**MEMORANDUM**

**TO:** Kieng Iv, Director at the Department of Arts and Culture

**FROM:** Celinda, Business Analyst at the Department of Arts and Culture

**DATE:** March 17th, 2023

**SUBJECT:** Recommendations for Language and Cultural Programs in the United States

I am writing to you to summarize the findings from the analysis we have done using the 2013 US Census data to recommend language and cultural programs in the United States. I understand that you have asked me to examine the English proficiency of individuals depending on the age of the individual, the state they are from, and the languages spoken at home to understand how to improve the English education we have, where, how, and what language resources we require to implement these services. These findings will help us determine how best to implement translation, English education, and other additional cultural programs such as language education (especially Spanish) across the country.

The dataset is the 2013 US Census data where a sample of 10,000 individuals is chosen from the first people dataset. The sample was generated randomly in R and filtered to provide us with the necessary columns of data we require to perform our analysis. The guide to using the visualization (in the order below) is as follows. We note that not all visualizations are displayed.

Our first task was to understand the demographic of the sample being used. For this analysis, the state was recorded as a categorical numeric variate so I created a calculated column to turn the states into state abbreviations to be able to turn it into a geographical variate. We notice that our data mostly comes from California, a couple states on the East Coast, and Alaska. This makes sense since the dataset is sorted alphabetically by state and we are using the first dataset. I also took a look at the ancestry breakdown to understand the ancestry of individuals.

Our first ask is to learn about the English proficiency by state. This can be done through a bar chart of the English proficiency compared. We note that since the observations by state differ, we do not want to look at the absolute count but a percentage of the total by state. We also want to exclude all the observations that are “NA” since these results aren’t useful to us. From the results, it seems like the top three states that have the lowest English proficiency are Minnesota, Hawaii, Indiana. This can be seen in **Figure 1**.

**Figure 1:** English Proficiency by State Graph

Chart, bar chart

Description automatically generated

We can also break down the results even further by creating a calculated column where individuals that can speak English very well or well are grouped together as “proficient” and individuals who cannot speak English at all or do not speak it well are grouped together as “not proficient”. This can give us greater insights on the breakdown of English proficiency and understand which states require the greatest amount of English education. The results can be seen in **Figure 2**.

**Figure 2:** Breakdown of English Proficiency by State

Chart, bar chart

Description automatically generated

We can define states that require English education as states where English proficiency is less than 75% (i.e. the percentage of individuals that are “not proficient” as defined above is greater or equal to 25%). We can see states that do require greater English education are Minnesota, Hawaii, Indiana, Louisiana, Idaho, Kansas, and the District of Columbia. We can also filter the data by age category. What we do find is that English proficiency is significantly lower for senior citizens than any other category of individuals.

We can further examine this by comparing the age category of individuals and their English proficiency. We first note that infants and toddlers all have “NA” as their English proficiency so this data is not included. We notice that the English proficiency is the lowest among senior citizens at 67.05% followed by adults at 77.36%. English proficiency is highest among teenagers at 97.92% followed by children at 91.81%. Thus, we notice that the most pressing age groups are adults and senior citizens. The results are shown in **Figure 3**.

**Figure 3:** English Proficiency by Age

Chart, bar chart

Description automatically generated

We can further dig into the states where English proficiency is lowest among senior citizens and adults from our visualization shown in **Figure 2**. We note that from filtering the results for senior citizens, most states have English proficiency lower than 75%. What is most pressing are the results in Minnesota with an English proficiency of 0%, Kansas with an English proficiency of 0%, Georgia with an English proficiency of 25%, California with an English proficiency of 61.26%, Hawaii with an English proficiency of 60%, and Massachusetts with an English Proficiency of 61.11%. For adults, the results are less pressing with many states having an English proficiency between 60% to 80%. The states with the lowest adult English proficiency are Minnesota, Indiana, Louisiana, Hawaii, and the District of Columbia.

We can further break this down by noting that for senior citizens that are not proficient in English, the language they speak most at home is Spanish followed by Vietnamese, Mon-Khmer, Chinese, and Russian as seen in **Figure 4**. This tells us that the US Department of Arts and Culture needs to implement translation services from languages above to English for services that senior citizens use more frequently such as the healthcare system, public transportation, government-run senior homes, etc.

**Figure 4:** English Proficiency by Language at Home

Chart

Description automatically generated

Overall, we have determined that we want to focus our English language programs in Minnesota, Georgia, Kansas, California, Hawaii, and Massachusetts for senior citizens specifically. We note that this includes English-language learning programs geared towards senior citizens and translation services for services that are accessed frequently by senior citizens such as government senior/retirement homes, grocery stores, hospitals, other medical centers, etc. We also note that the English-language learning programs and translation services should be primarily focused on Spanish, Vietnamese, Mon-Khmer, Chinese, and Russian to English. We note that there are also many other states and age categories that have low English proficiency. However, due to the limitation in the resources of the Department of Arts and Culture, we want to focus our efforts in improving English proficiency on our largest problem area, as demonstrated through the visualization and analytics done with the sample data.

We note that some limitations of our data is that it is from 2013 so the English proficiency may have changed since then, as well as the demography of individuals surveyed. Additionally, we only used one dataset so data in some states are unavailable. Our next step would be firstly to perform the same analytics on a sample of the other people dataset.