Rosalie Day

Introduction to Data Science, Springboard, February 2018

Ideas for the 3 proposals

*MODIFIED*

*I did not do any of these. The data sources were too unwieldy. I ended up doing the natural language processing project that I wanted to do initially when I signed up for the course.*

PROBLEM - DATA SOURCES – mostly public data but I have no idea how to acquire them. Because these data sets are so large, I don’t know:

How to find out the granularity for which the data is available (the data that went into creating tables on the websites for these giant public data bases)

How to specify the data for acquisition,

The process for acquisition

1) TOPIC: Is the prime working age category (25-54) obsolete considering the extended life expectancies? Seniors (65 and older) have experienced growth in full-time employment. However, there is not much research on employment in the gap ages (“Gap”) between so-called Prime Working Age and Seniors, 65 or older.

DATA SOURCES:

Maybe - [https://lehd.ces.census.gov/data/#qwi](https://lehd.ces.census.gov/data/)

Using US census data – 2006 -2016

13 industry codes, ALL US pop?,

age groups- 35-54, 55-64, 55-59, 60-64, >= 65

education (< 4-yr degree, 4 yr degree, grad/professional degree)

gender

(optional urban- metropolitan statistical area vs national)

Using US gov – Bureau of Labor Statistics (BLS) data

UNKNOWN – data availability for target age group and demographics in BLS? [https://www.bls.gov/cps/demographics.htm#older](https://www.bls.gov/cps/demographics.htm) - Employer data for NAICS industry codes –

QUESTIONS:

What is the employment status for the Gap?

Demographics of employed status

Which industries are likely to employ the Gap population?

Education attainment levels?

Urban/rural?

Gender?

Should prime working age be extended – at least until 60? Or the US retirement age: early benefits at 62 and full benefits at 67 (for those born after 1960)?

There is a precipitous fall in employment from 55 to 59 at 71.9% of the Labor Force to 56.5% for ages 60 to 64? (The Labor Force category in BLS does not include the portion who dropped out of the labor force and desire a full-time job.)

2.TOPIC: How likely is stock market performance to affect finances of individual US households? Economic well-being is determined by quality of life and Maslow’s hierarchy of needs but the majority of households do not own stocks or even have pensions that own stocks. Stock prices are not a good indicator of economics for policy purposes.

DATA SOURCE: <https://www.census.gov/ces/dataproducts/index.html>

(NYT article source)

Boston Census Research Data Center (BRDC). Public Use Data Archive

[james.c.davis@census.gov](mailto:james.c.davis@census.gov), 617-613-1215, www: <http://www.nber.org/~jdavis>

QUESTIONS:

Percentage of households owning stock

Percentage of stock owned by individuals

Various demographic questions

Age group

Education level – college degree Y/N

Urban vs national?

Income in (recent data collection) year realized from market increase as an average? (which is meaningless because of the skew of the distribution but distribution is not available for median)

**OTHER TOPICS?**

TOPIC: Real-time Pricing or congestion pricing impacts

DATA??? CES - <https://www.census.gov/ces/dataproducts/index.html>, or EIA data for energy?

Real-time pricing impact demand for electricity?

Real-time or congestion pricing traffic data from Stockholm, Singapore and London

* How large a price per trip not taken?
* Distribution – not linear
* Permanent (5 years?) or ineffective when expectation incorporated?

TOPIC: Chicago versus 8-10 suburbs of Chicago, are property taxes progressive, i.e. if larger median income, then larger tax rate?

DATA: Census Data

Zillow?

QUESTIONS

Incomes by zipcode?

Property values by zipcode? – mean vs median

Tax rates

Average effective rates? (value, 2015 property taxes)

TOPIC: Diversity on Corporate Boards for US Fortune 500 Companies – gender discrimination on Boards of Directors even though there is a positive correlation between women’s presence on corporate Boards of Directors and financial performance.

DATA:

I have made an inquiry about access to the ISS Voting Analytics Database –

<https://www.issgovernance.com/solutions/iss-analytics/voting-analytics/>

and to Wharton where they do research on that data base.-

ANSWER: need requests to come from education institution email address or .edu ending

QUESTIONS

How likely are you to find a woman on the Board of Directors of a US corporation?

Does the probability vary by industry?

Depending on the data available

Age?

Education level of Board composition?

Occupation of Board composition?

Industry (NAICS codes) by gender employment

**Data sources**

**IPUMS**

**American Community Survey – census data**

**EX:**

Justin Wolfers, Economist (Univ of MI) NYT contributer

<https://usa.ipums.org/usa-action/variables/DEGFIELD#codes_section>

2009 – 2016 – phD academic fields of study, gender split