**GFDP: Data & Analytics**

**Chinese Company Research**

* Absolute deadline to turn in data: Monday March 20 at 9pm US pacific time
* Next week: R session / tutorial
* Following week (after that): group research & turn in your report
* Then: presentations
* If you did not present US research, you must present China research!
* Goals
  + Predict financials
    - Revenue (size)
    - Employment (size)
    - EBITDA (profit)
    - Market capitalization (value)
  + Using online metrics: primarily global rank since country rank is usually NA
* Data cleaning
  + Be careful: do not eliminate too many observations
  + Look for inconsistencies
* Correlations and analysis run separately by
  + Primary industry
  + Stock exchange
    - Could create country groups for exchanges: US, mainland China, Asia…
  + Size range (employment, revenue)
    - 3 groups: small (< 1000 employees), medium (1000 – 5000), large (5000+)
  + Age range (based on year founded)
    - Internet was created in early 1990s, around 1994
    - Pre-Internet age, post-Internet age
    - Internet might have become popular in China around 2000, so can consider that cutoff…
* Classify industries / companies based on whether they are
  + B2B (business to business: sell to businesses) or B2C (business to consumer: sell to consumers)
    - Some companies are B2B2C: sell to businesses, who sell to consumers
      * These are still B2B
    - Expect B2C companies get more Internet traffic on average
  + Depend on Internet viewership or not
    - Internet business or not
  + Services vs. Manufacturing vs. Technology vs. Construction vs. Retail etc… (broad sectors)
* Transform variables (financials & online metrics): reciprocal (e.g., 1 / global ranking), natural log (revenue, market cap, employees; LN function in Excel), rank (revenue, market cap, employees)
* Predict ratios of financial metrics: run regressions or look at correlations involving ratios
  + Revenue / Employees
  + Market cap / Employees
  + EBITDA / Revenue: EBITDA % (measure of profitability)
  + FOR EXAMPLE:
    - Run regression: ln(revenue / employees) against 1 / global rank
    - Correlation of revenue / employee versus global rank

**U.S. Company Research**

* Data to collect: 100 companies in your industry, 1,000 – 5,000 employees
* Variables to collect (copy the exact **SAME** variable names into your excel)

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| subj\_industry | industry code for your subjective industry of the company if the industry listed is different from what you think the company’s real industry is. (list of industries on BQ Prospect IRS Form 5500: https://bizqualify.com/list\_of\_industries) |
| parent\_comp | name of the parent company if the company is a subsidiary |
| correct\_web | correct official website of the company if BQ website is wrong |
| global\_rank | Alexa global ranking |
| country | Alexa country |
| country\_rank | Alexa country ranking |
| category | Alexa category |
| category\_rank | Alexa category ranking |
| web\_creation\_year | website creation year on Whois |
| web\_reg\_org | Name of website registrant organization |
| privacy\_protect\_indic | website privacy protection indicator: = 1 if protected (i.e., no information found), =0 if not protected (i.e., organization name is valid) |

**Analysis**

* BQ metrics (to be explained / predicted, dependent variables – provide variable names):

|  |  |
| --- | --- |
| **Name of metrics** | **Variable Name** |
| Employment | bq\_current\_employees\_plan |
| Revenue | bq\_revenue |
| BQ growth score | bq\_growth\_score |
| BQ risk score | bq\_risk\_score |
| BQ profitability score | bq\_profitability\_score |
| The 3-year CAGR growth rate of employee plus company contributions to retirement plans (which is an indicator of revenue.) | bq\_growth\_tot\_contrib\_pens\_amt\_a |
| The 3-year CAGR growth rate of company contributions to retirement plans (which is an indicator of profitability.) | bq\_growth\_co\_contrib\_pens\_amt\_a |
| The 3-year CAGR growth rate of the number of employees in the U.S. eligible for a retirement and/or welfare benefit plan. | bq\_emp\_growth\_rate |
| The number of years the company has been offering an employee benefit plan. This is usually fewer years than the overall age of the firm. | bq\_age\_oldest\_plan |

1. Employment
2. Revenue
3. BQ growth score
4. BQ risk score
5. BQ profitability score
6. 3 growth rates
7. Company age

* Online metrics (sources of explanation, independent variables):

|  |  |
| --- | --- |
| **Name of metrics** | **Variable Name** |
| Alexa global ranking | global\_rank |
| Alexa country ranking | country\_rank |
| Alexa categorization indicator (= 1 if the website is categorized, =2 if the website is not categorized) | categorization\_indic |
| Alexa category ranking | category\_rank |
| WHOIS privacy protection indicator | privacy\_protect\_indic |
| WHOIS year of website creation 🡪 BQ company age | web\_creation\_year |

1. Alexa global ranking
2. Alexa country ranking
3. Alexa categorization indicator (= 1 if it was categorized)
4. Alexa category ranking
5. WHOIS privacy protection indicator
6. WHOIS year of website creation 🡪 BQ company age

* Relationships can vary by (factors):

1. Industry
2. Sub-sector
3. Sector
4. State
5. Employment range: 1 – 100, 101 – 500, 501 – 1000, etc.

**Need 5+ volunteers to present your preliminary findings based on this:**

**To do:**

*BQ Employment against Alexa global ranking*

Revenue against global ranking

Employment against country ranking

Revenue against country ranking

**Control factors:**

Sector dummy variables (around 20 sectors)

* Accommodation\_dummy = 1 if sector = accommodation; 0 otherwise
* Finance\_dummy = 1 if sector = finance; 0 otherwise
* Etc… for all sectors

Category indicator = 1 if the company website has an Alexa category; 0 otherwise

Privacy indicator = 1 if the company website has privacy protection; 0 otherwise

Correlation(employment, global ranking) = ? Might be 0.6 (high & positive) for Finance companies, but it might be 0.2 (small & positive) for Healthcare

**Methods:**

Simple correlation

OLS regression

Panel regression

Other…

**Problem: “over-fitting”**

**Presentation**

* Use whatever program you want to do the analysis (R would be best, anything is fine)
* Excel file with your output / results
* Simple verbal explanation of what you did (Word format)
* PowerPoint (5 – 10 slides) summarizing your findings