Entrepreneurship Scores Report Draft

Authors:
Junfeng, Jason, Devrim, Kyrylo, Grant.

Table of Contents

Introduction

It is no secret that Texas is becoming a hub for entrepreneurs in the United States. During the pandemic, an overwhelming number of companies and entrepreneurs relocated to cities like Dallas, Houston, and Austin. Whether it be the tax-friendly business laws or the Southern hospitality, the number of people moving to Texas from other states nearly doubled during the first year of the pandemic. While many entrepreneurs recently moved to the state, Texas has long been a source of innovation with companies like Dell Technologies and ExxonMobil founded here.

This project set out to identify how each Texas county fared in entrepreneurship and what factors contributed to some being more successful than others. We used three main components to meaningfully measure entrepreneurial outcomes in counties: number of new companies per capita, number of small business loans (SBL) per capita, and number of patents awarded per capita. We collected many kinds of data from the past 30 years for each Texas county. We used an exponential regression machine learning model to analyze the statistical significance of the data. Separately, we created a web tool for people to explore our data, visualized in geographic, time-series, and correlational terms. After a thorough analysis of both tools, we came up with a multitude of patterns and observations across Texas.

Through our analysis, we have found that the counties located in large cities, such as Travis County with Austin or Collin County with Dallas, were the leaders in entrepreneurship when equally weighting the three variables across all 28 years of our data. It is interesting to note that counties in West Texas like Midland County and Martin County were leaders regarding pure market entry. Additionally, we saw a trend in high amounts of SBLs being taken out in Texas's northern and central regions.

This research highlights the differences in entrepreneurship across the state of Texas. It is apparent that large cities and a select few rural counties tend to lead the charge regarding entrepreneurship measures. Policymakers may be able to use this visualization to understand better which counties and regions need additional support to improve entrepreneurship ecosystems and make Texas the leading state of innovation across the country.

With this paper, we set out to find and create a solution for local Texan leaders and government officials to cultivate entrepreneurship in their counties. We know that every county desires to boost the local economy and create more jobs for its residents, and therefore small

local entrepreneurs are their best asset. We aim to bring objective truth about counties and regions' progress in accomplishing any of their economic growth targets, and potential reasons behind discrepancies compared to counties across Texas.

Data

The data collected for this research fell into two overarching categories: socioeconomic and living conditions measures, and economic data. Health-related socioeconomic metrics (unhealthy days, percent uninsured, primary care physicians per capita, and violent crimes per capita) were obtained from the Robert Wood Johnson Foundation. Transport-related living conditions were obtained from the Environmental Protection Agency. Federal Communications Communication's data on residential internet access was also included in our work. Finally, the economic data per county level was taken from the Texas Comptroller of Public Accounts, Community Reinvestment Act, Bureau of Labor Statistics, County Business Patterns, and U.S Patent and Trademark Office.
[More information here on the data cleaning before the ML model]

Methodology

This analysis was conducted using several steps. First, data from multiple sources was obtained and cleaned for duplicates, errors, and missing data. After all annual data for each county in Texas was deemed to be accurate and relevant to our analysis, a weighted average was created considering the three variables measuring entrepreneurship: number of new businesses per capita, number of patent filings per capita, and the number of SBLs per capita. Then, we ran a machine learning model on the data. [More information on the model here]. The model sought to uncover relationships between transit, socioeconomic, criminal data, and entrepreneurship outcomes. Setting a threshold of significance at 5%, we found a correlation of the data columns with each of the three components. Remarkably, the only column in our dataset reliably significant across all three components is road network density, which has a negative correlation to all of them.

In the second part of our analysis, the data was uploaded to Tableau to visualize the weighted entrepreneurship score across each Texas county from 1990 to 2019. Additionally, we created a sliding weight scale to allow users to conduct their own analysis of each of the variables by applying different weights to the Tableau map. For example, a weight of 1 for Loans and 0 for both Patents and New Businesses would tell the map to only consider Loans in the score and visualization, meanwhile setting all weights to 1 would cause them to be weighted equally heavily. Users are also able to select the year for which to see the data, allowing them to witness the progression patterns across years.

To investigate time progression on a scientific scale, we created a second graph. It can display a progression of each of the three entrepreneurship score measures over the years for the selected county compared to the Texas average county line.

We believe that the best way to correct for historical and societal factors that put some groups at a disadvantage is to explicitly see their effect. Therefore, we created a third customizable dynamic graph plotting entrepreneurship measures against socio-economic ones. For example, it provides the opportunity to explore how % of white or % of male residents affect all these factors. We provide the following socioeconomic metrics to explore: *working age population, white residents rate, male residents rate, bachelor's degree rate, high school diploma rate, county population, unemployment rate, uninsured rate, violent crime rate, and median household income.* We hope that by finding correlations between any of the metrics versus the three entrepreneurship score components (new companies founded, SBL received, or patents awarded) local leaders will be able to find their next policy targets. Additionally, this map can be cross-referenced and closely studied to connect information from charts 2 and 3 as the entrepreneurship score measures share colors between the two visualizations.

Results and observations

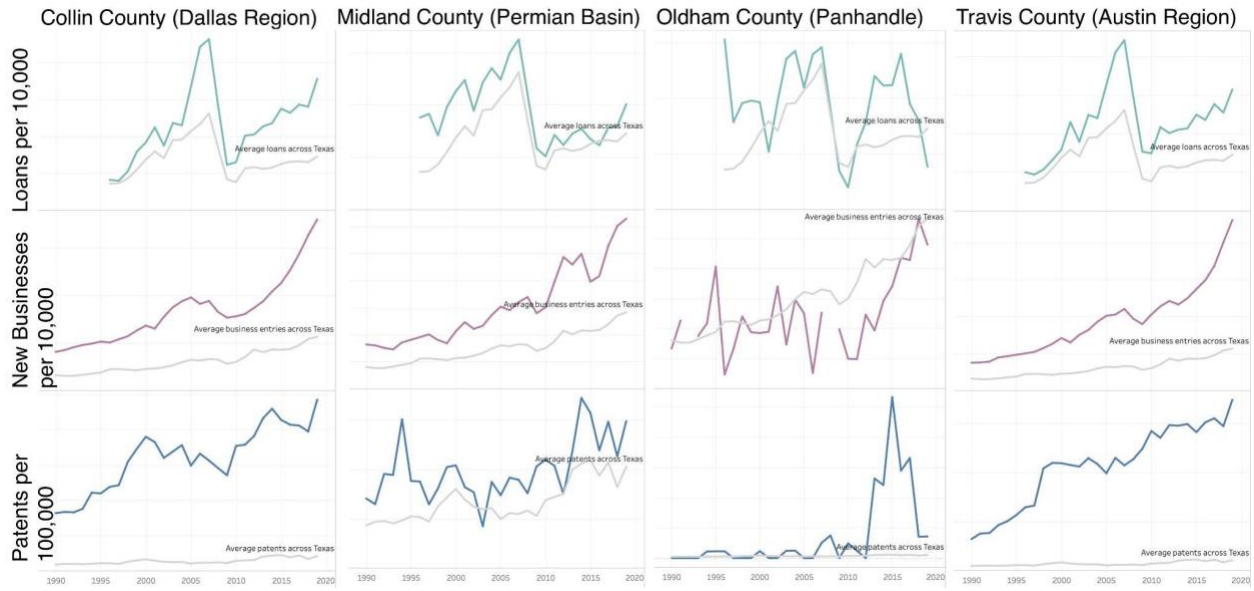


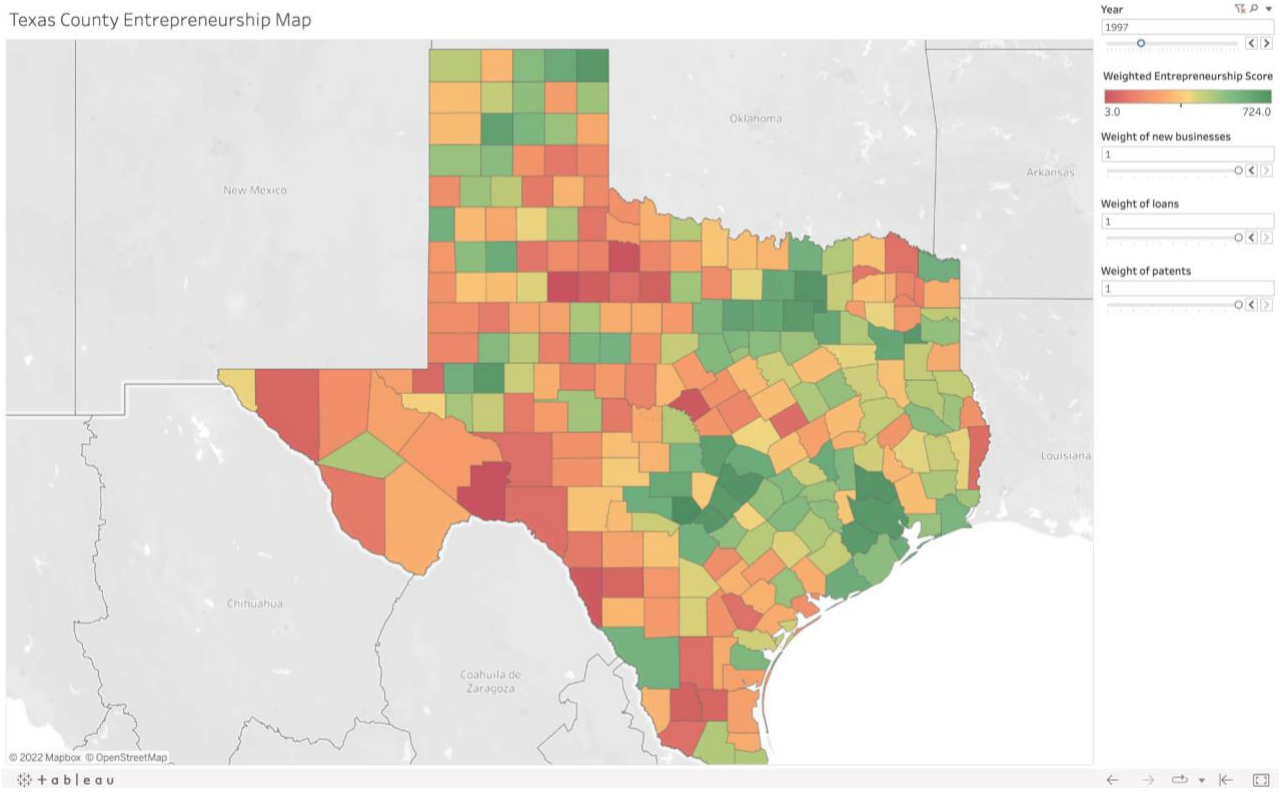
Figure 1

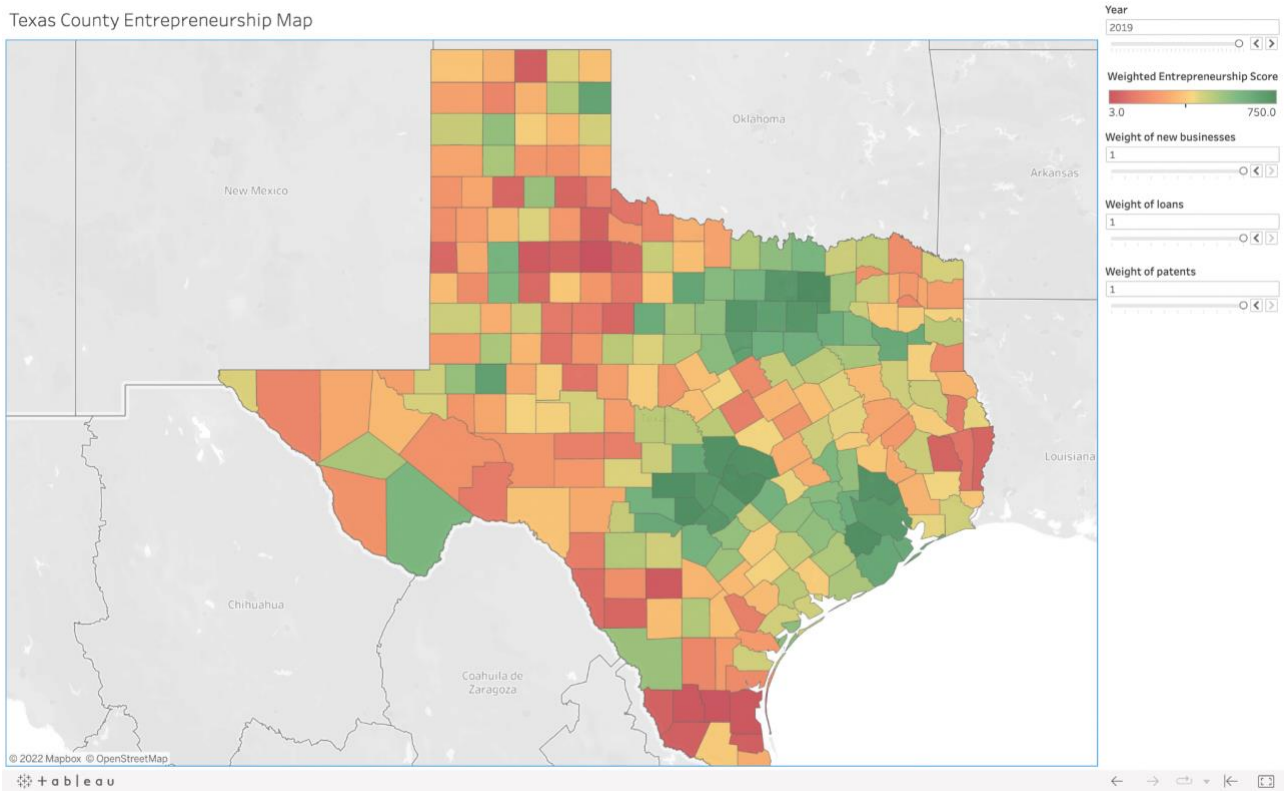
A comparison time series of four counties from different regions of Texas: Collin County - Dallas Region, Midland - Permian Basin, Oldham - Panhandle, and Travis - Austin Region. Displayed are each of the three entrepreneurship measures from 1990 to 2019.

Through our analysis, we have found that counties associated with larger cities, such as Travis County with Austin or Collin County with Dallas, were the leaders in respect to our entrepreneurship scores after equally weighting each of the metrics. This means that when

considering loans, number of new businesses, and patent filings equally, counties with a larger population seemed to have more entrepreneurial activity on average over the 28 years of our data. This trend is relatively consistent across all the years of our data, however, the 1990s saw some rural counties in northwest Texas perform relatively well to other regions across the state.

The Texas Panhandle region was far exceeding average Texas metrics for loans and new businesses per capita up until 2008. The 2008 global financial crisis led to a vast economic downturn for all counties across Texas, not circumventing the Panhandle. Due to the nature of its agriculture-centric economy, the Panhandle counties' rebound lagged behind the Texas average. The Panhandle has a very high rate of self-employment, and sole proprietorships are less likely to attract major investments. Perhaps, timely investments could have brought more capital to the region and helped it induce growth.

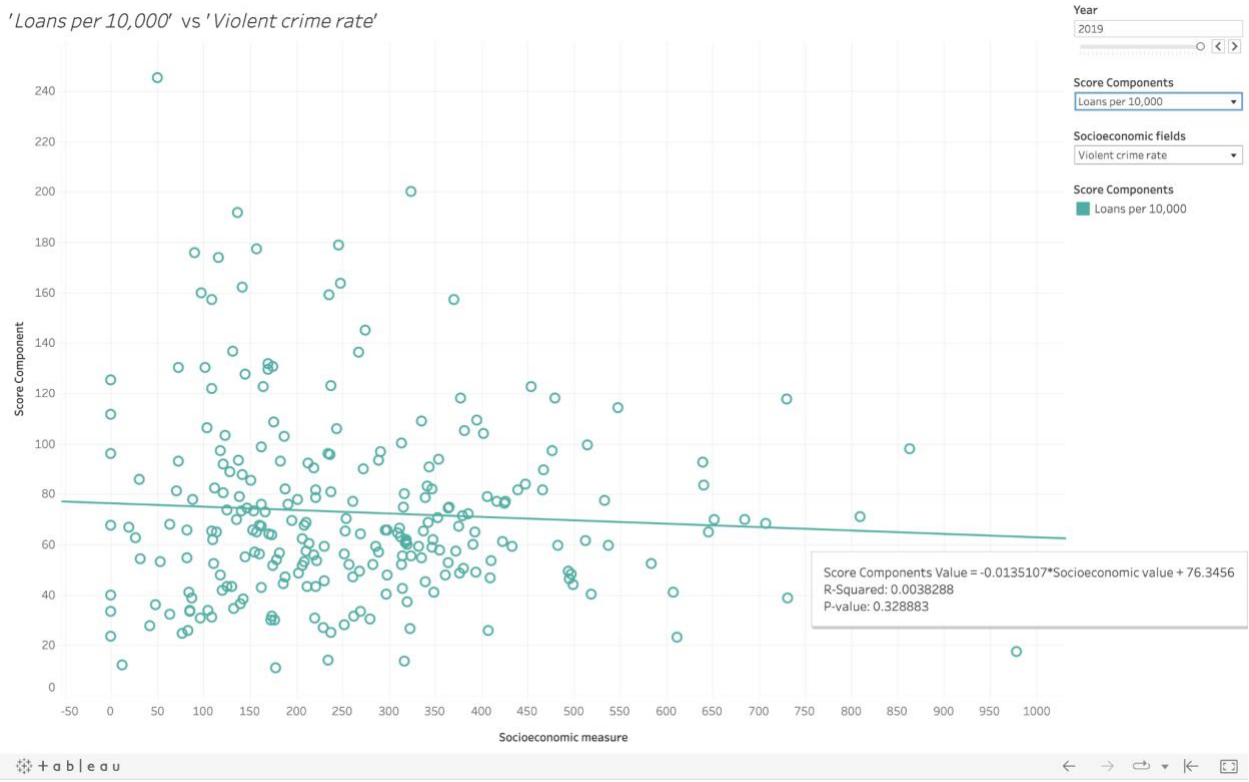


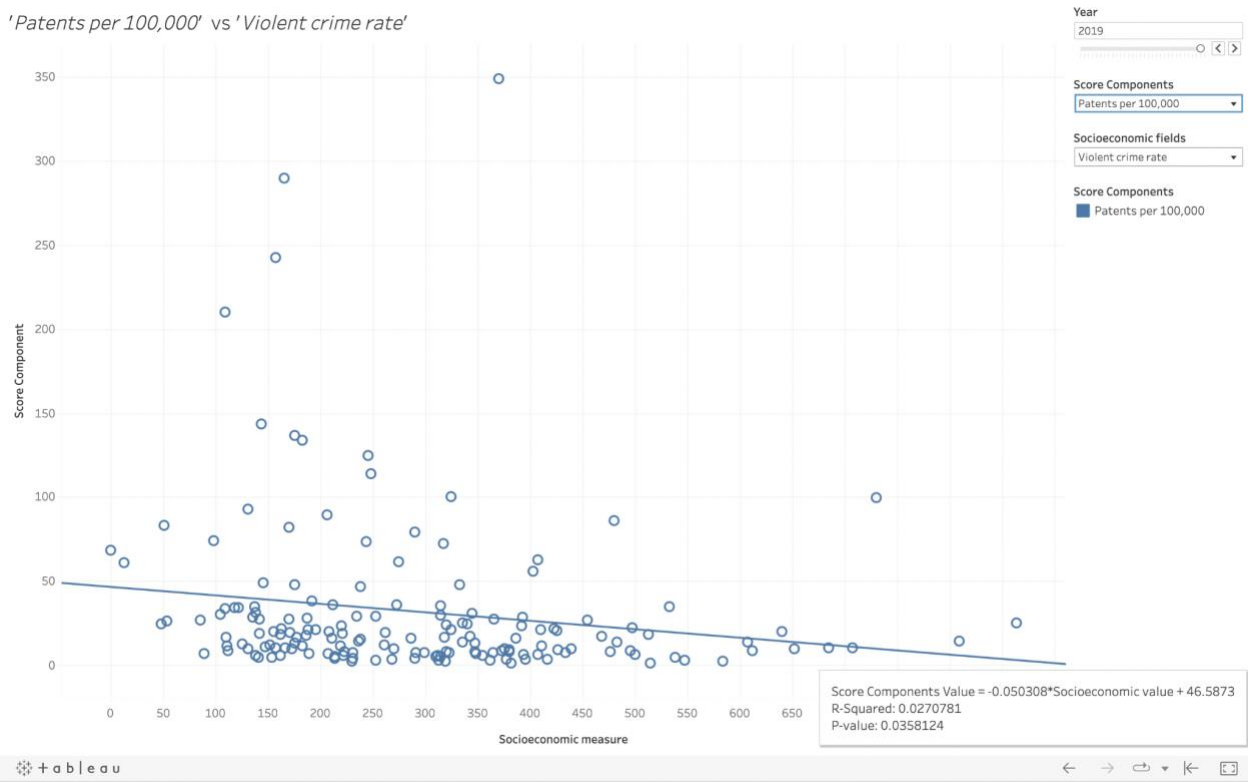


Figures 2 and 3
Texas entrepreneurship in 1997 (top) and 2019 (bottom)

It is also important to note that several rural counties scored high in other criteria outside of the aggregate calculation we used for our scoring criteria. For instance, when evaluating new companies registering in the state of Texas alone, counties in the Permian Basin showed strong signs of business growth. Similarly, rural counties in the Panhandle region of North Texas scored particularly well when evaluating SBLs per capita. The lack of patent filings is what ultimately dragged down rural West Texas counties when compared with their counterparts in the east.

Another significant observation is that impressively, the violent crime rate is not as much of a factor in either the number of loans or new businesses in a county (p-values .329 and .407). It is still a major factor for patent creation though (p-value .036), most likely because patent creation requires the availability of high-cost innovation centers, such as universities, research & development companies, etc., that are unwilling to risk their investments in dangerous areas.





Figures 4 and 5

Graphs measuring the correlation between loans per 10,000 (top) and patents per 10,000 (bottom) against the violent crime rate in Texas counties.

We were not surprised to learn that counties with a larger white population usually receive more loans but were shocked to see that advantage does not transfer into more new businesses or patents and creates a slight negative correlation instead. Sometimes, it is exactly social or cultural differences that are the difference.

Conclusion

In this paper, we desired to create the data foundation for local Texas leaders to be the driving force of entrepreneurial transformation, on both local and state legislative levels. We accomplished this by putting forth our analysis of the state and growth of Texas entrepreneurship. We were able to create a dynamic three-component visualization web tool the parameters of which can be manipulated to create truly important conclusions. We also put worth example usages of our web tool, from a shallow observance of which we were able to extract many axioms.

This study reveals multiple conclusions regarding the condition of entrepreneurship in the state of Texas. First, we looked for geographic patterns of growth and success around Texas, as we believe that location and natural surroundings can be determining factors for counties' growth. Unsurprisingly, counties in and surrounding business hubs like Austin, Houston, and Dallas scored exceptionally well in our evaluation criteria. This is to be expected, as entrepreneurs likely rely on resources and talent offered by major economic hubs to grow their small businesses. Rural counties in West Texas scored well in the SBLs and new business entries per capita, however, there was a major lapse in the number of patents filed per capita. This can be attributed to a lack of access to the technology required for innovation and patent filing.

With our second visualization, we were able to find growth spikes and slowdowns in the three entrepreneurship measures over the last 30 years. For example, the Texas Panhandle region was far exceeding average Texas metrics for loans and new businesses per capita up until 2008. The 2008 global financial crisis led to a vast economic downturn for all counties across Texas, but the Panhandle counties' rebound lagged behind the Texas average. Using our open-access project, local governments can also find the exact effects of local and global events on their economies and therefore find appropriate solutions or improvements that would address those exact factors. Our third visualization provides our users with further context on each county's social and economic situation, allowing stakeholders to conduct their own analysis. The overall trend line also shows an all-Texas perspective on how that measure correlates with the socioeconomic variable across the state.

Works Cited

https://www.theprpc.org/Programs/EconomicDevelopment/Texas%20Panhandle%20CEDS%20-%20FINAL.pdf