Capstone I Proposal

October 22, 2020

Chris Sanchez

Background/Context

I currently volunteer as the Head of Analytics at a non-profit Veteran Service Organization (VSO) called Elite Meet (EM). EM is a unique VSO in that it caters its services specifically to a very small segment of the overall transitioning military population, namely Special Operations veterans. In this context "Special Operations" refers to those members of the Army, Navy, Air Force, and Marine Corps who are SEALs, Green Berets, Explosive Ordnance Disposal, Fighter and Helicopter pilots, among other specialized communities. EM provides mentoring, job search/job skills workshops and direct access to hiring managers and recruiters at top companies, to it's members which currently number 1,000+. EM is also a social network for the Special Operations community writ large that allows it's members to stay in-touch post-military.

The Question/Proposal

To date, there is not a clear picture of what this transitioning Special Operations cohort looks like. From a hiring perspective, I'd be interested to know:

- How educated is this population? What type of degrees do they have?
- What are their target industries/companies?
- Where do they live and where are they looking for jobs? How mobile are they?
- What are their previous skillsets and how mature (in terms of age) are they?

From an organizational perspective I'd be interesting in knowing:

- At what point in their transition are these veterans connecting with EM?
- What are the recruitment trends of time (time series analysis)?
- At what point in their careers are they transitioning?
- What is the community breakdown (SEALs Green Berets, etc)?
- How prepared are they for Corporate America based on education level, experience, and informational interviews conducted?

The Data

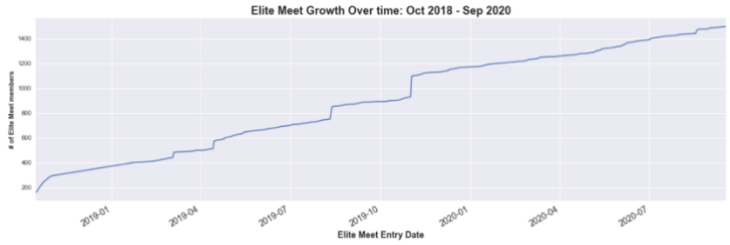
I have in my possession a relatively small (1,496 entries x 44 columns), but comprehensive, anonymized dataset that will allow me to answer all of the questions posed above (among others). This dataset was obtained from the EM Salesforce instance that is used to collect all information from EM applicants at their time of entry into the organization. Pictured below is a snapshot of what this dataset looks like using the pandas.info( ) method:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1496 entries, 0 to 1495
Data columns (total 44 columns):
Title                                          236 non-null object
Description                                    11 non-null object
CreatedDate                                    1496 non-null object
CreatedById                                    1496 non-null object
LastModifiedDate                               1496 non-null object
LastModifiedById                               1496 non-null object
SystemModstamp                                 1496 non-null object
npe01__HomeEmail__c                            1178 non-null object
Preferred_First_Name__c                        831 non-null object
What_is_your_current_location__c               1097 non-null object
LinkedIn_account_link__c                       771 non-null object
How_did_you_hear_about_Elite_Meet__c           758 non-null object
Are_you_currently_seeking_to_join_Elite__c     699 non-null object
Your_transition_to_civilian_life_begin__c      576 non-null object
Undergrad__c                                   779 non-null object
Undergraduate_School__c                        882 non-null object
Grad_Degree__c                                 572 non-null object
Graduate_School__c                             381 non-null object
Years_of_Service__c                            865 non-null object
Military_specialty_in_real_words__c            933 non-null object
Whichcourse__c                                 850 non-null object
When_did_you_graduate__c                       597 non-null object
What_was_your_first_unit_after_grad__c         744 non-null object
What_locations_do_you_WANT_to_live_in__c       729 non-null object
Where_do_you_NOT_WANT_to_live__c               547 non-null object
What_would_be_your_ideal_next_job__c           670 non-null object
Target_Industries_Pls_pick_up_to_Three__c      722 non-null object
of_informational_interviews_completed__c       541 non-null object
of_job_formal_interviews_completed__c          515 non-null object
of_counting_getting_a_job_through_EM__c        620 non-null object
What_s_more_important_to_you__c                715 non-null object
How_mobile_are_you_and_your_family__c          650 non-null object
Current_Industry__c                            83 non-null object
I_d_like_to_participate_in__c                  24 non-null object
Are_you_actively_recruiting_for_a_job__c       79 non-null object
What_positions_are_you_recruiting_for__c       37 non-null object
Did_you_serve_in_the_military__c               76 non-null object
If_so_what_was_your_military_job__c            23 non-null object
Undergrad_major__c                             834 non-null object
Grad_Degree_Type__c                            139 non-null object
Company__c                                     399 non-null object
state__c                                       658 non-null object
city__c                                        633 non-null object
Lead_RecordType__c                             804 non-null object
dtypes: object(44)
memory usage: 514.3+ KB
```

Figure 1. Snapshot of Elite Meet dataset

At first glance one can see that this analysis will not be a numerical analysis, but will rather rely on the multitude of categorical variables, to paint a picture of what this unique veteran cohort looks like. The data set is fairly clean but will require some touchups and as well as ensuring that open responses are standardized. Initial cleaning will start in Excel and migrate to pandas.

As an example of the possible here's a down and dirty snapshot of Elite Meet membership growth over time over the last 2 years from Oct 2018 – Sep 2020:



Value

I think there is immediate value from conducting this analysis from two points of view. The Elite Meet management team will get a holistic view of the members that they are currently serving. The information gleaned can help tailor future services based on what the veterans themselves are actually saying they are interested in. There is also a marketing/recruiting component to this analysis that the management team may find useful to help target their efforts across their currently available marketing channels (web, social media, email, word of mouth, etc.).

Secondly, from a general hiring perspective, this analysis can provide hiring partners who are associated with Elite Meet a robust snapshot of the talent pool that this elite community represents. This analysis will allow them to move beyond the banner headline of "Hire a Spec Ops vet because it's a cool thing to do" to "Hire a Spec Ops vet because not only do they have experience leading in complex, high stress environments, but they are also generally educated at X level, possesses this general X skillset, and are interested in the following X industries/companies."

Weaknesses

There are several weaknesses with this Capstone proposal as it currently stands. In order to showcase my current data science skillset I wanted to work on a project that utilizes the following components:

- Original dataset creation through webscraping. (I already have the dataset of interest)
- Analytics pipeline development:

  Raw data from internet → PostgresSQL → Extraction into notebook → Publish results

  (Raw data extraction from internet is not required nor is storing the data in a database required)

- Working on a question requiring Hypothesis Testing and p-values. (Given the descriptive statistics nature of this project I am currently hard-pressed to see how this component would fit into the analysis, open to suggestions)

- In depth analysis of a problem (While there will be some analysis required to answer some of the questions posed above, this project, as it currently stands, is largely descriptive in nature)

Minimum Viable Product (MVP)

The MVP for this project is self-defined as a beautifully designed and comprehensive report in the form of a README.md file hosted on Github that showcases the story of the Elite Meet population. This will involve several visualizations to likely include a time series analysis, choropleth maps of the United States, bar chart comparisons and other descriptive tools. Recommendations for improving the quality of the incoming Elite Meet membership data will be provided as well as general suggestions for future EM services.

MVP+: Interactivity with the various charts/graphs. Publishable reports for EM hiring partners. Comparisons of the data of this cohort with other population groups.