Git Package Summary

Team 7

December 13th, 2018

INFM 600

**5.7.6.3 A: Your audience and the decisions your analysis targets**

Our target audience starts off with low-income women of color. We want to specifically invest in this demographic because of the higher ROI as well as social impact. Our next target audience and stakeholders are local, state and federal government, local nonprofits, and boot camp companies like Revature. The decision that we concluded through the analysis target is that there should be more investment for low-income WOC to get technical education so that there would be no long-term need for government assistance.

**5.7.6.3 B: Brief description of the source data and processing**

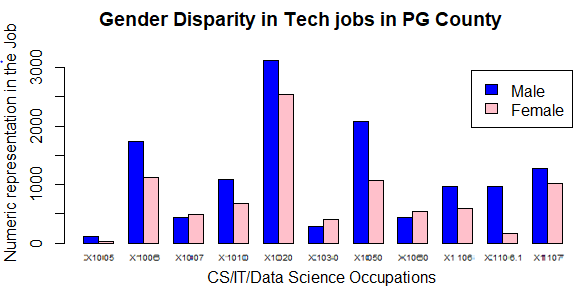
The first data source is from the Maryland Department of Planning (2018). The two datasets we have first highlighted the socioeconomic characteristics of the 24 counties while the second one is the Equal Employment Opportunity (EEO) Detailed Report highlighting PG County’s civilian labor force. The second data source is from Montgomery County (2018) regarding education and tuition assistance. The dataset, which is updated twice a year, includes a list of classes approved and paid through the County tuition assistance program, the degree this class curriculum is a part of, the Department, and the amount for the class.

The third data source is from the government site, the Bureau of Labor Statistics

For our project, we collected 3 datasets from occupations employment projections, labor force participation, and educational certificates and licenses. These datasets are huge and very encompassing because it was done on a national level by the federal government. They highlight the labor force, by demographic, employment by occupational group, fastest growing occupations, worker characteristics, as well as education and training assignments by detailed occupation. Our final data source comes from PG County’s 2018 fiscal year budget, spotlighting the Department Of Social Services budget alongside the approved budget for grants.

The metadata for the data sources ranges from a simple overview of the data described in a sentence to a paragraph summary. Some of the data sources had metadata that explained the frequency the data was being updated, that type of information in the data source and/or limitations of the data. Most of the metadata explains any sources and any specification when reading the dataset. The data sources from the state government had much more metadata than data sources from the federal government which we found to be interesting.

Luckily because our data sources are from government entities, we have not encountered any major issues with the data like a lot of missing values, unstandardized content, entity matching, etc. From our analysis, the datasets available are fairly cleaned and the data sources address any potential misunderstandings in the metadata. The only remediation that we will be doing on our part is filtering the dataset to highlight the scope of our project. For example, highlighting the data that reflects race, class, education, jurisdiction etc. Because the data we seek might be the only a part and not the sum of the in these very large datasets, we will be doing some small data cleaning to narrow the scope.

**5.7.6.3 C & D: A figure (plot) from your analysis and our interpretation of the plot** 

It is clearly visible that in eight out of eleven tech jobs with a focus in CS/IT/Data Science in Prince George’s County more men are employed than women. Also, we know from our dataset that the proportion of men and women for all races in PG County is a nearly 50-50 split. Since our focus has been women of color we can infer from the above graph that women of color are underrepresented in these tech roles and therefore our idea of making accessible bootcamps for them with help of government and NGO’s can be of significant help. We have made use of ggplot2 to get this visualization. Also, a lot more goes into this because we were asked to include a key/legend. We have therefore added a legend for showing the color code for men and women.

**5.7.6.3 E: A persuasive argument for a decision, our audience should make based on our results**

The decision for our target stakeholders (local, state and federal government, local nonprofits, and boot camp companies) is quite clear after our analysis. Stakeholders can sustainably invest in short-term education for WOC that would decrease the need for social dependency from the government, effectively increasing the standard of living as well of the quality of life for the whole DMV region. With the high poverty rate, large population with some college education, and the extremely high ROI on technical education, this practical, cost-effective, and sustainable option is the best way to the most social and financial impact for the region as a whole. With PG County’s unique socioeconomic demographic, it is clear that this is the perfect community to invest in to receive the most return, effectively decreasing long-term government assistance.

**5.7.6.5: A brief document summarizing contributorship to the individual project deliverables for the entire project**

|  |  |  |
| --- | --- | --- |
| 5.7.1 | Work Plan: | ET |
| 5.7.2 | Data Preparation & Documentation: | ET PB DY |
| 5.7.3 | R Script Draft: | DY |
| 5.7.4 | R Plot Draft: | PB |
| 5.7.5 | Presentation: | ET |
| 5.7.6 | Git Package: | ET DY |

Throughout the project, it was clear what roles everyone played based on skill and past experience. Parth acted as the technical coordinator, Ebanezare the project coordinator and Dan as the support. For the Data Preparation & Documentation, Parth and Dan did the intellectual policy, licensing, and the textual description of the data cleaning process part of the assignment. For the Git Package, Dan handled the final data cleaning documentation and scripts with the data source location as well as an analysis script and any outputs that it generates section.

**Word Count: 984**