Date: 08:56 AM CDT, 07/06/2018

From: Kenan Fikri <kenan@eig.org>

To: dzg943@my.utsa.edu, max.kilger@utsa.edu, danny.tagle@gmail.com, steven.j.alexander@usa.net, tut549@my.utsa.edu, uaa784@my.utsa.edu, gabriel.tellez18@gmail.com, nicholas.zamora@my.utsa.edu

Subject: Fwd: UTSA - City of San Antonio Project (Additional Info Request)

Hi team,

My sincere apologies for the late reply! I hope the information can still be useful in your project. Wonderful to see you incorporating the DCI into your work! I'm really glad it's serving as a model to educate the next generation of analysts and data scientists.

I've put my responses in red below. I'm super impressed that you were able to find all the exact ACS tables; we normally keep the exact metrics confidential but that's moot here and I'm pleased to affirm that you've identified them all correctly.

Again, I'm really sorry for the delay in getting back to you. You were in great shape without my input though, so hopefully you've been plowing full steam ahead!

All best,

Kenan

--
Kenan Fikri
Director for Research
Economic Innovation Group | 616 H Street NW | Washington, DC 20001
Desk: 202-839-3712 | Cell: 202-681-7697 | kenan@eig.org | @kenanfikri

Disclaimer: This message and any attached documents are provided by the Economic Innovation Group for informational and educational purposes only. They are not intended to constitute legal or professional advice and should not be relied upon or treated as a substitute for consultations with professional, legal, or other competent advisers.

---------- Forwarded message ----------
From: **Nicholas Zamora** <nicholas.zamora@my.utsa.edu>
Date: Tue, Jun 26, 2018 at 3:15 PM
Subject: UTSA - City of San Antonio Project (Additional Info Request)
To: info@eig.org
Cc: dzg943@my.utsa.edu, max.kilger@utsa.edu, danny.tagle@gmail.com , steven.j.alexander@usa.net, tut549@my.utsa.edu uaa784@my.utsa.edu , gabriel.tellez18@gmail.com

Kenan Fikri
Director for Research and Policy Development
Economic Innovation Group
616 H Street NW
Washington DC 20001

Dear Mr. Fikri,

We are master's degree students in our last semester at The University of Texas at San Antonio in the Data Analytics program. One of our summer courses, Data Analytics Practicum II, assigns us to address a real-world problem for an actual client. Eight of us are working on a project requested by the City of San Antonio, Texas.

San Antonio is a city of 1.5 million people, the seventh largest in the United States. It is notable, sadly, for economic segregation and poverty. City leaders and staff have been studying your Distressed Communities Index with great concern and urgency. They have asked our class to use our new data analytics skills to help them understand the insights of your work and to help them develop policies to improve life in our most distressed communities. Since most of us live in San Antonio, this is a very meaningful project to us.

First, we are awestruck with the DCI section of your website. We've learned statistical analysis and visualization, but this presentation sets an impressive standard for which we'll strive in our imminent careers.

We want to begin our work by repeating your analysis, to ensure that we understand it and to create a foundation from which we can give our city hall clients the tools to explore the details of DCI and evaluate potential remedies.

We recognize that you have a busy schedule, that you cannot tutor every correspondent around the country, and that some of your methods may be confidential. We will appreciate any guidance and correction you can provide

Is the following correct, or should we modify our approach? Our understanding is that you used counts and percentages from the American Community Survey of 2011-2015, Business Patterns 2011, and Business Patterns 2015. You use the survey estimates, not the margins of error, from the ACS.

All correct

- "No high school diploma" is from census table S1501 on the line "Percent high school graduate or higher." You subtract those percentages from 100 to get the percent with no high school diploma.

Yes

- "Housing vacancy rate" is from census table B25002 ("total" number of housing units) and B25004 ("total" vacancy minus "for seasonal, recreational, or occasional use"). The percentage is the vacancy calculation divided by the number of units.

Yes

- "Adults not working" is from census table S2301 on the line "Population 25 to 64 years" in the "EDUCATIONAL ATTAINMENT" section, and the column "Employment/Population Ratio." You subtract those percentages from 100 to get the percent not working.

Yes but table S23001 I think

- "Poverty rate" is from census table S1701 on the line "Population for whom poverty status is determined" and the column "Percent below poverty level."

Yes

- "Median income ratio" requires median income for each zip code and for each state, which are from census table S1903 on the line "Households" and the column "Median income (dollars)." It also requires a table that identifies the state that contains each zip code. The ratio is the number of each zip code divided by the number of the corresponding state.

Yes

- "Change in employment" is from census tables CB1100CZ11 and CB1500CZ11 in the column "Paid employees for pay period including March 12 (number)." The change in employment is the ratio of the 2015 number minus the 2011 number, divided by the 2011 number, multiplied by 100.

Yes

- "Change in business establishments" is from census tables CB1100CZ11 and CB1500CZ11 in the column "Number of establishments." The change in business establishments is the ratio of the 2015 number minus the 2011 number, divided by the 2011 number, multiplied by 100.

Yes

We had 33,120 zip codes from the American Community Survey and 39,032 zip codes from the Business Patterns survey. We eliminated any zip code for which one or more statistic was missing. This left us with 24,645 zip codes. Your website mentions that you ranked over 26,000 zip codes. **Can you tell us how to recover the 1500+ zip codes difference, or give us a list of the zip codes you used?**

We eliminate all zip codes with populations (from the ACS) *below* 500, and we circumvent employment suppressions by using the establishment growth rate as a proxy for job growth as well. We will actually change that methodology in the next DCI coming out in the fall to improve accuracy. You are perfectly justified in eliminating these zip codes and proceeding with a smaller sample set; your methods are otherwise the same and you'll have successfully replicated the DCI.

Our next step was to convert the DCI raw values (e.g. percentage without high school diploma) to a DCI score. **Was this the right approach? Would we get the same DCI scores if we had the same number of zip codes?**

Yes, solving the zip code count should leave you in good shape. We have a total of 26,126 zip codes.

- For each DCI component, we sorted the zip codes so that the most desirable score (e.g. lowest percentage without high school diploma) was first and the least desirable score was last. We assigned each zip code an integer from 1 to 24645.

Yes

- For each zip code, we averaged the seven DCI rankings.

Yes

- We sorted the zip codes by the average of the seven DCI rankings. We assigned each zip code an integer from 1 to 24645.

Yes

- We divided each zip code ranking by 246.45 and rounded to one decimal place, giving DCI scores from 0.0 to 100.0.

Yes, but let me know if your results differ; I might need to dig into the code to see if there's another step.

We would be happy and honored to share our work in progress with you, if you would like. On behalf of The City of San Antonio, The University of Texas at San Antonio, and our class, thank you for your time, your attention, and your assistance.

Please do keep me posted!

Yours,

Meredel Ababa, Steven Alexander, Mike Mahoney, Nathan Shepherd, Austin Somlo, Daniel Tagle, Gabriel Tellez, Nicholas Zamora