**The Business Advisor**

**Jaebin Park, Pooja Karla, Janit Modi, Sandeep Kumar Moparthy**

**Abstract**

Deciding a place for one’s new start up or establishing a new unit of an existent business can be a difficult task. A significant number of factors have to be considered in making such an important decision. This project gives an ability, for decision makers, to arrive at an informed decision on which industry yields the best in which state, in which county and in which city.

**Introduction**

Business Patterns Overview

County Business Patterns (CBP), ZIP Code Business Patterns (ZBP) are an annual series that provide sub-national economic data by industry. These programs cover most of the U.S. economy and feature industry and geographic statistics which supplement those published in the [Economic Census](http://www.census.gov/programs-surveys/economic-census.html). Data are published at the U.S. level and by State, County, Metropolitan area, ZIP code, and Congressional District. All data are classified by an industry code ([NAICS](http://www.census.gov/eos/www/naics/)) and can be viewed with employment-size class breakouts by establishment, and by legal form of organization at some geographic levels. See the [County Business Patterns Web site >>](http://www.census.gov/programs-surveys/cbp.html)

**Scope**

The CBP annual series provides information that is critical for understanding the Nation's changing economic structure and performance. The series is used to study the economic activity of small areas, analyze economic changes over time; and as a benchmark for statistical series, surveys, and databases between economic censuses. Businesses use the data for analyzing market potential, measuring the effectiveness of sales and advertising programs, setting sales quotas, and developing budgets. Government agencies use the data for administration and planning. Statistics from these surveys are widely used by policy officials, economic analysts, business decision-makers, and the news media.

**Data Analytics Problem**

To build a data analytics model solution based on the County Business Patterns data that can help decision makers in arriving at potential business locations. The solution gives a high, medium and low potential location for a business based on its industry, the annual payroll specific to employee count in the establishment. Feature creation of the target variable for this approach is necessary. ‘Average annual payroll’ for a specific business industry would be the determinant.

Given a county or a state,

Formula 1:

Avg annual Payroll of an industry = Annual Payroll for the industry / No. of establishments in the industry

Given a county or a state and the employee size of a business organization,

Formula 2:

Avg annual Payroll of an industry = Annual Payroll for the industry / (No. of establishments in the industry)

The **dependent variable** would categorize the created feature,

Avg Payroll of an Industry into high, medium and low avg annual payrolls.

**Data Set**

Since 1998, County Business Patterns has been tabulated based on the North American Industry Classification System (NAICS). For more information on NAICS, see the [NAICS information page](https://www.census.gov/naics). Data were tabulated according to the Standard Industrial Classification (SIC) System for prior periods. For more information on the SIC system, see the [SIC Information Page](https://www.census.gov/programs-surveys/cbp/technical-documentation/reference/naics-descriptions/sic-code-descriptions.html).

* 2012 to 2016 data use NAICS 2012
* 2008 to 2011 data use NAICS 2007
* 2003 to 2007 data use NAICS 2002
* 1998 to 2002 data use NAICS 1997
* 1988 to 1997 data use 1987 SIC
* 1974 to 1987 data use 1972 SIC

Prior to 2012, County Business Patterns lagged by one year in the adoption of the classification system employed in the Economic Census. Starting in 2012, the classification system was changed in the same year.

There are eight different datasets that considered for this project. These datasets, when integrated and work with intellectually can yield in interesting insights.

### Datasets

[Complete Congressional District File](https://www2.census.gov/programs-surveys/cbp/datasets/2016/cbp16cd.zip?" \o "Complete Congressional District File)  [<1.0 MB]

[Complete County File](https://www2.census.gov/programs-surveys/cbp/datasets/2016/cbp16co.zip?" \o "Complete County File)  [15.6 MB]

[Complete Metropolitan Area File](https://www2.census.gov/programs-surveys/cbp/datasets/2016/cbp16msa.zip?" \o "Complete Metropolitan Area File)  [7.5 MB]

[Complete State File](https://www2.census.gov/programs-surveys/cbp/datasets/2016/cbp16st.zip?" \o "Complete State File)  [10.9 MB]

[Complete U.S. File](https://www2.census.gov/programs-surveys/cbp/datasets/2016/cbp16us.zip?" \o "Complete U.S. File)  [<1.0 MB]

[Complete ZIP Code Industry Detail File](https://www2.census.gov/programs-surveys/cbp/datasets/2016/zbp16detail.zip?" \o "Complete ZIP Code Industry Detail File)  [28.2 MB]

[Complete ZIP Code Totals File](https://www2.census.gov/programs-surveys/cbp/datasets/2016/zbp16totals.zip?" \o "Complete ZIP Code Totals File)  [<1.0 MB]

[CBP and NES Combined Report](https://www2.census.gov/programs-surveys/cbp/datasets/2016/combine16_txt.zip?" \o "CBP and NES Combined Report)  [14.3 MB]

The CBP and NES Combined Report is CSV file with 822289 rows and 19 variables.

The Complete County File is a CSV file with 2124893 rows and 26 variables.

The complete metropolitan area file is a CSV file with 936105 rows and 23 variables

The complete state file is a csv file with 448310 rows and 84 variables

The complete US file is a csv file with 13002 rows and 83 variables

The complete ZIP code industry Detail file is a csv with 8418283 rows and 12 variables

The Complete Zipcode totals File is a csv with 38722 rows and 13 variables

The variables and the description of the variables can be found in the below links

[County Record Layout](https://www2.census.gov/programs-surveys/rhfs/cbp/technical%20documentation/2015_record_layouts/county_layout_2015.txt?" \o "County Record Layout)  [<1.0 MB] 2015-2016

[Metro Area Record Layout](https://www2.census.gov/programs-surveys/rhfs/cbp/technical%20documentation/2015_record_layouts/metro_area_layout_2015.txt?" \o "Metro Area Record Layout)  [<1.0 MB] 2015-2016

[State Record Layout for Puerto Rico & Island Areas](https://www2.census.gov/programs-surveys/rhfs/cbp/technical%20documentation/2015_record_layouts/state_pr_layout_2015.txt?" \o "State Record Layout for Puerto Rico & Island Areas)  [<1.0 MB] 2015-2016

[State Record Layout](https://www2.census.gov/programs-surveys/rhfs/cbp/technical%20documentation/2015_record_layouts/state_layout_2015.txt?" \o "State Record Layout)  [<1.0 MB] 2015-2016

[U.S. Record Layout](https://www2.census.gov/programs-surveys/rhfs/cbp/technical%20documentation/2015_record_layouts/us_layout_2015.txt?" \o "U.S. Record Layout)  [<1.0 MB] 2007-2016

[ZIP Code Industry Detail Record Layout](https://www2.census.gov/programs-surveys/cbp/technical-documentation/records-layouts/noise-layout/zip_detail_layout.txt" \o "ZIP Code Industry Detail Record Layout)  [<1.0 MB] 2015-2016

[ZIP Code Totals Record Layout](https://www2.census.gov/programs-surveys/rhfs/cbp/technical%20documentation/2015_record_layouts/zip_totals_layout_2015.txt?" \o "ZIP Code Totals Record Layout)  [<1.0 MB] 2015-2016

[CBP and NES Combined Report Record Layout](https://www2.census.gov/programs-surveys/cbp/technical-documentation/records-layouts/combined-record-layout/record_layout.txt?" \o "CBP and NES Combined Report Record Layout)  [<1.0 MB]

Related Data Products

CBP covers most of the country's economic activity. The series excludes data on self-employed individuals, employees of private households, railroad employees, agricultural production employees, and most government employees. For information on businesses with no paid employees, see [Nonemployer Statistics](https://www.census.gov/programs-surveys/nonemployer-statistics.html" \t "_blank). [See a complete list of economic surveys >>](http://www.census.gov/econ/survey.html)

**Methods and Models**

Designing the model is divided into five phases,

1. Exploratory phase
2. Preprocessing phase
3. Exploratory and visualizations
4. Modelling phase
5. Improvisation and insight phase

**Useful features**

ST - State FIPS Code

SDSCR - State Description

CTYDSCR - County Description

NAICS - Industry Code: 2- through 4-digit NAICS code

EMP - CBP Employment, including March 12th

AP - CBP Annual Payroll (in thousands)

FIPSTATE - FIPS State Code

FIPSCTY - FIPS County Code

NAICS - Industry Code - 6-digit NAICS code.

EST - Total Number of Establishments

N1\_4 - Number of Establishments: 1-4 Employee Size Class *to*

N1000\_4 - Employment Size Class: 5,000 or More Employees

CENSTATE - Census State Code

CENCTY - Census County Code

MSA - Metropolitan or Micropolitan Area Code

ZIP - ZIP Code

EST - Total Number of Establishments

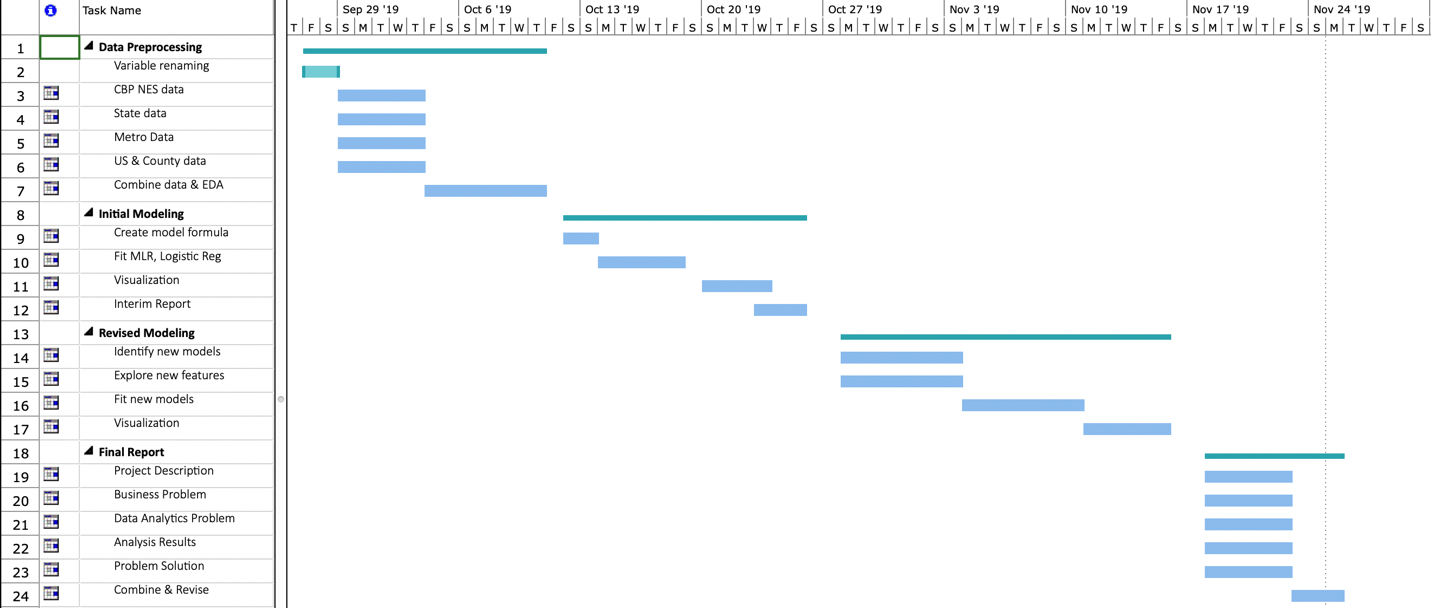
CITY - ZIP City Name

STABBR - ZIP State Abbreviation

CTY\_NAME - ZIP County Name

**Roles and Schedule**





**Bibliography**

1. <https://www.census.gov/programs-surveys/cbp/technical-documentation/methodology.html#par_textimage_379462313>
2. <https://factfinder.census.gov/faces/affhelp/jsf/pages/metadata.xhtml?lang=en&type=program&id=program.en.BP>
3. <https://www.census.gov/data/datasets/2016/econ/cbp/2016-cbp.html>
4. <https://www.census.gov/programs-surveys/economic-census.html>
5. <https://www.census.gov/eos/www/naics/>
6. <https://www.census.gov/programs-surveys/cbp.html>
7. <https://www.census.gov/programs-surveys/nonemployer-statistics.html>
8. <https://www.census.gov/econ/survey.html>
9. <https://www.census.gov/programs-surveys/cbp/data/datasets.html>
10. <https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t>
11. <https://factfinder.census.gov/faces/affhelp/jsf/pages/metadata.xhtml?lang=en&type=survey&id=survey.en.BP_CBP>
12. <https://factfinder.census.gov/faces/affhelp/jsf/pages/metadata.xhtml?lang=en&type=survey&id=survey.en.BP_ZBP>
13. [North American Industry Classification System (NAICS) Descriptions](https://www2.census.gov/programs-surveys/cbp/technical-documentation/reference/naics-descriptions/naics2012.txt)  [<1.0MB]2012-2016
14. [State and County Geography Reference](https://www2.census.gov/programs-surveys/cbp/technical-documentation/reference/state-county-geography-reference/georef12.txt)  [<1.0 MB]2012-2016
15. [Metro Area Geography Reference](https://www2.census.gov/programs-surveys/cbp/technical-documentation/reference/metro-area-geography-reference/msa_county_reference12.txt)  [<1.0 MB]2008-2014
16. [CBP and NES Combined Report Reference](https://www2.census.gov/programs-surveys/cbp/technical-documentation/reference/combined-reference/reference_file.xlsx?)  [<1.0 MB]