For the Python bridge course team project, we plan to use a data set from the US Census Bureau which contains statistics for all U.S. firms by Industry, Veteran Status, and Ethnicity for the U.S., States, Metro Areas, Counties, and Places based on the 2012 Survey of Business Owners.   The universe for the 2012 Survey of Business Owners (SBO) includes all U.S. firms operating during 2012 with receipts of $1,000 or more which are classified in the North American Industry Classification System (NAICS).

Natallia proposes to:

1. Modify the original csv data frame: filter the columns **VET\_GROUP.display-label** and **ETH\_GROUP.display-label** that contain “All firms”[[1]](#footnote-1) in order to find answers for the questions below:
   1. What are the predominant industries in USA (2012) by number of companies in this industries, % of total number of companies. In other words : What are the most represented industries in USA 2012? Data frame and bar graph will be created.
   2. What are the RICHEST industries in USA (2012) by sales, receipts, or value of shipments of ALL firms ($1,000), % of total sales, receipts, or value of shipments?
   3. Which industry's average percentage of companies with paid employees is the highest/the lowest?
   4. What is average pay across all industries in US in 2012? & What is average paycheck per industry/employee?
2. Data frames and bar graphs will be created for a through d.

1. Do not need to sum fields☺ [↑](#footnote-ref-1)