ADVANCED BUSINESS ANALYTICS

HIRE HEROES – CLIENT MANAGEMENT

Kishore Raj
Sai Priya V
Spoorthy Ravi
Subiksha Srivatsav
Venkatesh U

Table of Contents

Company Background:

Hire Heroes USA provides free, expert career coaching and job sourcing to more than a thousand transitioning U.S. military members, veterans and military spouses each week. As a national non-profit, their programs are funded exclusively by private grants and public donations. Hire Heroes USA maintains a singular focus on employment. Quite literally, veteran and military spouse employment is all they do, and they have been doing it with best-in-class results for more than a decade. Mission of the company is empowering U.S. military members, veterans and military spouses to succeed in the civilian workforce. Vision of the company is Be the Nation's preferred veteran employment service organization through a relentless focus on personalized career coaching that improves clients' quality of life and strengthens the U.S. economy. Core values of the company include Integrity, Passion, Effectiveness, Collaboration.

There are more than 17,000 veterans and military spouses move into new careers, generating an economic impact of $200 million. Hire Heroes USA has built a national reputation of excellence for helping veterans find jobs: now at the rate of more than 60 veterans confirmed hired every week. Hire Heroes USA is founded in 2005 by John Bardis, having headquarters located at Alpharetta, Georgia, United States with seven additional Offices throughout the United States. Serving 15,000 clients per year is a monumental task which the company alone cannot do. As company have grown, so has the ecosystem of incredible partners who advance their work. Funding Partners primarily support operating budget; Program Partners primarily collaborate in service delivery; Referral Partners primarily infuse program with eligible clients (job-seeking military members, veterans and military spouses).

HHUSA Services:

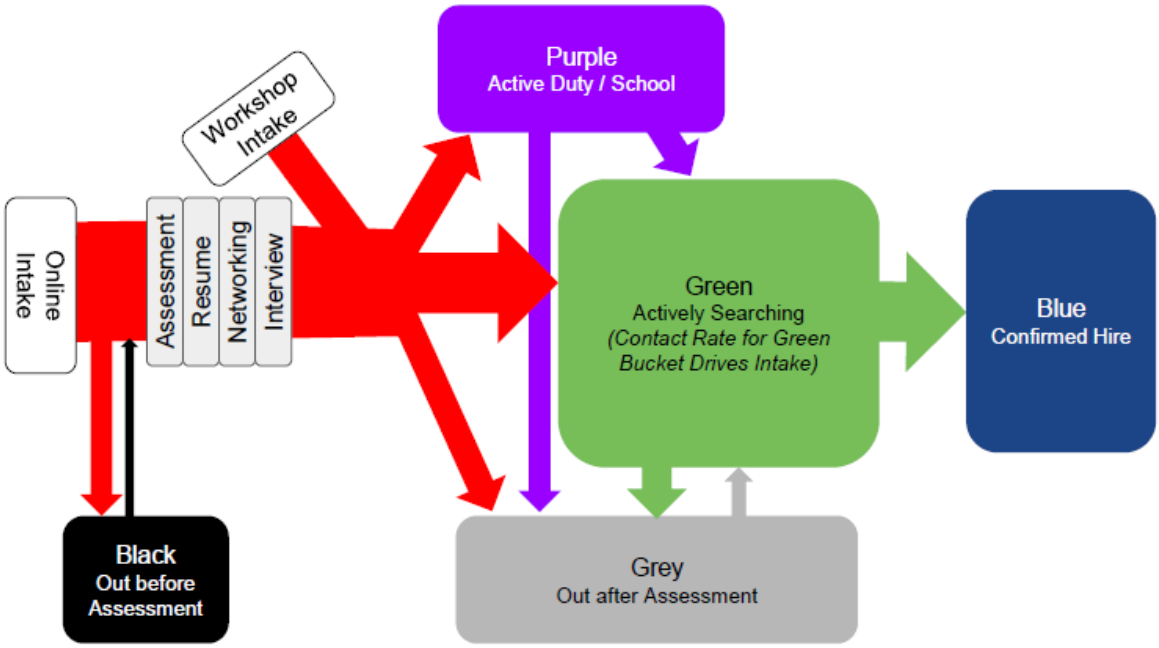
HHUSA's business professionals train veterans in the skills that are desired in the corporate world. Hire Heroes USA has built a national reputation of excellence for its success helping U.S. military members, veterans and their spouses find civilian employment. They partner with more than 200 veteran friendly companies to offer relevant and up-to-date job postings on the Hire Heroes USA Job Board. The hallmark of the program is a personalized approach where each client is paired with a Transition Specialist to:

- Create a civilian resume that effectively highlights skills and achievements
- Translate military experience into civilian terminology
- Learn effective job search, networking and interviewing techniques
- Get connected with companies who want to hire veterans

Hire Heroes USA Client Flow:

The Veterans get acquainted with Hire Heroes USA through 'Online Intake'. The veterans who gets registered with HHUSA are trained thoroughly after an 'initial assessment'. Their 'resume' is been developed and their 'networking skills' and 'interview skills 'are been developed. To further improve the

skills, veterans are trained through workshops so that they get ready for civilian employment. All these together constitute a phase. Some veterans just register for HHUSA services and black out before assessment. These people are put into a different bucket. Veterans who are actively doing their duty also register for HHUSA services. They constitute a different bucket(violet). Once the candidate is ready to be hired, the job search phase begins. His resume is being forwarded to the companies that are willing to hire veterans. Veterans who opt out after assessment stage are been put into a different phase. Thus, HHUSA trains veterans in the skills that are desired in the corporate world and help them get a job.



Problem Statement:

HHUSA's program successfully helps more than 60 veterans get hired for every week. The significant worry for HHUSA is that there are 500,000 unemployed veterans that exist at any given time. They need to connect with this unemployed population by breaking down and analyzing the existing information to determine if there are opportunities for further improvements to the existing systems.

Below are the questions that we are curious about:

Is there any relationship between the amount of time spent working with individual clients (time to complete an assessment, time to complete resume, # of logged activities, etc.) and how quickly they are employed?

What measures should we utilize to measure process efficiency and program performance relevant at the Veteran Transition Specialist, office, and regional levels, and are there alternative measurement techniques to maximize reporting accuracy and efficiency?

Is it possible to correlate client and employer demographic data (candidate qualifications, geography, industry, desired career fields etc.) and our current method of matching clients with potential employers to identify new effective ways to help clients connect with companies we've built relationships with?

Does working with a volunteer raise the probability of a client getting hired or getting hired sooner? Are there certain volunteer activities that are more effective than others? Do volunteer activities increase the quality job obtained – perhaps determined through salary level? (Essentially, can we prove quantitatively that our volunteers are increasing the overall effectiveness of our program?)

Data Description:

The dataset provided to us, has 7 different tables of various size. Our analysis of the data warehouse reveals that Hire Heroes has a star schema configuration with the contact table as the fact table.

Below is a table showing a brief description of each tables:

Interesting findings of the review of the data given to us are mentioned below.

The **contact table** has 65,549 clients. The different clients can be Candidates, Workshop participants, onward to opportunity, and AVR (attempted veteran registrant). The AVR are individuals that attempted to register for services, but were unable to complete REGISTRATION due to capacity.

The **Account table** has 12,967 distinct companies recorded. Companies, businesses, and other corporate partners are listed under the Accounts file.

The **record type table** appears to be for internal use. It has a 120 observation. It has 4 dimensions as shown in the picture below.

The **User table** has all the information regarding the employee, their region, position etc are stored.

The **User-role table** has all the information regarding role of the user for example recruiter, Accountant, CEO.

Tasks and **Cases** objects include activities that are associated with contacts and Accounts objects.

Summary Statistics of Contact file:

There are totally 65,549 clients having around 653 variables. Almost 70% of these variables have missing values. Only 125 out of 653 variables which doesn't have any missing values. 7347 of the clients were confirmed hired clients.

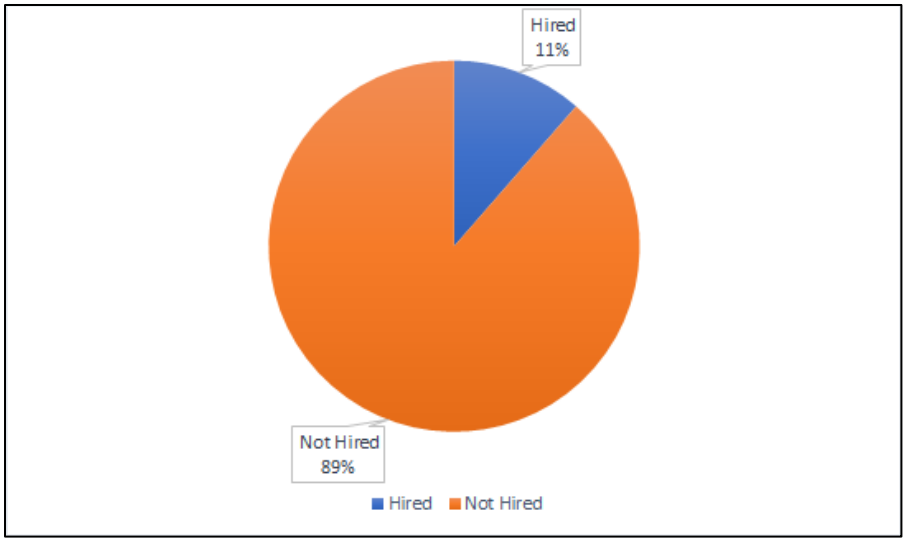


Fig 1. Hired vs Not Hired

Out of the total, only 11% of the clients are hired clients.

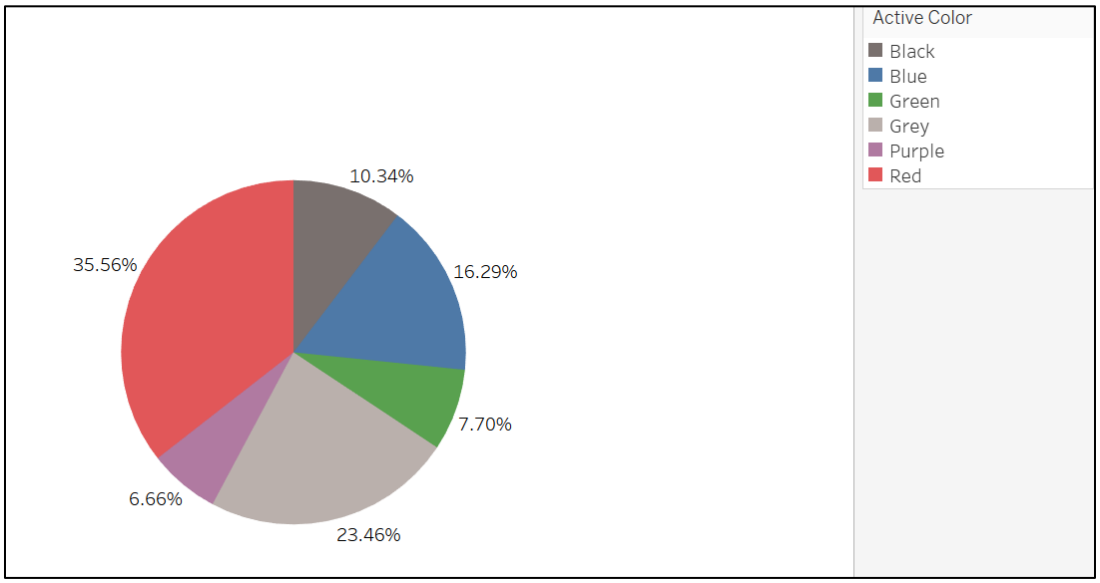


Fig 2. Percentage distribution of each Active color

Out of the total distribution, 35.56% of the clients come under the red bucket, which indicates that they are currently in the assessment and resume preparation process. Almost 10% of the total clients have blacked out after initial registration.

Business Problem 1:

What measures should we utilize to measure process efficiency and program performance relevant at the Veteran Transition Specialist, office, and regional levels, and are there alternative measurement techniques to maximize reporting accuracy and efficiency?

To measure the process efficiency and program performance; the percentage of people hired, the time taken for the veterans to get hired, the time taken by the employees to finish the various tasks allotted to them to make the client ready for employment have been considered.

Data preparation:

Each client in the contact file was mapped to the employee they were assigned to. The hire heroes' employees were listed in the user file. Further, information regarding the employee's location, position etc. were also obtained from the User file. From the contact file, the clients hired status, time to get hired, time taken to finish the hire heroes' processes were taken. Client who had a confirmed hire date were considered as hired.

There were around 1100 employees under Hire Heroes. But around 900 odd records had missing values for regions. Hence only the handful records which had region wise information were considered for the region wise performance analysis.

- The number of clients in each region was visualized through a bubble chart

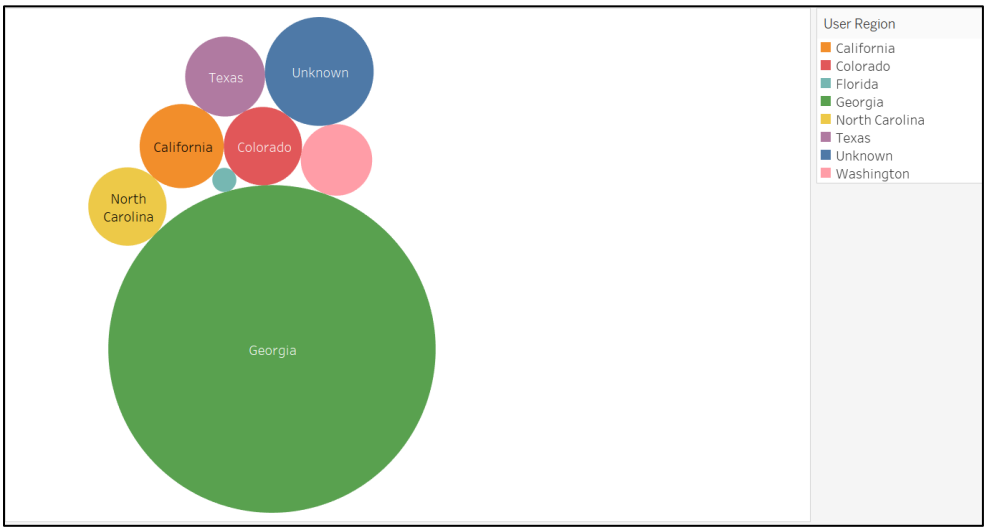


Fig 3 Number of Clients hired in each region

Majority of the clients were from Georgia. almost 70% of the clients belonged to Georgia. Florida region had very few clients.

- The number of veteran transition specialist per region was also visualized with the help of a bubble chart.

The number of veteran transition specialists were almost equal in the regions Georgia, Colorado and California. Florida has the least number of veteran transition specialists. From the previous visualization, we gathered that Georgia had the maximum number of clients, hence reallocation of the veteran transition specialists could be done to keep a balanced employee to client ratio in all the regions.

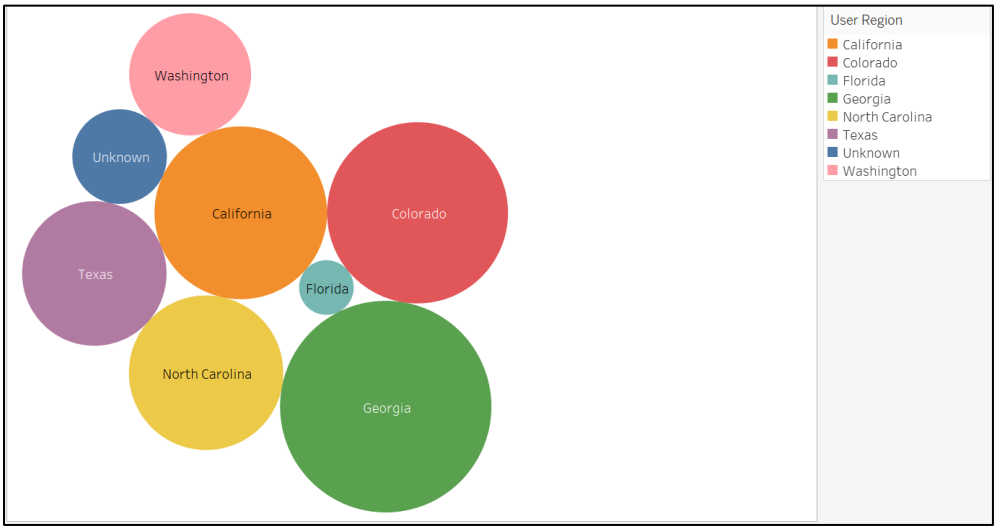


Fig 4 Number of Veteran Transition Specialist in each region

- The percentage of veterans hired in each region were studied and plotted using a stack chart.

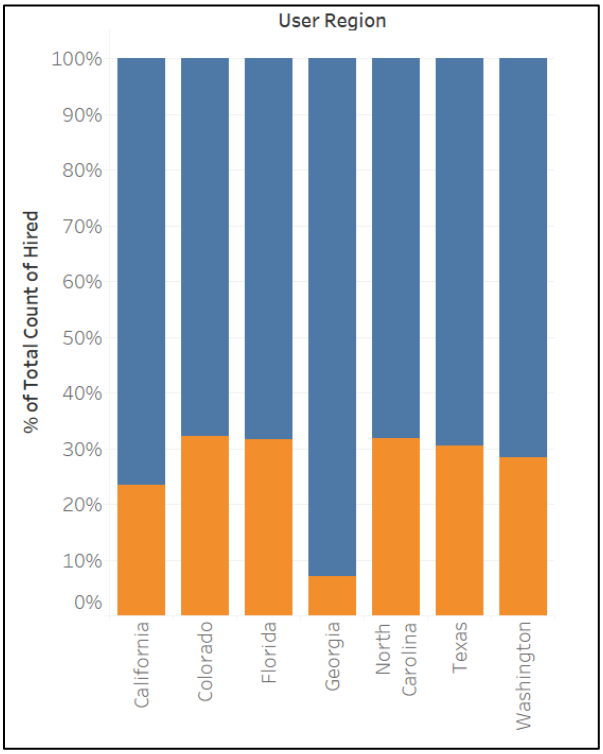


Fig 5 Employee region and percentage hired

From the stack chart, it can be inferred that Georgia has only around 10 % of veterans hired out of the lot applied in the program. All the other Hire heroes' regional offices have an average of around 30% of clients getting hired.

- The average time for the clients to get hired was considered next for each region.

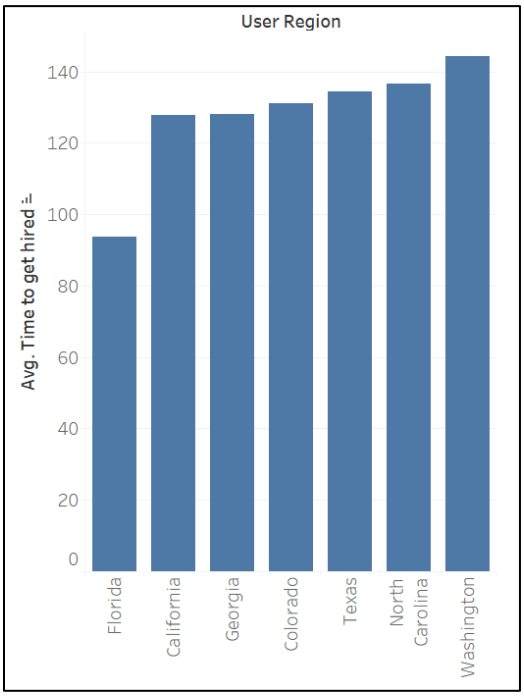


Fig 6 Avg time to get hired by region

In the average time to get hired, Florida ranks first with less than hundred days to get hired. Washington is last in the list with an average greater than 140 days to get hired.

- The average time taken for the various processes like time to get assigned, time to prepare the resume ready were considered.

  Interestingly, in both the bar graphs considered below (Fig 7 and Fig 8), it could be clearly seen that Georgia region office on an average takes very less time for the client to be assessed and also to prepare the resume of the client after assessment; though Georgia had a very less conversion rate of clients getting hired.

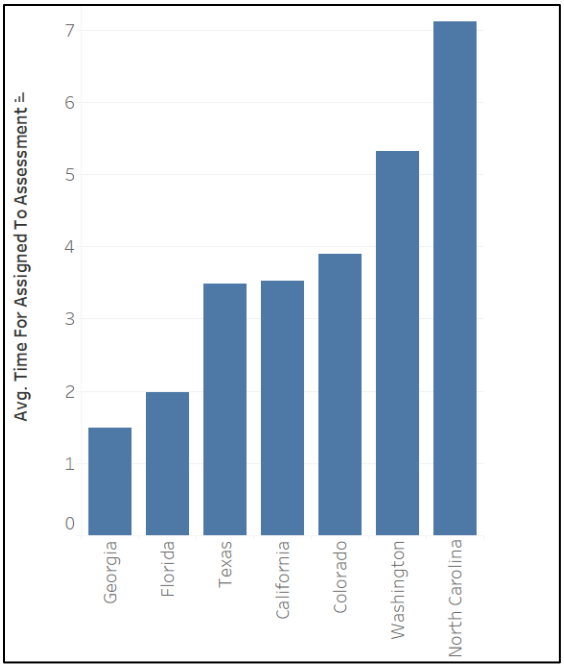


Fig 7 Avg time taken between assigned and assessment by employee region

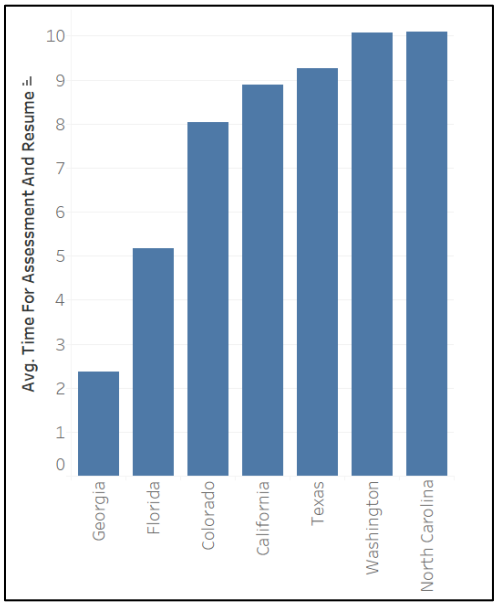


Fig 8 Avg time taken between assessment and resume preparation by employee region

- A tree map was visualized to see if there was any correlation between the amount of time involved in the process and the time taken by the veteran to get hired for the various regions. In the tree map, the intensity of the color represents the time to get hired and the size of the blocks represent the time taken for the various processes by the employees.

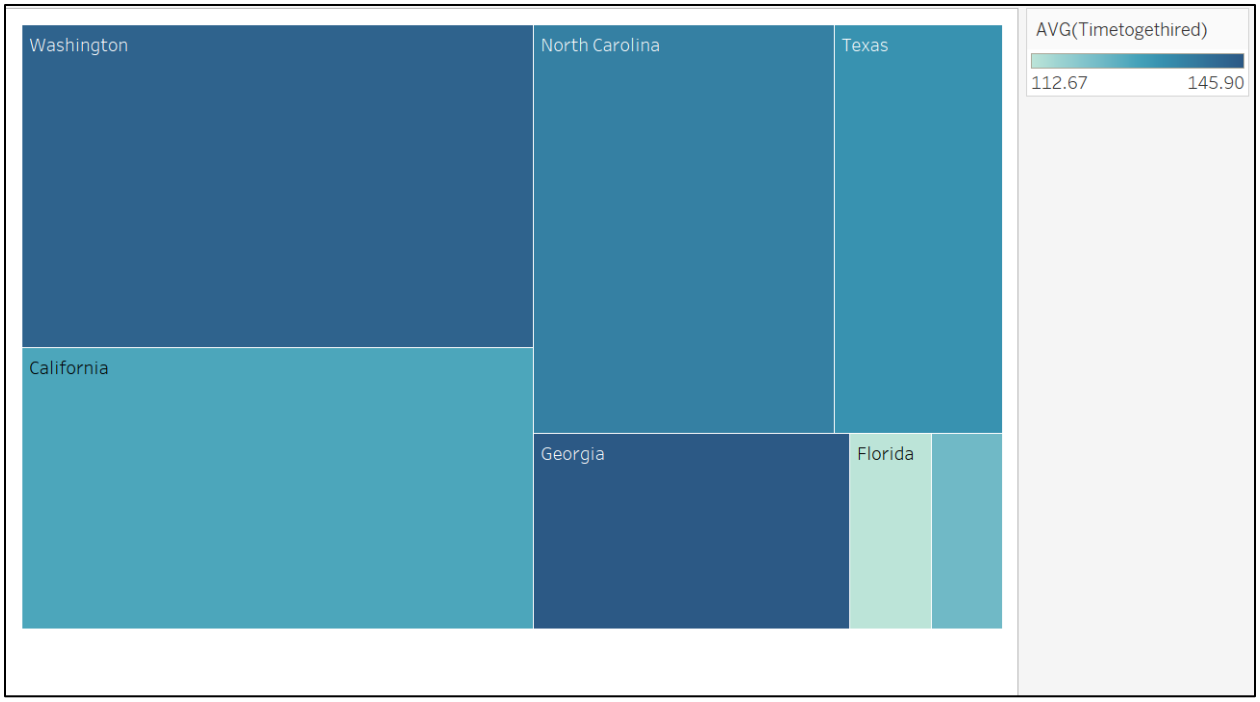


Fig 9 Tree map showing the time to get hired and the time from assigned to resume preparation

An exact relationship could not be inferred from the tree map. It can be seen from Fig 5 that in Florida region, the time taken for the process is less and also the clients get hired quickly. Whereas in Georgia region, clients take time to get hired and the time taken from assigned to resume preparation is less.

- A bar chart below shows each veteran transition specialist and the number of veterans hired who were assigned to them.

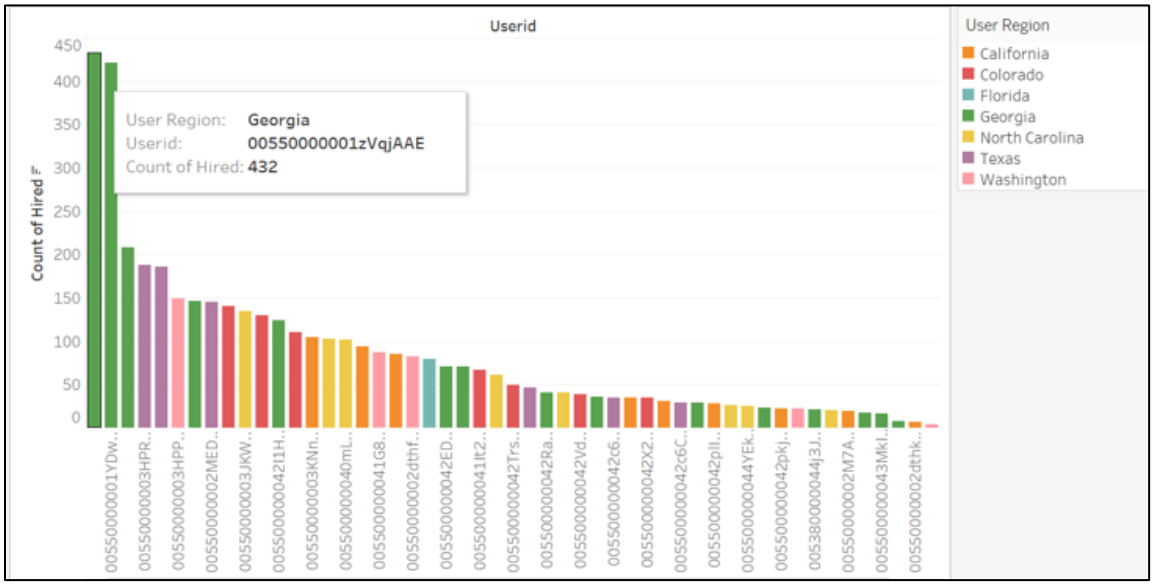


Fig 10 Number of veterans hired by Veteran Transition Specialist

From the above chart, the top three veteran transition specialist who have the highest number of veterans hired under them are from Georgia.

Inference and recommendations:

- Systematic maintenance of Hire Heroes employee's records is required since majority of the employee's region data were missing.
- Georgia has the maximum number of clients and the conversion rate for Georgia was less than average. More veteran transition specialists can be deployed here and a deep dive analysis could be carried out to find the reason for the same.
- The number of Veteran transition specialist in Georgia were less despite of its high client's population. The number of veteran transition specialist wasn't distributed proportionally an all the regions. Reallocation of veteran transition specialists is required to maintain a balanced employee to client ratio in all the regions

Business Problem 2:

Is it possible to correlate client and employer demographic data (candidate qualifications, geography, industry, desired career fields etc.) and our current method of matching clients with potential employers to identify new effective ways to help clients connect with companies we've built relationships with?

The current method of matching client wasn't available in the data. Hence, it was decided to look into the client's demographics data and see what qualities of the client gets them hired.

Data Preparation:

Clients who had a confirmed hired date were considered as hired.
The dataset had to be cleaned first, column which had more than 70% of missing values were removed.
From the final data set prepared the following were selected as independent and target variables.

Target Variable – Hire or Not

Independent Variables:

- Areas of Experience
- Disability Rating
- Education Summary
- Summary of Qualification
- Desired geographic location
- Desired industry to work in

Text mining of text data:

Out of these independent variables, three of them; namely Areas of Experience, Education Summary and Summary of Qualification were text. Hence text mining was chosen to be done to obtain text topics and categorize the text into the text topics.

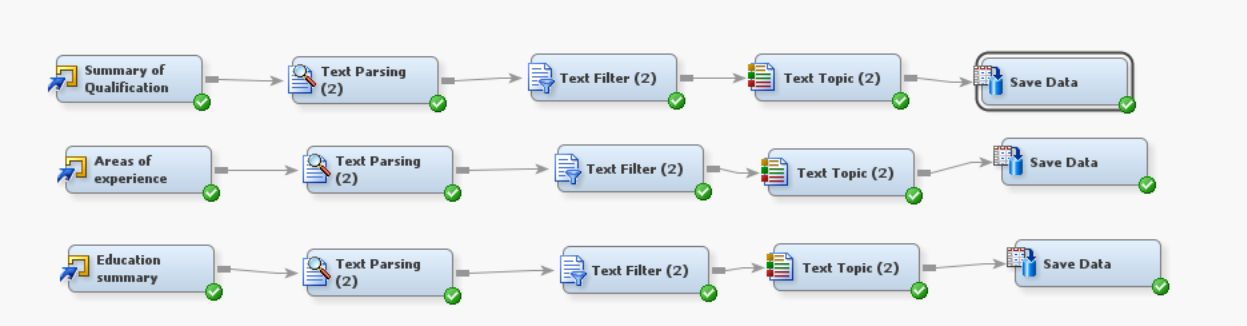


Fig 11 Process flow of text mining

Five topics were identified for each of the three text inputs.

The 5 topics identified for Summary of Qualification is attached below:

Category	Topic ID	Document Cutoff	Term Cut off	Number of Terms	Topic	# Docs
Multiple	1	0.151	0....	221	+fast,+fast - pace environment,+pace,+accomplished measurable result,+dynamic	6942
Multiple	2	0.064	0....	464	+logistic,+risk,critical thinking,+critical think	7429
Multiple	3	0.070	0....	207	+audience,+interact,+diverse cultural audience,+multiple level,cultural	3657
Multiple	4	0.045	0....	567	maintenance,+system,marine,corps,+maintain	5242
Multiple	5	0.054	0....	537	human,+resource,resources,+human resource,+master	7291

Fig 12 Text topics of Summary of Qualification

Based on the topics identified by SAS, the clusters were named as follows:
Cluster 1 – Fast and Dynamic
Cluster 2 – Logical and critical thinking
Cluster 3 – Culturally Interactive Quality
Cluster 4 - System Maintenance
Cluster 5 – Human Resource

Similarly, 5 topics were identified for Areas of experience:

Topics

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms
Multiple	1	0.096	0.025	+data analysis,+analysis,data,implementation,policy	41
Multiple	2	0.082	0.025	+strategic plan,strategic,planning,management,operations	81
Multiple	3	0.083	0.025	quality,assurance,quality assurance,+control,risk	55
Multiple	4	0.051	0.026	service,customer,development,training,resolution	95
Multiple	5	0.071	0.025	management,logistics,improvement,process,control	73

Fig 13 Text topics of Areas of experience

Based on the topics identified, the clusters were broadly classified as:
Cluster 1 – Data Analytics
Cluster 2 – Strategic planning and operations Management
Cluster 3 – Quality Assurance
Cluster 4 – Customer Service
Cluster 5 – Logistics and Process control

Finally, 5 topics of Education Summary were also identified:

Topics
Log

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms
Multiple	1	0.096	0.027	+high,+school,school,diploma,high school	7
Multiple	2	0.089	0.030	business,+bachelor,management,administration,science	42
Multiple	3	0.094	0.027	+college,+year,+degree,+credit,+complete	7
Multiple	4	0.086	0.027	hs,diploma,grad,business,+bachelor	2
Multiple	5	0.082	0.030	justice,criminal,+bachelor,science,+associate	43

Fig 14 Text topics of Education Summary

The clusters were identified as:
Cluster 1 – High school graduate
Cluster 2 – Management Business Administration
Cluster 3 – College graduate
Cluster 4 – Diploma or bachelor degree
Cluster 5 – Law related degree

Models considered:

Logistics regression and decision trees were considered appropriate for this case as the target variable is binary. Stepwise logistic regression was chosen. These two models were run and compared to predict the likelihood of being hired. Since the number of clients hired was less than 10% of the entire data considered, the data was sampled first. Data was portioned into train and validate in the ratio 3:2.

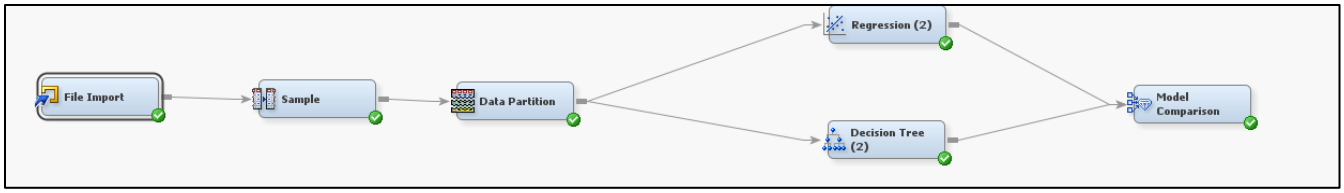


Fig 15 Process flow of modelling

The model comparison results were obtained and they are as follows:

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Akaike's Information Criterion	Train: Average Squared Error	Train: Average Error Function	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Tra To De of Fre m
Y	Tree3	Tree3	Decisio...	Hired	Hired	0.1679...	.	0.1336...	.	.	.	
	Reg3	Reg3	Regres...	Hired	Hired	0.1973...	3610.5...	0.1420...	0.4628...	3859	13	

Fig 16 Fit Statistics of model comparison

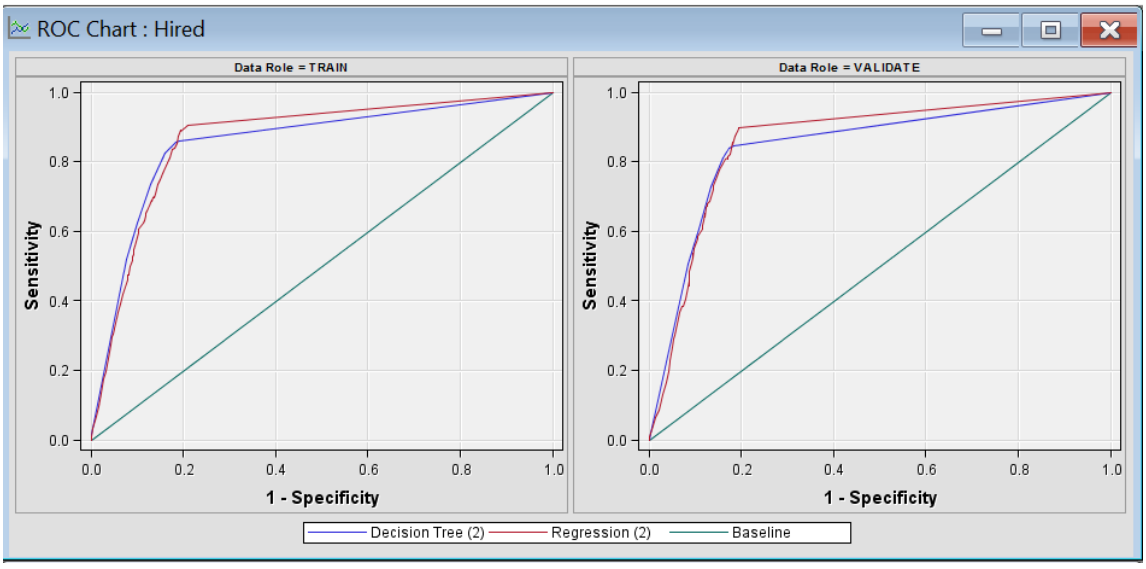


Fig 17 ROC Curves of model comparison

From the output results it could be inferred that decision tree has a validation misclassification rate less than logistic regression. Hence it was picked to further study the qualities of the client which gets them hired.

Output of the decision tree:

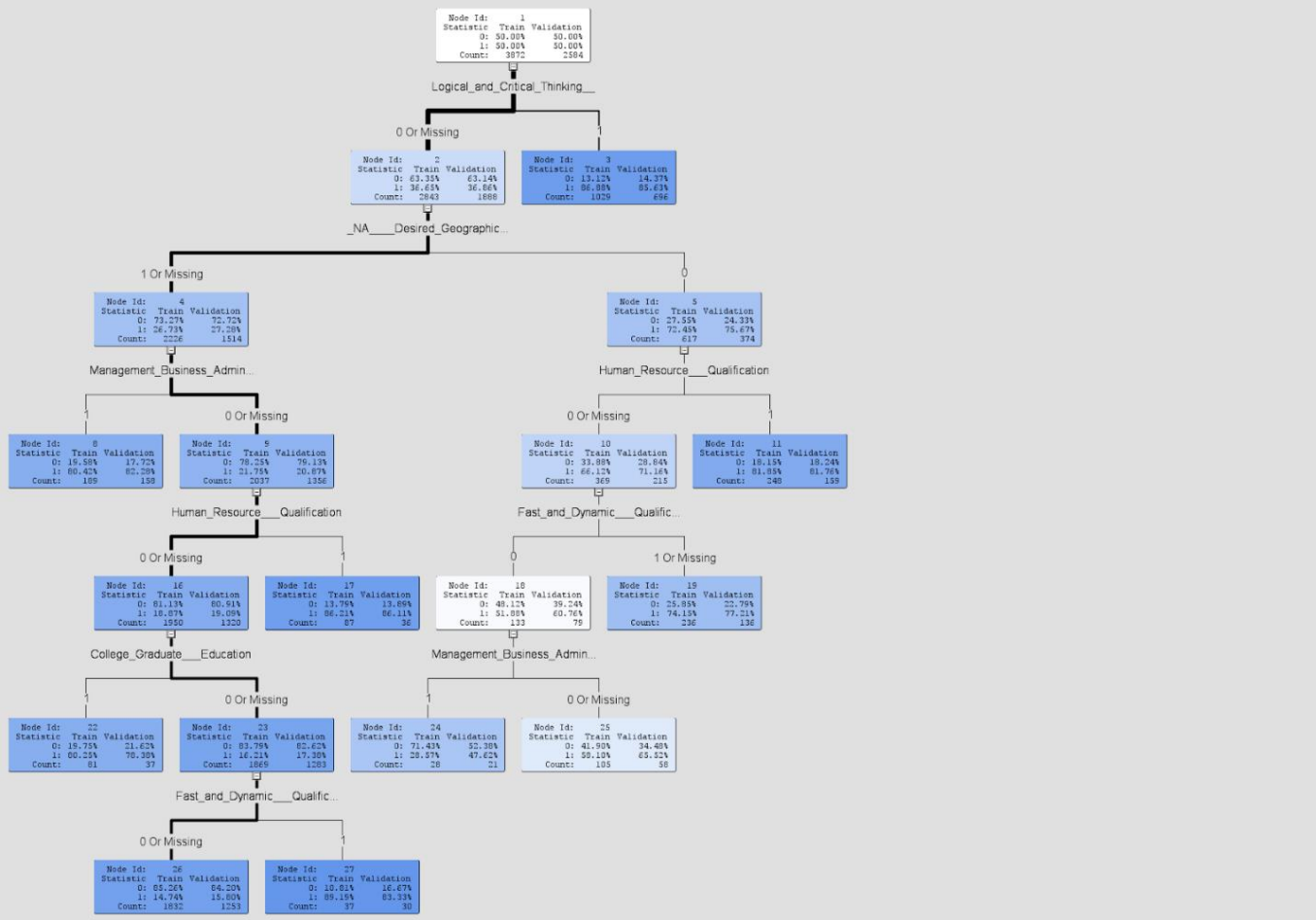


Fig 18 Decision Tree output

Interpretation of the decision tree:

From the decision tree, it can be inferred that a client with logical and critical thinking get hired 80% of the time. Even if they don't have much of logical and critical thinking, clients who don't have a desired geographic location preference with a business management administration degree also have a high probability of getting hired. On the other hand, even if the client doesn't have a master in business management, if he is qualified in human resources, there is still a high chance of getting hired.
Else even a college graduate client finds it easy to find a job.

Overall the education summary of the candidate plays a key role in a veteran getting hired. The qualities logical and critical thinking, Fast and Dynamic and Human resource qualification are essential for a client to get hired. Also, candidates with a management degree plays a vital role.

Business Problem 3:

Does working with a volunteer raise the probability of a client getting hired or getting hired sooner? Are there certain volunteer activities that are more effective than others? Do volunteer activities increase the quality job obtained – perhaps determined through salary level? (Essentially, can we prove quantitatively that our volunteers are increasing the overall effectiveness of our program?)

Data Preparation:

- Case data only appears if the client requested or responded to an offer for volunteer services. So, if a client has no case data in their record, they did not want to speak with a volunteer (or weren't offered the service).
- Case Data has 1,826 records, out of which 958 records has "completed" status meaning that only these veterans completed their volunteer request while others had withdrawn or did not respond to request.
- Contact Data has 64,549 records. Both these datasets were joined by matching their contact Ids using R.
- After removing duplicates, 842 veterans were identified to have used the volunteer service.
- The columns that are of interest are:
    - volunteer_c (Veteran used volunteer service or not; binary)
    - Blue_c (Hired or Not; binary)
      if(Active_color_c = "Blue") then 1 else 0 ;
    - Salary_Range__c
    - Volunteer_Services__c
    - Days_to_get_hired (Date_Turned_Blue__c - Date_turned_green__c)
- Salary_Range__c column contains the min-max range of salary for each veteran (ex : $40,000 - $55,000). Average salary for each veteran was calculated from the given range.
- Subsets were formed for covering different scenarios

Effect of Volunteering on getting hired:

Statistical Tests:

t-test was performed to compare the proportions of veterans who got hired without using any volunteer services vs proportion of veterans who got hired with the help of volunteers. Below results shows that there is a significant difference between the two groups (p-value < 2.2e-16). Around 50% of the veterans who took the help of volunteers got hired, whereas only 15% got hired among the veterans who did not get any volunteering service.

```
2-sample test for equality of proportions with continuity correction
data:  c(6948, 426) out of c(44430, 843)
X-squared = 736.29, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
-0.3834806 -0.3144339
sample estimates:
   prop 1    prop 2
0.1563808 0.5053381
```

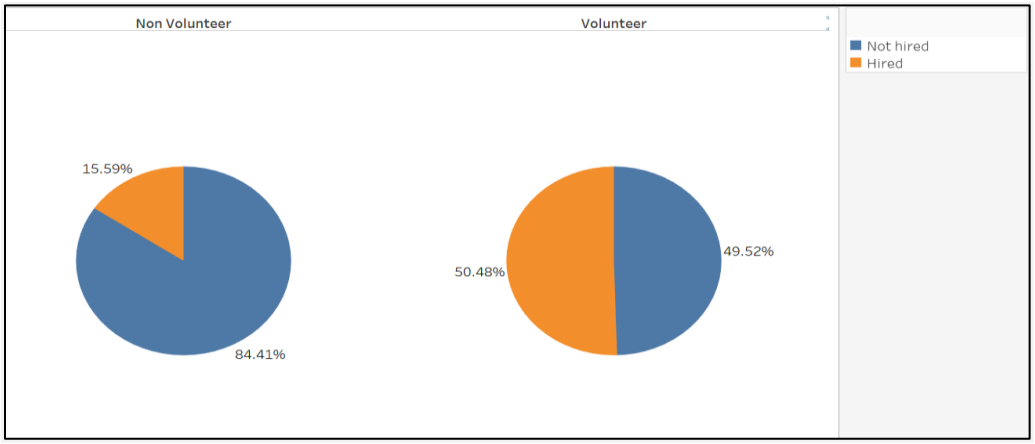


Fig 19 Comparison of Hired Proportion with and without volunteer help

Modelling:

The logistic regression model built for the above problem gives the below relationship between the log odds of getting hired and the effect of volunteering.

Log (odds of getting hired) = -1.68541 + volunteer_c * 1.70676

This indicates that logs of getting hired increases by 1.7 times when the veteran is offered volunteering service.

Effect of Volunteering on Days to get Hired:

Similarly, 2-sample t-test was performed to compare the means of number of days to get hired with and without volunteer service. On an average, it took 152 days to get hired for a veteran who used a volunteer, whereas it took 188 days to get hired without volunteering (see fig 20).

```
                Welch Two Sample t-test
data:  date_diff_with_volunteer and date_diff_without_volunteer
t = -4.4069 days, df = 496.36, p-value = 1.286e-05 days
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-52.54307 days -20.13895 days
sample estimates:
Time differences in days
mean of x  mean of y
152.1086  188.4496
```

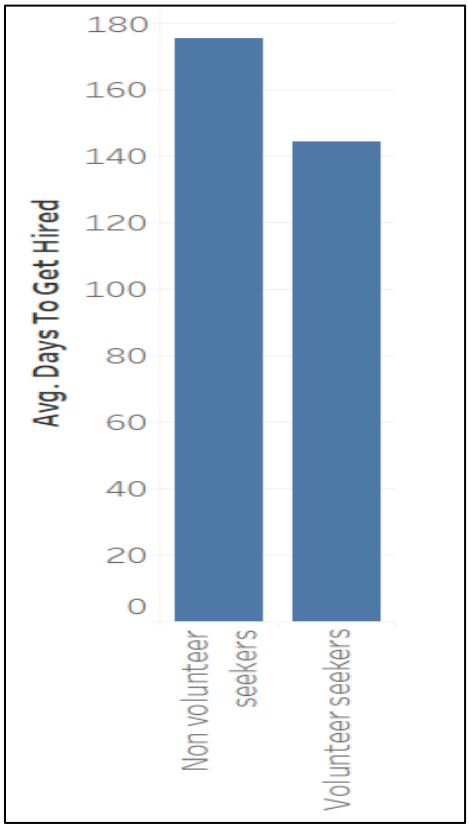


Fig 20 Comparison of Days to get hired with and without volunteer

Below linear regression model was built to find if working with volunteer increases the probability of getting hired sooner.

Days_to_get_hired = 188.45 - volunteer_c * 36.34

This shows that days to get hired decreases by 36 times when the veteran works with a volunteer.

Effect of Volunteering on Quality of Job (in terms of salary):

2- sample t-test for comparing the means of salary shows that average salary of veterans who got hired with volunteering service is more than the ones who did not get any volunteer service offered

```
Welch Two Sample t-test

data:  salary_with_volunteer and salary_without_volunteer
t = 4.2842, df = 376.47, p-value = 2.331e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2699.303 7278.863
sample estimates:
mean of x mean of y
 52916.19  47927.11
```

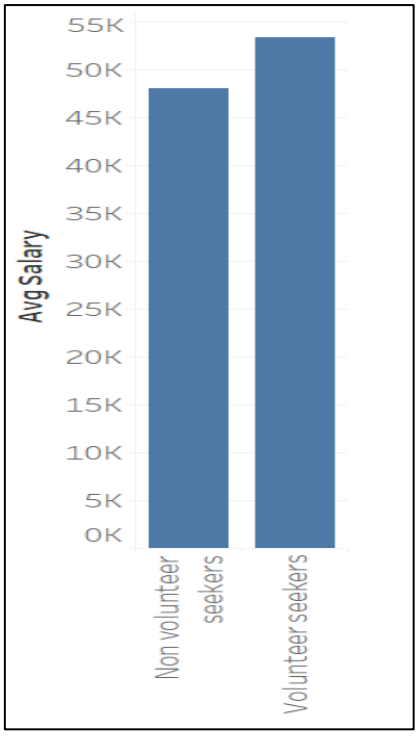


Fig 21 Comparison of Average Salary of Hired Veterans with and without Volunteers

The below linear regression model shows the relationship between average salary and the effect of volunteering. Average salary of a candidate increases by 4,989 times if volunteering service was offered.

Avg_salary = 47927 + volunteer_c1 * 4989

Are there certain volunteer activities that are more effective than others?

To find out whether certain volunteer activities helps the veterans to get hired sooner, we plotted the count of clients hired under various volunteer services (fig 22). Below plot shows that most of the veterans who got hired attended mock interview sessions and the veterans who attended these mock interview sessions got hired sooner when compared to other services that are offered (Fig 23).

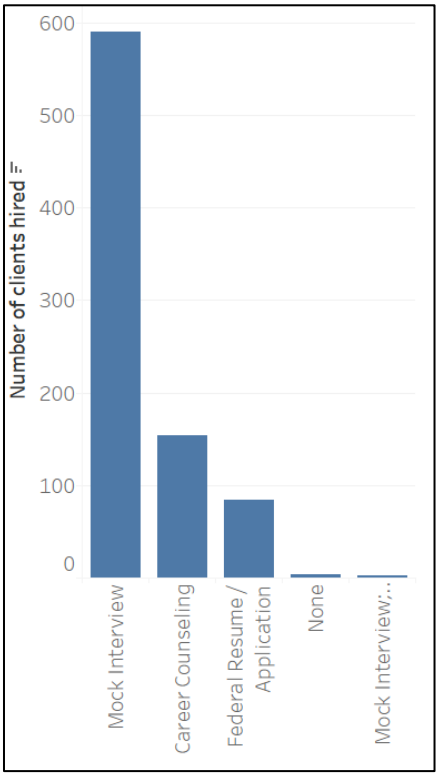


Fig 22 Number of veterans Hired under each Volunteer activity

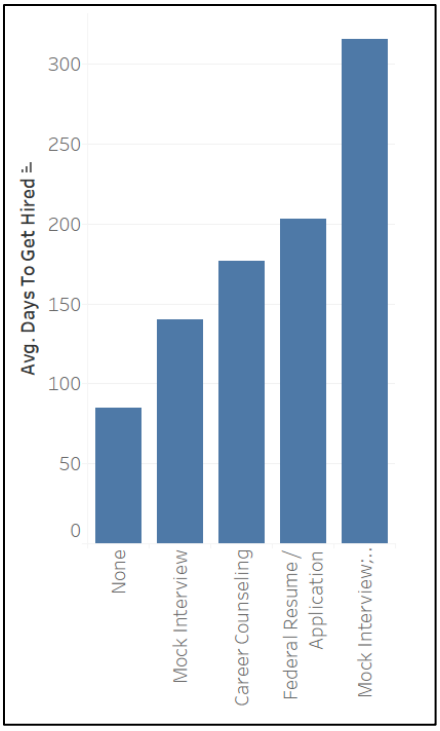


Fig 23. Days to get Hired versus Volunteer activities

Logistic regression to predict relationship between volunteer services and probability of getting hired

First a logistic regression was carried out to find the probability of getting hired under each volunteer activity. Volunteer_Services__c is a categorical independent variable with many levels. The target variable here is if the client is hired or not. The Regression Results from R also shows that most of the people who got hired have attended mock interviews and career counselling sessions

```
Call:
glm(formula = Blue_c ~ Volunteer_Services__c, family = "binomial",
    data = contacts_with_volunteer, na.action = na.omit)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.268  -1.268   1.090   1.090   1.665

Coefficients:
                                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                                       -0.7922     0.1740  -4.554 5.27e-06 ***
Volunteer_Services__cFederal Resume / Application  0.9353     0.2795   3.346 0.000819 ***
Volunteer_Services__cMock Interview                1.0025     0.1926   5.205 1.94e-07 ***
Volunteer_Services__cMock Interview;Career Counseling 14.3583   378.5929   0.038 0.969747
Volunteer_Services__cNone                         -0.3064     1.1677  -0.262 0.793039
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1158.8  on 835  degrees of freedom
Residual deviance: 1125.8  on 831  degrees of freedom
  (6 observations deleted due to missingness)
AIC: 1135.8

Number of Fisher Scoring iterations: 12
```

Fig 24 Logistic Regression results

Linear Regression to predict relationship between volunteer services and days to get hired:

Linear regression was carried out to see if the various volunteer services had an impact on the days to get hired. The target variable is days to get hired. The independent variable is the volunteer services. The data was partitioned in the ratio 7:3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	237301	59325	2.44	0.0462
Error	395	9588195	24274		
Corrected Total	399	9825496			

Model Fit Statistics

R-Square	0.0242	Adj R-Sq	0.0143
AIC	4043.8315	BIC	4045.9577
SBC	4063.7888	C(p)	5.0000

Type 3 Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
Volunteer_Services__c	4	237301.214	2.44	0.0462

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	184.2	38.7860	4.75	<.0001
Volunteer_Services__c Career Counseling	1	-7.5581	42.6729	-0.18	0.8595
Volunteer_Services__c Federal Resume / Application	1	19.0607	43.1229	0.44	0.6587
Volunteer_Services__c Mock Interview	1	-44.0385	39.3870	-1.12	0.2642
Volunteer_Services__c Mock Interview;Career Counseling	1	131.8	93.7365	1.41	0.1606

Fig 25 Linear regression results

None of the above terms are significant. Hence, we can say that, although certain type of Volunteer services has a positive effect on getting hired, it does not seem to help veterans in getting hired sooner. Also, the above model does not explain much of the variance in the response variable. The R-Squared value is very low (0.02).

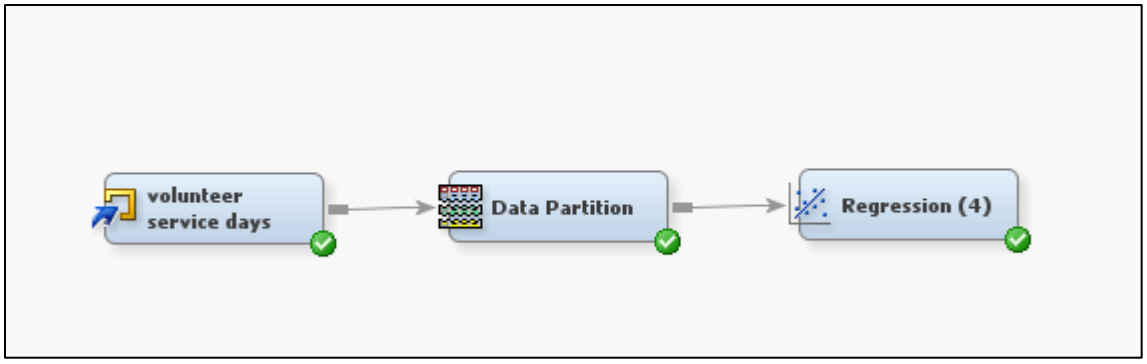


Fig 26 Process flow of regression model

Improved Regression Model:

To improve the R-squared value, the below factors were also considered to build the linear regression model that was used to predict the overall effect on days to get hired in presence of volunteering:

- Geographic region of veteran
- Areas of Experience of the veteran
- Educational Summary of the veteran
- Qualification Summary of the veteran
- Interview Skills
- Responsiveness of the veteran
- Days Between Assessed and Assignment
- Days Between Assessment and Resume
- Logged activities

Modelling Results:

After incorporating the above variables in our model, there is an improvement in the R-squared value (around 23%). Again, now also none of the volunteer activities seems to have an effect on the days to get hired(p-value>0.05)

Model Fit Statistics

R-Square	0.2313	Adj R-Sq	0.2215
AIC	4060.6176	BIC	4062.9910
SBC	4100.6565	C(p)	12.5639

Type 3 Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
Data_Analytics___Areas_of_Exper	1	403637.077	18.29	<.0001
High_School_Graduate___Educatio	1	172143.520	7.80	0.0055
Logistics_and_Process_Control__	1	188246.250	8.53	0.0037
Management_Business_Administrat	1	98744.8743	4.47	0.0350
Volunteer_Services__c	5	351537.044	3.19	0.0078

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept		1	225.5	36.6313	6.16	<.0001
Data_Analytics___Areas_of_Exper		1	-65.0435	15.2082	-4.28	<.0001
High_School_Graduate___Educatio		1	149.4	53.5083	2.79	0.0055
Logistics_and_Process_Control__		1	-45.0370	15.4197	-2.92	0.0037
Management_Business_Administrat		1	32.3745	15.3044	2.12	0.0350
Volunteer_Services__c	Career Counseling	1	-14.3524	39.4177	-0.36	0.7160
Volunteer_Services__c	Federal Resume / Application	1	9.5454	40.1057	0.24	0.8120

Fig 27 Linear regression results of improved model

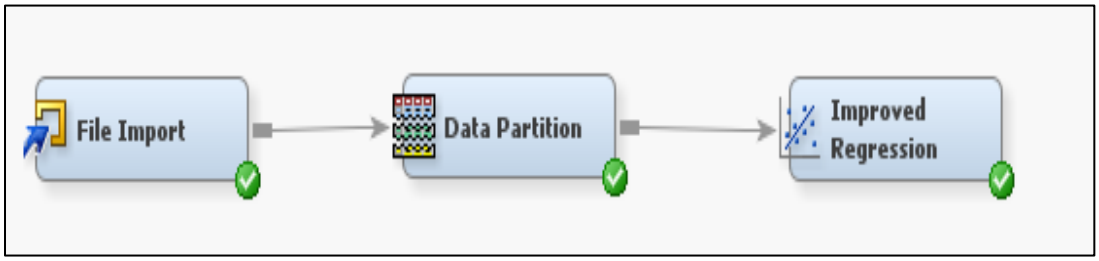


Fig 28 Process flow model of improved linear regression

Inferences:

- Based on the above results, we can conclude that offering veterans with volunteering service increases their probability of getting hired.
- Volunteer services also plays a role in helping the veterans to get hired sooner.
- Mock interview and career counselling sessions seems to help the veterans in getting hired. So, the HireHeroes USA must make it feasible for every veteran to attend these sessions that will help them gain interview skills and perform well during the hiring process.

Business Problem 4:

Is there any relationship between the amount of time spent working with individual clients (time to complete an assessment, time to complete resume, # of logged activities, etc.) and how quickly they are employed?

To find the relationship between the amount of time spent working with individual clients and how quickly they are employed, we first consider business days between Assigned and Assessment, business days between Assessment and Resume and Number of logged activities by the veteran.

Data Preparation:

Contact Data has Total of 64,549 records. Out of which only 7347 clients had a confirmed hire date. So, firstly 7347 records were filtered out of the original dataset. Then, Assessment time, Resume time and days to get hired are calculated for each of the records. After analyzing the dataset after calculations, we see that there are ambiguous records that have negative count of days for days to get hired, Assesment days and Resume days. So, these records are omitted. So, finally we had 2401 records to build a model with. Below are the formulas that were used for calculation.

- Bus_Days_between_Assigned_and_Assessed__c = (Dat_Initial_Assessment_was_Completed__c - Date_assigned_to_staff__c) – Interval Input.
- Bus_Days_between_Assessment_and_Resume__c = (Date_Resume_Completed__c-Dat_Initial_Assessment_was_Completed__c) – Interval Input.
- Days_to_get_hired = Date_Turned_Blue__c-Date_turned_green__c – Target.

Below are some of the logged activities by veterans – Nominal Input.

- Attended_Day_1__c
- Attended_Day_2__c
- Compass_Participant__c
- Career_Counseling_Sessions_Number_given__c
- Date_Attended_HHUSA_Workshop__c

- Career_Opportunity_Day_Participant__c
- Created_LinkedIn_account__c

The total number of logged activities were calculated by taking the count of each activity mentioned above.

Regression Model:

After Data preparation, regression was run upon the data with following variables using SAS E-Miner.

Bus_Days_between_Assigned_and_A	**Input**	**Interval**	**No**
contact_log_logged_activities	**Input**	**Interval**	**No**
Bus_Days_between_Assessment_and	**Input**	**Interval**	**No**
days_to_get_hired	**Target**	**Interval**	**No**

Fig.29. Process Flow Diagram

- Data Partition was set 60:40.
- Step-wise regression was run on the data.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4896614	1632205	39.54	<.0001
Error	1437	59318415	41279		
Corrected Total	1440	64215029			

Model Fit Statistics

R-Square	0.0763	Adj R-Sq	0.0743
AIC	15319.1117	BIC	15321.1340
SBC	15340.2041	C(p)	4.0000

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	157.8	8.2148	19.21	<.0001
Bus_Days_between_Assessment_and	1	0.7070	0.0812	8.70	<.0001
Bus_Days_between_Assigned_and_A	1	-0.1888	0.0946	-2.00	0.0461
contact_log_logged_activities	1	45.0081	8.0318	5.60	<.0001

Fig.30. Linear Regression Results

Interpretation of the Results:

From the above results, we can interpret that number of days to get hired for a veteran decrease as the Assessment time increases and Resume time decreases. But, surprisingly it is seen that Number of logged activities increases the number of days to get hired for a veteran. We see that all the variables taken into the consideration are significant since they have a p-value less than 0.05. The overall model is also significant. When examining the R-value, it is seen that this model unfortunately explains only close to 7% of the variance. This tells us that there are other reasons which influences the clients to get hired.
So, the next step was to deeply examine the contact dataset and find out what are the other factors that influences the client to get hired.

Improved Regression Model:

To improve the R-squared value and to make a more meaningful model, factors that might effectively influence veteran's success in getting hired quickly have been analyzed and has been added as input to the model. They are as follows

- Geographic region of veteran
- Areas of Experience of the veteran
- Educational Summary of the veteran
- Qualification Summary of the veteran
- Volunteer help
- Interview Skills
- Responsiveness of the veteran

In addition to the existing variables, these variables have been added to the model as input.

Fig 31 Process flow of regression

- Data partition was set 60:40.
- Step-wise regression was run on the dataset.

Below are the results of the regression model(Fig.31).

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	16963303	1211664	36.57	<.0001
Error	1426	47251726	33136		
Corrected Total	1440	64215029			

Model Fit Statistics

R-Square	0.2642	Adj R-Sq	0.2569
AIC	15013.3842	BIC	15015.5989
SBC	15092.4806	C(p)	19.7550

Type 3 Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
Bus_Days_between_Assessment_and	1	2262044.64	68.27	<.0001
Bus_Days_between_Assigned_and_A	1	267290.377	8.07	0.0046
Geo_NA	1	4156432.59	125.44	<.0001
areas_of_experience_data_analys	1	230786.567	6.96	0.0084
contact_Interview_Skills__c	1	934040.179	28.19	<.0001
contact_Responsive__c	1	293128.445	8.85	0.0030
contact_log_logged_activities	1	495166.593	14.94	0.0001
qualification__fast_pace	1	1038367.72	31.34	<.0001
qualification_hr	1	558333.708	16.85	<.0001
qualification_logistic_critical	1	681283.130	20.56	<.0001

Fig.31. Improved Regression Results

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept		1	124.4	17.5202	7.10	<.0001
Bus_Days_between_Assessment_and		1	0.6084	0.0736	8.26	<.0001
Bus_Days_between_Assigned_and_A		1	-0.2417	0.0851	-2.84	0.0046
Geo_NA	0	1	68.9862	6.1596	11.20	<.0001
areas_of_experience_data_analys	0	1	16.0848	6.0948	2.64	0.0084
contact_Interview_Skills__c	0	1	29.4398	5.5450	5.31	<.0001
contact_Responsive__c	0	1	-17.8706	6.0084	-2.97	0.0030
contact_log_logged_activities		1	32.7098	8.4616	3.87	0.0001
qualification__fast_pace	0	1	34.2073	6.1107	5.60	<.0001
qualification_hr	0	1	21.6991	5.2862	4.10	<.0001
qualification_logistic_critical	0	1	22.6466	4.9945	4.53	<.0001

Interpretation of the Results:

In addition to the existing takeaways from the results we see that,

- Geography of a veteran's place doesn't matter much in finding them a job. Out of all the various geographical locations of veteran's place, none of them were significant in the model. So, this proves that veteran's geographical location doesn't matter much in helping them finding a job.

- When the Veteran is responsive to HHUSA clients, then the number of days to get hired significantly decreases. It is seen that when there exists an effective communication and rapport between veteran and HHUSA employee, it helps the veteran to get hired quickly.

- Qualification Summary of veteran: Fast pace, HR, Logical and Critical Reasoning. Only these three were significant for a veteran to get hired.

- We see that we have an R-squared value close to 27%. In general, even though it is a poor percentage to explain the variance in the model, it is much better than the previous regression model. The reason for a low R-squared value can be accounted to very bad data quality. There were a lot of data discrepancies, missing values which made a dataset that had close to 65000+ values to be cleaned to just 2400 records so that it can be used for building a model.

Recommendations:

Based on the results, we have come up with the following recommendations

- It is recommended that HHUSA employees take sufficient amount of time to assess the veterans. Because, it would help them to get a better idea of the person and would help them in finding a relevant job sooner.

- Once this is followed, it would obviously decrease the number of days taken for writing a resume for the veteran. So, it is recommended that HHUSA employees take less number of days for Resume Writing for a veteran.

- Veterans should be more responsive to HHUSA people. So, HHUSA can take necessary measures to increase the communication and responsiveness between them and the veterans.

Conclusion and Final Recommendations:

- A balanced employee to client ratio must be maintained at a regional level through Reallocation of veteran transition specialists. Company should focus on deploying more Veteran Transition Specialist to regions where there are more number of clients.
- Company should also focus on deploying more number of volunteers for a faster rate of employment.
- It is recommended that HHUSA employees take sufficient amount of time to assess the veterans. Because, it would help them to get a better idea of the person and would help them in finding a relevant job sooner.
- Mock interviews and Career counselling sessions can be made mandatory for every veteran to attend once the resume preparation is done.
- Based on the initial assessment, Hire Heroes must maintain a log of the skills that every veteran is lagging which might be required for a particular job position.
- Logical thinking and HR skills seems to be the top most skills that every hiring company is looking for in a veteran. HHUSA should also consider conducting both onsite and online training sessions that can hone these foundational and strategic competencies.
- Majority of the Factors that are very critical for mapping a client to an employee were missing. An Overall Systematic maintenance of both client and employee records is required.

Tools Used:

- SAS for modelling
- R and Excel for Data Preparation
- Tableau for Visualization

Packages Used in R:

- Dplyr, Stringr, Data.table, reshape2 - Used for data cleaning
- Caret - Used for doing the data partition