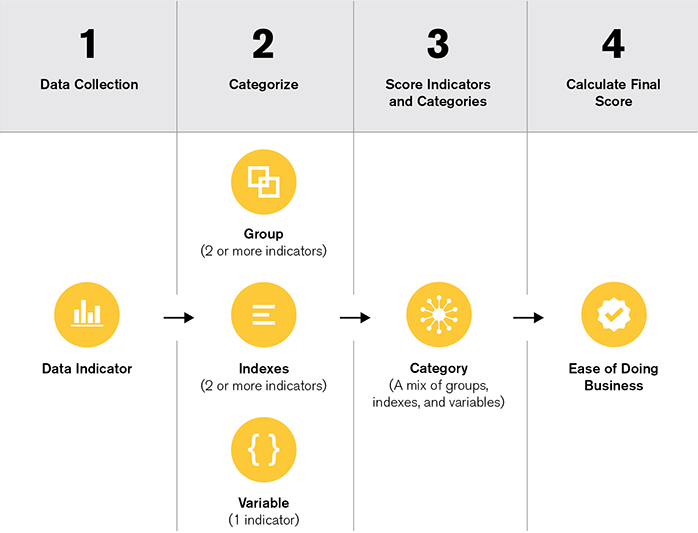
**Methodology - Ease of Doing Business**

**What is the Ease of Doing Business Score?**

The primary objective of the Doing Business North America report is to provide a measure of the ease of doing business for cities in North America. At least one city per state or province was scored. The District of Columbia was included as well. For states with much larger populations than average, more cities (up to four of the largest in the state or province) were included. The Ease of Doing Business Score is derived from a summation of the scores awarded in each of the six categories measured by this report. The Ease of Doing Business Score focuses on the regulatory burdens a small- to medium-sized business would face from the birth of the business to the death of the business in cities across North America. The following explains the process of arriving at the final score.



**Steps 1 and 2: Collecting and Categorizing the Data**

The Doing Business North America team collected data on 63 different regulatory and economic indicators across six different categories. The data collected came entirely from publicly available sources, including many published by the municipal governments we studied. Each observation is termed an “indicator.” Once collected, these indicators were then classified into one of three types: (i) a variable, (ii) a group, or (iii) an index.

**Variable**

A variable is the most common classification found throughout a dataset. These types of indicators are generally considered the most important, define the theme of the category, and are not combined into larger groups of indicators. To put it another way, a variable consists of only one indicator. For example, in the Starting a Business category, indicators that show the number of procedures to start a business, the time to establish a business, or the cost to start a business are all treated as variables. Although these three indicators have the same general theme, they each reflect an important and unique aspect related to the process of starting a business and make up the foundation of the category. They also use three different units of measurement — the procedures indicator measures the number of steps, the time indicator measures the number of days, and the cost indicator measures the number of dollars.

**Group**

Groups consist of two or more indicators that share similar attributes that can be combined. When combined, that group is considered a variable in the scoring process. An example of this can be seen in the Employing Workers category, in which there is a group of indicators measuring severance pay for workers with different tenure periods: There is an indicator that measures severance pay for workers with one year of tenure, an indicator for severance pay with five years of tenure, and an indicator for severance pay with 10 years of tenure. Because these three indicators all measure the same thing (but simply measure different time ranges) and all are measured using the same units, these three indicators are grouped together (in this case, averaged together) and then treated like a variable for scoring purposes. A group is the rarest collection of indicators found throughout the data.

**Index**

Indexes are a collection of indicators (similar to a group) that share a theme. However, there are two important conditions that must both be met for a collection of indicators to be treated as an index instead of as a group: (i) the indicators are binary in nature; and (ii) on their own, they do not rise to the importance of a variable but grouped together they describe an important element of the Ease of Doing Business analysis. An example of this can be seen in the Registering Property category, where there are several indexes used in the Quality of Land Administration Index, which itself is comprised of the Reliability of Infrastructure Index and the Transparency of Information Index. This transparency index measures various attributes of the land administration system, such as whether information on land ownership is made publicly available (a “yes” or “no” answer) or whether the agency in charge of immovable property registration commits to a specific time frame (a “yes” or “no” answer). These two indicators are grouped together to form an index. Additionally, indexes can be combined to form an even broader index as described above for the Registering Property category.

Finally, by the end of this process the data was categorized as either a variable (consisting of one indicator), a group (consisting of two or more continuous or binary indicators), or an index (consisting of two or more binary indicators). Each of them has equal weight in the category score (as we shall see soon).

**Step 3: Scoring the Indicators and Categories**

The indicators are scored following one of two procedures, depending on whether the data type is binary in nature (which is collapsed into indexes) or continuous in nature (which becomes variables on their own or combined in groups). Indexes and groups are also scored.

**Binary Indicators**

Binary indicators are the simplest of all the indicators. These indicators measure whether a policy exists. Locations where that policy exists are marked as 1 or “yes,” and locations where the policy does not exist are marked as 0 or “no.” Next, we determine whether the policy is harmful or beneficial to the ease of doing business — in other words, we decide whether a “1” or “yes” should be considered the best score or the worst score. So, for instance, if a policy is considered beneficial, locations with this policy (observations that are labeled with a 1 or “yes” in the data indicator) are awarded one point, and locations without this policy (locations that were labeled with a 0 in the data indicator) are awarded zero points. A collection of binary indicators can then be combined into an index by simply adding the ones and zeros to create the index score.

**Continuous Indicators**

The scoring of continuous indicators requires a little more work. The first step for analyzing a continuous indicator is to determine the default assumption about what is considered most favorable for the ease of doing business as described above. However, due to the large range of some indicators (indicators that either involve ratios or costs are notorious for having large variance) and the need to avoid outliers skewing the overall score, a threshold of two standard deviations greater than the mean is sometimes used to define the upper boundary for an indicator and a threshold of two standard deviations less than the mean is sometimes used to define the floor.

Once a specific numeric range and directionality have been established, a city’s score for that indicator is determined using the equation ((A-B)/(A-C))\*10), where A is the lowest observation (or “lowest performer”) for an indicator, across all locations, B is the observation for the location being scored, and C is the highest observation (or “highest performer”) for an indicator across all locations. This calculation produces values within a range of zero to 10, where the location(s) with the lowest regulatory performance (or those outside two standard deviations from the mean) are awarded no points, and the location(s) with the best regulatory environment are awarded 10 points. Cities in between the top and bottom receive scores based on their relative position as defined by the equation — where they end up in relation to the “frontier” of the highest-ranked indicator value. This “distance to frontier” equation and scoring method for continuous indicators can be applied to all types of indicators.

**Groups**

This “distance to frontier” method is also used when scoring groups. The difference is that, after deciding the directionality, the sum of all indicator scores within the group is first divided by the number of indicators included in the group, producing an average for that group. Then that average is scored based on the group’s distance to the frontier using the formula above.

**Indexes**

Indexes are scored in a similar fashion. The top and bottom values are assigned and the numeric score is calculated using the “distance to frontier” formula.

**Scoring the Categories**

To generate a score for a particular category, the scores among all variables, groups, and indexes within the category must first be aggregated. That aggregate score is then divided by the total number of indicators. This creates an average for the entire category. The average value for a category is used so that locations with missing or incomplete data are still able to be ranked and scored in a similar fashion to those with a complete dataset and not penalized by lack of data. That average value is the category score.

**Step 4: Calculating the Ease of Doing Business Score**

The Ease of Doing Business Score is derived by summing the scores from all categories. That sum is then divided by the total number of indicators across all categories. (Because that denominator is usually between 27 and 30 unique variables, groups, or indexes across all categories, the effect of missing data is significantly reduced at this level of aggregation.) This number is then multiplied by 10 to allow the final score to resemble a percentage (i.e., out of 100%). For example, the highest-scoring city (Oklahoma City) is 8.47, but when multiplied by 10, this becomes 84.70, which approximates an Ease of Doing Business percentile score of around 85%.

# Employing Workers

## WHAT DOES EMPLOYING WORKERS MEASURE?

Doing Business North America records a myriad of variables related to the flexibility and regulation of employing workers, specifically as it relates to the areas of hiring, working hours, laying off workers, and leave. There are nine indicators used to represent the legal requirements or fiscal burdens necessary to comply with an economy’s labor laws.

## ASSUMPTIONS ABOUT THE BUSINESS

### THE BUSINESS:

* Is a limited liability company.
* If there is more than one type of limited liability company in the economy, the limited liability form most common among domestic firms is chosen.
* If the country does not have a limited liability company option, the company structure most similar to that of an LLC is chosen.
* Operates in the economy’s largest business city.
* Is 100% domestically owned and operated.
* Performs general retail activities, such as the production or sale to the public of goods or services (NAICS Code: 4523).
* Does not qualify for financial incentives or special benefits.
* Has 50 employees, including the owner / entrepreneur.
* Is not subject to collective bargaining agreements.
* Abides by every law and regulation but does not grant workers more benefits than those mandated by law or regulation.

### THE WORKER:

* Is a full-time employee (works 2,080 hours per year).
* Is in their second year of employment and is eligible for all employment benefits.
* Is not a member of a labor union, unless membership is necessary.
* Earns minimum wage.

## INDICATORS

### Annual Minimum Wage as a Percent of Income per Capita

The Annual Minimum Wage (as a Percent of Income per Capita) is calculated by using a location’s hourly minimum wage, multiplied by the number of work-hours in a year (2,080), then dividing that number by a location’s income per capita.

#### Additional Notes:

Rather than simply using the minimum wage values, choosing annual minimum wage relative to income per capita gives context to the magnitude of the wage. Two different cities with the same minimum wage may feel the impact of paying that wage to different extents.

### Maximum Length of Probationary Period (in Calendar Months)

The maximum length of probationary period measures how long new employees are eligible to be classified under a probationary period. The length is measured in calendar months.

### Average Paid Annual Leave for 1, 5 and 10 Years of Tenure (in Working Days)

The paid annual leave average is a group of three indicators: (i) paid annual leave for a worker with one year of tenure; (ii) paid annual leave for a worker with five years of tenure; and (iii) paid annual leave for a worker with 10 years of tenure. Paid annual leave is measured in working days. These indicators measure the number of paid leave days a worker with different employment tenures is eligible for.

### Average Notice Period for 1, 5 and 10 Years of Tenure (in Weeks)

The notice period average is a group of three indicators: (i) notice period for redundancy dismissal for a worker with one year of tenure; (ii) notice period for redundancy dismissal for a worker with five years of tenure; and (iii) notice period for redundancy dismissal for a worker with 10 years of tenure. Notice period requirements are measured in calendar weeks. These indicators measure the time an employer must provide an employee before dismissal due to redundancy.

### Average Severance Pay for 1, 5 and 10 Years of Tenure (in Weeks)

The severance pay average is a group of three indicators: (i) severance pay for redundancy dismissal for a worker with 1 year of tenure; (ii) severance pay for redundancy dismissal for a worker with 5 years of tenure; and (iii) severance pay for redundancy dismissal for a worker with 10 years of tenure. Severance pay requirements are measured in calendar weeks. These indicators measure the amount of weeks of pay an employer must provide an employee before immediate dismissal due to redundancy.

### Length of Paid Maternity Leave (in Weeks)

The length of paid maternity leave measures how much paid leave an employer must provide an eligible worker who has taken maternity leave. It is measured in working weeks.

### Length of Unpaid Maternity Leave (in Weeks)

The length of Unpaid maternity leave measures the number of days per year an employer must provide an employee with unpaid maternity leave. It is measured in working weeks.

### Length of Paid Sick Leave (in Working Days)

The length of paid sick leave measures the number of days per year an employer must provide an employee with paid sick leave. It is measured in working days.

### Length of Unpaid Sick Leave (in Working Days)

The length of unpaid sick leave measures the number of days per year an employer must provide an employee with unpaid sick leave. It is measured in working days.

## How the ‘Employing Workers’ Category is Ranked and Scored

The ‘Employing Workers’ category was ranked and scored using the following nine indicators:

Topic and Indicator | Best Performer | Worst Performer

For each indicator, there is a top performer and a bottom performer. Economies with the best performance for a given indicator are awarded 10 “points,” or a score of 10. Cities at the level of bottom performance, or cities at or below two standard deviations from the mean, are awarded a score of 0. All the cities in between are scored based on their distance to the frontier. For each city, the number of awarded points across all indicators is aggregated, then divided by the number of indicators for which we had data. This is done because not all locations have complete data across all indicators, and doing so allows for all locations to be included in comparison.

For more information on how indicators, indexes, and groups are scored, how the ranking and scoring system for a category works, or how the overall Ease of Doing Business rank and score were derived, please read the first section of this methodology.

# Land and Space Use

'Land and Space Use' is a new, experimental category for the 2nd Version of the DBNA Project. It hopes to capture information related to zoning requirements, how land and space are used, and the difficulties a person would face when requesting changes to how this land and space is used.

Due to the aggressive timeline to collect these final pieces of information, we have broken down the indicators we're interested in collecting into three different tiers: Tier 1 (the highest priority indicators), Tier 2 (a list of medium priority indicators), and Tier 3 (the lowest priority indicators). While working on this section, we are hoping to collect all indicators in Tier 1, ideally all indicators in Tier 2, and if time allows, as many indicators in Tier 3 as possible.

### RELIABILITY OF INFRASTRUCTURE INDEX (SCALE: 0 - 2)

The Reliability of Infrastructure Index has two components:

1. How land titles are kept at the registry. A score of 1 is assigned if the majority of land titles are fully digital; a score of 0.5 is assigned if the majority are scanned; a score of 0 is assigned if the majority are kept in paper format.
2. How immovable property is identified. A score of 1 is assigned if there is a unique number to identify properties for the majority of land plots; a score of 0 is assigned if there are multiple identifiers.

The index ranges from 0 to 2, with higher values indicating a higher quality of infrastructure for ensuring the reliability of information on property titles and boundaries.

### TRANSPARENCY OF INFORMATION INDEX (SCALE: 0 – 6)

The Transparency of Information Index has six components:

1. Whether information on land ownership is made publicly available. A score of 1 is assigned if information on land ownership is accessible by anyone; a score of 0 is assigned if access is restricted.
2. Whether the list of documents required for completing the registration of property transactions is made publicly available. A score of 1 is assigned if the list of documents is accessible online or on a public board; a score of 0 is assigned if it is not made available to the public or if it can be obtained only in person.
3. Whether the fee schedule for completing the registration of property transactions is made publicly available. A score of 1 is assigned if the fee schedule is accessible online or on a public board free of charge; a score of 0 is assigned if it is not made available to the public or if it can be obtained only in person.
4. Whether the agency in charge of immovable property registration commits to a specific time frame for delivering a legally binding document that proves property ownership. A score of 1 is assigned if the service standard is accessible online or on a public board; a score of 0 is assigned if it is not made available to the public or if it can be obtained only in person.
5. Whether there is a specific and independent mechanism for filing complaints about a problem that occurred at the agency in charge of immovable property registration. A score of 1 is assigned if there is a specific and independent mechanism for filing a complaint; a score of 0 is assigned if there is only a general mechanism or no mechanism.
6. Whether the deed required to legally transfer title on immovable property can be processed online. A score of 1 is assigned if the deed is able to be processed online (such as through eRecording); a score of 0 is assigned if the deed must be processed in person.

The index ranges from 0 to 6, with higher values indicating greater transparency in the land administration system.

## Tier 1 Indicators (Top Priority for Data-Collection)

### Indicator 1: Parking Lot Minimums

This indicator aims to measure whether there is a local ordinance for the minimum requirement for parking lot sizes (a floor for how small a parking lot is allowed to be). This is generally a requirement for local businesses and is determined by city ordinances.

One year ago, we collected data for nearly every U.S. city in the first version of DBNA. Our objective is to make sure that this data has not changed within the last year (highly unlikely to have changed), make sure the original sources are up to the standards we've created this year (Official Source Name, Source Link/URL), and to collect data for the missing / new locations added in DBNA Version 2.

Our data from last year can be found [here](https://github.com/PaulBernert/DBNA/blob/master/_data/2020-report/Parking%20Requirements%20-%20Zoning%20Permits%20and%20Land%20Use.xlsx). It's not going to show up, but if you click the Download button, you should have a local copy. We will want to transition the data from this spreadsheet onto QuickBase. While doing this, please make sure to check to make sure things are still up-to-date, and make sure the sourcing is up to our new standard!

### Indicator 2: Wharton Index Survey Question #2

Access to the Wharton Index can be found [here](http://realestate.wharton.upenn.edu/wp-content/uploads/2017/03/558.pdf). Survey Question #2 is on Page 49 (Page 51 in PDF Viewer).

Question 2 of the Wharton Index Survey asks which of the following are required to approve zoning changes in your local jurisdiction. There is a second part to the question (by what vote), but we are not interested in that portion of information. We are simply checking which entity is responsible for approving zoning changes.

#### Example of How To Collect This Data

I imagine the simplest way to find this information is to search for a city's municipal codes (for example, a Google search for 'Phoenix Zoning Municipal Codes' returns [this](https://phoenix.municipal.codes/ZO) as it's first result. On this site, we can click on Chapter 3, titled 'Decisionmaking and Administrative Bodies'.

We can see that there is a Board of Adjustment (Section 303) and a Zoning Administrator (Section 307). After briefly reading over, we can see that the Board of Adjustment is responsible for appeals and tasks delegated by the Zoning Administrator, so that's **NOT** necessarily a step required to approve zoning changes. However, the Zoning Administrator is in charge of approving zoning permits. This administrator is comissioned by the Planning and Development Director. Therfore, I would consider Phoenix AZ to require 'Local Council, Managers, and Commissioners'.

The definitions between the different options aren't that clear, so I'm working on a better write-up of the differences with the help of Steve Slivinski. I hope to have this section finished soon!

### Indicator 3: Wharton Index Survey Question #3

This also comes from the [Wharton Index](http://realestate.wharton.upenn.edu/wp-content/uploads/2017/03/558.pdf). Survey Question #3 is on Page 50 (Page 52 in PDF Viewer).

Question 3 of the Wharton Index Survey asks which of the following are required to approve a new project that does not need rezoning. There is a second part to the question (by what vote), but we are not interested in that portion of information. We are simply checking which entity is responsible for approving zoning changes that don't require rezoning.

#### Example of How to Collect This Data

This question is very similar to the previous indicator, with the only difference being that this one is looking for approval changes that don't require rezoning. The contents will generally be found in the same location as the previous indicator (the city's [municipal codes](https://phoenix.municipal.codes/ZO/3)). If there is a mention of anything related to exceptions surrounding rezoning specifically, then that information is the correct value for the previous indicator. I suspect a lot of the values for this indicator and the previous indicator will be identical, but the goal is to measure if there is any difference between the approval process of things that require rezoning and things that do not.

**RESOLVING INSOLVENCY**

**WHAT DOES RESOLVING INSOLVENCY MEASURE?**

Doing Business North America studies the time, cost, and outcome of insolvency proceedings involving domestic entities as well as the strength of the legal framework applicable to judicial liquidation and reorganization proceedings. Indicators related to time and the strength of the legal framework index are used to calculate the regulatory performance for resolving insolvency.

To make the data on the time, cost, and outcome of insolvency proceedings comparable across economies, several assumptions about the business and the case are used.

**ASSUMPTIONS ABOUT THE BUSINESS AND PARTIES**

**The business:**

* Is a limited liability company.
* If there is more than one type of limited liability company in the economy, the limited liability form most common among domestic firms is chosen.
* If the country does not have a limited liability company option, the company structure most similar to that of an LLC is chosen.
* Has a 10-year loan agreement with a domestic bank secured by a mortgage over the real estate property.
* Has a market value, operating as a going concern, of five times income per capita or $200,000, whichever is greater.

**The parties:**

The bank wants to recover as much as possible of its loan, as quickly and cheaply as possible. The unsecured creditors will do everything permitted under the applicable laws to avoid a piecemeal sale of the assets. The majority shareholder wants to keep the company operating and under their control. Management wants to keep the company operating and preserve its employees’ jobs. All the parties are local entities or citizens; no foreign parties are involved.

**INDICATORS**

**Time (in Calendar Years)**

Time for creditors to recover their credit is recorded in calendar years. The period of time measured by Doing Business North America is from the company’s default until the payment of some or all of the money owed to the bank.

**Strength of Insolvency Framework Index (Scale: 0 – 16)**

The strength of insolvency framework index is constructed using four other indexes: (i) the commencement of proceedings index, (ii) the management of debtor’s assets index, (iii) the reorganization proceedings index, and (iv) the creditor participation index.

The index ranges from zero to 16, with higher values indicating insolvency legislation that is better designed for rehabilitating viable firms and liquidating nonviable ones.

**Commencement of Proceedings Index (Scale: 0 – 3)**

**The commencement of proceedings index has three components:**

1. Whether debtors can initiate both liquidation and reorganization proceedings. A score of 1 is assigned if debtors can initiate both types of proceedings; a score of 0.5 is assigned if they can initiate only one of these types; a score of 0 is assigned if they cannot initiate insolvency proceedings.
2. Whether creditors can initiate both liquidation and reorganization proceedings. A score of 1 is assigned if creditors can initiate both types of proceedings; a score of 0.5 is assigned if they can initiate only one of these types (either liquidation or reorganization); a score of 0 is assigned if they cannot initiate insolvency proceedings.
3. What standard is used for commencement of insolvency proceedings. A score of 1 is assigned if a liquidity test is used; a score of 0.5 is assigned if the balance sheet test is used; a score of 1 is assigned if both the liquidity and balance sheet tests are available but only one is required to initiate insolvency proceedings; a score of 0.5 is assigned if both tests are required; a score of 0 is assigned if a different test is used.

The index ranges from 0 to 3, with higher values indicating greater access to insolvency proceedings.

**Management of Debtor’s Assets Index (Scale: 0 – 6)**

**The management of debtor’s assets index has six components:**

1. Whether the debtor can continue performing contracts essential to the debtor’s survival. A score of 1 is assigned if yes; a score of 0 is assigned if continuation of contracts is not possible or if the law contains no provisions on this subject.
2. Whether the debtor (or an insolvency representative on its behalf) can reject overly burdensome contracts. A score of 1 is assigned if yes; a score of 0 is assigned if rejection of contracts is not possible or if the law contains no provisions on this subject.
3. Whether transactions entered into before commencement of insolvency proceedings that give preference to one or several creditors can be avoided after proceedings are initiated. A score of 1 is assigned if yes; a score of 0 is assigned if avoidance of such transactions is not possible or if the law contains no provisions on this subject.
4. Whether undervalued transactions entered into before commencement of insolvency proceedings can be avoided after proceedings are initiated. A score of 1 is assigned if yes; a score of 0 is assigned if avoidance of such transactions is not possible or if the law contains no provisions on this subject.
5. Whether the insolvency framework includes specific provisions that allow the debtor (or an insolvency representative on its behalf), after commencement of insolvency proceedings, to obtain financing necessary to function during the proceedings. A score of 1 is assigned if yes; a score of 0 is assigned if obtaining post-commencement financing is not possible or if the law contains no provisions on this subject.
6. Whether post-commencement financing receives priority over ordinary unsecured creditors during distribution of assets. A score of 1 is assigned if yes; a score of 0.5 is assigned if post-commencement financing is granted super-priority over all creditors, secured and unsecured; a score of 0 is assigned if no priority is granted to post-commencement financing or if the law contains no provisions on this subject.

The index ranges from 0 to 6, with higher values indicating more advantageous treatment of the debtor’s assets from the perspective of the company’s stakeholders.

**Reorganization Proceedings Index (Scale: 0 – 3)**

**The reorganization proceedings index has three components:**

1. Whether the reorganization plan is voted on only by the creditors whose rights are modified or affected by the plan. A score of 1 is assigned if yes; 0.5 if all creditors vote on the plan, regardless of its impact on their interests; 0 if creditors do not vote on the plan or if reorganization is not available.
2. Whether creditors entitled to vote on the plan are divided into classes, each class votes separately and the creditors within each class are treated equally. A score of 1 is assigned if the voting procedure has these three features; 0 if the voting procedure does not have these three features or if reorganization is not available.
3. Whether the insolvency framework requires that dissenting creditors receive as much under the reorganization plan as they would have received in liquidation. A score of 1 is assigned if yes; 0 if no such provisions exist or if reorganization is not available.

The index ranges from 0 to 3, with higher values indicating greater compliance with internationally accepted practices.

**Creditor Participation Index (Scale: 0 – 4)**

**The creditor participation index has four components:**

1. Whether creditors appoint the insolvency representative or approve, ratify or reject the appointment of the insolvency representative. A score of 1 is assigned if yes; 0 if no.
2. Whether creditors are required to approve the sale of substantial assets of the debtor in the course of insolvency proceedings. A score of 1 is assigned if yes; 0 if no.
3. Whether an individual creditor has the right to access financial information about the debtor during insolvency proceedings. A score of 1 is assigned if yes; 0 if no.
4. Whether an individual creditor can object to a decision of the court or of the insolvency representative to approve or reject claims against the debtor brought by the creditor itself and by other creditors. A score of 1 is assigned if yes; 0 if no.

The index ranges from 0 to 4, with higher values indicating greater participation of creditors.

**How the ‘Resolving Insolvency’ Category is Ranked and Scored**

The ‘Resolving Insolvency’ category was ranked and scored using the following two indicators:

Topic and Indicator | Best Performer | Worst Performer

Time (in Calendar Years) | 0.80 years | 1.80 years

Strength of Insolvency Framework Index | 15 points | 11 points

For each indicator, there is a top performer and a bottom performer. Economies with the best performance for a given indicator are awarded 10 “points,” or a score of 10. Cities at the level of bottom performance, or cities at or below two standard deviations from the mean, are awarded a score of 0. All the cities in between are scored based on their distance to the frontier. For each city, the number of awarded points across all indicators is aggregated, then divided by the number of indicators for which we had data. This is done because not all locations have complete data across all indicators, and doing so allows for all locations to be included in comparison.

For more information on how indicators, indexes, and groups are scored, how the ranking and scoring system for a category works, or how the overall Ease of Doing Business rank and score were derived, please read the first section of this methodology appendix.

# STARTING A BUSINESS

## WHAT DOES STARTING A BUSINESS MEASURE?

Doing Business North America records all procedures officially required for an entrepreneur to start up and formally operate a limited liability company, as well as the time and cost to complete these procedures. These procedures include the processes entrepreneurs undergo when obtaining all necessary approvals while completing any required notifications, verification steps, or inscriptions for the company and employees with relevant authorities. It does **not** include steps related to keeping a business compliant. For example, it does not include steps related to obtaining licenses and permits (which is generally a necessary step to continue legal operation).

After a study of laws, regulations, and publicly available information on business entry, a detailed list of the most common procedures was developed, along with the time and cost to comply with each procedure under normal circumstances.

Information is also collected on the sequence in which procedures are to be completed and whether procedures may be carried out simultaneously. It is assumed that any required information is readily available and that the entrepreneur will pay no bribes.

## ASSUMPTIONS ABOUT THE BUSINESS

To make the data comparable across economies, several assumptions about the business are used.

### THE BUSINESS:

* Is a limited liability company.
* If there is more than one type of limited liability company in the economy, the limited liability form most common among domestic firms is chosen.
* If the country does not have a limited liability company option, the company structure most similar to that of an LLC is chosen.
* Operates in the economy’s largest business city.
* Is 100% domestically owned and operated.
* Has start-up capital of two times income per capita.
* Performs general retail activities, such as the production or sale to the public of goods or services (NAICS Code: 4523).
* The business does not perform foreign trade activities and does not handle products subject to a special tax regime.
* Does not qualify for any financial incentives or special benefits.
* Has 50 employees, including the owner / entrepreneur.
* Has a company deed that is 10 pages long.

## INDICATORS

### Number of Procedures to Start a Business

A procedure is deﬁned as any interaction of the company founder with external parties. Procedures that must be completed in the same building but in different offices or at different counters are counted as separate procedures. If the founder has to visit the same office several times for different sequential procedures, each is counted separately. The founder is assumed to complete all procedures themselves, unless the use of such a third party is mandated by law. If the services of professionals are required, procedures conducted by such professionals on behalf of the company are counted as separate procedures. Each electronic procedure is counted as a separate procedure. Only pre-incorporation procedures that are officially required or commonly done in practice for an entrepreneur to formally operate a business are recorded.

Procedures required for official correspondence or transactions with public agencies are also included. For example, if a company seal or stamp is required on official documents, such as tax declarations, obtaining the seal or stamp is counted. Similarly, if a company must open a bank account in order to complete any subsequent procedure, this transaction is included as a procedure.

Only procedures required for Limited Liability Companies is being accounted for in Starting a Business. Procedures that the company undergoes to connect to electricity, water, gas, and waste disposal services are **not** included in the Starting a Business indicators.

#### Additional Information:

After a study of laws, regulations, and publicly available information on business entry, a detailed list of the most common procedures was developed, along with the time and cost to comply with each procedure under normal circumstances. That list is composed of the following eight steps for the United States:

1. Reserve/Register the Name of LLC
2. Choose/Assign a Registered Agent
3. File the Articles of Incorporation/Organization/Formation (or any similar name)
4. Complete State LLC Publication Requirements
5. File the Initial Statement of Information
6. Create a State LLC Operating Agreement
7. Obtain an Employment Identification Number for your States LLC
8. Additional County / City Level Requirements

### Time (in Working Days) to Start a Business

Time is recorded in calendar days. It is assumed that the minimum time required for each procedure is one day. Procedures that can be fully completed online are also considered to take one day. Although procedures may take place simultaneously, they cannot start on the same day, unless procedures can be completed entirely online. The registration process is considered completed once the company has received the final incorporation document or can officially commence business operations.

It is assumed that the entrepreneur does not waste time and commits to completing each remaining procedure without delay. The time that the entrepreneur spends on gathering information is not measured. It is assumed that the entrepreneur is aware of all entry requirements and their sequence from the beginning but has had no prior contact with any of the officials involved. If there is an option to expedite processing times, that option is always taken (even if it comes with an additional fee, which will be captured in the Cost (as a Percent of Income per Capita) to Start a Business indicator).

#### Additional Information:

This indicator measures the amount of time it takes to finish all procedures at both the city and state level. Ideally, information related to processing time is listed somewhere within the official sources, but very rarely is that the case. Most locations choose not to commit to a specific time frame in order to protect themselves from additional liability to complete things in a timely manner.

### Cost (as a Percent of Income per Capita) to Start a Business

Cost is recorded as a percentage of the economy’s income per capita. It includes all official fees and fees for legal or professional services if such services are required by law or commonly used in practice. Fees for purchasing and legalizing company books are included if these transactions are required by law. In all cases the cost excludes bribes.

## How the ‘Starting a Business’ Category is Ranked and Scored

The ‘Starting a Business’ category was ranked and scored using the following three indicators:

Topic and Indicator | Best Performer | Worst Performer

Number of Procedures | 3 procedures |13 procedures

Time (in Calendar Days) | 1 day | 51+ days

Cost (% of Income per Capita) | 0.14% | 30.59%

For each indicator, there is a top performer and a bottom performer. Economies with the best performance for a given indicator are awarded 10 “points,” or a score of 10. Cities at the level of bottom performance, or cities at or below two standard deviations from the mean, are awarded a score of 0. All the cities in between are scored based on their distance to the frontier. For each city, the number of awarded points across all indicators is aggregated, then divided by the number of indicators for which we had data. This is done because not all locations have complete data across all indicators, and doing so allows for all locations to be included in comparison.

For more information on how indicators, indexes, and groups are scored, how the ranking and scoring system for a category works, or how the overall Ease of Doing Business rank and score were derived, please read the first section of this methodology.

# University of Arizona - Coding Bootcamp - Final Project

## Research Objectives

This project aims to take the [raw data](https://github.com/PaulBernert/DBNA/blob/master/_data/2020-report/data_with_msas.csv) provided from the Doing Business North America report, calculate a rank and score to determine the 'Ease of Doing Business' (see the [Methodology](https://github.com/PaulBernert/DBNA/wiki/Ease-of-Doing-Business-Methodology) for explanation of what Ease of Doing Business measures). The effectiveness of the ranking/scoring process will be tested in an applied setting--using the calculated 'Ease of Doing Business' ranks to see if it correlates with business activity relative to the local population.

Another primary goal of the project is to use Machine Learning algorithms (through Scikit-learn) to test multiple clustering algorithms to see which locations are the most similar in nature. The two clustering algorithms to be tested are KMeans and Affinity Propagation. After comparing these two clustering structures, the next step is to test the relationship between clusters and the 'Ease of Doing Business' ranks to see if there clusters are formed around ranks (whether ranks are a good representative of how locations are clustered).

These tests were chosen not only to test the effectiveness of the data-set in an analytical environment, but to also see whether regulatory burdens truly have an impact on business starts.

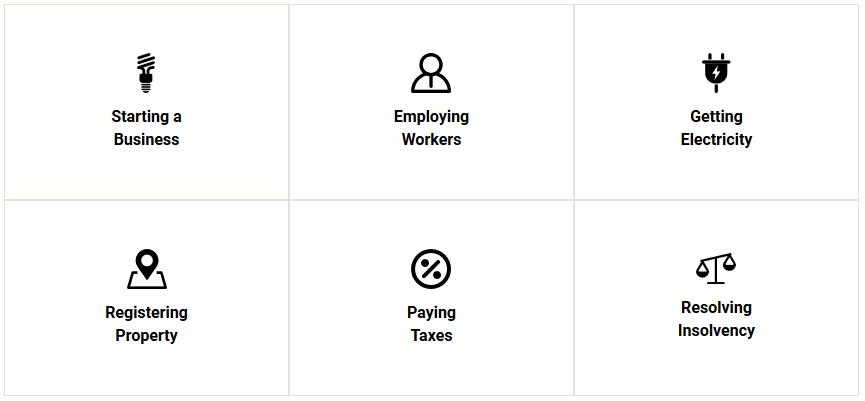
## About the Project

The Doing Business North America (DBNA) project annually provides objective measures of the scale and scope of business regulations in 130 cities across 92 states, provinces, and federal districts of the United States, Canada, and Mexico. It uses these measures to score and rank cities in regard to how easy or difficult it is to set up, operate, and shut down a business.

Over the years, researchers have begun to understand how robust measurement and ranking of regulations that either enhance business activity or constrain it can provide substantial insight into economic outcomes. Objective measurements of those regulations have been vital in this understanding. Unlike many studies that measure regulations at the state level, this annual study measures the impact at the city level and does so for over 100 municipal jurisdictions across North America.

The Doing Business North America team collected data on 63 different regulatory and economic indicators across six different categories. The data collected came entirely from official and publicly-available sources.

## About the Data



This project manually collects data on primary regulatory burdens businesses face throughout the entire life-cycle of a business, ranging from Starting a Business to eventually Resolving Insolvency if the business were to shut down / go bankrupt. The report contains data on 63 regulatory indicators within the following six categories:

1. Starting a Business
2. Employing Workers
3. Getting Electricity
4. Paying Taxes
5. Land and Space Use
6. Resolving Insolvency

These six categories are then combined to create a catch-all value known as the 'Ease of Doing Business'. This is the value used in all analysis for this project.

#### Special Thanks to Arizona State University and the Center for the Study of Economic Liberty

**Research Question #1**

**Question: Does the *'Ease of Doing Business'* Score Correlate with Business Activity?**

In order to determine whether the *'Ease of Doing Business'* Score correlates with relative business activity for a given location, we must first determine what the *'Ease of Doing Business'* Score measures and how it is calculated.

The *'Ease of Doing Business'* Score is a metric created that focuses on the regulatory burdens a small- to medium-sized business would face from the birth of the business to the death of the business in cities across North America. It takes these regulatory burdens (over 60 included in this report) and creates a single number used to represent the regulatory climates in these different locations.

The *'Ease of Doing Business'* Score is calculated by using the raw data and applying a basic linear transformation equation of: ((W-C)/(W-B))\*10, where W is the Worst Regulatory performance for a given indicator, B is the Best Regulatory performance for a given indicator, and C is the performance for the current observation (the city being calculated). For example, if the lowest minimum wage is $7.25 across the U.S., the highest minimum wage is $15.00 across the U.S., and I want to know the value for Phoenix AZ (a $12.00 minimum wage), that would be: (($15.00-$12.00)/($15.00-$7.25))\*10 => ~3.87. A location with a $15 minimum wage would get a 0.00 and a location with a $7.25 minimum wage would get a 10.00, where 10 is granted to the "best regulatory performance", and 0 is granted to the "worst regulatory performance".

This process is repeated for all indicators in a category, and then across all categories to get a final *'Ease of Doing Business'* Score. The locations with the highest scores are granted the highest rank, and the locations with the lowest scores are granted the lowest rank.

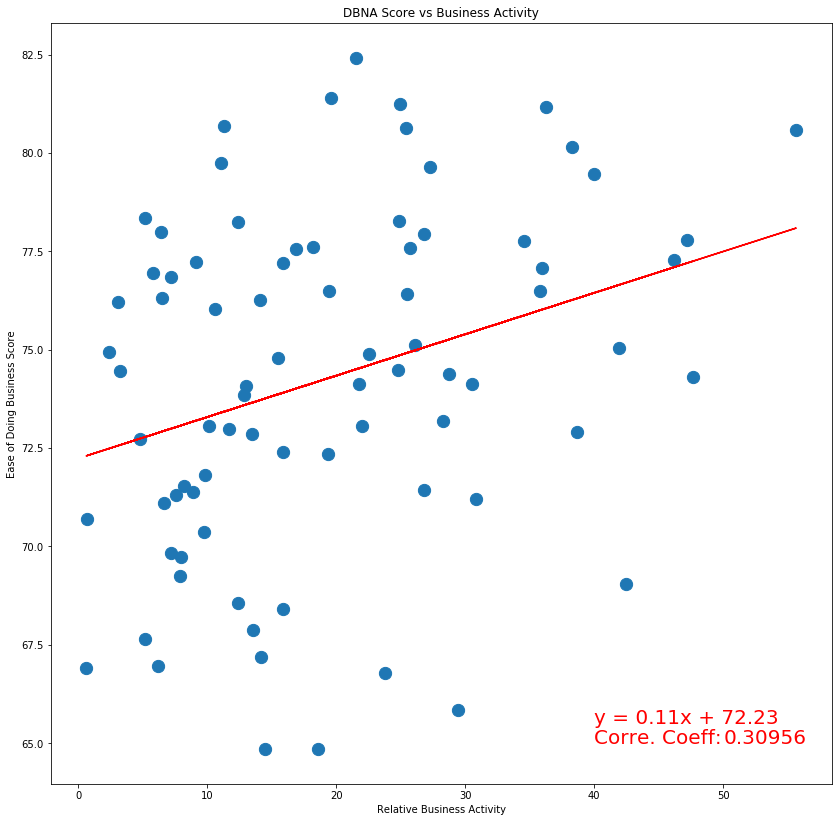
With the *'Ease of Doing Business'* Scores calculated, the top 10 cities are as follows:

1. Raleigh - North Carolina (82.42)
2. Jackson - Mississippi (81.39)
3. Tulsa - Oklahoma (81.25)
4. Sioux Falls - South Dakota (81.17)
5. Charleston - South Carolina (80.69)
6. Houston - Texas (80.64)
7. San Antonio - Texas (80.58)
8. Colorado Springs - Colorado (80.16)
9. Cincinnati - Ohio (79.75)
10. Cheyenne - Wyoming (79.65)

The assumption is that there should be relatively higher amounts of business activity in locations that have lower regulatory burdens, because lower burdens / barriers to entry incentivize taking higher risk and being involved in entrepreneurial activity. To test this theory, we need to find a metric of relative business activity.

Because the DBNA data specifically collects information geared towards small- to medium-sized businesses, data for this particular category of business type from the Census and other government websites was used. The Census provides the number of businesses with 25 or fewer employees at the city level, which is exactly what we need. The number of businesses is then divided by the city population, to get a number that reflects "the Number of Businesses relative to the local population". Cities with more businesses relative to population indicate that the population is, on average, more likely to be involved in entrepreneurial activity.

We can now begin to ask whether the DBNA *'Ease of Doing Business'* Score correlates with relative business activity. To do this, we use Numpy, Scipy and Matplotlib to produce the following results:



The results of our calculations tell us that: ***The R-Value between DBNA Scores and Relative Business Activity is 0.309564***, meaning the *'Ease of Doing Business'* Score and *'Relative Business Activity'* calculation have fairly strong, positive correlation at roughly 31 percent.

**Conclusion**

The results of this first research question were very satisfying. This data-set has never undergone any form of analysis before, so I was quite pessimistic in its ability to produce coherent results. Not only did it produce coherent results, but it also confirmed our initial hypothesis that regulatory burdens may indeed play a role in relative business activity. The magnitude of that correlation isn't perhaps as large as initially anticipated, but it does provide some context to where businesses start. It also allows us to now incorporate additional, non-regulatory indicators to see if we can complete a bigger picture on answering the question "What are the determinants when choosing where to open a business". The next steps forward are clear, and these are questions we plan to answer in time.

# Research Question #2

## Question: What New Information Can Clustering Provide?

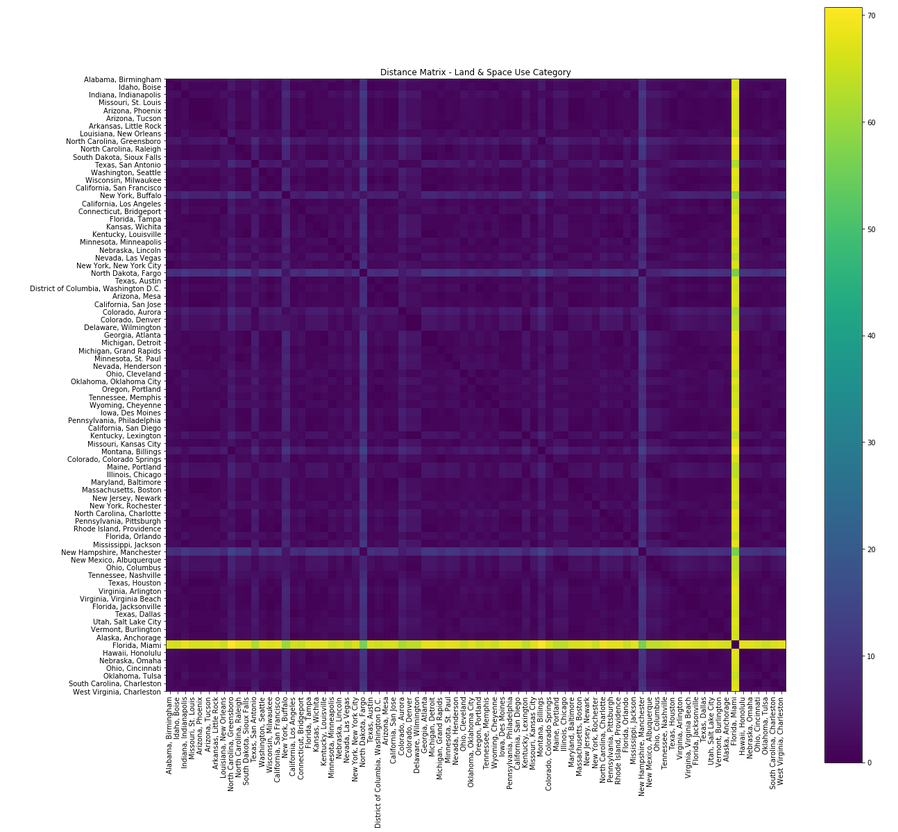
This question is intentionally open-ended, because it allows us to tackle multiple problems. First and foremost, does the DBNA Data cluster in a way that is coherent? Does clustering the data or clustering the scores (post-linear transformation) provide more interesting results? What clusters provide the most interesting results (by category, by all indicators, etc.)? If the data is able to be clustered, are clusters formed around the 'Ease of Doing Business' Scores/Ranks, or at least correlate with them? Do different clustering methods produce the same results, and if not, which clustering algorithm is best? There are many different questions that can be be answered, and my goal is to tackle as many of them as possible. Due to time constraints, we will focus on general outcomes and answering the questions proposed above.

### Clustering Method - KMeans vs Affinity Propagation

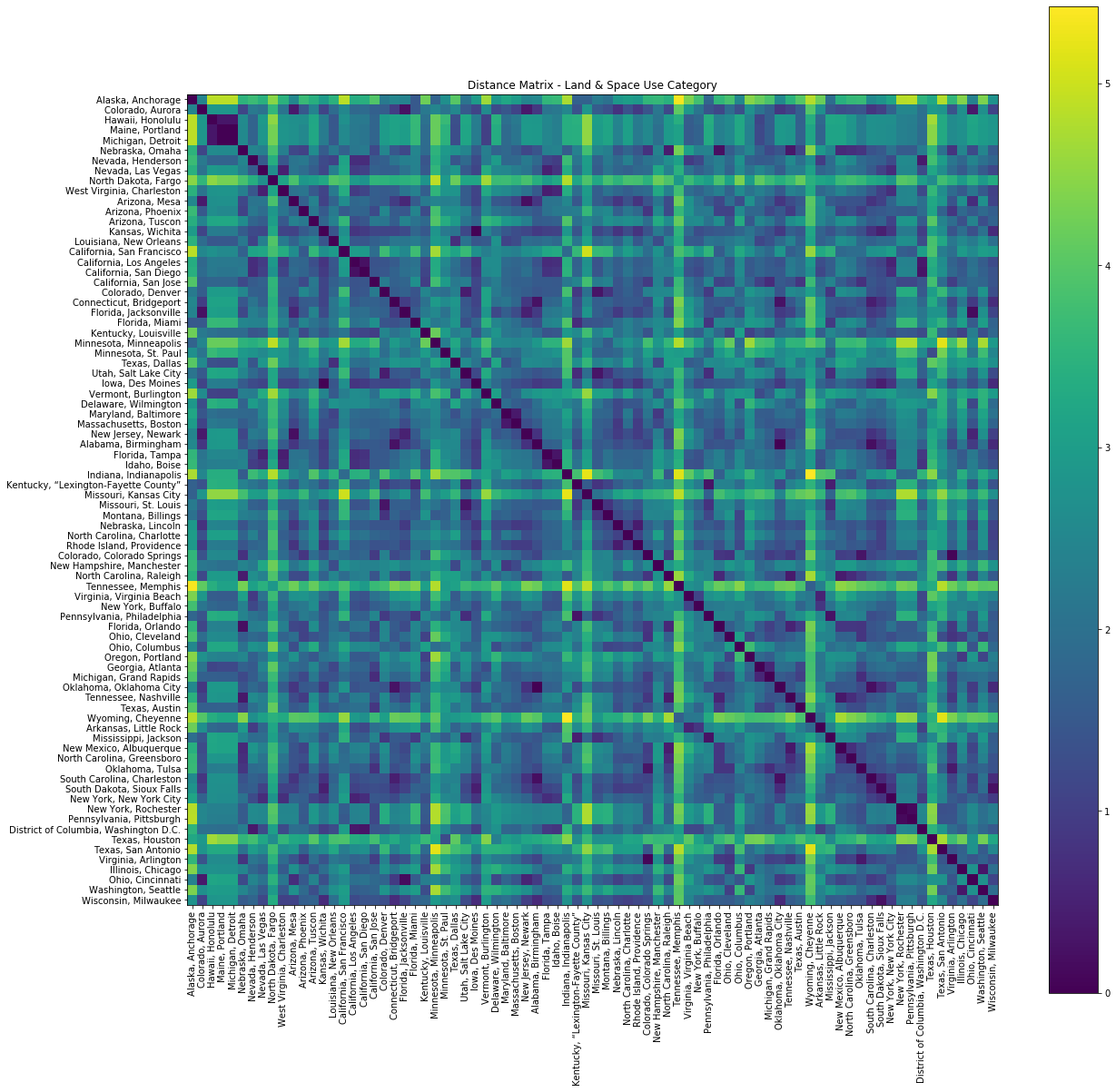
Before beginning a deep dive into the clustering, I wanted to test different algorithms to see which one produced the most accurate results. The two methods I was primarily interested in testing were KMeans clustering (a clustering method we've talked about in class) and Affinity Propagation (a different method of clustering that Scikit-learn can handle). KMeans is a nice introductory method of clustering because it simply calculates the Euclidean distances between different points. After assigning the number of clusters (the bulk of the "art" for this clustering method), you can get some interesting results. However, I believe that Affinity Propagation is a better clustering method for data that's as dynamic and varied as what's found in this particular data-set. Some indicators simply calculate number of procedures (ranging from values such as 3 to 13), while others are dollar values with massive distances ($10 to $100,000). Affinity Propagation easily allows for the inclusion of a weighting system, allowing data to be slightly more normalized. As a result, it is the method that I plan to use for all further clustering analysis.

### Clustering Raw Data vs Normalized Data

The Ease of Doing Business data is very dynamic, and as a result, even after switching to Affinity Propagation to use weighted values in clustering, there is still issues with the scales on some values. Take the 'Land and Space Use' category for example. In this category, there are some indicators with values that are averaged around 100 (with most values between 0 and 200). However, the Affinity Propagation method uses the Maximum Value when calculating the distance between points, and the distance matrix ends up looking something like this:

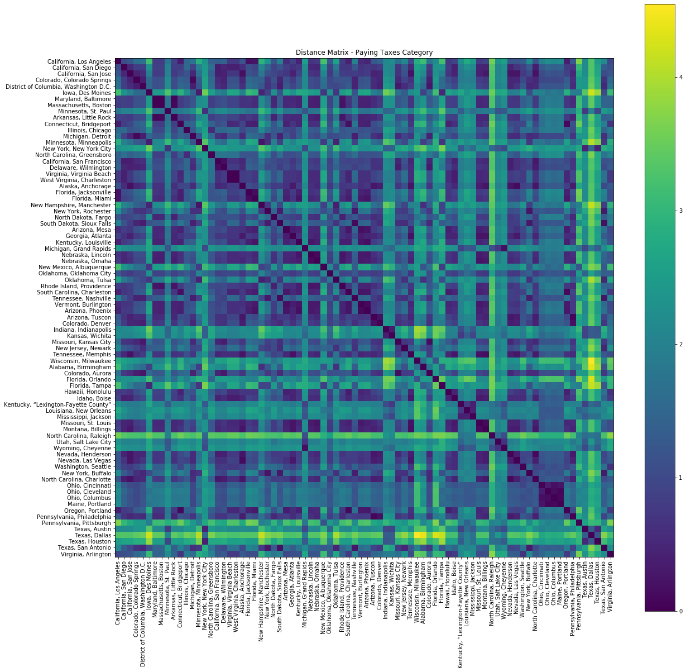
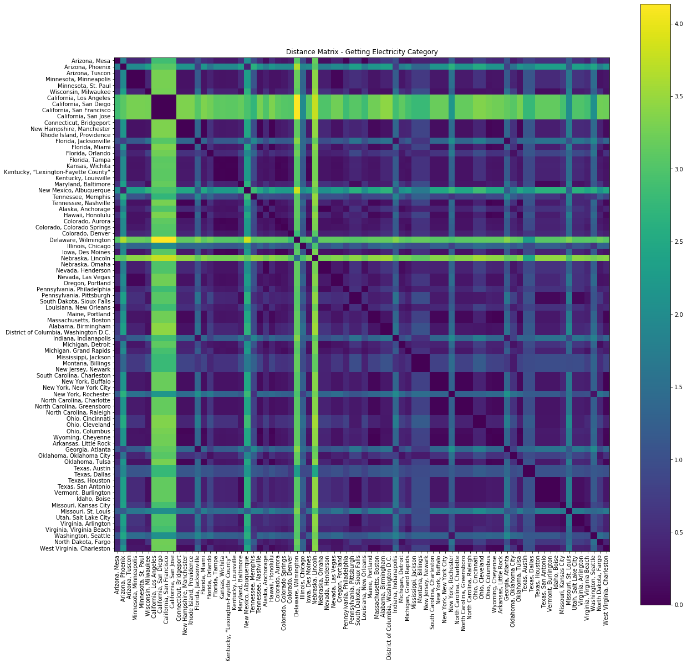
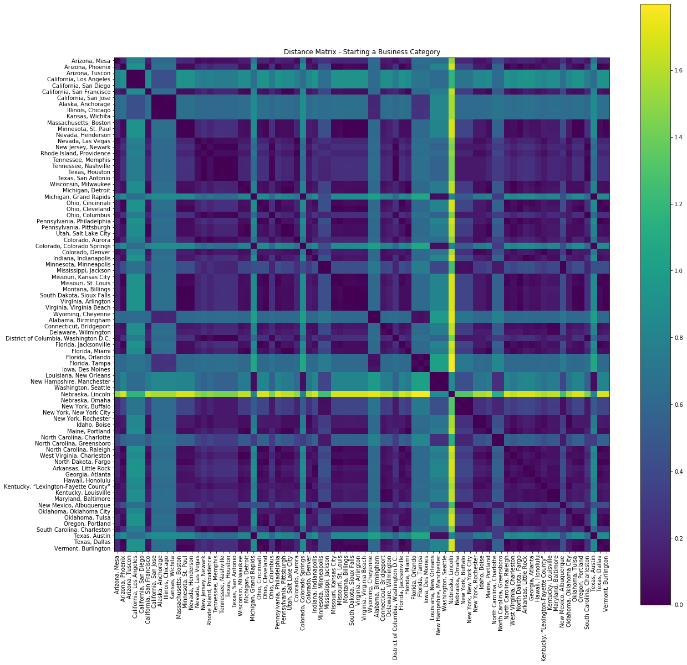


Using the Maximum Values has an obvious limitation in that outliers can have a dramatic impact on the scale in which the distance matrices are calculated. One of the most intuitive ways to fix this problem is to omit the outliers and adjust the scale in which Affinity Propagation is calculating values. Fortunately, we have already done this when answering Research Question #1. We have taken the data and applied a basic linear transformation to get the values to always be between 0 and 10 (where 0 is a bad outcome and 10 is a favorable outcome). When doing this linear transformation, we also removed outliers, so the data is not only more accurate but also easier to work with. When applying the same clustering technique on the **exact** same data (but after doing a simple transformation on the data), the results are a lot more interesting and digestible:



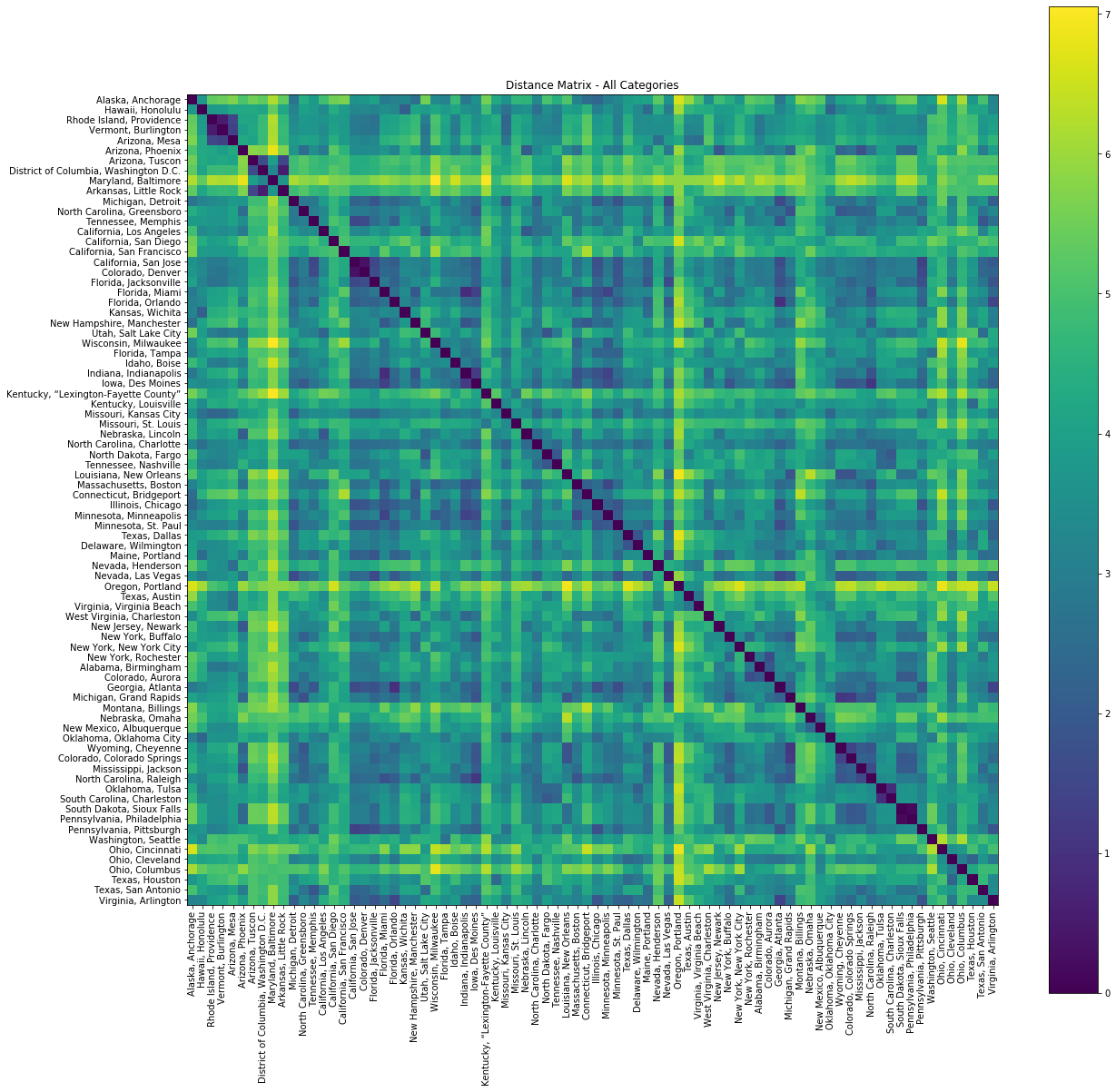
This graph produces results that are a lot more interesting. You can now begin to see some clusters forming within the distance matrix (areas with dark blue values), and begin to see which locations do not cluster well together at all (areas with yellow values). This process is then repeated for the other five categories (you can see those images [here](https://github.com/PaulBernert/DBNA/tree/master/images)).

##### Other Category Previews:

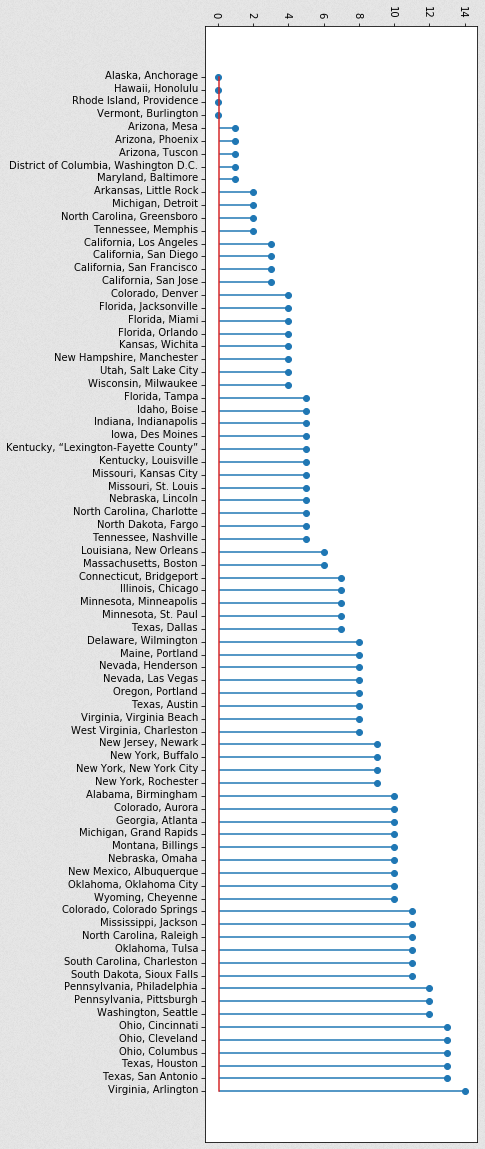


### Clustering Normalized Data Across ALL Categories

With every category individually clustered, the next test was to see whether clusters can be formed across all categories. The distances for each individual category generally had a small range (sometimes 0-1, sometimes 0-4), so my initial hypothesis was that the categories would be able to cluster (albeit some of the clustering would look forced due to high distances between points). However, after using Affinity Propagation to do clustering, these were the results:



I think the results are quite interesting. It's clear that clusters do indeed form, and the average distance between locales isn't as bad as initially predicted. The range of the distances goes from 0 to 7, which isn't absurd given the high variance for distances between some of the categories. Affinity Propagation produces the following clusters:

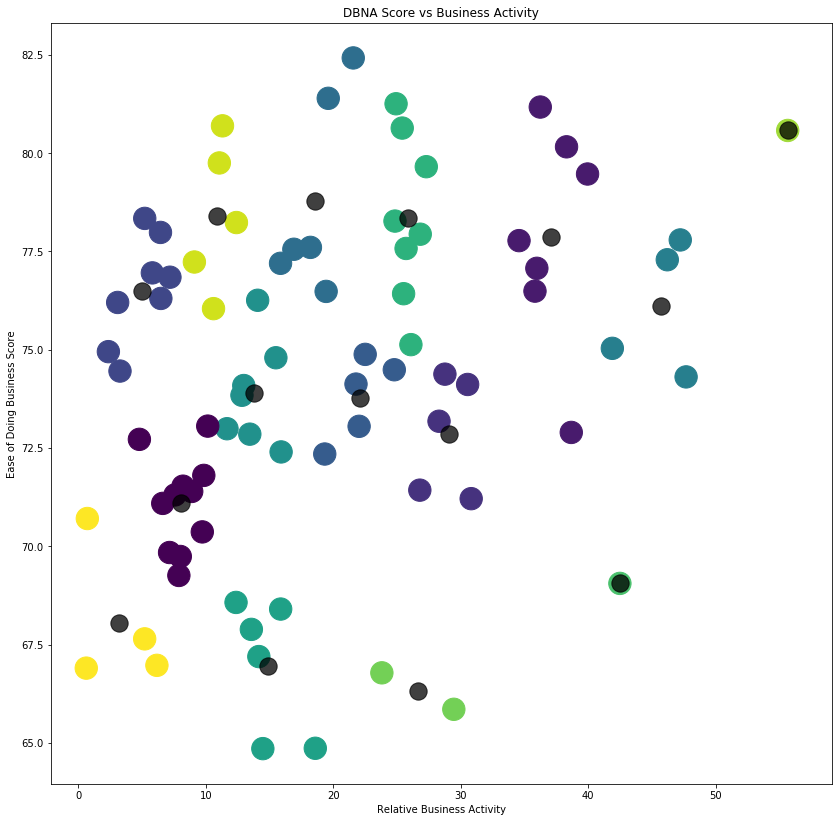


### Cluster Analysis

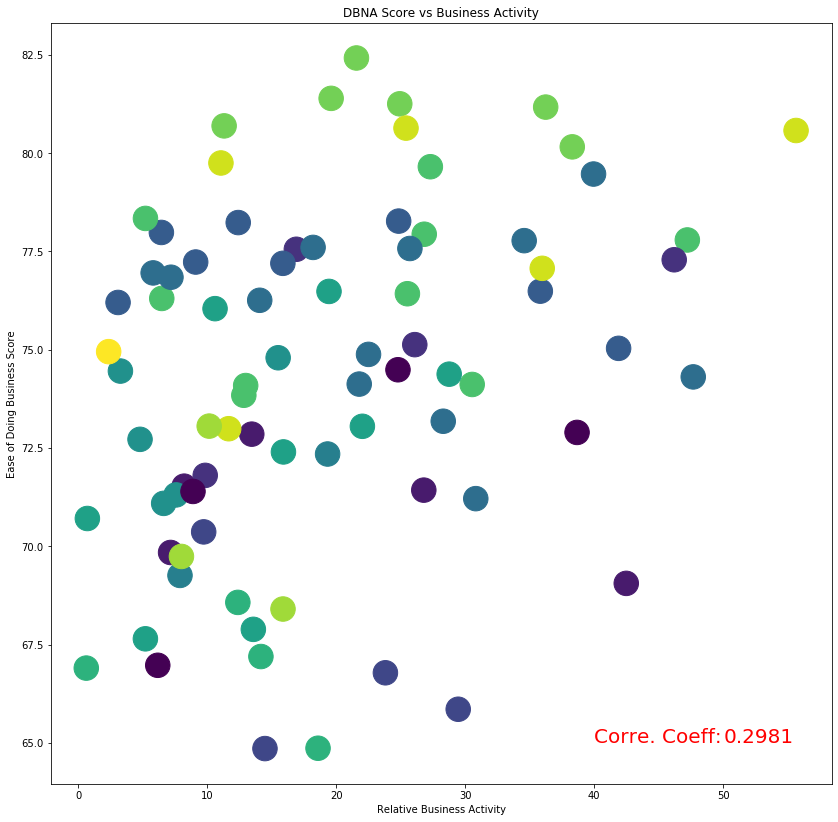
With the clusters assigned, one of the objectives was to solve whether there was a relationship between clusters and the 'Ease of Doing Business' Rank and Score. Ideally, the clusters are created around the 'Ease of Doing Business' Scores, where Cluster 1 contains cities with the highest Score, Cluster 2 the 2nd highest Scores, etc. Previous analysis confirmed that the DBNA data does indeed reflect where there is relatively higher amounts of business activity, so by transfer-ability, we should expect to see clusters form around the calculated Ranks and Scores. To test this, we need the following parameters:

1. Location (State, City)
2. 'Ease of Doing Business' Score
3. 'Ease of Doing Business' Rank
4. Cluster Number
5. Relative Business Activity

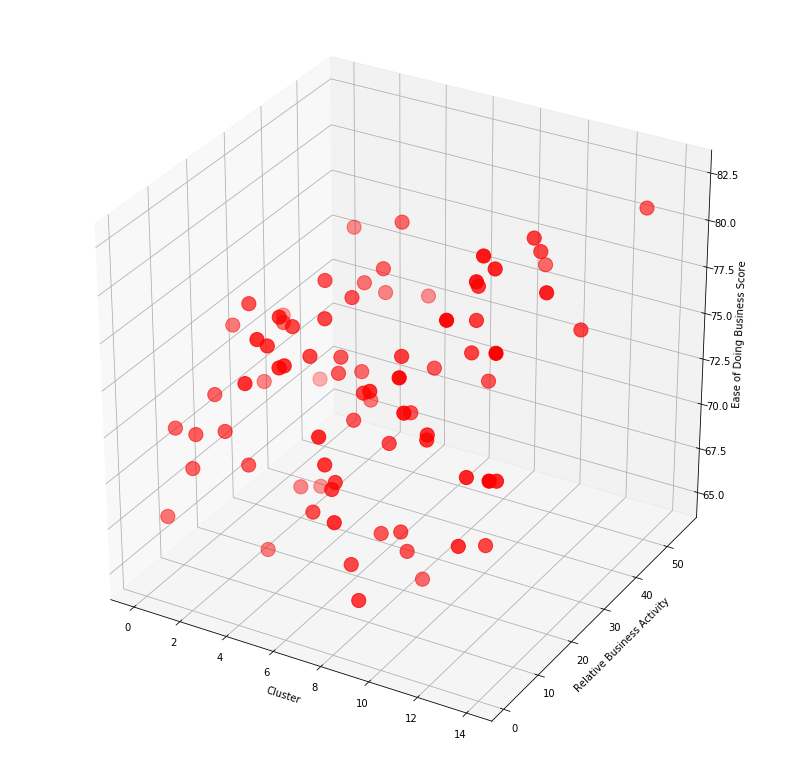
Using these five parameters, we now have everything to begin testing clustering methods. However, it is important to recognize that our clustering is around pre-determined groups based on the distance matrices calculated in the analysis above. It is **NOT** clustering around our previous research question, where we calculated points for the 'Ease of Doing Business' Score against Relative Business Activity. That clustering would produce these results:



While an interesting graph, it creates a new set of clusters that contradict and conflict with our previously-determined clustering. However, it does create the foundation for some analysis we can do using the correct clusters. We can continue to use our correct set of clusters to determine whether clusters form around ranks and scores like we initially anticipated. That results in the following visualization:



With this visualization, have reached the final results of analysis on this particular project. All further analysis is done using the results found here. Notice how it appears clusters are formed not around local dots, but across the horizontal spectrum (the top 10% of the graph is Green, the next 10% is Yellow, the following 10% is Dark Blue, etc.). This behavior indicates that clusters do indeed form around 'Ease of Doing Business' Scores. However, it is apparent that it's not a perfect correlation between Rank and Cluster. Observe the following two visualizations, using the exact same data but split into two different methods:



It's difficult to take three massive sections of data and assume that these things are the sole cause of one another, as there can simply be correlation without causation. Like with the previous research question, this is simply the beginning of the potential analysis that can be done with the Doing Business North America dataset, and I encourage the reader to take this dataset and conduct your own research to see what interesting results you can find!