Melinda Fraser

11/5/2020

**Capstone Proposal**

For my capstone project I am accessing two datasets from the open data on the TucsonAZ.gov website provided by the City of Tucson. My two data sets both include Neighborhood names and Ward areas (1-6). The data set on racial demographics in neighborhoods include population count of ethnic background per neighborhood for White, Hispanic, Black, Asian, Native American, and others. In the education achievement data set we also have count of population that did not complete HS, completed HS, GED achieved, some college, completed Associates, completed Bachelors, and post graduate degrees per neighborhood. These data values are presented in strings and integers. I will be merging the two data sets and grouping this information by wards.

My proposal is to review the data from these two data sets to see which neighborhoods or ethnicities need more attention for education funding, or social outreach programs to be developed.

My hypothesis is that the Wards (collection of neighborhoods in pre-defined areas of the city) with a higher percentage of white ethnicity will have higher education completion attainment.

I am going to test this hypothesis by finding the correlations between ward and ethnic background, ward and education attainment, and ethnicity and education attainment.

My findings, whether my hypothesis proves true or is unfounded, will be invaluable for identifying problem areas that impede the citizens of Tucson, AZ from completing HS and post-secondary education. The City of Tucson, non-profits for under-represented groups, and school districts would find this information useful. These users can review this data and use it to make decisions for funding, for creating programs to mobilize certain neighborhoods, to create programs to encourage certain ethnicities to complete more high school and find funding for them to pursue post-secondary education.