Data Analytics Practicum II

June 19, 2018

One approach to using our time together more efficiently would be:

1. We each give a short update on our work in progress
2. We discuss upcoming milestones (once we set milestones)
3. We talk about our problems and challenges and help each other with solutions
4. We set new goals and assignments as the data reveals new lines of inquiry
5. We “adjourn” as a group and confer one-to-one on shared tasks

Any thoughts on this structure?

We reviewed Steven & Mike’s “Ideas for DCI Project,” and we each picked some tasks:

Ababa, Meredel:

* Work with Austin to plot the San Antonio zip codes. Here are some suggestions for the plots. Please modify as you both see fit.
  + Each zip code has a bar chart with eight bars: the seven DCI rankings and the ranking of the averages
  + Color code the bars and use the legend to identify the color of each DCI component
  + Taller bars are good; calculate Height = 24645 – ranking
  + We’ll have to produce multiple charts to include all zip codes:
    - Sort by zip code and include the zip codes in the chart title
    - All charts should be to the same scale
  + Use notable images for each zip code in San Antonio (Alamo, Tower of Americas, Six Flags, SeaWorld, Toyota, etc.) if there’s room

Alexander, Steven:

* Draft second email to EIG for class review
* Publish data dictionary for “DCI Components v2.csv”

Kilger, Max:

* Does UTSA has guidelines for fonts, colors, and logos, particularly applying to ggplot and PowerPoint?
* Does UTSA has any copyright rules about intellectual property created by students on a faculty-led project?
* Could you ask CSA (City of San Antonio) whether they have guidelines for fonts, colors, and logos?

Dr. Kilger’s answers:

1. I asked the SA city fellow about colors, logos, etc. and have received no answer.  Because this is a city of SA project I would not worry about UTSA colors, logos, etc. Design it the way that works best.
2. The practicum attracts interested organizations and companies partly because the terms of the practicum give the organization the rights to the IP.  This also harmonizes the external policy where IP created by the student belongs to the company or organization.

Mahoney, Mike:

* Provide QA support to Nathan on the plot of the seven DCI components.

Shepherd, Nathan:

* Plot the seven DCI component ranks and the rank of the average of the DCI ranks. Here are some suggestions for the plots. Please modify as you see fit.
  + Do it two ways: all U.S. zip codes and Texas-only zip codes (San Antonio might look relatively better among Texas-only)
  + X-axis: 24645 (worst position) to 1 (best position) in the country
  + X-axis: 1935 to 1 in Texas
  + Y-axis: the value of the component (e.g. vacancy percent from 0 to 100) – with the “bad” value at the bottom (high vacancy, low job growth, etc.)
  + The line or the scatter plot will go from the bottom left to the top right of the grid
  + Have rectangular areas (like the presidential terms in Dr. Malshe’s assignment colored according to the color code used by EIG
  + Highlight relevant zip codes in two ways (on two sets of charts):
    - Have San Antonio’s zip codes shown with large dots (or squares or whatever) – this will require a categorical variable (“SA” = 0 or 1)
    - Color code all zip codes in Texas (Texas vs. U.S. or Texas by region)
  + Place a San Antonio logo (The Alamo, Towers of America, or the San Antonio skyline?) in the upper left and a UTSA logo in the lower right (roadrunner?)

Somlo, Austin:

* Investigate additional statistics beyond the seven DCI components. Let Gabriel follow along and validate interim results.
* Work with Meredel to plot the San Antonio zip codes. Here are some suggestions for the plots. Please modify as you both see fit.
  + Each zip code has a bar chart with eight bars: the seven DCI rankings and the ranking of the averages
  + Color code the bars and use the legend to identify the color of each DCI component
  + Taller bars are good; calculate Height = 24645 – ranking
  + We’ll have to produce multiple charts to include all zip codes:
    - Sort by zip code and include the zip codes in the chart title
    - All charts should be to the same scale
  + Use notable images for each zip code in San Antonio (Alamo, Tower of Americas, Six Flags, SeaWorld, Toyota, etc.) if there’s room

Tagle, Daniel:

* Test for correlations and clusters. Here are some suggestions along those lines. Please add, change, and delete as the data suggests and as you see fit. Let Gabriel follow along and validate interim results.
  + Between each pair of raw DCI components (especially No\_HS vs. No\_Job, Chg\_Job vs. Chg\_Biz, and Med\_Inc vs. Poverty)
  + Between each pair of DCI component ranks
  + Between consecutive zip codes (can we use autoregressive tools?)
  + Between zip codes in San Antonio with similar DCI rankings
  + Explore clustering with K nearest neighbors and hierarchical cluster analysis

Tellez, Gabriel:

* As people report and upload results (code, data, documentation) try to replicate their work, letting them know (privately) if there are variances between your work and theirs.

Zamora, Nicholas:

* Check that our data includes every zip code in San Antonio.
* Compare DCI raw data and rankings from EIG and from our team for each zip code in San Antonio.