Assignment 1: Analysis of Bank Filings Data – Data Wrangling and Exploratory data analysis using Tableau Case Study

You have been hired by QuantUniversity LLC. as a Financial Sector analyst. QuantUniversity specializes in financial research and uses quantitative tools like R, MATLAB, Python, Tableau for data analysis. QuantUniversity's client has requested designing a dashboard to evaluate banking data for Peer 1 Banks. i.e, banks that have a net asset value of greater than $10B. The client wants QuantUniversity to use publicly available filings data from NIC. The National Information Center (NIC) is a central repository of data about banks and other institutions for which the Federal Reserve has a supervisory, regulatory, or research interest, including both domestic and foreign banking organizations operating in the United States. The following are the tasks your manager Sri has asked you to work on. To ensure you get the best possible results, your manager has asked you to work in teams of three. In addition, your manager has also asked you to submit the following deliverables by **October 14th 11.59.pm**

1. A Powerpoint deck with your process workflow and insights from the 3 parts
2. Python/R scripts to scrape the data
3. CSV files for the scraped data.
4. Finished Tableau dashboards – 2 files for parts 1 and 3

Note: All data and files must be submitted through **github** and shared with analyticsneu@gmail.com

Also, the Powerpoint deck should be uploaded to the discussion board by the deadline.

Part 1: Analyzing Banks with asset values greater than $10 Billion

Data and reference:

See https://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf for different charts/graphs you can use in Tableau.

Data: https://www.ffiec.gov/nicpubweb/nicweb/HCSGreaterThan10B.aspx

In this part, you will be scraping data for "all" 15 quarters for the data on this site using R or Python.

You will then:

1. Consolidate the data into a single csv. Do data cleansing if needed
2. You need to have a stacked version and unstacked version of all data in csv
3. Bring this data in Tableau for Exploratory data analysis
4. Plot all banks on the US Map using location information
5. Dashboard 1: Use Histograms and Pie Charts to describe the banks by value
6. Dashboard 2: Do a Tree map by State
7. Dashboard 3: Rank the top 10 banks, bottom 10 banks
8. Dashboard 4: Using 15 quarters worth of data plot Trend charts – per company to show asset growth; Display actual value and growth rates (Rate = (Value from current quarter /Value from prior quarter -1) *100 Ex: (101/100 – 1) *100 = 1% growth rate)
9. Dashboard 5: Provide the ability to compare 2 banks. Ex: compare trend charts BAC vs JPMorgan
10. Dashboard 6: Show a pivot table of the total asset values by state and quarter

Discuss the results in a Powerpoint slide deck.

Assignment 1: Analysis of Bank Filings Data – Data Wrangling and Exploratory data analysis using Tableau Case Study

Part 2: Data scraping

Reference: See https://automatetheboringstuff.com/chapter13/ and https://www.r-bloggers.com/introducing-pdftools-a-fast-and-portable-pdf-extractor/ for examples

Data:

https://www.ffiec.gov/nicpubweb/content/BHCPRRPT/BHCPR_Peer.htm

Task:

1. Scrape this page
2. Get PDFs for all Peer 1 banks and extract into csv format for all years
3. Name files as Peer1_Year_Quarter.csv
4. Review data and comment on how clean the data is
5. Clean data and normalize format across all files using Python/R

Part 2: Analysis of Banking Organization Systemic Risk Reports for the years 2012, 2013 and 2014

Data:

https://www.ffiec.gov/nicpubweb/nicweb/Y15SnapShot.aspx

As-of Date	Indicators	All Line Items	Other Formats
2014-12-31	2014 Indicators	2014 Line Items	2014 XLSX
2013-12-31	2013 Indicators	2013 Line Items	2013 XLSX
2012-12-31	2012 Indicators	2012 Line Items	

User Guide:

https://www.ffiec.gov/nicpubweb/content/Y15SnapShot/FR%20Y-15%20SnapShots%20User%20Guide.pdf

Data Dictionary:

https://www.ffiec.gov/nicpubweb/content/DataDownload/NPW%20Data%20Dictionary.pdf

Tasks:

Refer to 2014XLSX and 2013XLSX. Review the Graphs and Time Series sheets. Your goals are:

1. Using 2012,2013,2014 (data Indicators, All Line Items), create dashboards in Tableau for the graphs presented in the Graphs and Time Series Sheets
2. What other charts/graphs could be used instead of plain old bar charts to represent information?
3. For the Time Series chart you should be able to do year to year comparison for all 3 years.
4. Discuss your dashboards and results in your Powerpoint presentation.