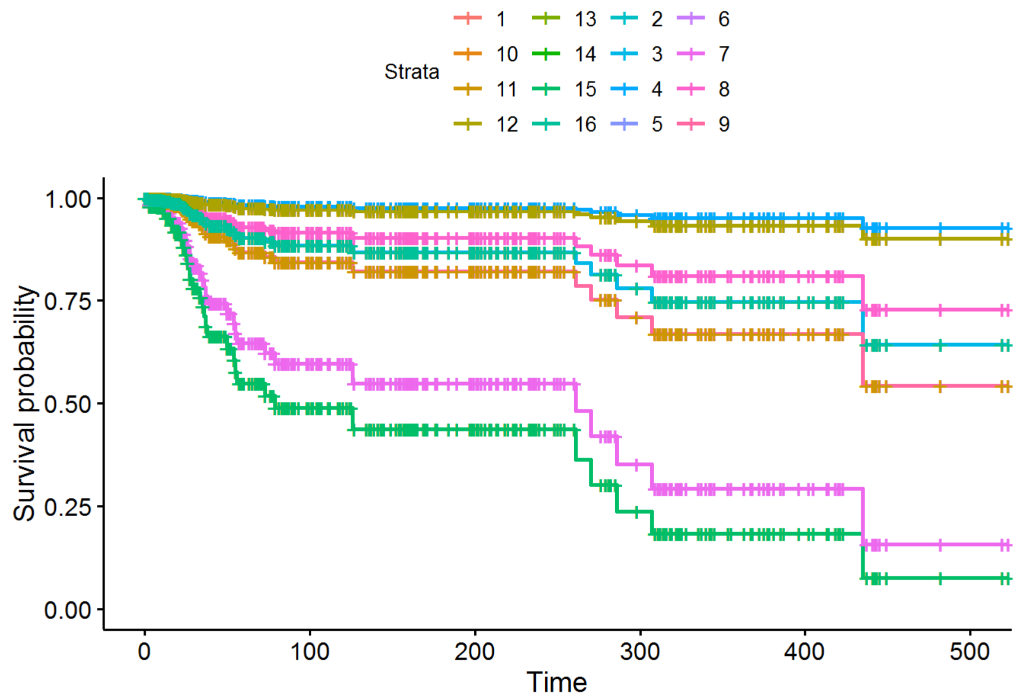


Exhibit 27: Probability of candidates not obtaining a job over time

Strata	Emails Sent to <REDACTED>	Number of In-Person Interviews	Number of Video Interviews	Pipelined
1	5	0	0	2
2	5	0	0	4
3	5	0	1	2
4	5	0	1	4
5	5	1	0	2
6	5	1	0	4
7	5	1	1	2
8	5	1	1	4
9	10	0	0	2
10	10	0	0	4
11	10	0	1	2
12	10	0	1	4
13	10	1	0	2
14	10	1	0	4
15	10	1	1	2
16	10	1	1	4

Exhibit 28: Strata Inputs

Interpretation

The higher a survival curve is on the chart, the less likely a candidate is going to get matched over time. Based on the chart, here's the relationships in the order of magnitude:

1. The **more** the number of jobs a candidate is pipelined for, the **greater** the chance of a match.
2. The **more** the number of emails sent to <REDACTED>, the **greater** the chance of a match **eventually.** However, in the early phase, the odds of a match is **greater** if the number of emails sent are lower.
3. The **more** the number of in-person interviews a candidate has, the **greater** the chance of a match.
4. The **more** the number of video interviews a candidate has, the **greater** the chance of a match.

Surprisingly, as a direct contradiction to a previous analysis, **in-person interviews** appear to be a better predictor of match from the candidates' perspective over **video interviews.**

Also, consistent with the perspective from the companies, the **number of jobs a candidate is pipelined for is more important than the number of applications or the number of interviews he/she gets.**

RECOMMENDATIONS

The recommendations in this section are not exhaustive or include all of the recommendations or points that are brought up during the body of the report. These are some points that we feel are the **most important** based on our analysis.

Regarding Clients

Problem: From **Part 1, ResQ, Destiny Solutions, and CaseWare** are among the companies with the worst matching odds with <REDACTED>, but they each have many requisitions with <REDACTED>.

Recommendation: <REDACTED> should sit down with these clients and the other clients in **Exhibit 5** to discuss the reasons why they are not satisfied with the candidates that <REDACTED> sent them. It could be due to internal reasons such that <REDACTED> is not matching candidates suitable for their needs, or external reasons such as these companies tend to cast a wide net and have multiple avenues to hire new employees. After seeking the responses, <REDACTED> would have a better idea on how to proceed. Also, <REDACTED> should **actively seek out opportunities with companies with high matching odds.**

Regarding Job Types

From **Part 1,** <REDACTED> is doing a good job at matching and obtaining requisitions on the job types (**engineers and programmers**) that they are best at matching and have the highest average salary (and in turn fees). <REDACTED> should continue to focus on these two types of jobs.

Regarding Candidates

Problem: From both **Part 1** and **Part 4**, the biggest problem regarding candidates is that they are not submitting job applications after being pipelined. This is a problem, as <REDACTED> spent resources screening the candidates and pipelined them anticipating that they will submit the application.

Recommendation: <REDACTED> should review their job matching process to see if the candidates are being matched with jobs they are not interested in, or if companies are being matched with candidates that they are not a good fit in. They should consider stricter rules regarding applying to a job after being pipelined for it.

Regarding Interviews

Problem 1: From **Part 2,** requisitions involving **video and phone interviews** take longer to fill, all else equal.

Recommendation: Find out the possible reasons (e.g. why those interviews are needed to begin with). If <REDACTED> is able to facilitate part of why those interviews are needed, they should do.

Problem 2: From **Part 3,** requisitions involving **phone interviews** are less likely to be filled, all else equal

Recommendation: Find out the possible reasons (e.g. why do a phone interview and at which stage the interview is typically for). Suggest that if possible, clients replace those interviews with video interviews (which are more likely to be filled) using applications like Skype or Zoom. <REDACTED> could facilitate if necessary, as it would be a good use of <REDACTED>’s time

Regarding Processes

Problem: From **Part 3,** <REDACTED> are losing requisitions even before pipelining candidates

Recommendation: <REDACTED> should focus on their early-stage execution. Consider fast-tracking existing candidates that have gone through screening to new requisitions as early as possible.

Regarding Pipeline

From **Part 3,** number of **pipelined candidates** do not matter as much as the **number of applicants. 8 to 10 applicants** per requisition is optimal. In many occasions, when many candidates are pipelined (well over 50), the matching odds are the same as when 30 candidates are pipelined. A high application rate from pipeline **could be an indication that the job is very competitive**, and thus <REDACTED> is less likely to fill that requisition.

LIMITATIONS

The analysis is based on a dataset that is very small (compared to the usual requisite datasets) and contains a disproportionate amount of requisitions that are not filled and candidates that are not matched. As a result, the results are not robust to randomness, and may not be representative of real-life operations. It should only be used as a reference and not as facts.