

# TEXAS COLLEGE ANALYSIS AND FOURSQUARE API DATA ANALYSIS

By: Frederick Orndorff

Date: January 28, 2020

*Coursera IBM Data Science  
Specialization*

A series of several parallel white lines of varying lengths, slanted diagonally from the bottom left towards the top right, located on the right side of the slide.



# Agenda

- Introduction
  - Business Problem
  - Stakeholders
- Data Sources
- Methodology
  - Visualizations
- Results
- Conclusion



# Introduction

## Business Problem:

There are over 5,300 Colleges and Universities in the United States and over 450 of these colleges are in the state of Texas. There are so many colleges in the Texas that a person could spend months (or even years) attempting to determine the best fit for a student -- or to determine which college is the best in Texas. There are many personal opinions on which college ranks first, but can we use data to determine the best college in Texas?

## Stakeholders:

There are many stakeholders in this problem, including: students, parents (who might pay for the student to attend college), transfer students, as well as, the college administration and presidents



# Data

## [2018 Texas Crime Data](#)

The Texas Department of Public safety compiles data from state and local law enforcement agencies across the country, as well as the FBI. Crime data is focused on violent crimes, hate crimes, and assaults.

## [Texas State Expenditures by County](#)

Texas State Expenditures by County shows where state dollars are spent. It lists state expenditures by agency and type of expenditure in each county. This data provides an overview of where the State is spending the local tax dollars for improvement(s).

## [Mixed Beverage Sales](#)

This file contains a list of taxpayers required to report mixed beverage gross receipts tax reports under Texas law. It provides total sales reported and taxes paid.

## [List of Texas Colleges](#)

This database provides the list of Colleges, Universities, and trade schools available in the state of Texas.



# Methodology

(1 of 3)

- **Part 1: County Analysis**

- Determine the best Texas county based on: 2018 Crime Rates; Alcohol Sales; and State Expenditures
- Use the county analysis as the basis for the college analysis

- **Part 2: College Analysis**

- Using the County Analysis – determine the best college from five preselected colleges

- **Part 3: Foursquare API data analysis**

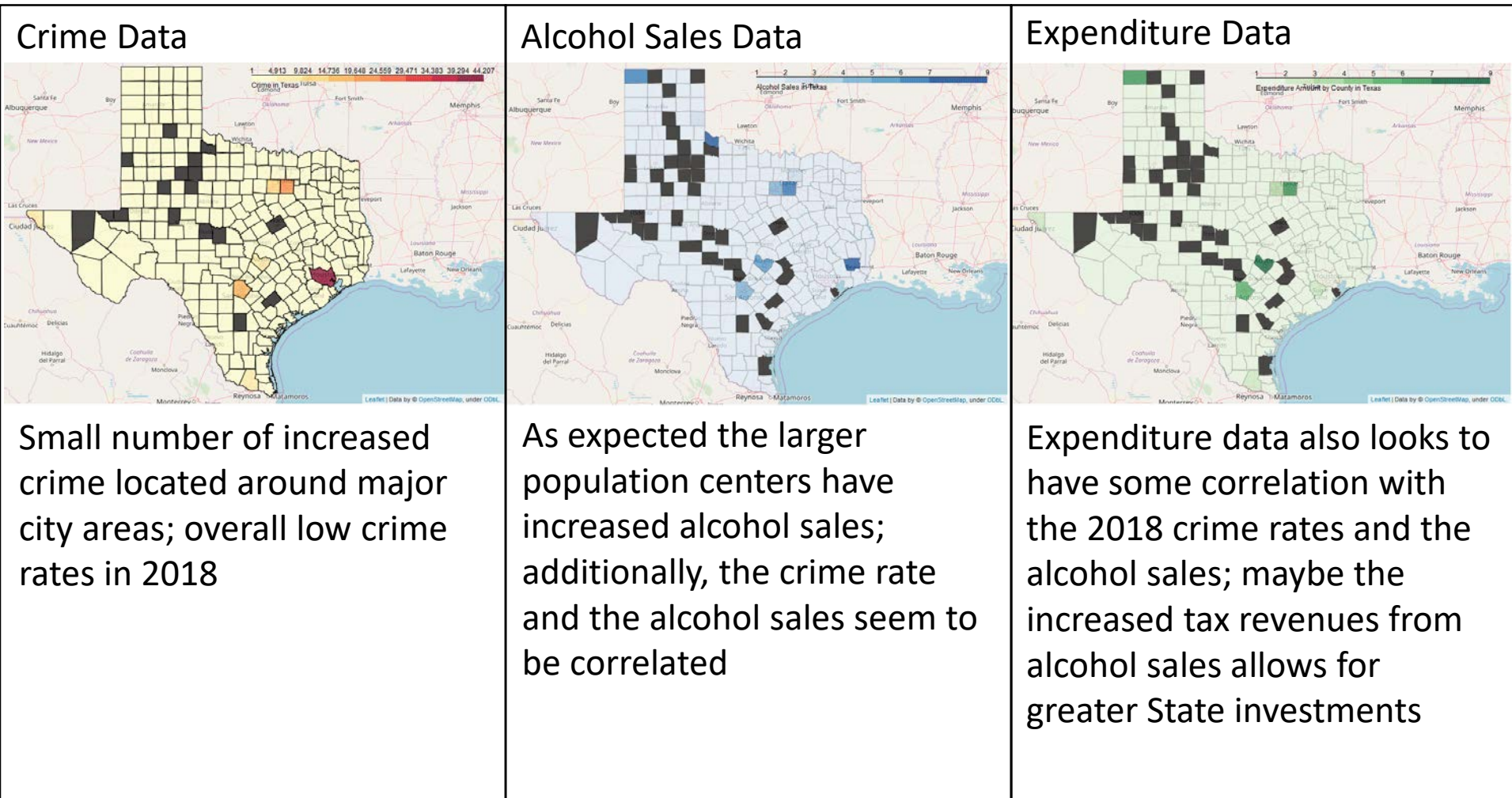
- Once an ideal college is determined – use the Foursquare API data to understand the college area
- Use clustering techniques to better analyze the area(s)



# Methodology

(2 of 3)

## Exploratory Data Analysis: Visualizations





# Methodology

(3 of 3)

## Exploratory Data Analysis:

Based on personal preference – the following colleges were used during the baseline assessment to compare counties and return the best college option

- University of Texas (UT) - Travis County
- University of Texas at San Antonio (UTSA) - Bexar County
- University of Houston (Houston) - Harris County
- Texas A&M (TAMU) - Brazos County
- Texas Tech University (TTU) - Lubbock County





# Results

(1 of 2)

## College Assessment:

Using our baseline options of colleges (below) and using non-weighted rankings we determined the best college was the University of Texas at Austin:

- University of Texas (UT) - Travis County
- University of Texas at San Antonio (UTSA) - Bexar County
- University of Houston (Houston) - Harris County
- Texas A&M (TAMU) - Brazos County
- Texas Tech University (TTU) - Lubbock County



University of Texas  
(UT)

	College	Crime_Rate_Rank	Alcohol_Sales_Rank	Expenditures_Rank	Total
0	UT	3.0	1.0	1.0	5.0
2	UTSA	4.0	2.0	3.0	9.0
3	TTU	2.0	3.0	4.0	9.0
1	TAMU	1.0	4.0	5.0	10.0
4	Houston	5.0	5.0	2.0	12.0



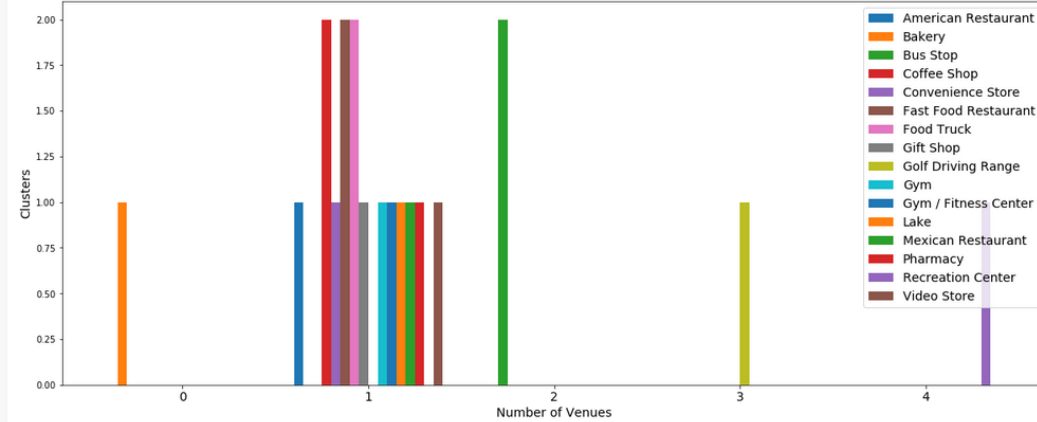


# Results

(2 of 2)

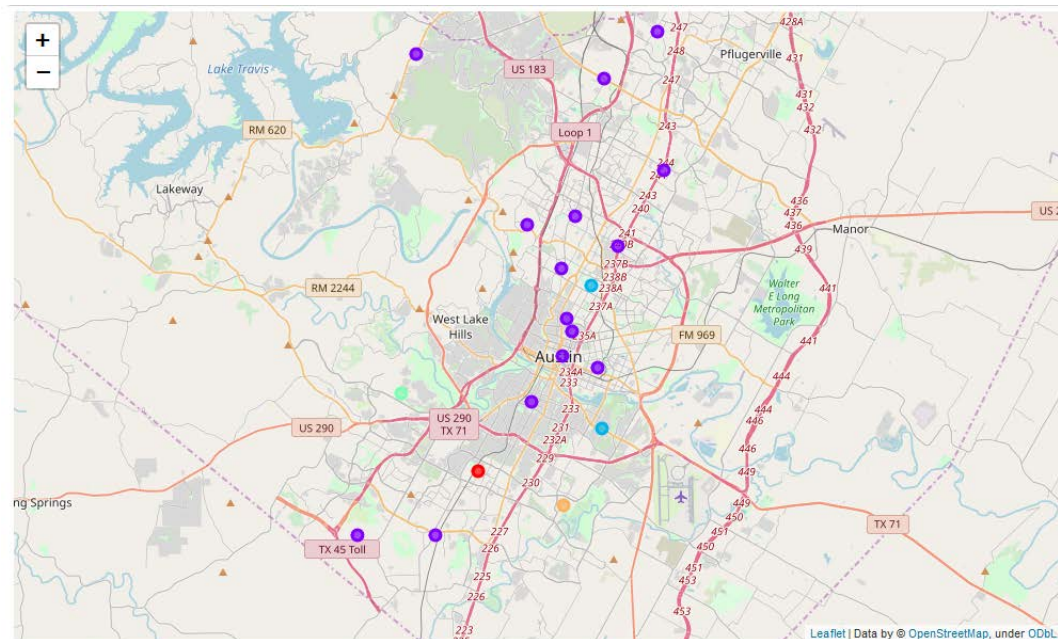
## K Means Clustering:

Number of Venues in Each Cluster



## Five Clusters:

1. Yoga & Bakery
2. Coffee & Fast Food
3. Transportation
4. TopGolf
5. Recreation Centers





# Conclusion

- Data cleaning and analysis is a large portion of a data driven analysis or project
- Not all programs are created equally. Understanding the limitations of programs allows for a faster completion time and less stress.
- Visualizations were key to understand the data
- I will need to further research the correlation between the variables in the county analysis
- Less populated counties were lacking data
- There are too many colleges in the United States