Learn IT Project: Analytics with Google Docs and Google Fusion Tables

You must complete the entire exercise to receive credit.
Check the last section of this document for the deliverables.

In this assignment, you will analyze data on changes in population and income in the United States using Google Docs and Fusion Tables. These tools allow you to easily combine multiple sets of data and plot them on a map.

Make sure you complete the entire assignment, including the questions at the end.

Note that the “Temple Google” accounts do not have access to Fusion Tables. You will need to create a separate Google account to complete the assignment.

Part 1: Upload the Data to Google Drive

Upload the data files to your (non-Temple) Google Drive account. You’ll need to upload all three of them:

- payroll_final.xlsx – Payroll data, by state, for 2009 (taken from the US Census Website)
- migration_final.xlsx – Data on net population growth, by state, for 2009 (taken from the US Census website)
- StateKML.csv – Geographic data for the outline of each US state

Part 2: Prepare the Data for Analysis

1) Open the payroll_final data set and create a new column. Label it PERCAP_PAY – it will contain the average annual salary per person (per-capita pay) for each state. That’s computed by dividing the total payroll by the number of people employed in each state (i.e., the EMPLOYMENT column).

Google Docs works pretty much like Excel, so you can just enter the formula to compute the average into a cell and then copy it down to the rest of the rows.

*Note that TOTALPAYROLL is in $1,000s, so the per capita amounts will also be in $1,000s. That won’t make a difference for this exercise.*

2) Create another new column and label it STDPAY – it will contain the standardized values for per capita annual salary across all states. These standardized values will tell you whether a state is way above or way below the national average. You can standardize the data using the following formula:

\[
\frac{(\text{PERCAP_PAY for the current state} - \text{Average of PERCAP_PAY for all states})}{\text{Standard Deviation of PERCAP PAY for all states}}
\]
Note that in Google Docs there is an AVERAGE() formula and a STDEV() formula.

**HINT: Don’t include the data for the United States (row 2) in your calculations!!**

3) Close the spreadsheet and open the migration_final spreadsheet. Note that there are more state entries in the migration_final data set than the payroll data set.

4) Remove the rows with the states from this sheet that aren’t included in the payroll sheet (there are two of them).

5) Create a new column in this sheet and call it STDMIGR – it will contain standardized values for the net migration for each state.

   Compute the values for that column just like you did in step 2.

**Part 3: Merge the Data into a Single Fusion Table Data Set**

1) Import payroll_final as into Fusion Tables by clicking on the Create button, and selecting More/Fusion Table (experimental).

2) At the Import New Table screen, select Google Spreadsheets and choose the payroll_final sheet. Follow the wizard defaults to import the spreadsheet.

3) Return to the main Google Drive page and do the same for the other two files.

4) Now merge data together by selecting the payroll_final Fusion Table (make sure it has the icon next to it, not the icon; it may also have a “Sheet 1” or “Sheet 2” as part of the name).

   Select File/Merge… and then choose the migration_final Fusion Table (again, it may have a “Sheet 1” or “Sheet 2” as part of the name).

5) At the next screen, make sure you are matching “AREA” with the “State” field of the migration_final table. Accept the rest of the defaults and select “View Table.”

6) When the merge is done, it will have a new name like “Merge of…” You should see both sets of data in one sheet, matched on state (AREA).

7) Now merge this new, merged sheet with the StateKML Fusion Table. Make sure you are matching “AREA” with the “name” field of the StateKML table.
8) It will ask you whether you want to Add columns to the current table or save the result as a new table. Select “Add columns to this table.”

**Part 4: Map and Analyze the Data**

1) Using your new merged data set, click on the “+” tab (after Cards 1) and select “Add Chart.”

2) Click on the icon for the second horizontal bar chart for the chart format ( ).

3) Choose STDPAY and STDMIGR as Values to plot (if you don’t see them Click the Choose… button). Uncheck everything else.

4) Change the Maximum categories value to 50.

5) Now plot the data on a map: click on the “+” tab (after Chart 1) and select “Add Map.”

6) The default map isn’t very helpful, so click on the “Map 1” tab and select “Change map styles.”

7) Under Polygons, select Fill Color. Then go to the Buckets tab. Plot the STDPAY variable by dividing the data into three buckets. You can divide the ranges equally by clicking on “use this range.”

*NOTE: Google Fusion Tables will sometimes pick odd lower and upper bounds, so make sure the lower bound is -1.5 (the first text box) and the upper bound is 3.0 (the last text box). Then click Save.*

8) You will then have a fancy, color-coded map.

9) Now create a second map the same way, but this time plot the STDMIGR variable

*NOTE: Make sure the lower bound is -1.8 and the upper bound is 3.3.*
Deliverables

Email your instructor a Word document containing the following:

1) The title of the assignment and your name at the top.

2) A screen shot of your two maps in a Word document.

3) Answer the following questions. For each question, provide an answer along with evidence to back up your conclusion:
   a) Which states appear to be doing particularly well (high per capita payroll and high net migration)?
   b) Assess the situation in Kentucky.
   c) Look at the maps. Do there appear to be regional patterns? Explain.
   d) Why is Texas “empty?”
   e) Why was it important to remove the two states from the migration_final sheet? (HINT: The merge would have still been successful, so that’s not the problem.)

Choose one variable that we haven’t yet analyzed and plot it, either on a chart or on a map. If you create new, computed measures (such as standardized values or a combination of two or more other variables), you can go back to your original worksheet, do the additional calculations, and then re-merge the data by following the steps in Part 3.

4) Take a screen shot of the new chart or map and insert it into the Word document.

5) Answer the following questions:
   a) What was the variable you chose to analyze? Why did you choose it and what additional insight does it give you?
   b) Write a few sentences about what you learn about the states’ economic activity based on analyzing the new variable.

Credits

This was based on an assignment originally created by Kartik Ganju.