8/5/2020

**HHUSA- Client Services**



DSBA 6211 – Summer 2020

Nityanand Kore

Sachin Varule

Soumyadip Mitra

Mrunmayee Kulkarni

Eric Rwabuhihi

Table of Contents

[About HHUSA 3](#_Toc47511733)

[HHUSA Services 3](#_Toc47511734)

[HHUSA Process Flow 3](#_Toc47511735)

[Problem Statement 4](#_Toc47511736)

[Detailed analyses processes and steps 5](#_Toc47511737)

[Data Description 5](#_Toc47511738)

[Interesting Fact 5](#_Toc47511739)

[Cleaning the data 5](#_Toc47511740)

[SQL Scripts to load data to Teradata Database 6](#_Toc47511741)

[SQL Scripts to get values and their distribution 6](#_Toc47511742)

[SQL Script to create the new dataset 6](#_Toc47511743)

[Relationship between a client's demographic profile and when that client registers for services? 7](#_Toc47511744)

[Data preparation 7](#_Toc47511745)

[Modeling and Visualization 7](#_Toc47511746)

[Conclusion 11](#_Toc47511747)

[Relationship between the amounts of time spent working with individual and how quickly they are employed? 11](#_Toc47511748)

[Data Preparation 11](#_Toc47511749)

[Modeling and Visualization 13](#_Toc47511750)

[Conclusion 17](#_Toc47511751)

[Effective Communication method for their success in job placement or some other outcome? 17](#_Toc47511752)

[Data preparation 17](#_Toc47511753)

[Modeling and Visualization 17](#_Toc47511754)

[Conclusion 18](#_Toc47511755)

[Relationship between a client's demographic profile and a client's likelihood to complete a survey? 19](#_Toc47511756)

[Data preparation 19](#_Toc47511757)

[Modeling and Visualization 21](#_Toc47511758)

[Conclusion 21](#_Toc47511759)

[Any Client's demographic profile that impacts to be a confirmed hire or any other outcome? 23](#_Toc47511760)

[Data preparation 23](#_Toc47511761)

[Modeling and Visualization 23](#_Toc47511762)

[Conclusion 24](#_Toc47511763)

[Tools and Technologies used 25](#_Toc47511764)

[Appendix and References 25](#_Toc47511765)

# About HHUSA

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, and 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 in2005 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

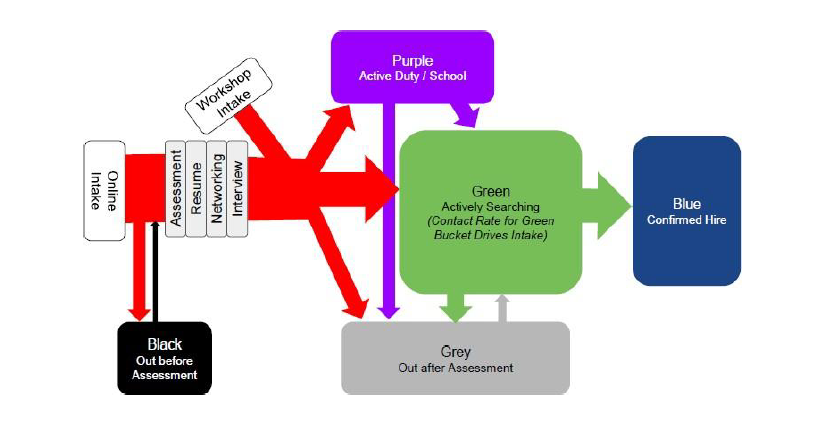
• Translate military experience into civilian terminology

• Learn effective job search, networking and interviewing techniques

• Get connected with companies who want to hire veterans.

# HHUSA Process 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 need to be answered for this project –

* Is there any relationship between the amounts 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?
* Is there a relationship between a client's demographic profile (rank, branch, time in service, spouse status etc.) and when that client registers for services?
* Is there a relationship between a client's demographic profile (rank, branch, time in service, spouse status etc.) and a client's likelihood to complete a survey?
* Is there a way to tell what communication method (call, texting or email) is more successful with clients, either based on their success in job placement or some other outcome?
* Is there anything in the client's demographic profile that indicates that a client is more likely to become a confirmed hire or any other outcome?

# Detailed analyses processes and steps

## Data Description

The dataset provided to us has 9 different Salesforce datasets and 4 Teradata extracts. To answer questions of our interest we used only 4 datasets highlighted in green below:

|  |  |  |
| --- | --- | --- |
| **Dataset** | **Total Records** | **Purpose of the dataset** |
| SalesForce\_Contact | 132445 | It has different type of client’s information. The different clients can be Candidates, Workshop participants, onward to opportunity, and AVR |
| SalesForce\_Hire\_Information\_\_c | 30754 | Successful hired client information is maintained in this dataset |
| Feedback\_\_c | 15822 | All feedback received from Clients are stored here. |
| SalesForce\_2018Activities | 451184 | All contact activities such as email, text, phone etc are maintained in this dataset. |
| SalesForce\_Account | 16852 | Companies, businesses, and other corporate partners are listed under the Accounts file |
| SalesForce\_Case | 14845 | activities that are associated with contacts and Accounts objects |
| SalesForce\_Opportunity | 10849 | Donation related information stored in this |
| Campaign | 1386 | All Campaign data maintained in this |
| vr\_\_VR\_Email\_History\_Contact\_\_c | 378080 | Summarized Email history for contacts |

## Interesting Fact

We had in total 1016 unique columns for all the datasets provided but about 70% of the columns were having missing values or were highly skewed or were not required for answering our problem statements. So finally we just used about 50-60 columns and prepared new dataset by joining all of the required dataset together.

## Cleaning the data

1. We loaded data to Teradata database to perform data distribution analysis. Most of the files were very small and as such they were loaded without any issues. However, the file for contact information had too many columns. For Contacts dataset we created two tables with 200 columns. We chose contact ID as the identifier for each of the tables for contact information.

1. We used the SQL queries to create sample dataset and did visualization through Tableau.

1. Those columns that showed up as heavily skewed were removed from the data files.
2. We did also cleanup through R for all the Variables which had all null/blank values and with constant values using reusable functions.

*Code snippet*

*dfSV <- dfSV[!myvars]*

*#1. A table showing the overall structure of the dataset, including variable names, data types*

*# find columns which has all missing values*

*allmisscols <- apply(dfSV,2, function(x)all(is.na(x)));*

*colswithallmiss <-names(allmisscols[allmisscols>0]);*

*print("the columns with all values missing");*

*print(colswithallmiss);*

*myvars <-*

*names(dfSV) %in% c(colswithallmiss)*

*dfSV <- dfSV[!myvars]*

*dim(dfSV)*

*str(dfSV)*

*summary(dfSV)*

*# remove columns with constants , otherrs*

*write.csv(dfSV, "newdfSV.csv")*

## SQL Scripts to load data to Teradata Database

****

## SQL Scripts to get values and their distribution

****

## SQL Script to create the new dataset



# Relationship between a client's demographic profile and when that client registers for services?

It is to find the relationship between client demographic information such as their gender, race, rank, branch, time in service, current job status and their spouse status and when they actually registered for HHUSA client services program.

## Data preparation

Each client information was available in Contacts dataset including their required demographic information. But Contacts dataset has about 400 columns, so we filtered the dataset and prepared new dataset with only 14 columns as below:

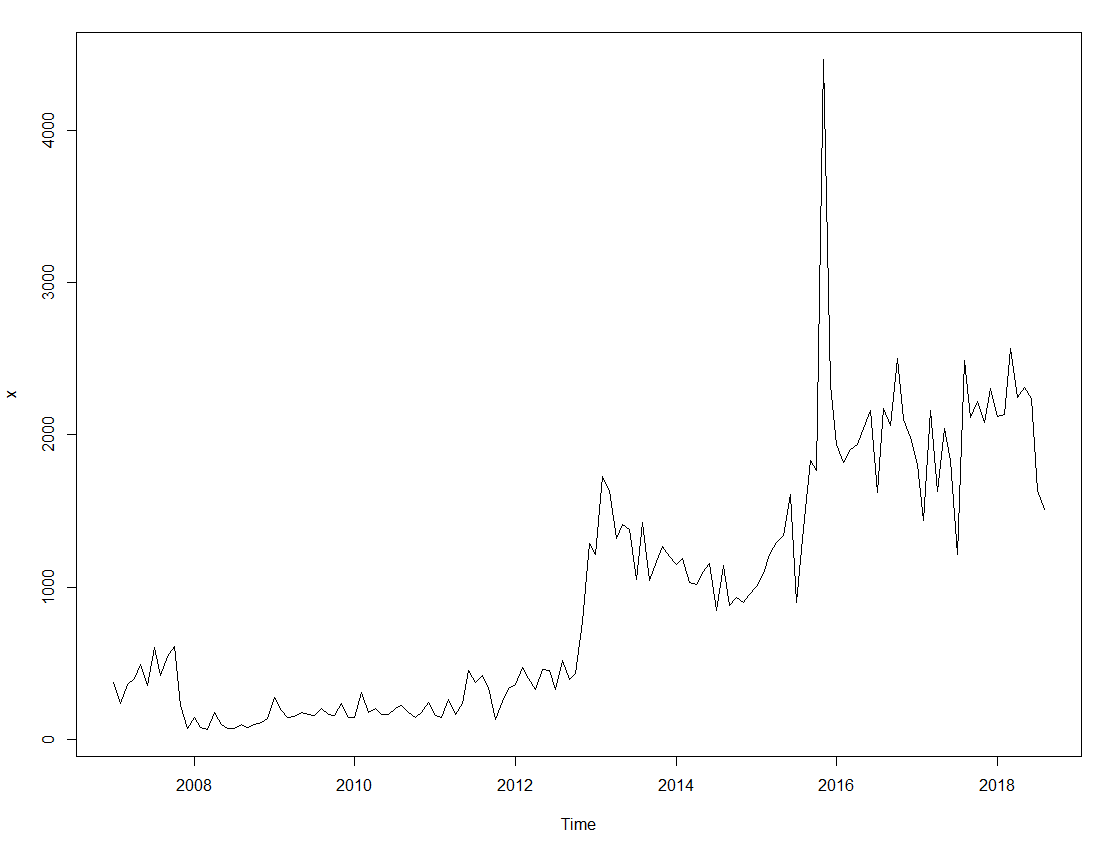
Id, CreatedDate, OwnerId, Gender\_\_c, Race\_\_c, Interview\_Skills\_\_c, Status\_\_c, Service\_Branch\_\_c, Service\_Rank\_\_c, Job\_Board\_Access\_\_c, Date\_of\_Service\_EntryNew\_\_c, Date\_of\_SeparationNew\_\_c, Years\_In\_Service, Military\_Spouse\_Caregiver\_\_c

Years\_In\_Service = Date\_of\_SeparationNew\_\_c - Date\_of\_Service\_EntryNew\_\_c

## Modeling and Visualization

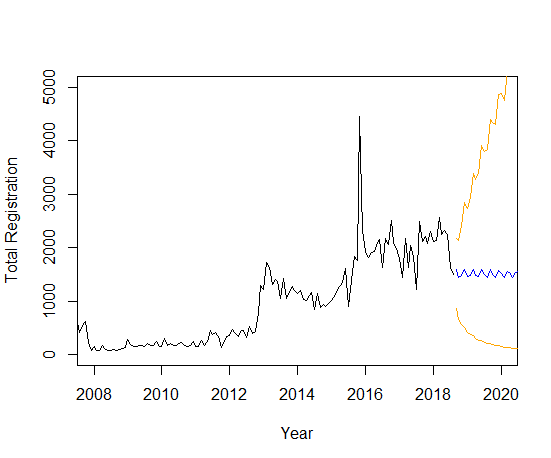
We were not able to find Time in Service because 45% date columns data were blank. So we focused on the remaining columns such as Gender, Race, Branch, Rank etc. To see the trend in client registration we tried to implement ARIMA forecasting model. Initial registration trend plot looks like below.

There was linear incremental trend observed in the total client registration and sometime huge upward spike also noticed.



Below is snippet of code for ARIMA model

|  |
| --- |
| library(forecast)  ARIMAfit <- auto.arima(z, approximation=FALSE,trace=TRUE)  summary(ARIMAfit)  # Use the best ARIMA model to forecast future scales  pred <- predict(ARIMAfit,n.ahead=24)  pred  # Plot the data  # Remember initial log-transformation?  par(mfrow = c(1,1))  plot(x,type='l',xlim=c(2008,2020),ylim=c(1,5000),xlab = 'Year',ylab = 'Total Registration')  lines(10^(pred$pred),col='blue')  lines(10^(pred$pred+1\*pred$se),col='orange')  lines(10^(pred$pred-2\*pred$se),col='orange') |

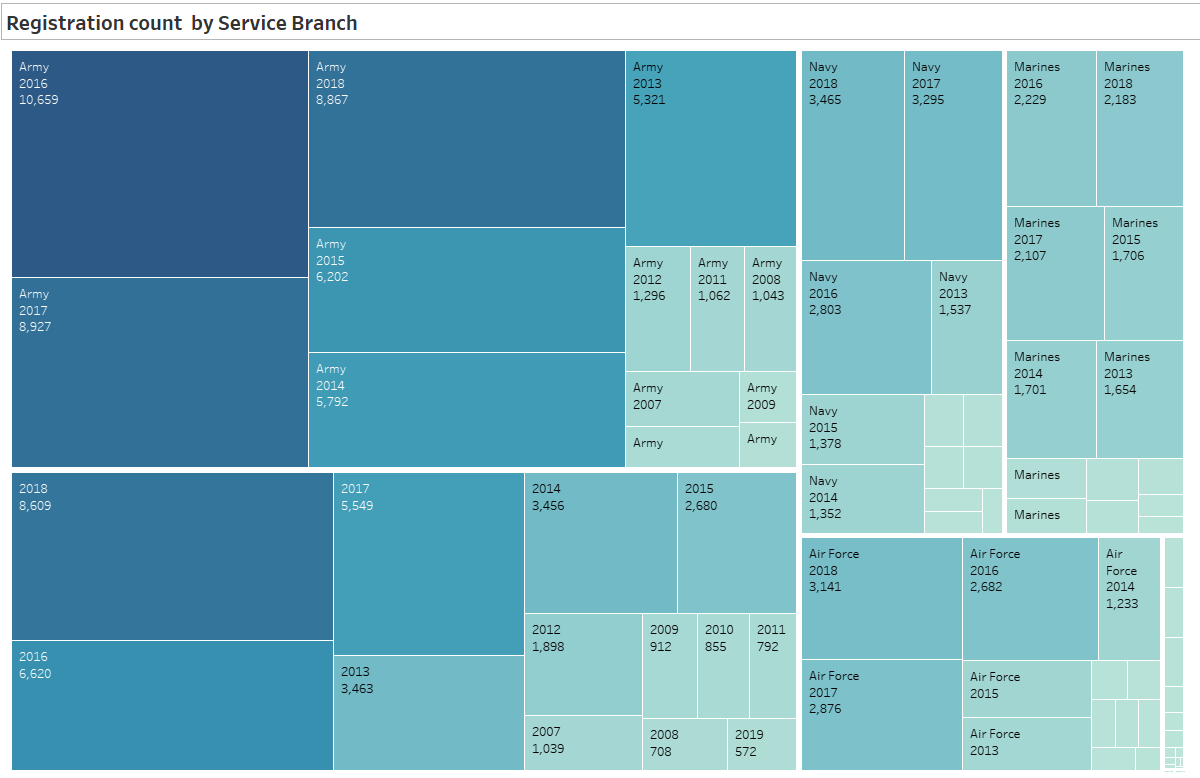


Entire code for ARIMA model can be found in below attached R script.



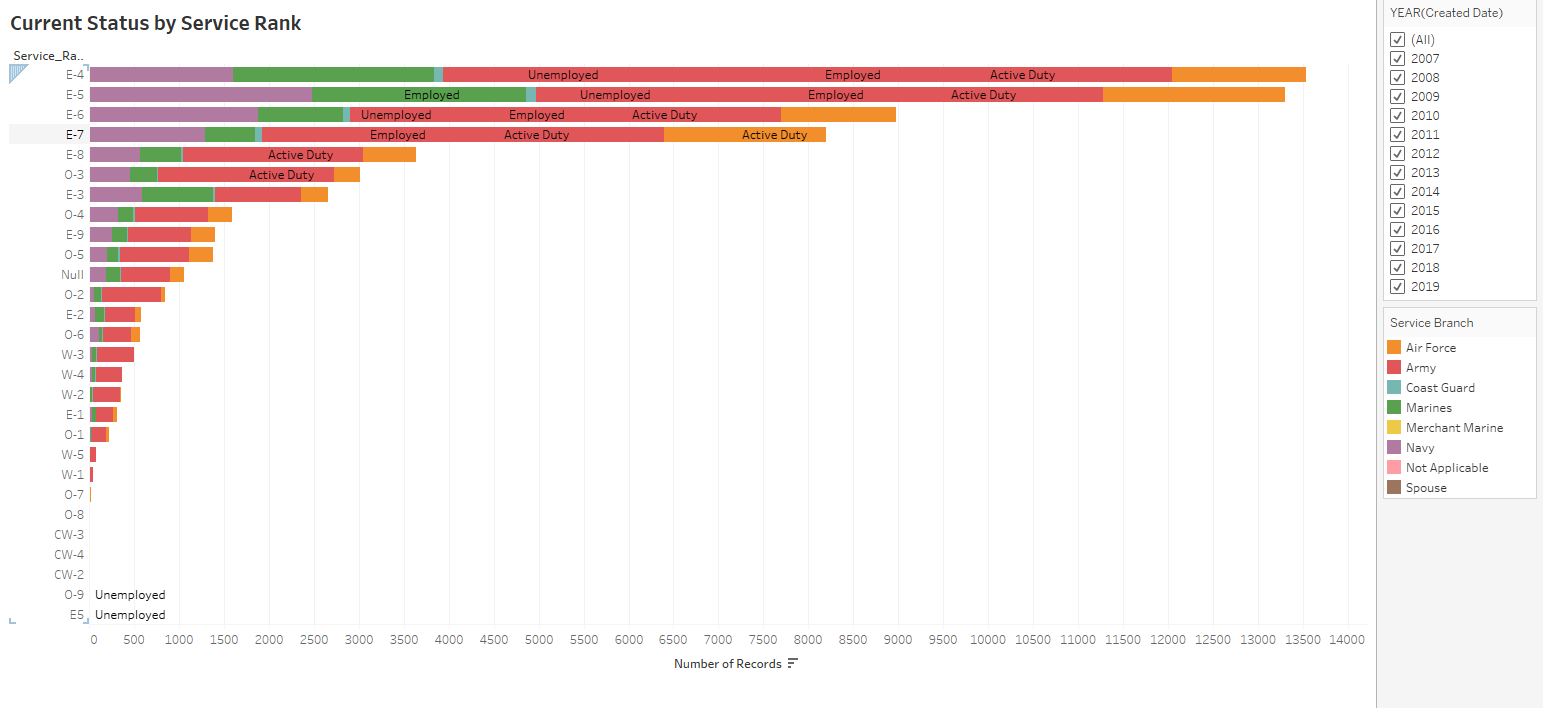
* **Registration count by Service Branch**

This tree-map plot shows distribution of clients registered based on their Service branch.



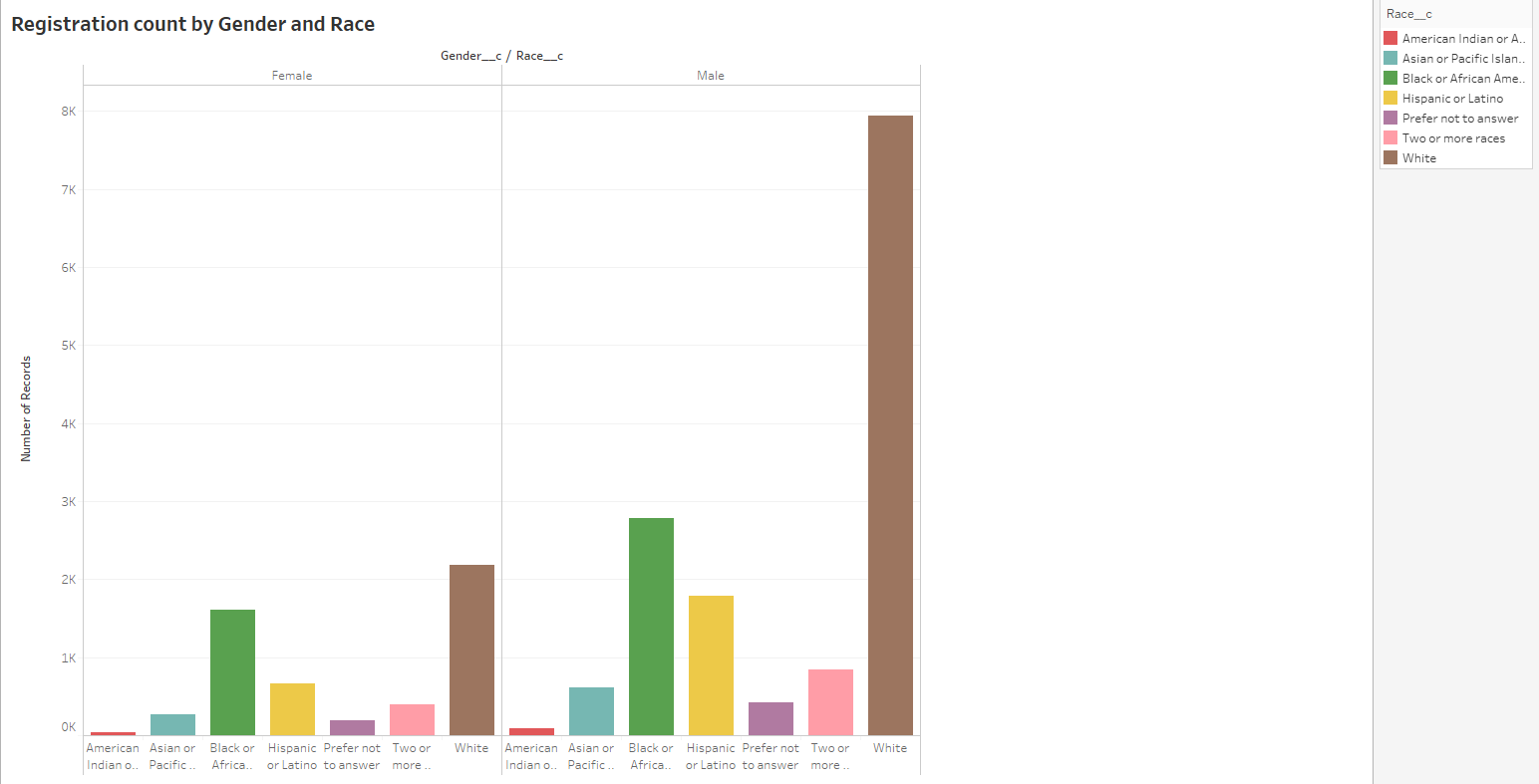
* **Current Status by Service Rank**

This bar chart shows all the registered HHUSA clients and their service rank along with their current employment status.



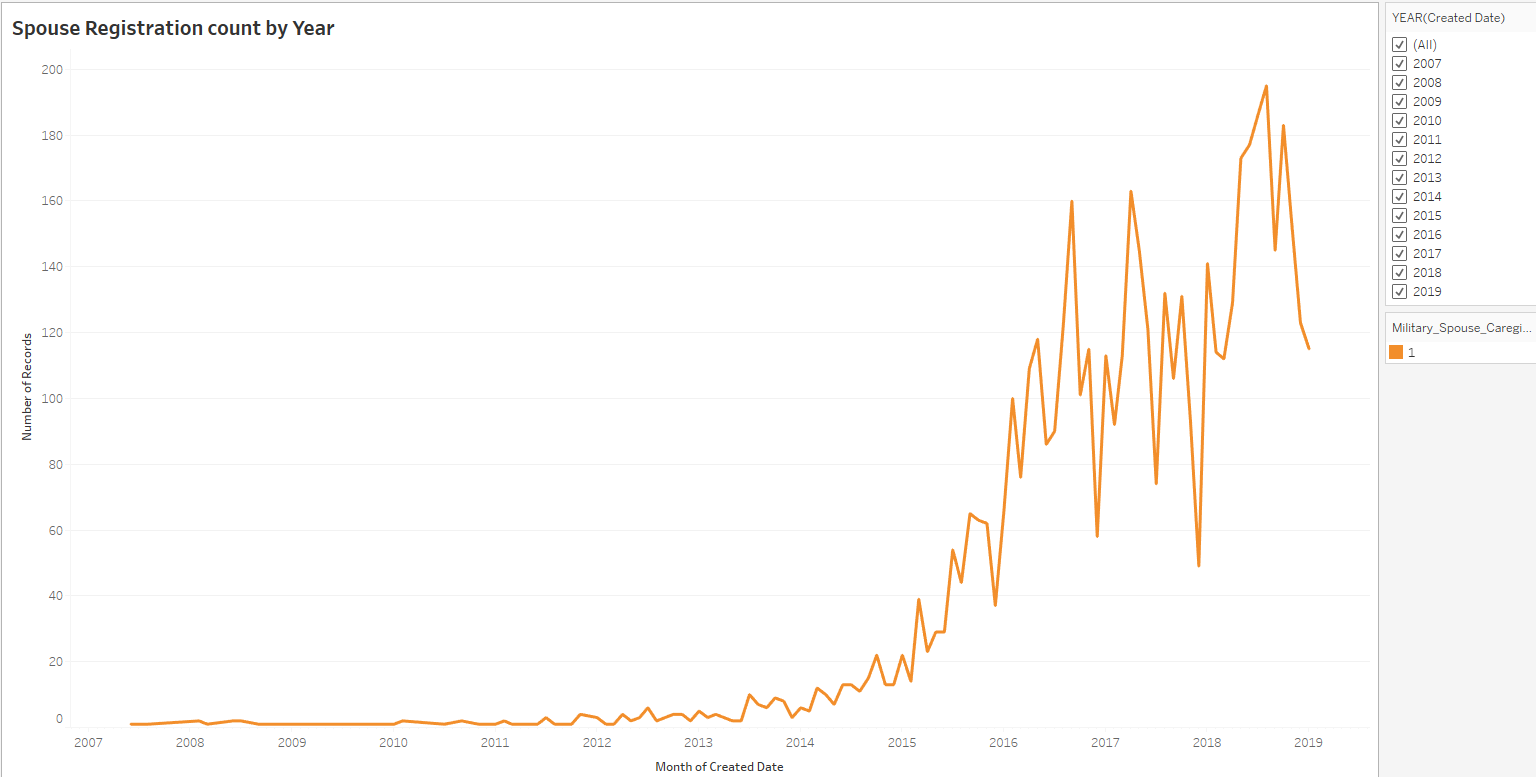
* **Registration count by Gender and Race**

This bar chart is to show the distribution of HHUSA clients by their gender and Race.



* **Spouse Registration count by Year**

This line graph indicates that overall trend of spouse registration increased over the year.



Complete dashboard can be found at: <https://public.tableau.com/profile/nityanand.kore#!/vizhome/HHUSA-ClientServices/HHUSA-ClientRegistrationByDemographic>

## Conclusion

Based on the visualization here are the conclusion on relationship between client demographic and when client registered for the HHUSA services:

* Majority of the veterans from Army (51K) branch are registering followed by Navy (15K) and Marine and Air force (13K each).
* Almost all branches veterans with Rank as E-4 are more interested in HHUSA services irrespective of their current job status (Active Duty, Employed, Unemployed etc.)
* White Male and Female are highest percentage of people registering for HHUSA followed by Black African or American Male/Female.
* Spouse registration also shows significant growth over the time as spouses are using services provided by HHUSA.

# Relationship between the amounts of time spent working with individual and how quickly they are employed?

We are going to investigate whether there is any relationship between the amount of time spent working with individual and how that affects a candidate’s ability to get employed.

## Data Preparation

For that we need to calculate the below 2 derived fields namely ‘**Days\_to\_get\_hired**’ and ‘**Time\_spent\_on\_clients**’ using the below formula.

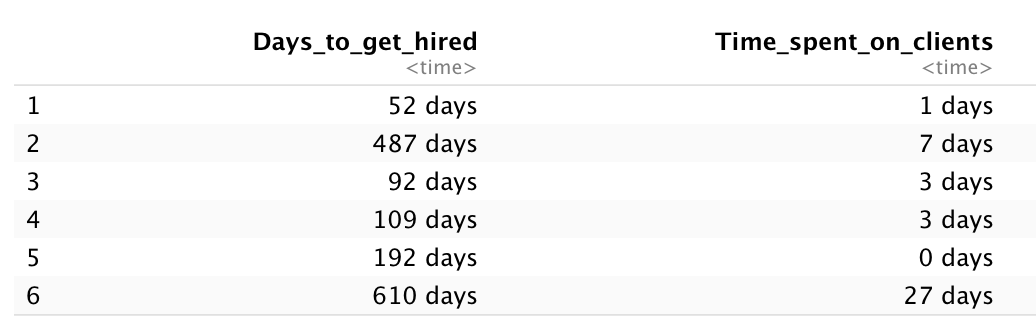
|  |
| --- |
| * **Days\_to\_get\_hired** = Date\_Turned\_Blue\_\_c - Date\_turned\_green\_\_c * Days\_between\_Assigned\_and\_Assessed = c\_Dat\_Initial\_Assessment\_was\_Completed\_\_c - c\_Date\_assigned\_to\_staff\_\_c * Days\_between\_Assessment\_and\_Resume = c\_Date\_Resume\_Completed\_\_c - c\_Dat\_Initial\_Assessment\_was\_Completed\_\_c * **Time\_spent\_on\_clients** = Days\_between\_Assigned\_and\_Assessed + Days\_between\_Assessment\_and\_Resume |

For that we first used ‘lubridate’ library to format the date fields as per the date time stamp using the below code.

|  |
| --- |
| library(lubridate)  df1$c\_Date\_Turned\_Blue\_\_c <- date(mdy\_hm(df1$c\_Date\_Turned\_Blue\_\_c))  df1$c\_Date\_turned\_green\_\_c <- date(mdy\_hm(df1$c\_Date\_turned\_green\_\_c))  df1$c\_Dat\_Initial\_Assessment\_was\_Completed\_\_c <- date(mdy\_hm(df1$c\_Dat\_Initial\_Assessment\_was\_Completed\_\_c))  df1$c\_Date\_assigned\_to\_staff\_\_c <- date(mdy\_hm(df1$c\_Date\_assigned\_to\_staff\_\_c))  df1$c\_Date\_Resume\_Completed\_\_c <- date(mdy\_hm(df1$c\_Date\_Resume\_Completed\_\_c)) |

And we then calculated the fields using the below code and the calculated fields looked as below.

|  |
| --- |
| library(dplyr)  myDF <- myDF %>%  mutate(Days\_to\_get\_hired = c\_Date\_Turned\_Blue\_\_c - c\_Date\_turned\_green\_\_c) %>%  mutate(Days\_between\_Assigned\_and\_Assessed =  c\_Dat\_Initial\_Assessment\_was\_Completed\_\_c - c\_Date\_assigned\_to\_staff\_\_c) %>%  mutate(Days\_between\_Assessment\_and\_Resume =  c\_Date\_Resume\_Completed\_\_c - c\_Dat\_Initial\_Assessment\_was\_Completed\_\_c) %>%  mutate (Time\_spent\_on\_clients = Days\_between\_Assigned\_and\_Assessed + Days\_between\_Assessment\_and\_Resume)  head(myDF) |



We then examined the range of values for these two variables to see if there are any missing values, zero or negative rows using summary().

|  |
| --- |
| finalDf$Days\_to\_get\_hired <- as.integer(finalDf$Days\_to\_get\_hired)  finalDf$Time\_spent\_on\_clients <- as.integer(finalDf$Time\_spent\_on\_clients)  summary(finalDf) |
| **OUTPUT:**  Days\_to\_get\_hired Time\_spent\_on\_clients  Min. :-2170 Min. :-1408.00  1st Qu.: 61 1st Qu.: 1.00  Median : 117 Median : 3.00  Mean : 174 Mean : 14.15  3rd Qu.: 218 3rd Qu.: 9.00  Max. : 2113 Max. : 2654.00  NA's :2673 NA's :1038 |

We then removed the zero, negative and missing values using the below code.

|  |
| --- |
| finalDf <- finalDf %>%  filter(Days\_to\_get\_hired > 0) %>%  filter(Time\_spent\_on\_clients > 0)  summary(finalDf) |
| **OUTPUT:**  Days\_to\_get\_hired Time\_spent\_on\_clients  Min. : 1.0 Min. : 1.0  1st Qu.: 63.0 1st Qu.: 2.0  Median : 119.0 Median : 4.0  Mean : 177.2 Mean : 16.6  3rd Qu.: 218.0 3rd Qu.: 10.0  Max. :2113.0 Max. :2654.0 |

## Modeling and Visualization

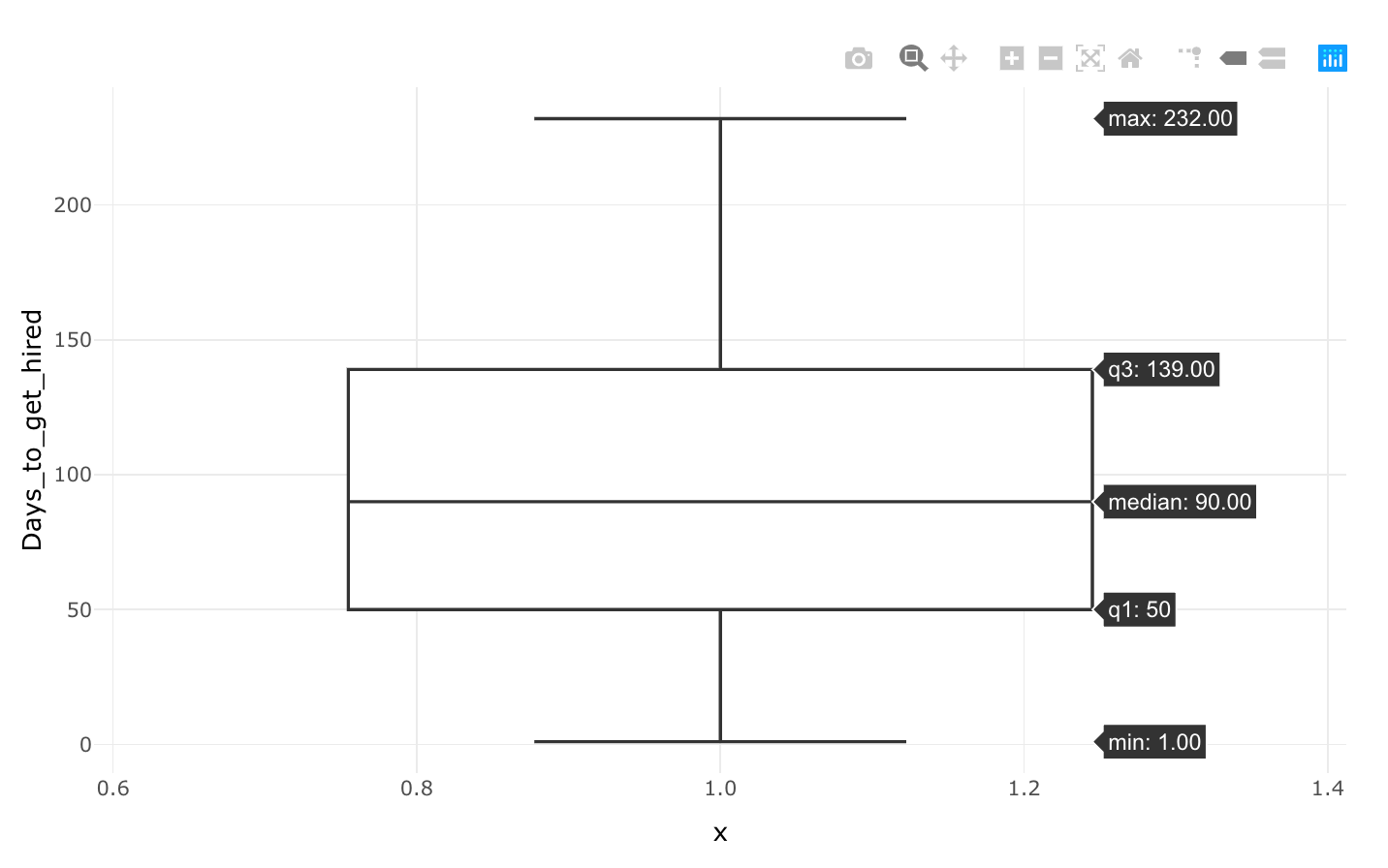
We then examined the outliers of “Days\_to\_get\_hired” using boxplot.

|  |
| --- |
| library(ggplot2)  library(plotly)  bp <- finalDf %>%  ggplot(aes(x=1,y=Days\_to\_get\_hired)) +  geom\_boxplot() +  theme\_minimal()  ggplotly(bp) |
|  |

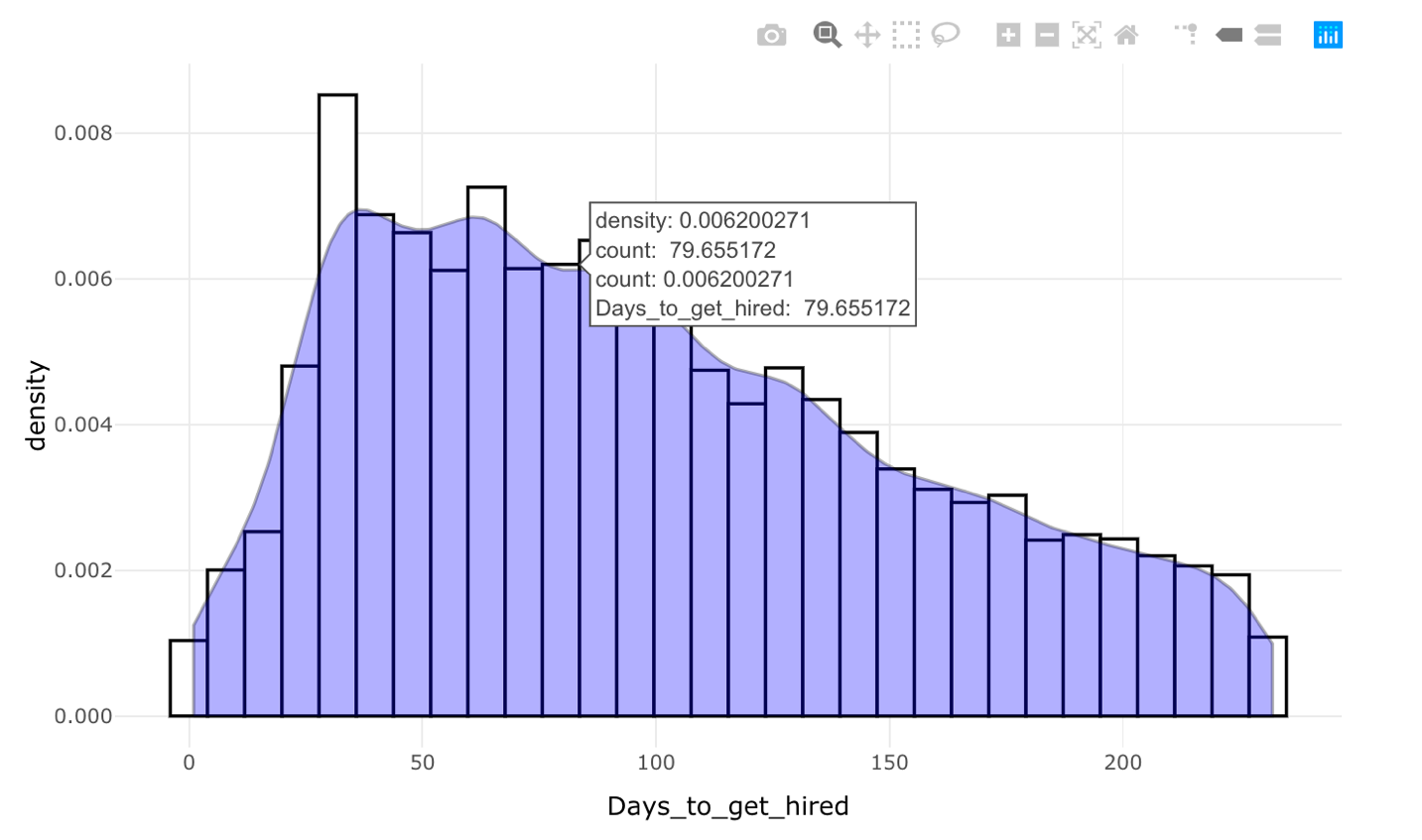
We removed the outliers using given code and the data looked as below.

|  |
| --- |
| finalDf <- finalDf %>%  filter(Days\_to\_get\_hired < (1.5\*(218-63))) %>%  filter(Time\_spent\_on\_clients < (1.5\*(10-2)))  summary(finalDf) |
| **OUTPUT:**  Days\_to\_get\_hired Time\_spent\_on\_clients  Min. : 1.00 Min. : 1.000  1st Qu.: 50.00 1st Qu.: 2.000  Median : 90.00 Median : 3.000  Mean : 98.11 Mean : 3.757  3rd Qu.:139.00 3rd Qu.: 6.000  Max. :232.00 Max. :11.000 |

The boxplot after removing outliers looked as below.



The density distribution of the “Days\_to\_get\_hired” looks as below.



The plot of the two variables namely “Time\_spent\_on\_clients” and “Days\_to\_get\_hired” shows that, for the same time spent on multiple clients, the number of days required to find a job has many records.

Plotting a Linear Regression also show that the line is almost flat and hence may be a

|  |
| --- |
| p <- finalDf %>%  ggplot(aes(x=Time\_spent\_on\_clients,y=Days\_to\_get\_hired)) +  geom\_point() +  geom\_smooth(method = "lm") +  theme\_minimal() |
|  |

Running a linear regression between them shows that the coefficient is almost flat with no variations.

|  |
| --- |
| model <- lm(Days\_to\_get\_hired ~ Time\_spent\_on\_clients, data=finalDf)  summary(model) |
| Call:  lm(formula = Days\_to\_get\_hired ~ Time\_spent\_on\_clients, data = finalDf)  Residuals:  Min 1Q Median 3Q Max  -105.258 -47.130 -8.381 39.745 136.995  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 93.8793 0.7679 122.261 < 2e-16 \*\*\*  Time\_spent\_on\_clients 1.1253 0.1624 6.931 4.35e-12 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 57.67 on 15285 degrees of freedom  Multiple R-squared: 0.003133, Adjusted R-squared: 0.003068  F-statistic: 48.04 on 1 and 15285 DF, p-value: 4.351e-12 |

R code used for this modeling and visualization can be found in below attached object.



## Conclusion

We see that there is no proper relation between amounts of time spent working with individual and how quickly they are employed.

# Effective Communication method for their success in job placement or some other outcome?

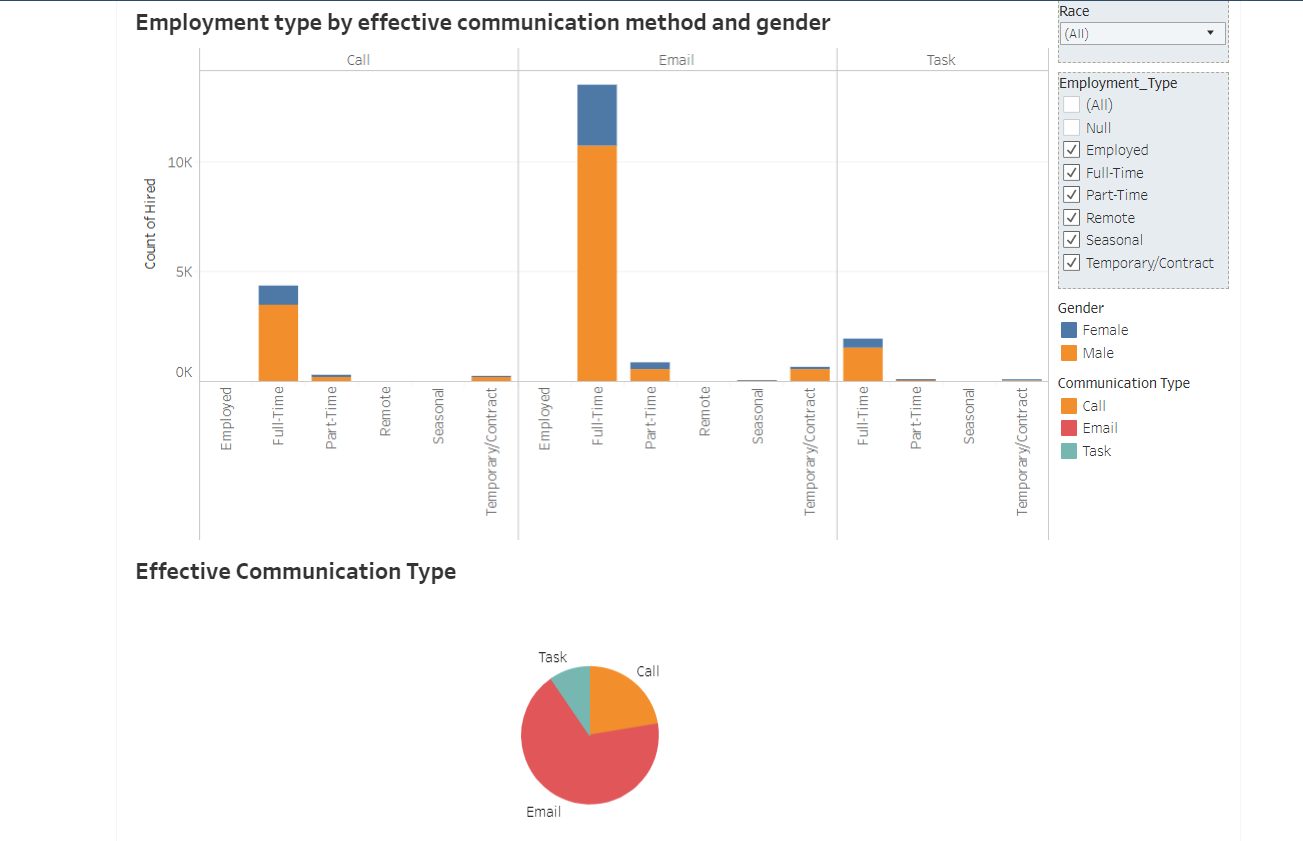
Primarily we are going to find the relationship between the Hired and the most effective communication method used by the hired candidates.

## Data preparation

* We use combined dataset with following variables: Variables considered for this: a\_TASKSUBTYPE, c\_Race\_\_c, c\_Gender\_\_c, h\_Confirmed\_Hired\_Date\_\_c
* Also, we have cleanup in one of values in ‘a\_TASKSUBTYPE’ in Sales Activities Dataset, we replace ‘Task’ to ‘Text’ categorical value.

## Modeling and Visualization

We Used combined datasets which had already performed data imputation and cleanup on insignificant variables



Complete dashboard can be found at:

<https://public.tableau.com/profile/nityanand.kore#!/vizhome/Q4_CommunicationMethods/HHUSA-EffectiveCommunicationtypeforsuccessfulhiring>

## Conclusion

* **Email** was the most effective communication method for the candidate to get hired.
* Also, we used visualization to compare data using different filter criteria like Race, Employment type.

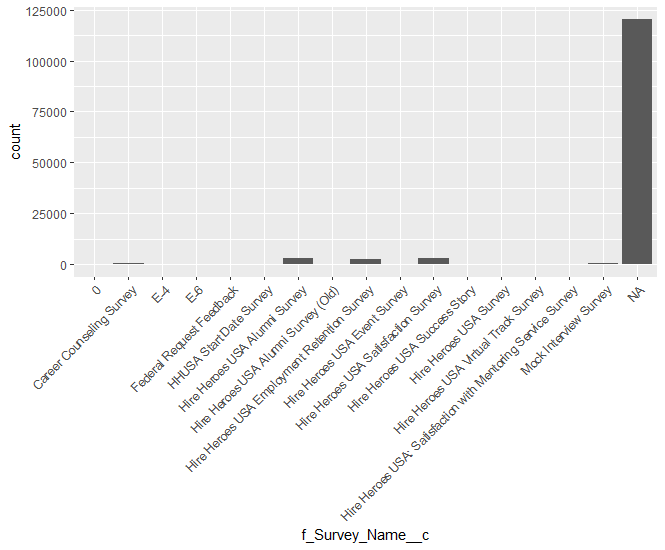
# Relationship between a client's demographic profile and a client's likelihood to complete a survey?

For this question we combined the contact and feedback datasets to achieve our analysis and modeling.

As with the other data sets in this project there was a huge issue with nulls and missing values. Especially the columns in X related to retention survey sent were almost all ‘NA’. Therefore, we decided to drop those for modeling. Our first aim was to have a dependent variable and we decided on “f\_Survey\_Name\_c”. It had entries for different surveys that a client has completed or else it is blank. We created a new variable “Survey” by transforming it with 0 for a blank cell and 1 for any survey name mentioned (which is undergone). Thus, creating our dependent variable for this analysis.

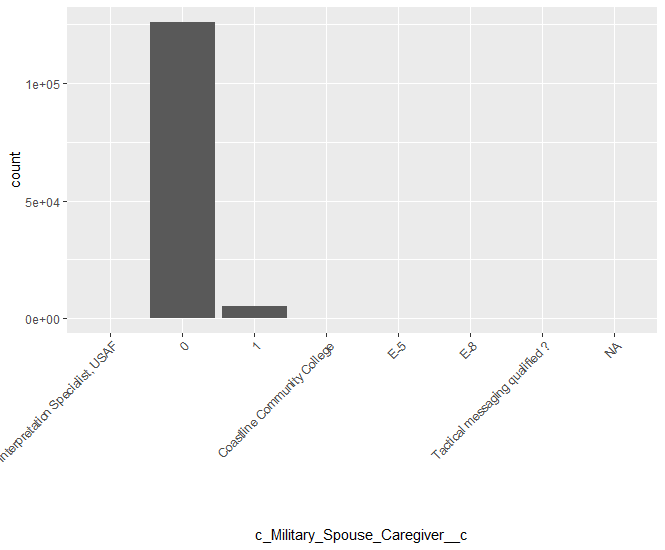
## Data preparation

The further we visualized our data with plots we got to understand the distribution by count, by categories within a variable as to how we would perform any further transformations.



We saw many NA for survey column which showed that many clients had not taken a survey. Therefore, this analysis would be a good opportunity to understand why a client does or does not take part in a survey? Are there any relationships that we can unearth during this analysis?

If a client is a military spouse caregiver or not was also an important variable to understand,



**Feature Selection**

There were still few categories within it which showed the level of support for the spouse if any. Thus, we transformed this to another binary column “Caregiver” with 0 for cells with 0 value and 1 for all other categories.

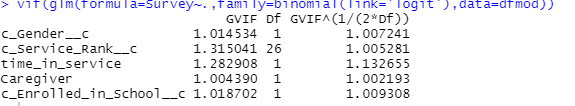
We calculated a new variable “time\_in\_service” by subtracting “c\_Date\_of\_Service\_EntryNew\_\_c” from “c\_Date\_of\_SeparationNew\_\_c”.

We created a subset of this data frame for our specific question and included the following columns

* Survey: Survey conducted or not,
* c\_Gender\_\_c: Gender Male and Female,
* c\_Service\_Rank\_\_c: Rank,
* time\_in\_service: time in service,
* Caregiver: military spouse caregiver or not,
* c\_Enrolled\_in\_School\_\_c: is the client enrolled in school?

Next, we converted the “time\_in\_service” to integer. Also filtered the column for only positive values as there were outliers. We converted all the columns to factor except “time\_in\_service”. We imputed the 376 missing values for service rank feature for max values.

We checked the multicollinearity for our dataset by running the vif function,

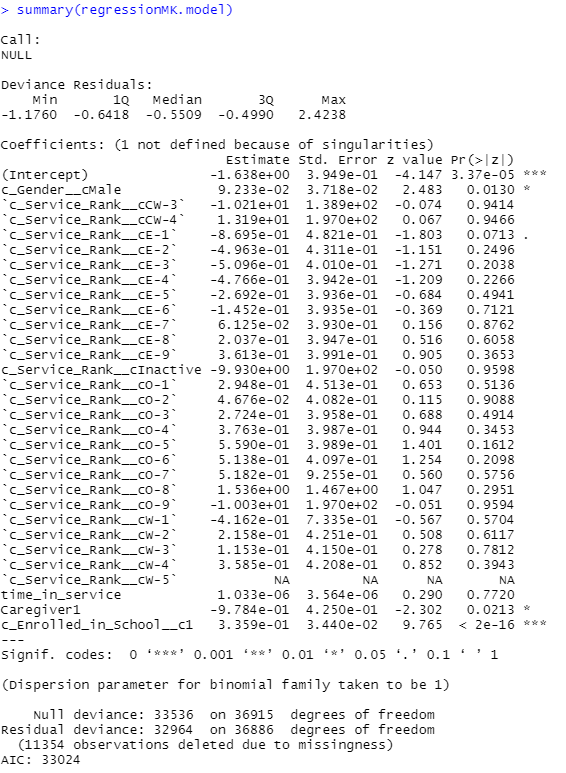


This showed us that there was no multicollinearity in our dataset.

## Modeling and Visualization

As our dependent variable was binary, we conducted a simple logistic regression to understand the features and their relationships better.

We split that data into train and validation with 70-30. Performed a regression on train to get the summary of the model,



## Conclusion

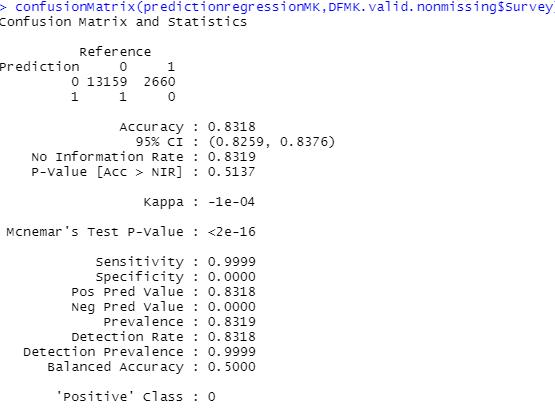
Our model shows us that “c\_Gender\_cMale”,“c\_Enrolled\_in\_School\_c”, “Caregiver1” are significant features.

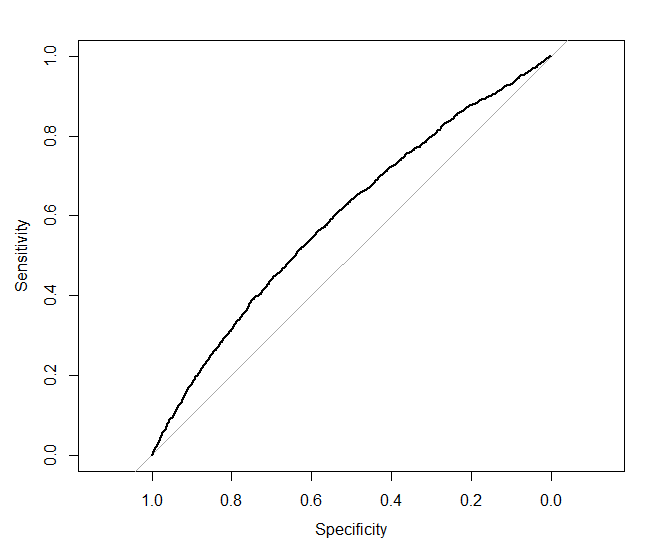
We can interpret that,

Keeping everything else constant female clients are less likely to complete a survey with respect to male clients with the log odds of 0.09233.

Keeping everything else constant clients’ who are caregivers for their spouses are less likely to complete a survey with respect to clients’ who aren’t with the log odds of 0.978.

Keeping everything else constant clients enrolled in school are more likely to complete a survey with respect to clients not enrolled in school with the log odds of 0.0335. The enrollment of a client in school is very significant indicator for relationship between a client completing a survey or not.





Accuracy for our model is 83.18% and Area under the ROC is 0.5954.

R code for this question is below:



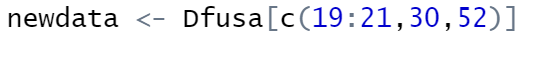
# Any Client's demographic profile that impacts to be a confirmed hire or any other outcome?

It is to find the relationship between client demographic information such as their gender, race, educational background and their ability to find employment through HHUSA client services program.

## Data preparation

We compile a dataset which had about sixty-two columns and carried out Exploratory Data Analysis (EDA) to try to understand the relationships between different independent and depend variables. To write codes to clean my dataset, normalize and create different visualizations. This required dropping and imputing missing values. After a thorough analysis, we filtered the dataset and prepared a new dataset with only five columns as below:

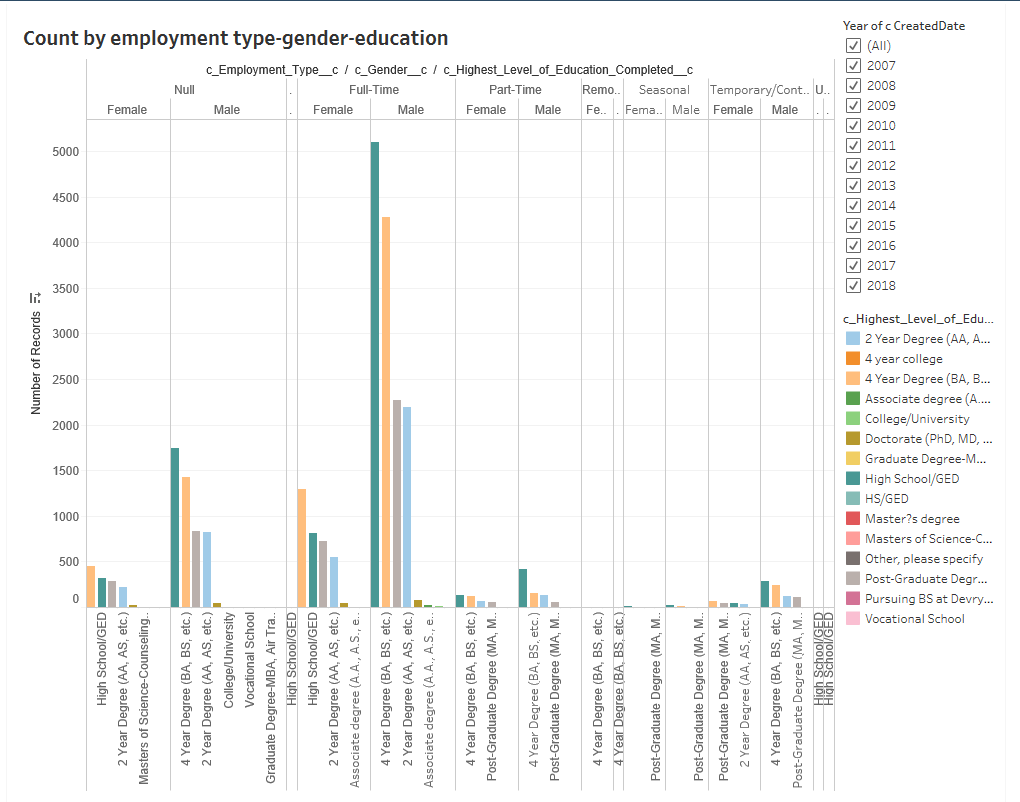
c\_Gender\_\_c, c\_Race\_\_c, c\_Highest\_Level\_of\_Education\_Completed\_\_c, c\_Enrolled\_in\_School\_\_c, Hired.



## Modeling and Visualization

We did many visualizations to better understand the relationships between different independent variables(c\_Gender\_\_c, c\_Race\_\_c, c\_Highest\_Level\_of\_Education\_Completed\_\_c, c\_Enrolled\_in\_School\_\_c) and dependent variable (Hired). We tried to build a model but were not able to because the dependent variable had only one outcome which was hired. Run other linear model to establish relationship but the results did not show any significant correlation. Throughout the data we have, the majority are male. For example, below visualization show that a male with a high school diploma is likely to find a full-time employment than his fellow female.

Online dashboard can be available at: <https://public.tableau.com/profile/nityanand.kore#!/vizhome/HHUSA-Emp-TypebyEducation-Gender/HHUSA-EmpTypeByGenderEducation>



## Conclusion

Based on the visualization here are the conclusion on relationship between client demographic and when client registered for the HHUSA services:

* A male with a high school diploma is likely to find a full-time job than his fellow female, followed a four-year college degree.
* A while male has a higher probability of securing a full-time employment than other races.
* White Male and Female are highest percentage of finding employment for HHUSA followed by Black African or American Male/Female.
* In all areas of employments (full time, part time and seasonal), men lead women and white male led all other races.

# Tools and Technologies used

* R Studio
* Tableau
* Teradata
* MS - Excel

# Appendix and References

<https://www.hireheroesusa.org/about/>

<https://en.wikipedia.org/wiki/Brian_Stann>

<https://r4ds.had.co.nz/> - R Basics

<https://socviz.co/> - R Visualization