WSU_Data_Analysis

UIC CUPPA - Capstone
Spring 2024

Libraries and Data

```
library(readr)
library(readxl)
library(tidyverse)
library(dplyr)
library(stringr)
library(purrr)
library(knitr)
library(kableExtra)

vendor <- read_csv("~/Desktop/Capstone/data/vendor.csv")
bea <- read_excel("~/Desktop/Capstone/data/table.xlsx")
qcew <- read_csv("~/Desktop/Capstone/data/qcew.csv")
```

Vendor Data

Vendor data = WSU provided data on vendor partners with existing contact information available.

```
vendor$capabilities <- tolower(vendor$CAPABILITIES)
vendor$category <- tolower(vendor$CATEGORY)
```

Search Keywords for Vendor Capabilities

Reviewing the current vendor list provided by WSU to understand where they already have capabilities in the local community.

```r
vendor <- vendor %>%
  mutate(it_vendor = case_when(
    str_detect(capabilities,
               "it|information|technology|software|computer|
               computers|computing|tech") ~ 1,
    str_detect(category,
               "it|information|technology|software|computer|
               computers|computing|tech") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(cust_vendor = case_when(
    str_detect(capabilities,
               "custodial|janitorial|maintenance|plumbing|
               cleaning|custodian|janitor|plumber") ~ 1,
    str_detect(category,
               "custodial|janitorial|maintenance|plumbing|
               cleaning|custodian|janitor|plumber") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(ground_vendor = case_when(
    str_detect(capabilities,
               "ground|grounds|landscape|landscaping|
               lawn|snow|garden|gardening") ~ 1,
    str_detect(category,
               "ground|grounds|landscape|landscaping|
               lawn|snow|garden|gardening") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(dei_vendor = case_when(
    str_detect(capabilities,
               "dei|diversity|equity|inclusion|diverse|inclusive") ~ 1,
    str_detect(category,
               "dei|diversity|equity|inclusion|diverse|inclusive") ~ 1,
    TRUE ~ 0
  ))
```

```r
vendor <- vendor %>%
  mutate(food_vendor = case_when(
    str_detect(capabilities,
               "food|catering|caterer|foods|cater|nutrition") ~ 1,
    str_detect(category,
               "food|catering|caterer|foods|cater|nutrition") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(secure_vendor = case_when(
    str_detect(capabilities,
               "secure|security|surveillance|building security|guard") ~ 1,
    str_detect(category,
               "secure|security|surveillance|building security|guard") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(law_vendor = case_when(
    str_detect(capabilities,
               "legal|lawyer|litigation|lawyers|law") ~ 1,
    str_detect(category,
               "legal|lawyer|litigation|lawyers|law") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(hvac_vendor = case_when(
    str_detect(capabilities,
               "hvac|heating|cooling|mechanics|mechanical") ~ 1,
    str_detect(category,
               "hvac|heating|cooling|mechanics|mechanical") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(electric_vendor = case_when(
    str_detect(capabilities,
               "electric|electrical|electrician|lighting") ~ 1,
    str_detect(category,
```

Table 1: Summary of Columns 30-39

it_vendor_total	253
cust_vendor_total	66
ground_vendor_total	34
dei_vendor_total	2
food_vendor_total	90
secure_vendor_total	17
law_vendor_total	3
hvac_vendor_total	17
electric_vendor_total	22
patient_vendor_total	288

```
                "electric|electrical|electrician|lighting") ~ 1,
    TRUE ~ 0
  ))

vendor <- vendor %>%
  mutate(patient_vendor = case_when(
    str_detect(capabilities,
                "census|ai|predictive|prediction") ~ 1,
    str_detect(category,
                "census|ai|predictive|prediction") ~ 1,
    TRUE ~ 0
  ))

summarized_data <- vendor %>%
  summarize(across(30:39, sum, .names = "{col}_total"))
summarized_data <- t(summarized_data)
```

```
kable(summarized_data, align = "c", caption = "Summary of Columns 30-39") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"),
                full_width = FALSE)
```

BEA Data

Bureau of Economic Analysis - U.S. Department of Commerce https://www.bea.gov/industry/input-output-accounts-data Total inputs by industry required (directly and indirectly) in order to deliver one dollar of industry output to final users

The value of the final goods and services produced in the United States is the gross domestic

product. The percentage that GDP grew (or shrank) from one period to another is an important way for Americans to gauge how their economy is doing. The United States' GDP is also watched around the world as an economic barometer.

"Total inputs by industry required (directly and indirectly) in order to deliver one dollar of industry output to final users," – this provides insights into the economic interconnectedness and dependencies between industries within an economy. This concept is analyzed through input-output tables, which understand the flow of goods and services between industries and sectors.

The values in the data represent the total inputs required by each industry, both directly and indirectly, to produce one dollar of industry output that eventually reaches final users. This concept is often referred to as "total requirements coefficients" or "total requirements per dollar of output."

Assuming that output for a hospital is equal to hospital spend, this assumes that the budget they are spending is the hospital profit. The total operational spend for all hospitals in Illinois is $17,000,000,000. There are about 210 hospitals - so the average operational spend is $80,952,381 per hospital. This will be used as a multiplier in understanding the input-output data.

Highest Inputs for Output

```
hospitals <- select(bea, "Industry", "Industry Description", "Hospitals")

top_inputs <- bea %>%
  select("Industry", "Industry Description", "Hospitals") %>%
  filter(!grepl("Hospitals", `Industry Description`)) %>%
  arrange(desc(Hospitals))

top_inputs$output <- top_inputs$Hospitals * 80952381
options(scipen = 999)

top_inputs
```

```
# A tibble: 70 x 4
   Industry `Industry Description`                                Hospitals output
   <chr>    <chr>                                                     <dbl>  <dbl>
 1 ORE      Other real estate                                       0.116  9.39e6
 2 561      Administrative and support services                     0.0862 6.98e6
 3 5412OP   Miscellaneous professional, scientific, and techni~    0.0646 5.23e6
 4 325      Chemical products                                       0.0513 4.15e6
```

```
 5 42       Wholesale trade                                      0.0452 3.66e6
 6 524      Insurance carriers and related activities            0.0340 2.75e6
 7 722      Food services and drinking places                    0.0275 2.22e6
 8 523      Securities, commodity contracts, and investments     0.0262 2.12e6
 9 339      Miscellaneous manufacturing                          0.0255 2.06e6
10 55       Management of companies and enterprises              0.0248 2.01e6
# i 60 more rows
```

For example, Hospitals require $0.086 worth of inputs from Administrative and support services to produce $1.00 worth of output to final users. These inputs can come directly from other industries or indirectly through intermediate stages across the economy. With the multiplier, it can be understood that on average, hospitals spend $6,977,220.95 on Administrative and support services from their budget annually.

Lowest Inputs for Output

```r
bottom_inputs <- bea %>%
  select(Industry, `Industry Description`, Hospitals) %>%
  filter(!grepl("Hospitals", `Industry Description`)) %>%
  arrange(Hospitals)

bottom_inputs$output <- bottom_inputs$Hospitals * 80952381
options(scipen = 999)

bottom_inputs
```

```
# A tibble: 70 x 4
   Industry `Industry Description`                     Hospitals  output
   <chr>    <chr>                                          <dbl>   <dbl>
 1 HS       Housing                                    0            0
 2 GFGD     Federal general government (defense)       0            0
 3 624      Social assistance                          0.0000002   16.2
 4 623      Nursing and residential care facilities    0.0000264 2137.
 5 445      Food and beverage stores                   0.000064  5181.
 6 525      Funds, trusts, and other financial vehicles 0.000074  5990.
 7 61       Educational services                       0.000110  8937.
 8 452      General merchandise stores                 0.000124 10006.
 9 483      Water transportation                       0.000344 27872.
10 213      Support activities for mining              0.000355 28730.
# i 60 more rows
```

For example, Hospitals require $0.000064 worth of inputs from Food and beverage stores to produce $1.00 worth of output to final users. These inputs can come directly from other industries or indirectly through intermediate stages across the economy. With the multiplier, it can be understood that investing in Food and beverage stores will result in an output for the hospital of $5,180.95.

QCEW Data

Quarterly Census of Employment and Wages - County level data https://www.bls.gov/cew/downloadable-data-files.htm

The Quarterly Census of Employment and Wages (QCEW) program publishes a quarterly count of employment and wages reported by employers covering more than 95 percent of U.S. jobs, available at the county, MSA, state and national levels by industry.

Filter data for Cook County for only metrics impacting the West Side neighborhoods:

```
cook <- qcew %>%
  filter(qcew$Area == "Cook County, Illinois")

cook <- cook %>%
  rename(Establishment_Count = `Establishment Count`)
```

Concentrated Businesses

```
top_bus <- cook %>%
  filter(Ownership == "Private") %>%
  select("Industry", "Establishment_Count") %>%
  filter(!grepl("10 Total, all industries", Industry)) %>%
  arrange(desc(Establishment_Count))
top_bus
```

```
# A tibble: 13 x 2
   Industry                                    Establishment_Count
   <chr>                                                     <dbl>
 1 102 Service-providing                                    124521
 2 1024 Professional and business services                   30288
 3 1021 Trade, transportation, and utilities                 29298
 4 101 Goods-producing                                       17161
 5 1025 Education and health services                        17130
 6 1027 Other services                                       16090
```

```
 7 1023 Financial activities                          14185
 8 1026 Leisure and hospitality                       14155
 9 1012 Construction                                  11406
10 1013 Manufacturing                                  5621
11 1022 Information                                    3024
12 1029 Unclassified                                    351
13 1011 Natural resources and mining                    134
```

Location Quotients

Location Quotient (LQ): The location quotient for total wages/employment in a specific industry or occupation in a given area is calculated as the ratio of the proportion of total wages/employment in that industry or occupation in the area to the proportion of total wages/employment in the same industry or occupation at the national level.

Interpretation:

LQ $>$ 1: Indicates that the industry or occupation has a higher concentration or specialization in total wages/employment in the local area compared to the national average. A location quotient greater than 1 suggests a comparative advantage or specialization in total wages/employment in that industry or occupation. LQ = 1: Indicates that the industry or occupation is proportionately represented in total wages/employment in the local area compared to the national average. LQ $<$ 1: Indicates that the industry or occupation has a lower concentration or specialization in total wages/employment in the local area compared to the national average. A location quotient less than 1 suggests a comparative disadvantage or lack of specialization in total wages in that industry or occupation.

Rename columns for ease in this analysis:

```
cook <- cook %>%
  rename(EmploymentLQ = "Employment Location Quotient Relative to U.S.") %>%
  rename(WageLQ = "Total Wage Location Quotient Relative to U.S.")
```

```
specialize <- cook %>%
  filter(Ownership == "Private") %>%
  select("Industry", "EmploymentLQ") %>%
  filter(!grepl("10 Total, all industries", Industry)) %>%
  arrange(desc(EmploymentLQ))
specialize
```

```
# A tibble: 13 x 2
   Industry                                    EmploymentLQ
```

```
   <chr>                                          <dbl>
 1 1023 Financial activities                       1.42
 2 1027 Other services                             1.25
 3 1024 Professional and business services         1.23
 4 1025 Education and health services              1.13
 5 102 Service-providing                           1.11
 6 1022 Information                                0.99
 7 1026 Leisure and hospitality                    0.99
 8 1021 Trade, transportation, and utilities       0.97
 9 1013 Manufacturing                              0.81
10 101 Goods-producing                             0.67
11 1012 Construction                               0.58
12 1029 Unclassified                               0.07
13 1011 Natural resources and mining               0.05
```

A relative employment location quotient greater than 1 suggests that, on average, industries or occupations in the area are more concentrated or specialized compared to the national average. Cook county is more specialized in financial activities, professional services, education, health, and service providing, etc.

```
wage <- cook %>%
  filter(Ownership == "Private") %>%
  select("Industry", "WageLQ") %>%
  filter(!grepl("10 Total, all industries", Industry)) %>%
  arrange(desc(WageLQ))
wage
```

```
# A tibble: 13 x 2
   Industry                                  WageLQ
   <chr>                                      <dbl>
 1 1023 Financial activities                   1.66
 2 1027 Other services                         1.35
 3 1024 Professional and business services     1.18
 4 102 Service-providing                       1.13
 5 1026 Leisure and hospitality                1.08
 6 1025 Education and health services          1.03
 7 1021 Trade, transportation, and utilities   0.94
 8 1022 Information                            0.84
 9 1013 Manufacturing                          0.7
10 101 Goods-producing                         0.63
11 1012 Construction                           0.62
12 1011 Natural resources and mining           0.05
```

```
13 1029 Unclassified                              0.04
```

The same industries appear when analyzing wages. A relative total wage location quotient greater than 1 suggests that, on average, industries or occupations in the area have higher total wages compared to the national average.

```
wage_low <- cook %>%
  filter(Ownership == "Private") %>%
  select("Industry", "WageLQ") %>%
  filter(!grepl("10 Total, all industries", Industry)) %>%
  arrange(WageLQ)
wage_low
```

```
# A tibble: 13 x 2
   Industry                                  WageLQ
   <chr>                                      <dbl>
 1 1029 Unclassified                           0.04
 2 1011 Natural resources and mining           0.05
 3 1012 Construction                           0.62
 4 101 Goods-producing                         0.63
 5 1013 Manufacturing                          0.7
 6 1022 Information                            0.84
 7 1021 Trade, transportation, and utilities   0.94
 8 1025 Education and health services          1.03
 9 1026 Leisure and hospitality                1.08
10 102 Service-providing                       1.13
11 1024 Professional and business services     1.18
12 1027 Other services                         1.35
13 1023 Financial activities                   1.66
```

A relative total wage location quotient less than 1 suggests that, on average, industries or occupations in the area have lower total wages compared to the national average. These industries show where Cook county is paying lower wages than national averages.

Tier II Vendors

WSU is interested in learning more about vendors in their Tier II classification. Find vendors in West Side Zip Codes that are part of these NAICS for businesses.

```
     industry            naics
 [1,] "security_services" "5616"
 [2,] "fire_services"     "922160"
 [3,] "food_distribution" "424490"
 [4,] "ad_marketing"      "5418"
 [5,] "records"           "518210"
 [6,] "bus_consulting"    "541611"
 [7,] "furniture"         "33712"
 [8,] "photography"       "541922"
 [9,] "video_av"          "512110"
```

West Side United 10 Community Areas and Zip Codes

```
 [1] 60634 60644 60622 60624 60612 60608 60606 60635 60639 60641 60647 60651
[13] 60607 60623 60642 60616 60610 60661
```