**Purchasing Power of Salaries by State and City Tool** Version Number: 1.1.0

Developer: William Holmes

Publication Date: 1/17/2019

About the application: This application is to allow the customer to select multiple cities to compare the purchasing power of their salaries between the different locations.

About the Data that drives it:

Every year, the American Statistical Association (ASA) hosts a competition called Datafest. As part of the competition, the ASA selects a new “data sponsor” each year from one of many competing companies; the selected company then provides real data from their business and challenges participants to generate useful insights using the data provided. In the 2018 Miami University event, the data sponsor was [Indeed.com](file:///C:\Users\Admin\Desktop\Datafest%20and%20R\William%20Holmes%20-%20Example%20of%20Work\Indeed.com), the job-hunting website. While certain aspects of the data were masked (tokenized and jittered) from the participants for legal reasons and to preserve trade secrets (for example, employers’ names and the job titles advertised), the data itself was still real.

However, participants are also encouraged to seek and incorporate outside data sources as they are available; to identify the real purchasing power of a job’s salary, data from the [US Bureau of Labor Statistics](https://www.bls.gov/) (BLS) was used to supplement the data provided by Indeed.

The data used in the application combines the BLS data mentioned above and a subset of that original Datafest data. In the Job Details by City and State table, each row represents a job advertised by a specific company on Indeed.com, and is uniquely represented by the combination of the CompanyID and the JobID. This app creates an interactive interface to simulate a job-seeker using Indeed’s platform, augmented by data the company already has and/or can freely acquire, to find jobs that are of greater value, thus generating value for both themselves and for Indeed and its customers.

Note: because some jobs are advertised with multiple locations, certain openings may correspond to multiple rows to display the disparate purchasing power that would result from choosing one location over the other.

Why it Matters:

As with [Monster.com](https://hiring.monster.com/solutions/pricing.aspx?intcid=PHP_LFT_CTA_PAJ_C), Indeed likely makes several hundred dollars for every employer that posts an opening on their site. This application would allow Indeed to draw more job-seekers to their site by highlighting the jobs that are most valuable and desirable; in turn, Indeed could selectively use this feature on employers’ behalf as part of a two-tiered advertising system, an expansion on their current model that would generate larger amounts of revenue and would require little to no maintenance or upkeep, thus adding tens or hundreds of thousands of dollars directly to their bottom line every year.

About the Code:

Attached in this folder is a Shiny App. [Shiny](https://shiny.rstudio.com/) is a package in R that translates R code into HTML, JavaScript, and CSS, which allows it to be viewed locally or placed on servers and viewed remotely; it also allows for an interactive analysis by exploiting the point-and-click nature of these web-based languages; finally, those with knowledge of web development can insert HTML, JavaScript, etc. directly into the Shiny app, allowing for greater flexibility during the creation process.

How to Open the Application:

If you have access to this file and the associated files locally on your machine, you can click on any of the R files presented to automatically open them in R, assuming it is installed. If R is not already open, the working directory will be set to the opened file’s location; if R is already running, then the working directory will need to be shifted to the file’s current location so it can execute properly. The included files are:

* app: the Shiny app itself, named as such by default
* AppData: the final dataset used, after all cleaning, joining, and subsetting has been performed; includes data from both Indeed.com and the BLS.
* CensusCity2010: data from the BLS recording the local cost of living for various cities in each of state in the US; collected once every 10 years as part of the census.
* CleaningAndJoining: a script that converts the Datafest data into the data used for the app by excluding irrelevant columns, cleaning the data, joining the BLS data to the Datafest data, and performing necessary calculations.
  + Requires the *dplyr* and *readxl* packages
* factoredData: the original dataset; modified by CleaningAndJoining (see above) to create the final dataset AppData (see above) used in the Shiny app.
  + Requires the *dplyr*, *shiny*, and *shinyjs* packages.

As a final note, because this app has been hosted on Github, it can be accessed any time by loading the shiny package and entering *runGitHub( "Purchasing-Power-of-Salaries-by-State-and-City", "wholmes105")* into the R console (this requires an internet connection to download the app, and will take a few moments before the app launches).

Metadata and Column Definitions:

|  |  |  |
| --- | --- | --- |
|  | **Class** | **Description** |
| CompanyID | Character | The unique identifier for the employer with the job opening. It was masked/tokenized for Datafest to be non-indicative of the employer’s true identity. |
| JobID | Character | The unique identifier for the job title of the job opening. It was masked/tokenized for Datafest to be non-indicative of the job’s actual title. |
| State | Character | The two-digit abbreviation of the US state in which the job opening is located |
| City | Character | The city in which the job opening is located |
| estimatedSalary | Numeric | The estimated annual salary of the position (no adjustments made). |
| CoL | Numeric | The annual cost of living for the city in which the job is located, expressed as a percentage of the national average.  ex. a value of 130 indicates that the local cost of living is 130% of the national average for a given location |
| adjustedSalary | Numeric | The estimatedSalary adjusted to account for the local cost of living; because the local cost of living has been accounted for, this shows the true value of the salary; the calculation is estimatedSalary / CoL \* 100.  ex. if the local cost of living for a city is 1.5 times the national average, then a job paying $60,000 in that city is equivalent to a job paying $40,000 ($60,000 / 1.5) in an “average” city. |
| CoLDollars | Numeric | The annual cost of living for the city in which the job is located, expressed as a dollar amount; the calculation is CoL \* 60,060 / 100. Note: the [2017 cost of living](https://www.statista.com/statistics/247455/annual-us-consumer-expenditures/) nationwide was $60,060.  ex. a value of 85000 indicates that the local cost of living is $85,000.00 |
| CoLRatio | Numeric | The percentage of the local cost of living that the job will pay, expressed as a percentage; the calculation is estimatedSalary / CoLDollars \* 100.  ex. if CoLRatio for a given opening is 87.49, then the opening is expected to pay 87.49% of the local cost of living. |
| Salary | Character | The formatted value of estimatedSalary; includes a dollar sign ($), thousands separator (,), and is rounded to the nearest cent. |
| Adjusted Salary | Character | The formatted value of adjustedSalary; includes a dollar sign ($), thousands separator (,), and is rounded to the nearest cent. |
| Cost of Living | Character | The formatted value of CoLDollars; includes a dollar sign ($), thousands separator (,), and is rounded to the nearest cent.  It is displayed in the Adjusted Average Cost of Living by City table as Annual Avg. Cost of Living. |
| Salary / Cost of Living | Character | The formatted value of CoLRatio; includes a percentage sign (%), and is rounded to 2 decimal places. |