Sports Bar Market Analysis:
Midland, Hidalgo, Victoria, Bowie

Team : Blue 1
Authors:
Daniels, Jerdan
Deshmukh, Arpita
González-Hermosillo, Monika
Li, Bolin
Mou, Shuqi

Contents:

Objective

Perform market analysis on the data for a hypothetical company named "Billy's Sports Bar and Grill". The objective of the team is to find a prospective market where opening a sports bar would be lucrative. The analysis is to be done between 4 counties and find out the investment potential of opening of a sports bar in one of the 4 counties.

Part - I Analysis
I. County Demographic Summary

Midland

Midland is the second largest county in our group, based on population, just after Hidalgo County. It is located in west Texas and is an oil-industry driven county. Even though its population is the second largest of the four counties, its per capita income is the highest among the group. The population growth rate is also the highest of the 4 and stands at 4.3%, more than 4x larger than the following county. Among our group, Midland is the county with the highest percentage of the population with a university degree, and the lowest percentage of the population that lies under the poverty line.

Population (July 1, 2018)	Population growth rate	% of population that is white	% of population that is non-white	High school or higher	Bachelor's or higher	Per Capita Income in the past 12 months	% of people in poverty
172,578	4.3%	88.1%	11.9%	83.7%	26.8%	$38,545	10.7%

Source: all data retrieved from the US Census Bureau.
https://www.census.gov/quickfacts/fact/table/victoriacountytexas,bowiecountytexas,hidalgocountytexas,midlandcountytexas/AGE775218

Victoria

Out of our 4 counties, Victoria is the one with the smallest population. Additionally, its population has a negative growth rate, which means that the population is actually decreasing over time. The county's per capita income is the second largest in our sample, as is the percentage of people that hold a Bachelor's degree or higher. The percentage of whites in this county is 89.9%, second highest in the sample.

Population (July 1, 2018)	Population growth rate	% of population that is white	% of population that is non-white	High school or higher	Bachelor's or higher	Per Capita Income in the past 12 months	% of people in poverty
92,035	-0.07%	89.9%	10.6%	87.4%	19.5%	$28,181	15.7%

Source: all data retrieved from the US Census Bureau.
https://www.census.gov/quickfacts/fact/table/victoriacountytexas,bowiecountytexas,hidalgocountytexas,midlandcountytexas/AGE775218

Hidalgo

Hidalgo is the county with the largest population of the 4 counties, yet it has the lowest per capita income and the highest percentage of the population in poverty according to the Census Bureau. Out of the sample, Hidalgo has the highest percentage of white people, but the lowest education rates.

Population (July 1, 2018)	Population growth rate	% of population that is white	% of population that is non-white	High school or higher	Bachelor's or higher	Per Capita Income in the past 12 months	% of people in poverty
865,939	0.89%	97%	3%	63.7%	17.8%	$15,883	29.5%

Source: all data retrieved from the US Census Bureau.
https://www.census.gov/quickfacts/fact/table/victoriacountytexas,bowiecountytexas,hidalgocountytexas,midlandcountytexas/AGE775218
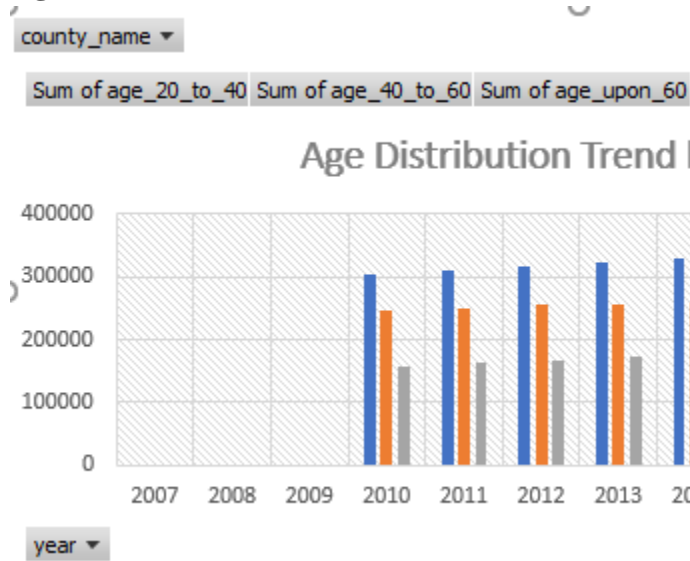
Bowie

The county of Bowie is the smallest one, with a population of approximately 94,324. Of that number 69.7% are white, the smallest percentage out of the sample. 30.3% of the population is non-white, which makes it the most diverse county out of the above mentioned. The population is growing at a very small rate of 0.44% in 2017.

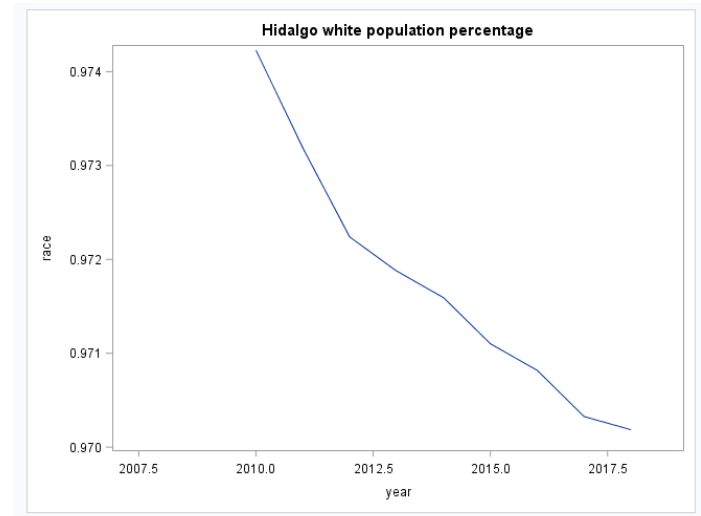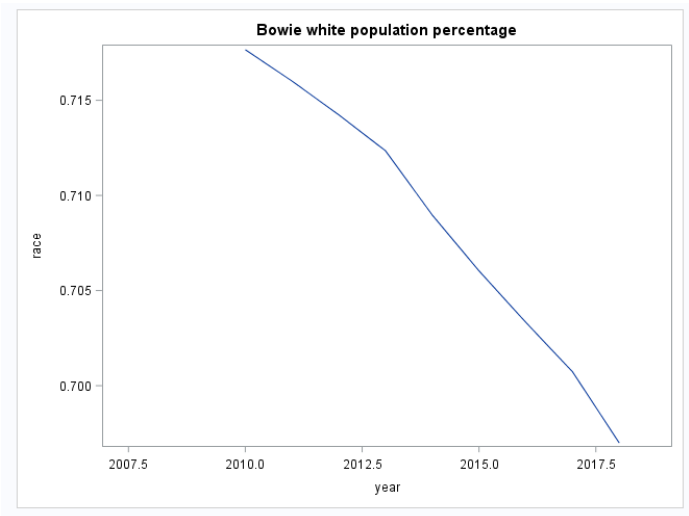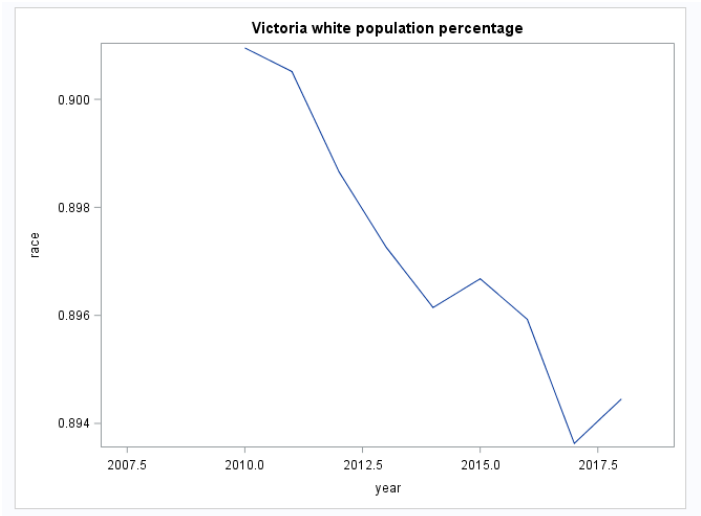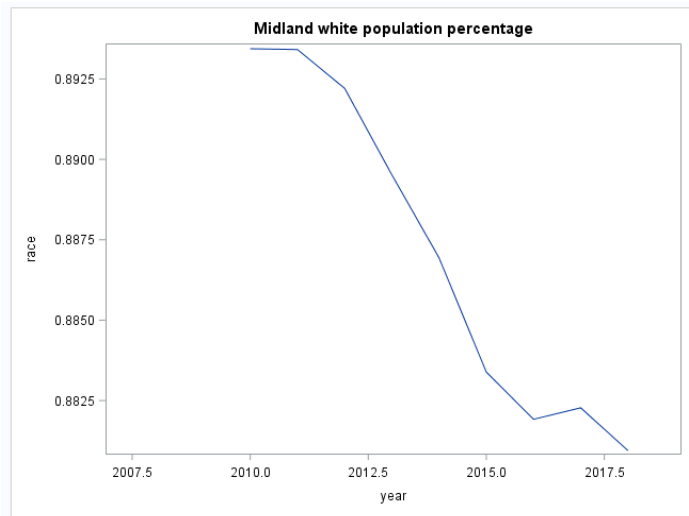| Population (July 1, 2018)	Population growth rate	% of population that is white	% of population that is non-white	High school or higher	Bachelor's or higher	Per Capita Income in past 12 months	% of people in poverty
94,324	0.44%	69.7%	30.3%	87.4%	19.2%	$24,761	15.9%

II. County Demographic and Income Growth Analysis
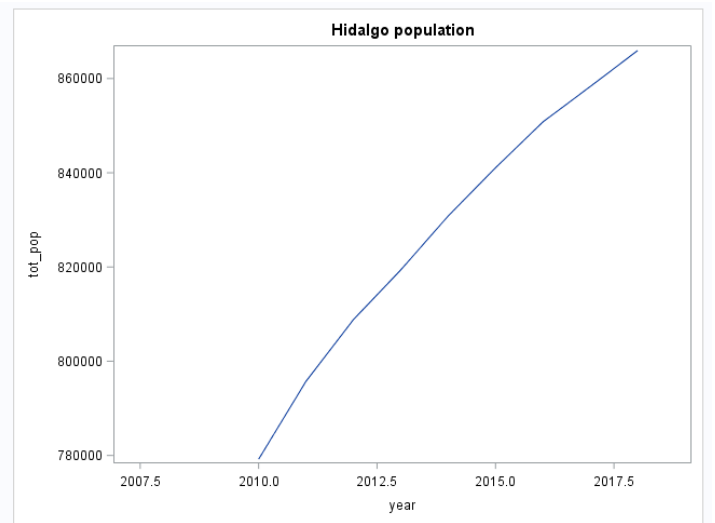
Age

The figure above is the age distribution for all four counties from 2010 to 2018. As we can observe from the graph, the age group that dominates in all four counties is the blue bar, age 20 to 40. The second most common age group is the orange bar, which represents age 40 to 60. Lastly the gray bar, age 60 and above, is the smallest proportion of the population in all 4 counties.

Race

Midland white population percentage

Victoria white population percentage

Bowie white population percentage

Hidalgo white population percentage

Population trends

Midland population

Victoria population

Bowie population

Hidalgo population
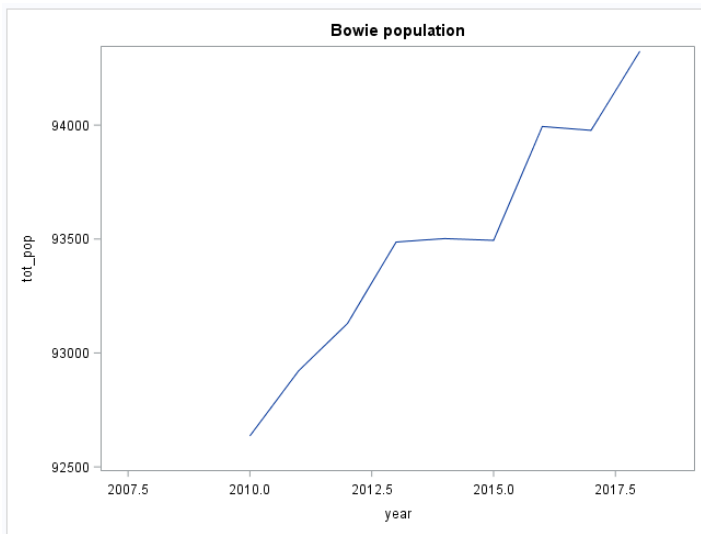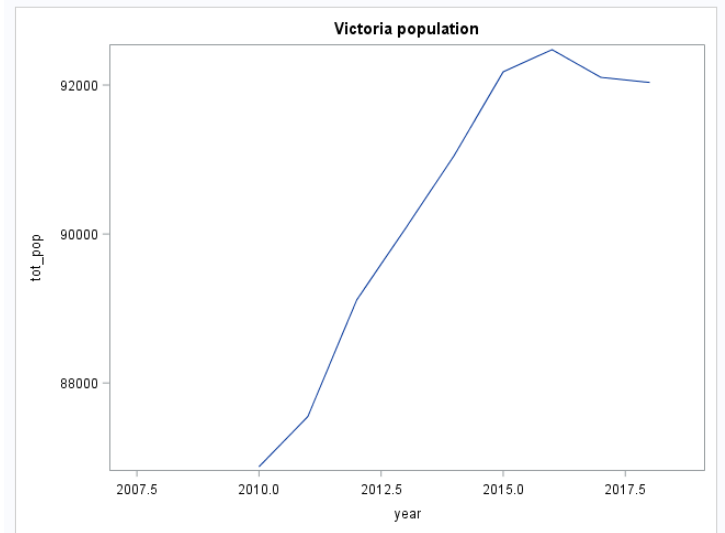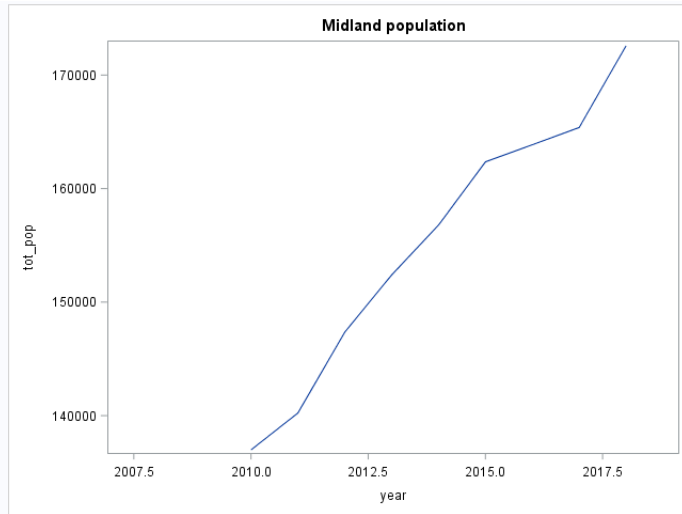
Personal Income Trends

Midland
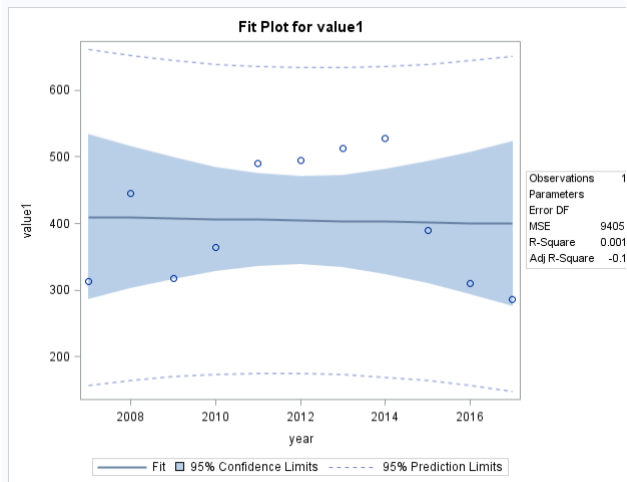
Hidalgo
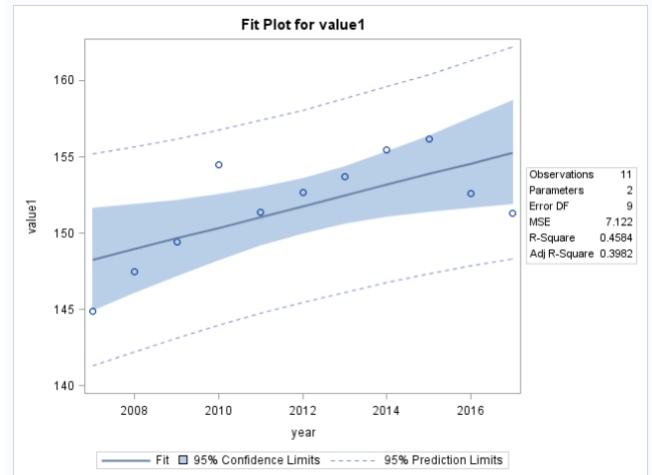
Victoria
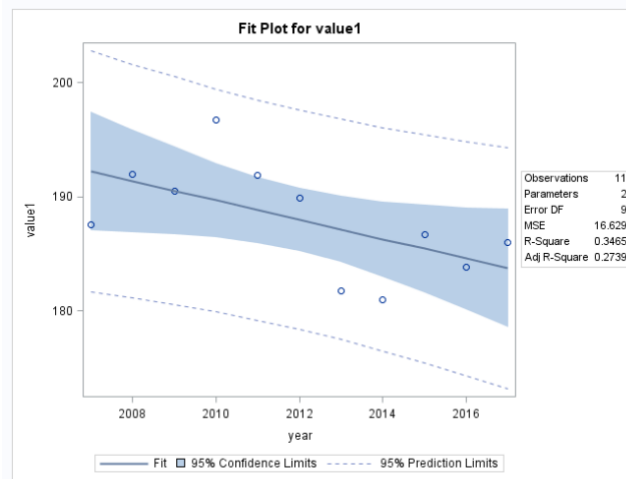
Bowie
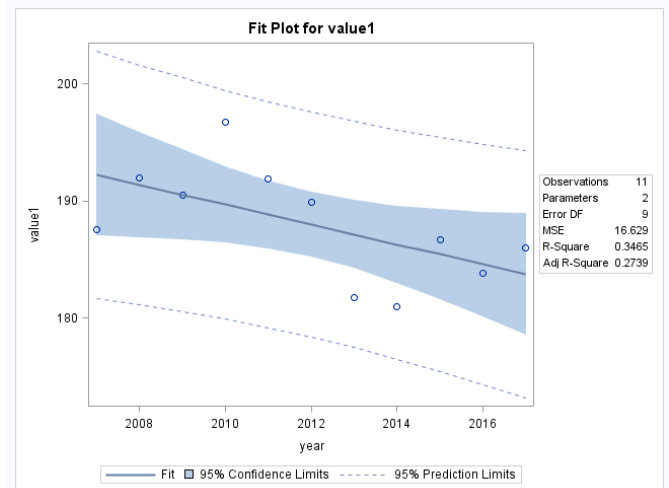
III.   Correlation Analysis

Correlation Analysis is usually done to understand the relationship between any two variables. This analysis helps in determining the strength and direction of this relationship. The relationship between any two variables could be positive or negative. The positive relationship is denoted by a + sign and a negative relation is denoted by a - sign. The correlation value falls in between 0 and 1. The absolute value which is closer to 1 represents that two variables are closely related and if the value is closer to 0, it means the values do not have high correlation between them.

The figure below is a correlation matrix generated in order to understand the correlation between the variables in our dataset. The results of this matrix will be discussed subsequently.
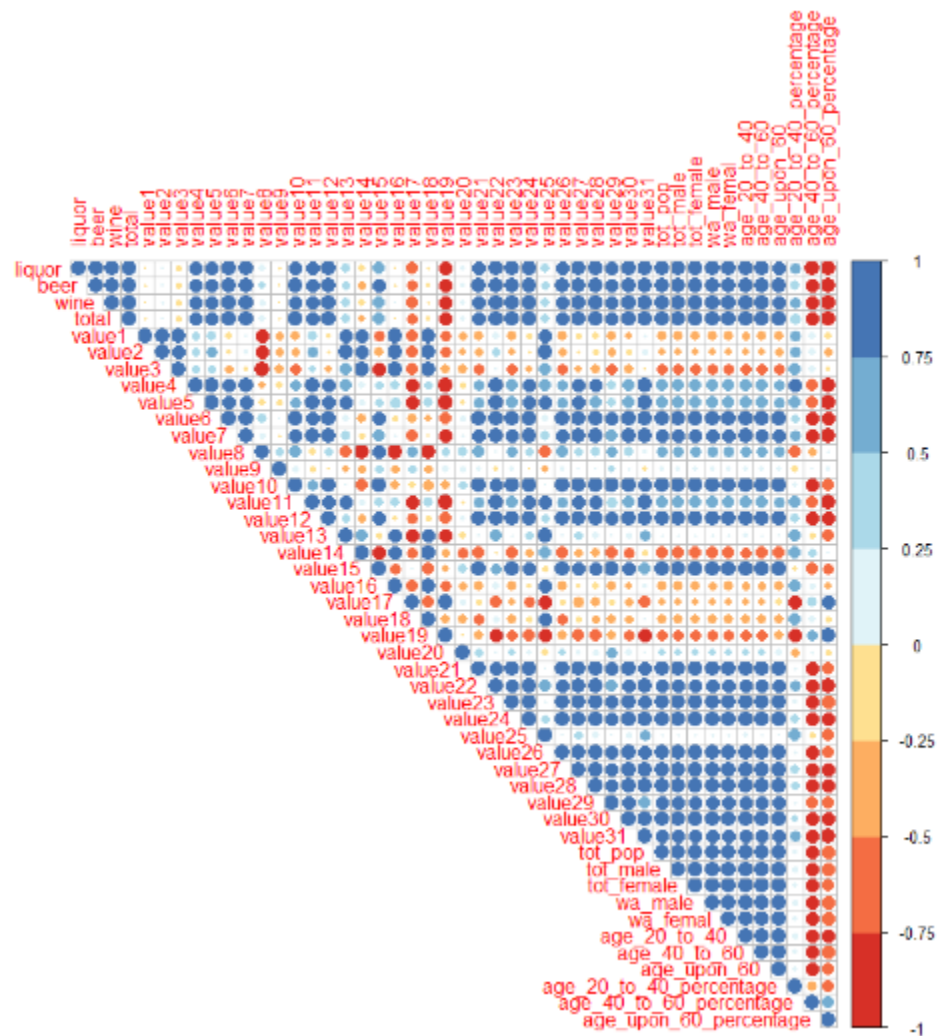
Value and type table

value1	Personal income (thousands of dollars)
value2	Net earnings by place of residence
value3	Personal current transfer receipts
value4	Income maintenance benefits 1/
value5	Unemployment insurance compensation
value6	Retirement and other
value7	Dividends, interest, and rent 2/
value8	Population (persons) 3/
value9	Per capita personal income 4/
value10	Per capita net earnings 4/
value11	Per capita personal current transfer receipts 4/
value12	Per capita income maintenance benefits 4/
value13	Per capita unemployment insurance compensation 4/
value14	Per capita retirement and other 4/
value15	Per capita dividends, interest, and rent 4/
value16	Earnings by place of work
value17	Wages and salaries
value18	Supplements to wages and salaries
value19	Employer contributions for employee pension and insurance funds 5
value20	Employer contributions for government social insurance
value21	Proprietors' income
value22	Farm proprietors' income
value23	Nonfarm proprietors' income
value24	Total employment (number of jobs)
value25	Wage and salary employment
value26	Proprietors employment
value27	Farm proprietors employment 6/
value28	Nonfarm proprietors employment
value29	Average earnings per job (dollars)
value30	Average wages and salaries
value31	Average nonfarm proprietors' income

The following table displays the rankings of the counties in relation to the highly correlated variables. In this table, each county gets assigned a rank for each variable, descending from 1-3. At the end, we sum up the values for all variables and determine the winner. The winner is the county that has the lowest total number, and will be considered the best option for opening a sports bar.

	Hidalgo	Victoria	Midland	Bowie
Value4	1	3	2	4
Value5	1	3	2	4
Value6	1	4	2	3
Value7	1	4	2	3
Value10	1	4	2	3
Value11	1	3	2	4
Value12	1	3	2	4
Value19	2	3	1	4
Value21	1	4	2	3
Value22	1	3	2	4
Value23	1	4	2	3
Value24	1	3	2	4
Value26	1	4	2	3
Value27	1	4	2	3
Value28	1	4	2	3
Value29	1	4	2	3
Value30	1	4	2	3
Value31	1	3	2	4
Tot_Pop	1	4	2	3
Tot_Male	1	4	2	3
Tot_Female	1	4	2	3
Wa_Male	1	3	2	4
Wa_Female	1	3	2	4
Age_20_to_40	1	4	2	3
Age_40_to_60	1	4	2	3
Age_Upon_60	1	4	2	3
Age_40_to_60_ Percentage	1	3	2	4
Age_upon_60_Percentage	1	3	2	4
Total	29	100	55	96

This table summarizes the rankings of the highly correlated variables and their rankings per county. We observe that Hidalgo with ranking total of 29 is currently the location with the best environment to obtain a higher amount of total receipts, due to the number of variables that are positively correlated to the total receipts variable. This analysis would lead us to believe that **Hidalgo** is the optimal location to implant a sports bar.

IV.   Survival Analysis

| Location_County	survive	total_pay_number	survive_percentage	County_name
19	43	81	0.53086	Bowie
108	237	790	0.3	Hidalgo
165	95	181	0.52486	Midland
235	51	132	0.38636	Victoria

Location_county - code for each of the counties
Survive - Number of stores that survived over a period of time i.e. until 2019
Total_pay_number - Total number of stores in each of the counties
Survive_percentage - The percentage of number of stores that didn't shut down over the period of time

In this analysis we use the opening and closing dates of the establishments in each county, and determine the closing date as the date the establishment "dies". The analysis provides the chances an establishment has to survive in the current business environment. As we observe from the table, Bowie has the highest chances of an establishment surviving at 53%, and Hidalgo has the lowest at 30%. However, in Hidalgo 237 establishments survived until 2019, and in Bowie only 81 survived. This can be attributed to the overall size of the county. The small number of establishments in Bowie could mean that the market is more concentrated, and the barriers to entry are higher for a new competitor.
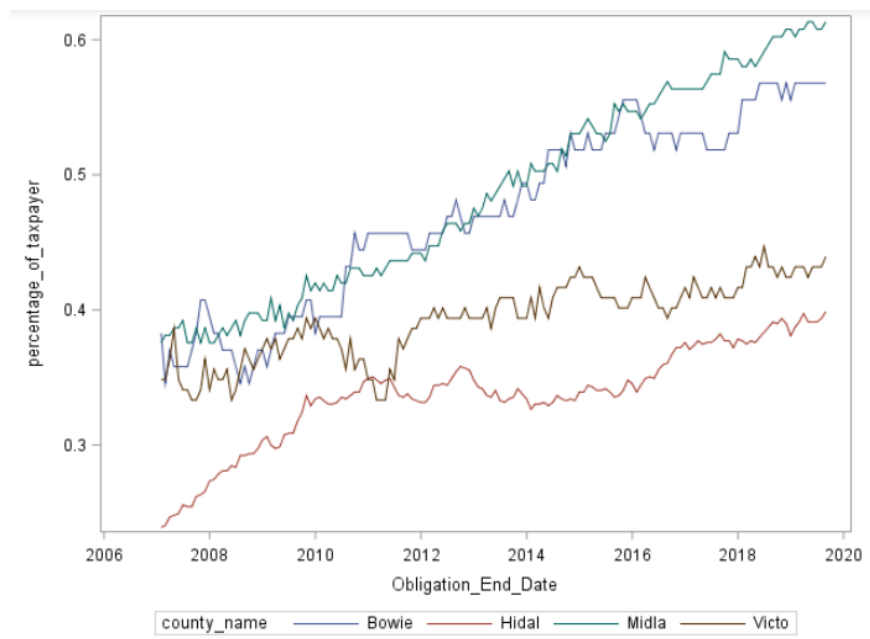
V.   Herfindahl-Hirschman Index

In order to assess the market concentration in all four counties, we calculated the Herfindahl-Hirschman Index (HHI), which indexes the number of relevant firms in the market for each specific location, and their market shares. According to The Federal Reserve Archival System for Economic Research, this index ranges from 0 to 10,000. 0 would represent a perfectly competitive market, and 10,000 represents a market in which one competitor has 100% of the market share, i.e. a monopoly. This index is calculated by using the squared market share value of each firm, and adding the squares. The formula used is $HHI = S_1^2 + S_2^2 + S_3^2 + ... + S_n^2$
For the purpose of our analysis, we changed the scale of the index, having it range from 0-1, with the same criteria as above where a lower index represents closer proximity to perfect competition, and an index of 1 represents a monopoly.

In the graphs depicted below we can observe each county's historical HHI. As represented in the graphs, all four countries have a relatively low HHI, which means that there is a relatively competitive environment. Ideally, for the establishment of a new "Sports Bar" we would choose the county in which the HHI is the lowest. A low HHI not only means less market concentration, but it also represents less barriers of entry and a better environment for a new competitor to enter the market. In 2019 **Hidalgo** was the county with the lowest HHI among our sample.
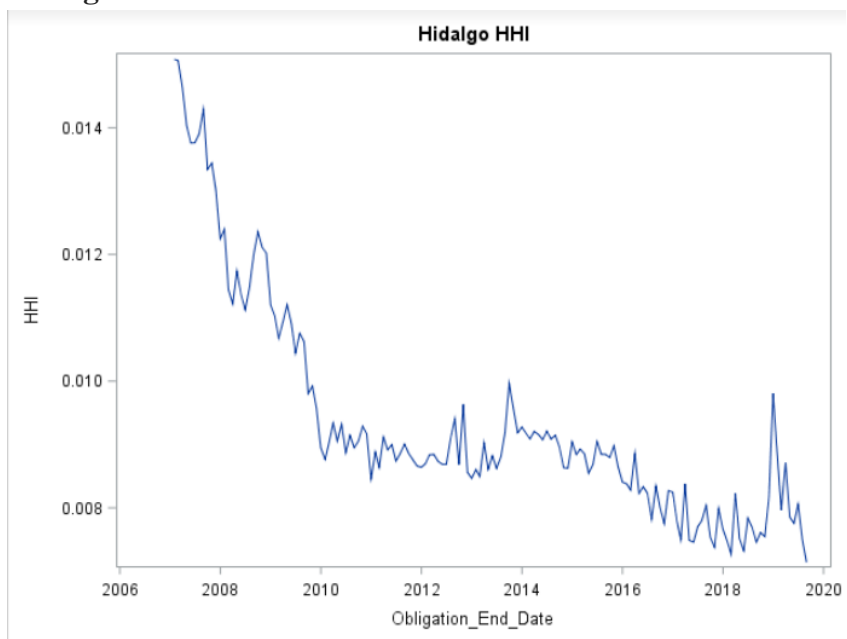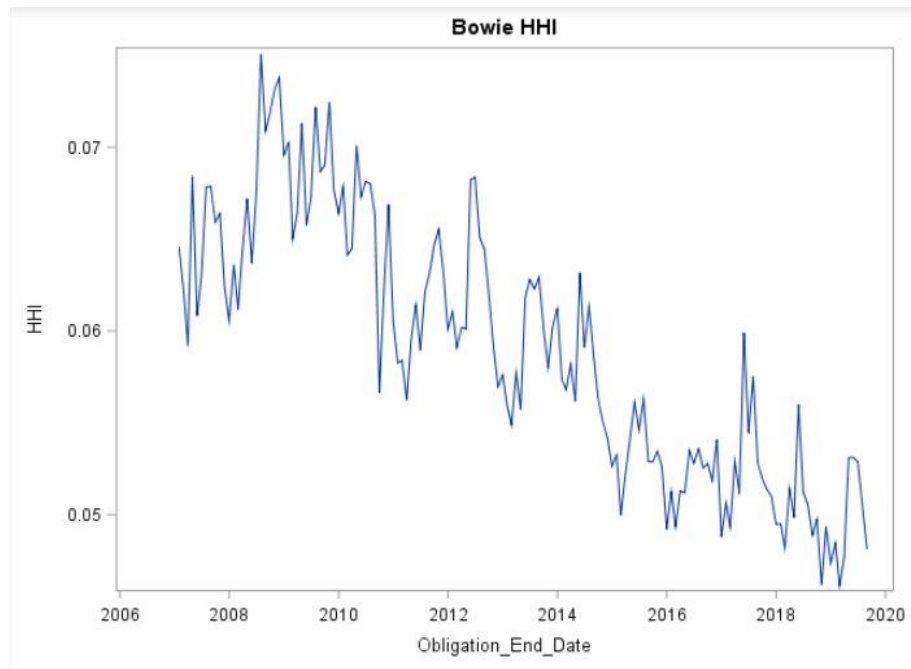
Midland
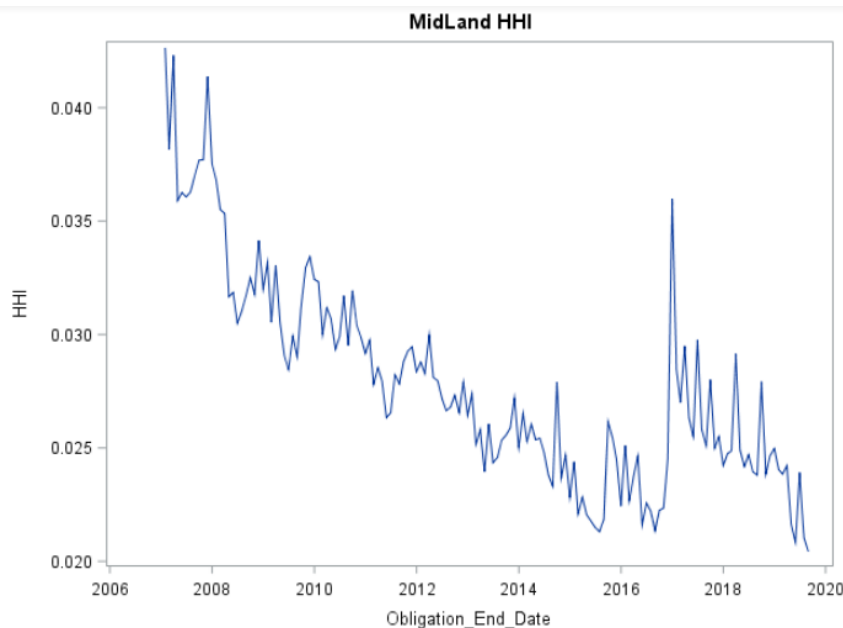
Victoria

Hidalgo

Bowie

Bowie HHI
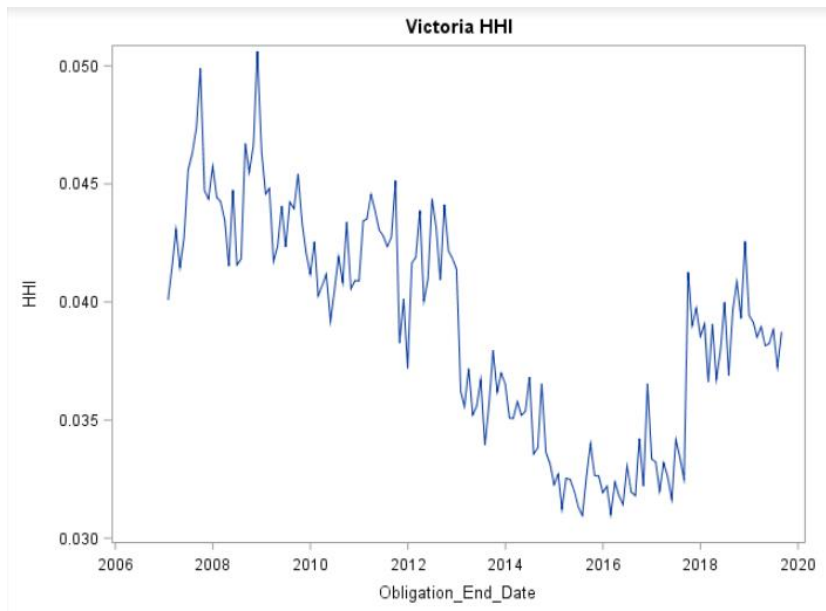
Part - II

Clustering Method

We use the clustering method for better analysis. It sorts similar counties into one group and we can predict the target counties by analyzing similar counties. We first got rid of the four target counties and we take 3 types of alcohol's receipts data in 2017 from tabc_data and the 31 variables from CAINC30 to create a new group with 34 vectors. We normalized these vectors and used K-mean to do the clustering.

After the clustering, all the counties are split into 6 groups.
As it is shown on the summary above: Victoria and Bowie both belong to type 4; Hidalgo belongs to type 6 and Midland is a special case. We then ran a regression model looks like this:

$$RPC = PI + PI^2 + GP + RP + AG_1 + AG_2 + AG_3 + \varepsilon$$

RPC means Receipt per capita. PI means personal income. GP means gender percentage, it is calculated by total_male_population/total_female_population. RP means race percentage, it is calculated by White_population/total_population. AG means different age groups which range from 20-40, 40-60 and 60 above.

The results are below:

Victoria
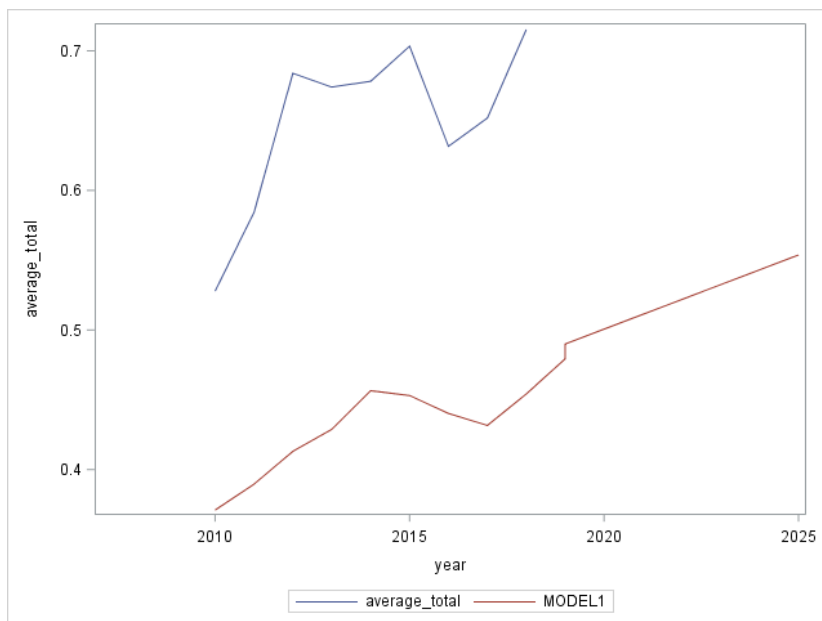
Bowie
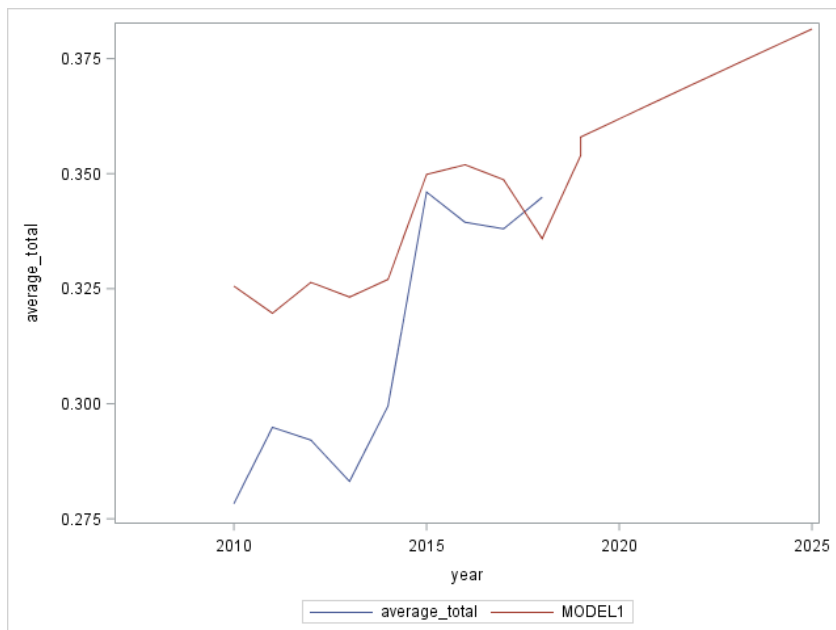
Hidalgo
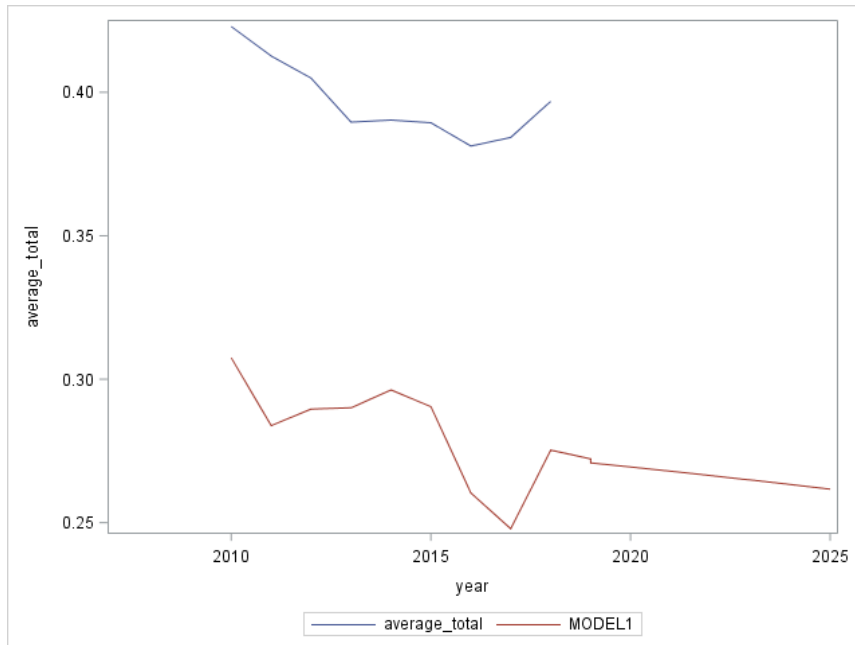
Midland
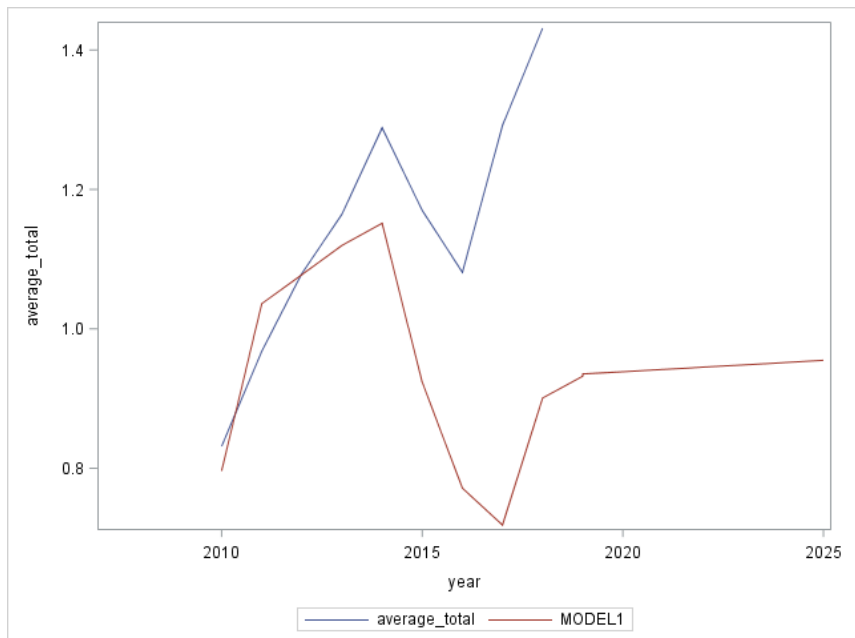
As shown on the graphs, the blue line is the real total receipt of alcohol for each county and the red line is the line predicted by the regression model. It fits well with the reality. As the population and income is unknown for the future, the lines remain straight.

Midland is a petroleum town where its biggest industry is working with petroleum and it shows difference with other counties. Its alcohol consumption is not influenced by income but only on

population. Also, there the Midland doesn't fit well in its group either. We ran the regression for Midland in the whole groups where it contains all the counties. The result seems very fit.

According to the analysis above and graphs, Bowie, Victoria and Midland's future alcohol consumption per capita is increasing and Hidalgo's future alcohol consumption per capita is decreasing. Among the increasing counties, Midland doesn't show a very significant increase. Therefore, the best candidates are Bowie and Victoria by conducting the clustering analysis.

Conclusion

We have presented two analysis. Part - I deals with analysis based solely on the variables and Part - II deals with predictive analytics based on the method of clustering. Since, this method would be more appropriate for forecasting, we can use this method to predict where the sports bar should be opened in about 5 years. Hence, we are going to base our results on Part - I. The ranking table display that Hidalgo is ranked 1 when compared with the highly correlated variables. Furthermore, HHI for Hidalgo is the lowest in 2019. A low HHI not only means less market concentration, but it also represents less barriers of entry and a better environment for a new competitor to enter the market. However, if a business wants to perform well in a local area i.e. be a "local boss" Bowie can be considered to be a good option with highest survival percentage of more than 50%.

Appendix
Bowie and Victoria

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: average_total

Number of Observations Read	700
Number of Observations Used	700

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	19.13742	2.73392	57.08	<.0001
Error	692	33.14155	0.04789		
Corrected Total	699	52.27896			

Root MSE	0.21884	R-Square	0.3661
Dependent Mean	0.27174	Adj R-Sq	0.3597
Coeff Var	80.53404		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.97857	0.34768	-8.57	<.0001
value1	1	0.00607	0.00121	5.04	<.0001
personal_income_2	1	-0.00001223	0.00000277	-4.41	<.0001
age_20_to_40_percentage	1	7.10058	0.46467	15.28	<.0001
age_40_to_60_percentage	1	3.64136	0.55520	6.56	<.0001
age_upon_60_percentage	1	3.04641	0.30616	9.95	<.0001
gender	1	-1.16081	0.07561	-15.35	<.0001
race	1	0.44046	0.14395	3.06	0.0023

Hidalgo

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: average_total

Number of Observations Read	187
Number of Observations Used	187

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	4.17851	0.59693	10.46	<.0001
Error	179	10.21357	0.05706		
Corrected Total	186	14.39207			

Root MSE	0.23887	R-Square	0.2903
Dependent Mean	0.33174	Adj R-Sq	0.2626
Coeff Var	72.00445		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.85595	0.76628	-2.42	0.0164
value1	1	0.01705	0.00446	3.82	0.0002
personal_income_2	1	-0.00003461	0.00001079	-3.21	0.0016
age_20_to_40_percentage	1	5.84187	1.31942	4.43	<.0001
age_40_to_60_percentage	1	0.11947	1.15545	0.10	0.9178
age_upon_60_percentage	1	1.85069	0.93023	1.99	0.0482
gender	1	-2.03914	0.40962	-4.98	<.0001
race	1	0.38130	0.28820	1.32	0.1875

Midland

The REG Procedure
Model: MODEL1
Dependent Variable: average_total

Number of Observations Read	1673
Number of Observations Used	1673

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	72.94356	12.15726	157.90	<.0001
Error	1666	128.26797	0.07699		
Corrected Total	1672	201.21152			

Root MSE	0.27747	R-Square	0.3625
Dependent Mean	0.30636	Adj R-Sq	0.3602
Coeff Var	90.57100		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.27162	0.25002	-13.09	<.0001
value1	1	0.00170	0.00014424	11.78	<.0001
age_20_to_40_percentage	1	9.74913	0.41200	23.66	<.0001
age_40_to_60_percentage	1	6.02436	0.41391	14.55	<.0001
age_upon_60_percentage	1	4.08489	0.24109	16.94	<.0001
gender	1	-1.76364	0.07458	-23.65	<.0001
race	1	0.26311	0.10467	2.51	0.0120

6 groups of clusters

Midland(tabc code = 165) belong to type 1 this type has:
['Hartley', 'Williamson', 'Kendall', 'Collin', 'Denton', 'Randall', 'Dallam', 'Fort Bend', 'Comal', 'Galveston', 'Hansford', 'Parker', 'Montgomery', 'Shackelford', 'Brazoria', 'Hemphill', 'Midland', 'Ochiltree', 'Chambers', 'Rockwall', 'Smith', 'Sutton']
and their tabc code:
['103', '246', '130', '43', '61', '191', '56', '79', '46', '84', '98', '184', '170', '209', '20', '106', '165', '179', '36', '199', '212', '218']

Hidalgo(tabc code = 108) belong to type 6 this type has:
['Presidio', 'Kleberg', 'Zavala', 'Liberty', 'Morris', 'San Patricio', 'Calhoun', 'Maverick', 'Jefferson', 'Zapata', 'Brooks', 'Harrison', 'Starr', 'El Paso', 'Matagorda', 'Val Verde', 'Hidalgo', 'Cameron', 'Webb', 'Nueces', 'Willacy', 'Jim Wells', 'Orange', 'Dimmit', 'Jim Hogg']
and their tabc code:
['189', '137', '254', '146', '172', '205', '29', '159', '123', '253', '24', '102', '214', '71', '158', '233', '108', '31', '240', '178', '245', '125', '181', '64', '124']

Victoria(tabc code = 235), Bowie(tabc code = 19) belong to type 4 this type has:
['Tom Green', 'Hutchinson', 'Brewster', 'Ellis', 'Freestone', 'Frio', 'Fannin', 'Dawson', 'Walker', 'Jackson', 'Madison', 'Jeff Davis', 'Kaufman', 'Callahan', 'Reeves', 'Swisher', 'Castro', 'Live Oak', 'Kinney', 'Cherokee', 'Taylor', 'Blanco', 'Hardin', 'Wise', 'Culberson', 'Deaf Smith', 'Oldham', 'Bowie', 'Grimes', 'Karnes', 'Medina', 'Pecos', 'Scurry', 'Grayson', 'Johnson', 'Gaines', 'Hockley', 'Childress', 'Guadalupe', 'Hays', 'Crockett', 'Andrews', 'Jack', 'Potter', 'Atascosa', 'Archer', 'Bandera', 'Nacogdoches', 'Burnet', 'Hood', 'Gonzales', 'Lubbock', 'Bell', 'Howard', 'Erath', 'Bailey', 'Cooke', 'La Salle', 'Moore', 'Angelina', 'Ector', 'Winkler', 'Austin', 'Clay', 'Hudspeth', 'Hunt', 'Titus', 'Brazos', 'Bee', 'Hale', 'Somervell', 'Wilson', 'Lee', 'Victoria', 'Hopkins', 'McLennan', 'Burleson', 'Coryell', 'Anderson', 'Cochran', 'Franklin', 'Waller', 'Camp', 'Rusk', 'Bastrop', 'Gray', 'Carson', 'Gregg', 'Ward', 'Caldwell', 'Jones']
and their tabc code:
['226', '117', '22', '70', '81', '82', '74', '58', '236', '120', '154', '122', '129', '30', '195', '219', '35', '149', '136', '37', '221', '16', '100', '249', '55', '59', '180', '19', '93', '128', '163', '186', '208', '91', '126', '83', '110', '38', '94', '105', '53', '2', '119', '188', '7', '5', '10', '174', '27', '111', '89', '152', '14', '114', '72', '9', '49', '142', '171', '3', '68', '248', '8', '39', '115', '116', '225', '21', '13', '95', '213', '247', '144', '235', '112', '161', '26', '50', '1', '40', '80', '237', '32', '201', '11', '90', '33', '92', '238', '28', '127']

other type information (useless):
type2:
['Harris'] ['101']
type3:
['Haskell', 'Mills', 'Sabine', 'Wood', 'Aransas', 'DeWitt', 'Marion', 'Lampasas', 'Lavaca', 'Falls', 'Knox', 'Hardeman', 'San Jacinto', 'Rains', 'Menard', 'Jasper', 'Limestone', 'Goliad', 'Comanche', 'Robertson', 'Baylor', 'Brown', 'Shelby', 'Collingsworth', 'Bosque', 'Real', 'Henderson', 'Mason', 'Cass', 'Red River', 'Eastland', 'McCulloch', 'Donley', 'Coleman', 'Runnels', 'Refugio', 'Tyler', 'Fayette', 'Delta', 'Uvalde', 'Wilbarger', 'Milam', 'Terry', 'Lamar', 'Houston', 'Gillespie', 'Wheeler', 'Llano', 'Young', 'Stephens', 'Kerr', 'Kent', 'Hall', 'Navarro', 'Throckmorton', 'Wharton', 'Upshur', 'Floyd', 'Palo Pinto', 'Leon', 'Colorado', 'Hill', 'Wichita', 'Hamilton', 'Foard', 'Washington', 'Panola', 'Trinity', 'Van Zandt', 'Polk', 'Montague', 'San Saba', 'Nolan']
['104', '167', '202', '250', '4', '62', '155', '141', '143', '73', '138', '99', '204', '190', '164', '121', '147', '88', '47', '198', '12', '25', '210', '44', '18', '193', '107', '157', '34', '194', '67', '160', '65', '42', '200', '196', '229', '75', '60', '232', '244', '166', '223', '139', '113', '86', '242', '150', '252', '215', '133', '132', '96', '175', '224', '241', '230', '77', '182', '145', '45', '109', '243', '97', '78', '239', '183', '228', '234', '187', '169', '206', '177']
type5:
['Bexar', 'Travis', 'Dallas', 'Tarrant'] ['15', '227', '57', '220']