Term Project: Comal County Liquor Sales Forecasting

Hamza Amin, Andy Bartolo, Monish Civunigunta, William Duong

Department of Economics, Texas A&M University

ECMT 475-500: Economic Forecasting

Dr. Craig Schulman

April 30, 2024

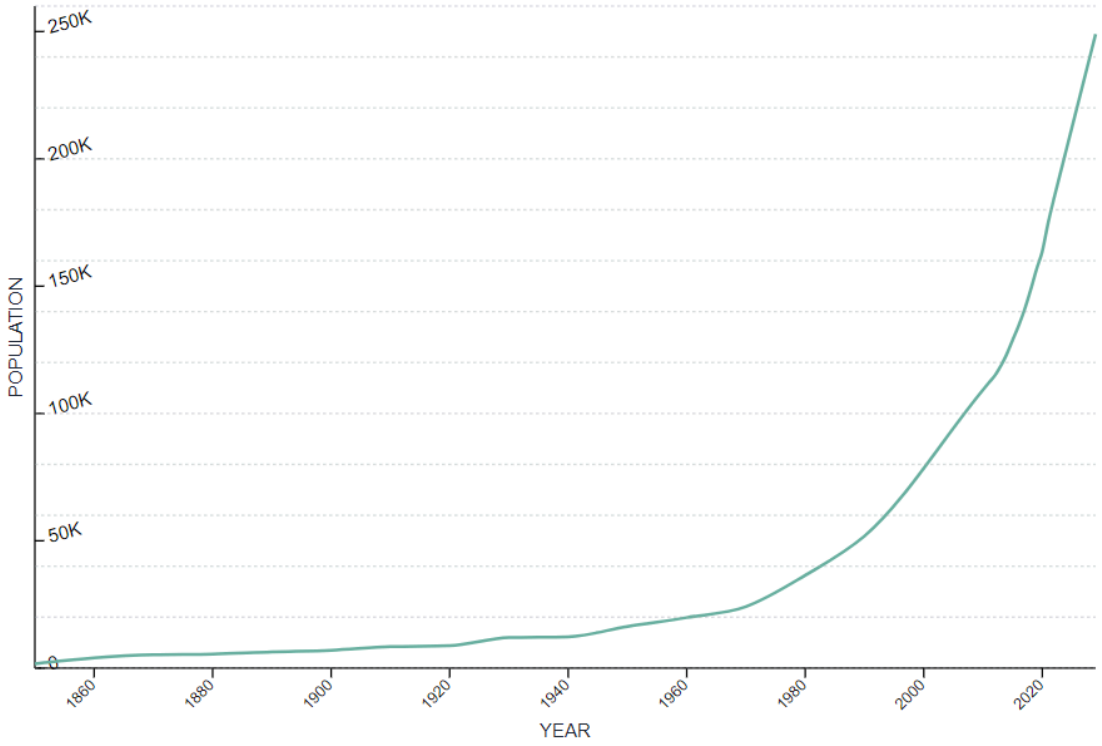
Comal County Socio-Economic Profile

Section A: Recent Evolution of Age and Ethnic Demographics

Over the past decade and a half, Comal County has experienced notable demographic shifts. Since 2010, the population has undergone a significant increase of 85.81%, climbing from approximately 109,311 to an estimated 203,107 inhabitants in 2024 according to World Population Review.

Figure 1

Comal County Population Over Time

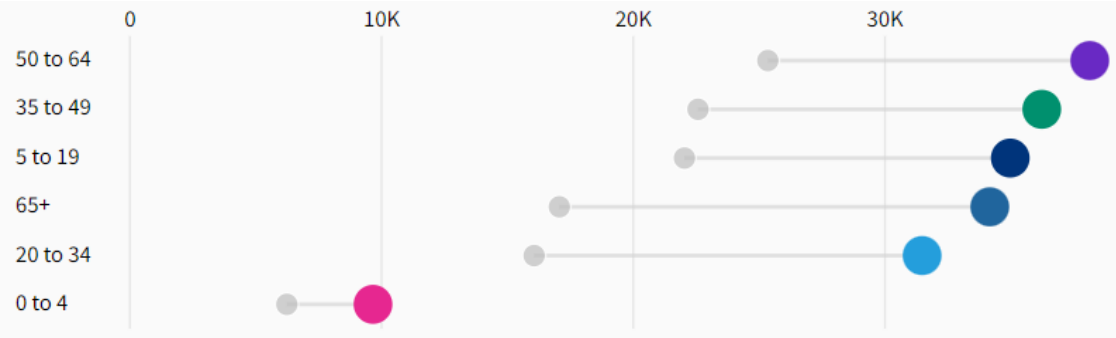


Note. Line graph depicting past, current, and projected population for Comal County, Texas from 1850 to 2029. From World Population Review (https://worldpopulationreview.com/us-counties/tx/comal-county-population).

With this increase in total population, the age demographics of the county have experienced a moderate shift as well, with a particular emphasis on the elderly population. In 2010, adults 50 to 64 years of age made up the largest age group in Comal County, followed by those 35 to 49 years of age, youths 5 to 19 years old, elderly 65 years and older, young adults 20 to 34 years of age, and young children 0 to 4 years old. The most recent data provided in 2022 by the U.S. Census Bureau reveals this order has since remained unchanged. However, the 65+ year age group has demonstrated a disproportionately significant rate of change, a staggering growth of 100.23%, indicating the possibility of Comal County holding a particular appeal to elderly and retired individuals. Still, as of 2024, data from the World Population Review reveals the median age is approximately 41.7 years, indicating a relatively balanced mix of youth and elderly populations.

Figure 2

Comal County Population by Age

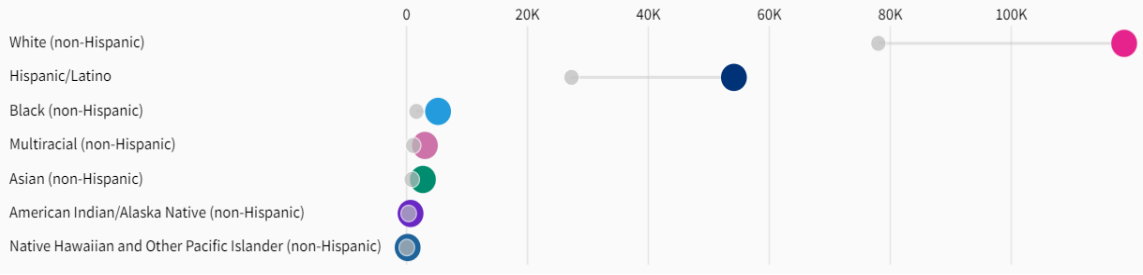


Note. Graphic depicting changes in the Comal County population by age group from 2010 to 2022. (1) 50 to 64 years: 25,345; 38,136. (2) 35 to 49 years: 22,571; 36,229. (3) 5 to 19 years: 22,034; 34,977. (4) 65+ years: 17,062; 34,164. (5) 20 to 34 years: 16,062; 31,476. (6) 0 to 4 years: 6,237; 9,660. From USA Facts (https://usafacts.org/data/topics/people-society/population-and-demographics/our-changing-population/state/texas/county/comal-county/).

In terms of racial and ethnic composition, according to USA Facts, over the last decade, Comal County has stayed relatively the same. In 2010, Comal County comprised a majority white (non-Hispanic) population, with Hispanic/Latino, Black, multiracial, Asian, American Indian/Alaska Native, and Native Hawaiian/Other Pacific Islander making up the remainder in that order. The most recent data, provided in 2022 by the U.S. Census Bureau, shows this order has since remained unchanged, similar to the order of age groupings in the county. Over this 12-year period, the Native Hawaiian/Other Pacific Islander population experienced the greatest percentage change, a 277.55% increase, followed by the Black and Asian populations, at 218.81% and 218.10% respectively. Next in line is the multiracial population, with a 177.48% increase, followed by Hispanic/Latinos, American Indian/Alaska Natives, and White Americans with 98.41%, 86.87%, and 52.09% increases respectively. However, a closer inspection reveals that the Native Hawaiian/Other Pacific Islander population actually experienced the least total population change–a measly increase of 136 individuals. The significant percentage change is the result of the initial Native Hawaiian population measure being only 49 people. The opposite is true for the White population, which experienced the greatest total change, an increase of 40,646, but the smallest percentage change, with the latter attributable to the large initial population measure of 78,031 individuals.

Figure 3

Comal County Population by Race and Ethnicity



Note. Graphic depicting changes in the racial and ethnic populations of Comal County from 2010 to 2022. (1) White (non-Hispanic): 78,031; 118,677. (2) Hispanic/Latino: 27,288; 54,141. (3) Black: 1,637; 5,219. (4) Multiracial: 1,097; 3,044. (5) Asian: 851; 2,707. (6) American Indian/Alaska Native: 358; 669. (6) Native Hawaiian/Other Pacific Islander: 49; 185. From USA Facts (https://usafacts.org/data/topics/people-society/population-and-demographics/our-changing-population/state/texas/county/comal-county/).

Section B: Population Concentrations

Comal County consists of the cities of Bulverde, Garden Ridge, and Spring Branch, all of which are located entirely within the county, as well as Fair Oaks Ranch, New Braunfels, San Antonio, Schertz, and Selma, which span multiple counties including Comal County. The unincorporated communities of Bracken and Fischer, the ghost towns of Canyon City and Wesson, and Canyon Lake, a census-designated place, round out the list of municipalities within Comal County. Due to sparse data existing for certain smaller municipalities, only the eight major places that are fully or partially contained within Comal County will be included in this discussion: San Antonio, New Braunfels, Schertz, Canyon Lake, Selma, Fair Oaks Ranch, Bulverde, and Garden Ridge.

Regarding population concentration, Comal County inhabitants are concentrated in a few key areas, with New Braunfels being the most populated location, followed by the smaller population centers of Schertz and Canyon Lake. Despite San Antonio appearing to dwarf the other places within the county in the included figure below, it is important to note that San Antonio is a large city that extends into multiple counties, and not all of its population resides within Comal County. Therefore, while San Antonio may have the largest population according

to the figure, New Braunfels is the largest urban center fully within Comal County, making it a focal point of the population within the county itself.

For context, according to the Texas State Historical Association, New Braunfels, located primarily in Comal County, Texas, presents a unique blend of urban growth and rural traditions. Established by German settlers in 1845, it has maintained a strong cultural heritage evident in its annual festivals and architectural landmarks. As the county seat, New Braunfels is a key urban center with a 2020 population of over 90,000, indicating significant growth and urban development, particularly as it was one of the fastest-growing cities in the United States from 2010 to 2020. Despite this expansion, the city retains a sense of its rural beginnings, especially in areas surrounding the Guadalupe and Comal rivers, which offer a plethora of outdoor recreational activities. The balance between its bustling city life and pastoral outskirts characterizes New Braunfels as a city that cherishes its origins while embracing modern growth.

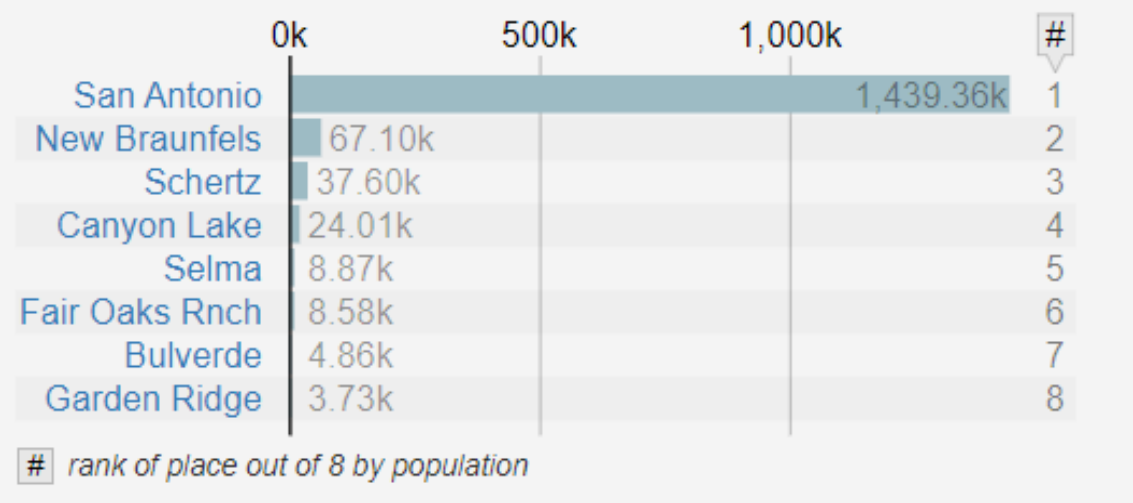
Schertz, another significant urban area partially located Comal County, contributes to the diversity and economy of the county as well. Part of the greater San Antonio metropolitan area with its deep roots stretching back to around 1843, Schertz spans across Guadalupe, Bexar, and Comal counties. According to Texas Almanac, its story begins in the mid-19th century when the first settlers engaged in wheat, oats, and corn cultivation. Schertz experienced substantial development with the arrival of the railroad in 1876 and further grew with the establishment of Randolph Air Force Base in the late 1920s. Despite its evolution into an incorporated city in 1958 and subsequent population growth—reaching over 42,000 by 2021—Schertz maintains a suburban character with close-knit community values. It combines the convenience of an urban setting with the tranquility of less densely populated areas, making it a balanced and attractive place to live within the dynamics of the San Antonio-New Braunfels corridor.

Lastly, Canyon Lake, situated solely in Comal County, offers a rural retreat with an emphasis on outdoor recreation and natural beauty. As a census-designated place with a population of over 31,000 people, Canyon Lake represents a community that enjoys the tranquility and scenic splendor of its environment while catering to the recreational needs of visitors and residents alike. The lake itself is a reservoir on the Guadalupe River, created by Canyon Dam, and managed by the U.S. Army Corps of Engineers. With 80 miles of shoreline, the area around Canyon Lake is less urbanized. The local economy is bolstered by tourism focused on the abundant recreational activities provided by the lake, such as fishing, boating, and hiking, with water conservation and environmental preservation playing significant roles in community life. Canyon Lake holds onto its rural essence but is far from isolated, given its reasonable proximity to cities like San Antonio and Austin, making it a desirable locale for those seeking a balance between rural peace and urban access.

Overall, Comal County residents are distributed primarily between the urban New Braunfels, suburban Schertz, and rural Canyon Lake, with the presence of other small cities and towns contributing to the overall socio-economic fabric of the county.

Figure 4

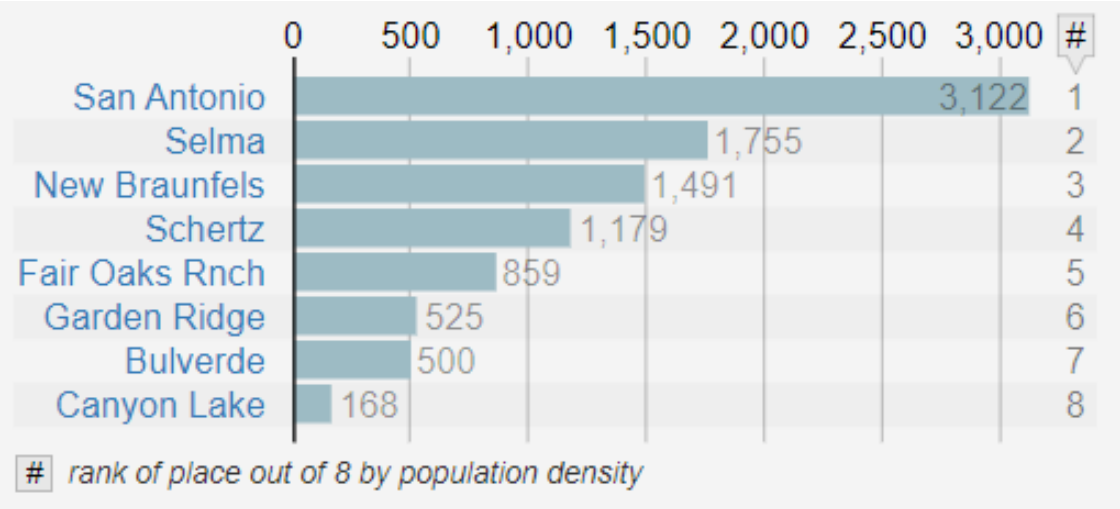
Comal County Total Population by Place



Note. Bar graph depicting the total population of eight major places fully or partially contained within Comal County (three fully and five partially). Fair Oaks Ranch was abbreviated to "Fair Oaks Rnch" on the source website to prevent issues with consistency and formatting. From Statistical Atlas (https://statisticalatlas.com/county/Texas/Comal-County/Population).

Figure 5

Comal County Population Density by Place



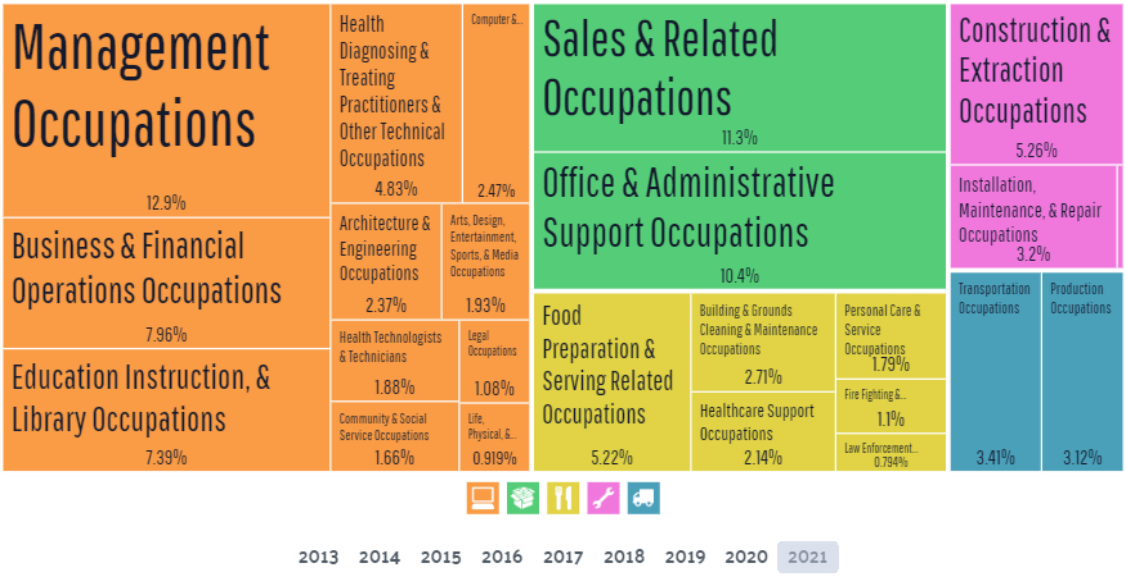
Note. Bar graph depicting people per square mile (excluding waters) for eight major places fully or partially contained within Comal County (three fully and five partially). Fair Oaks Ranch was abbreviated to "Fair Oaks Rnch" on the source website to prevent issues with consistency and formatting. From Statistical Atlas (https://statisticalatlas.com/county/Texas/Comal-County/Population).

Section C: Key Economic Industries

According to the latest 2021 data from Data USA, the economic landscape of Comal County is diversified, with significant contributions from educational services, health care, social assistance, finance, insurance, real estate, rental and leasing, and professional, scientific, management, and administrative services. The largest industries in Comal County are health care and social assistance, retail trade, and construction, which employ 9,159 individuals, 8,757 individuals, and 6,977 individuals respectively. The highest paying industries are mining, quarrying, and oil and gas extraction, company and enterprise management, and agriculture, forestry, fishing and hunting, and mining. These sectors indicate a robust economy with a variety of employment opportunities for residents. The county also benefits economically from its domestic trade, particularly in electronics, gasoline, and fuel oils, underscoring its integration into larger economic networks.

Figure 6

Comal County Occupations by Sector



Note. Graphic depicting various occupations by percentage within Comal County in 2021. Icons left to right: (1) management, business, science and arts occupations; (2) sales and office occupations; (3) service occupations; (4) natural resources, construction, and maintenance occupations; (5) production, transportation, and material moving occupations. From Data USA (https://datausa.io/profile/geo/comal-county-tx#economy).

Section D: Disproportionately Large Employers

Though Comal County contains major employers, no disproportionately large employer operates within the county. Despite a February 2022 report from Veramendi Homes, a planned community in New Braunfels, listing Comal ISD as the largest of the major employers in the Comal County area, the 3,100+ employees working for the school district make it nowhere near "disproportionately large." The latest 2021 employment data for the county provided by Data USA reports roughly 73,300 employees in Comal County. Determining what constitutes a "disproportionately large employer" is not strictly defined by a specific number, but rather by context and comparison to other local employers. If a single company employs a significant percentage of the total workforce–anywhere from a generous 5% to 10% or more–it could

reasonably be considered disproportionately large. However, the workforce employed by Comal ISD falls below the 5% threshold of 3,665 employees. Therefore, Comal ISD cannot be considered, and Comal County lacks a disproportionately large employer based on the criteria defined above.

Total Sales Variable Overview

Section A: Trend Elements

The trend component of a time series reflects the long-term progression of the series. It illustrates the direction in which the data moves over a long period, regardless of short-term fluctuations. Within the graph, the trend smooths out short-term fluctuations to show the long-term direction or movement in the dataset. The trend here shows a gradual upward movement over time, suggesting an increase in the data on alcohol sales in Comal County, Texas. This rising trend could indicate growing sales over time. Understanding the trend is crucial for making predictions about future sales, as it removes the seasonal and random variations to reveal the underlying direction of the data.

Section B: Seasonal Elements

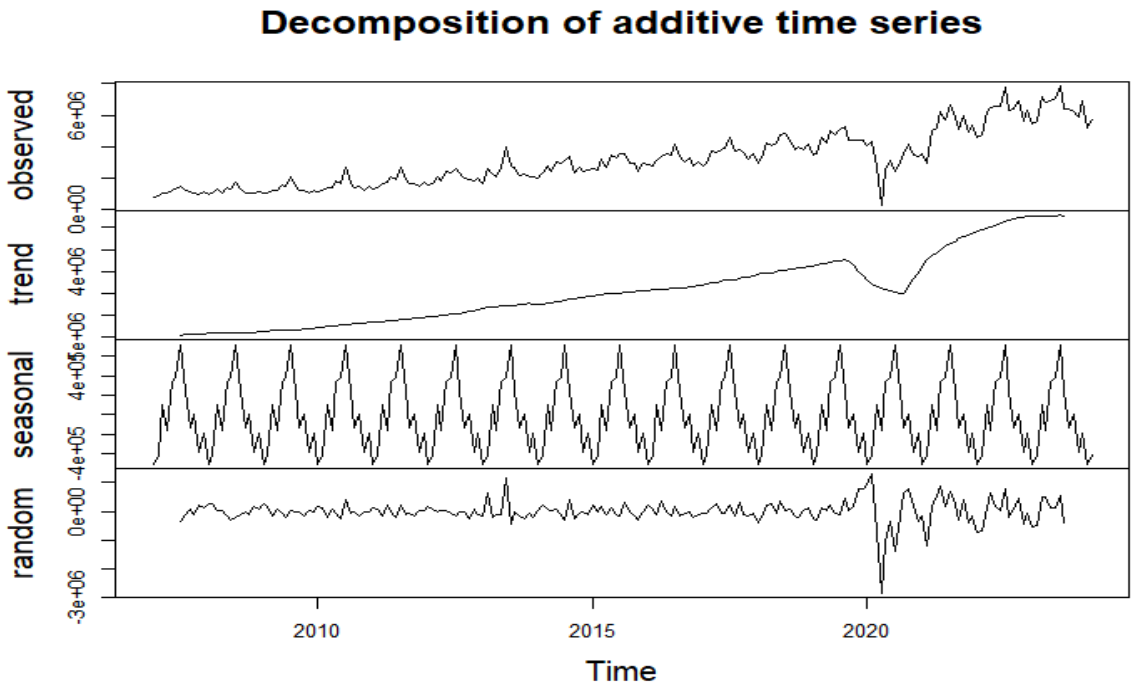
Seasonal elements refer to patterns that repeat intervals over time, typically within a year. This could be due to the seasons, quarters, months, weeks, days, or even parts of a day. Within both charts, the years are fixed over time. This seasonality could be due to various factors like holidays, weather seasons, or events specific to Comal County that influence alcohol sales. The seasonal effect is consistent, peaking and troughing at regular intervals. This information is vital for inventory and staff planning to meet seasonal demand variations.

Section C: Cyclical Elements

Cyclical elements represent the fluctuations that occur at irregular intervals influenced by external factors such as economic conditions rather than time. This would be the best among the four samples as time is not a factor in the increase or decrease of data. The “Random” portion of the plot would best illustrate the cyclical element. When observing the alcohol sales, the data appears stagnant besides the strong downward shift in 2021 where COVID-19 was the primary impact on alcohol sales, not just in Comal County, but all over the nation.

Figure 7

Comal County Alcohol Sales TS Overview



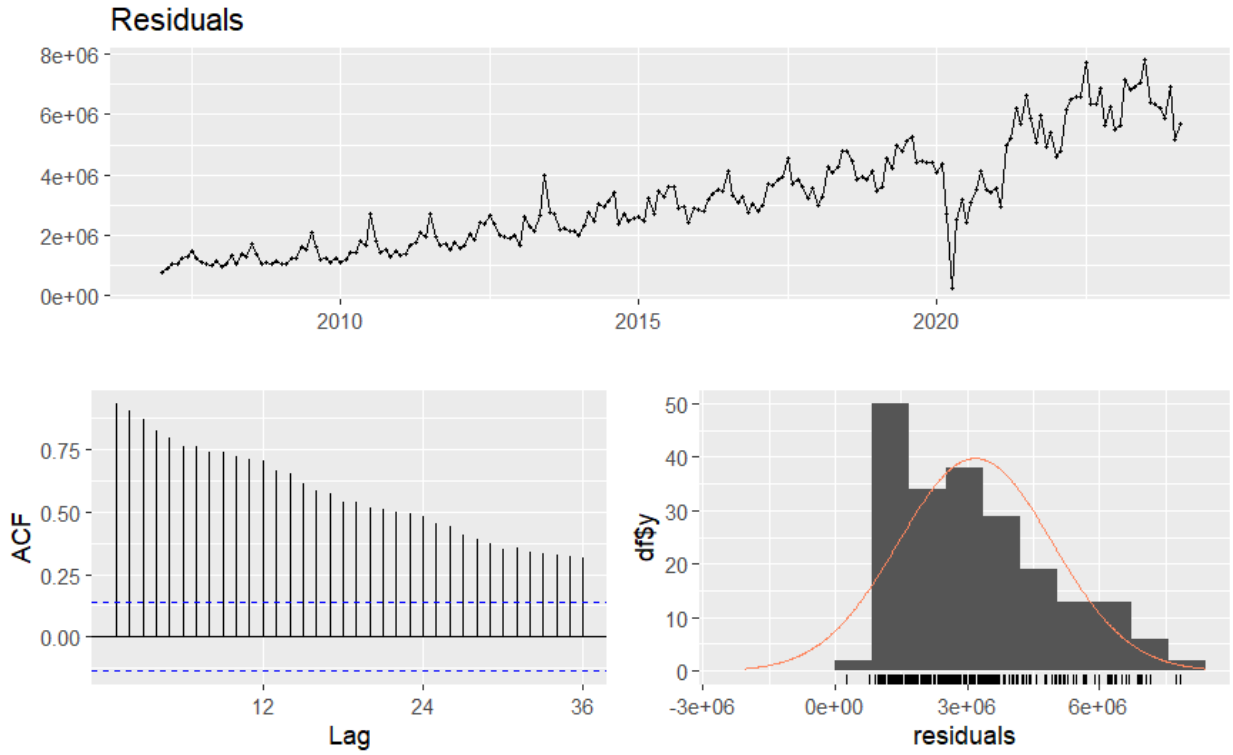
Univariate Forecasting Models

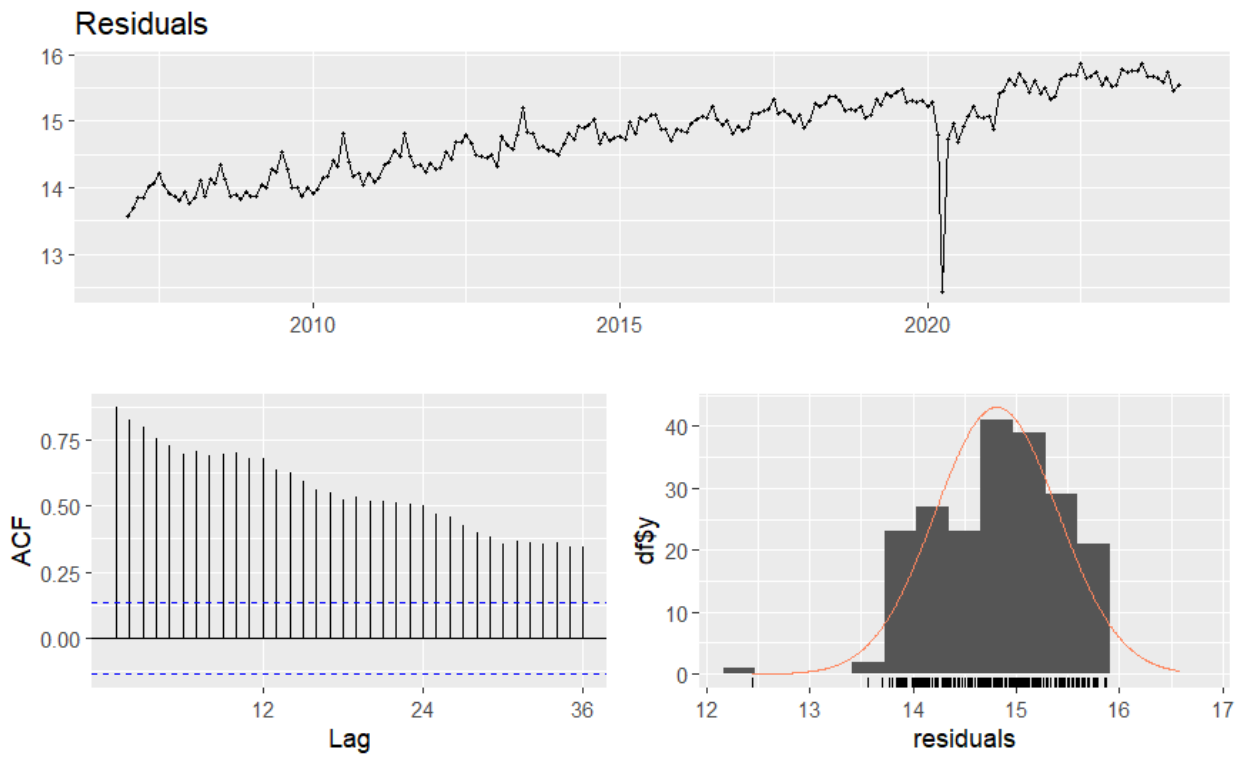
Section A: Total Alcohol Sales Variable Forecasting Models

Part I: Basic Model Description and Residual Diagnostics

To prepare for our 12 months ahead path forecast of total alcohol sales in Comal County, Texas we need to first identify the particular model that will be used. To determine which model fits best we will do a 6 month forecast with a training set. The models that will be tested are an ETS forecasting model, an ARIMA forecasting model, a One-Step Ahead ETS forecasting model, and a One-Step Ahead ARIMA forecasting model.

We also need to determine if a log transformation fits our data. To this, we can use the check residuals function in R and interpret the output. The two sets of plots are total alcohol sales in Comal County. The first of the two is the data without a log transformation and the second is the data with a log transformation.



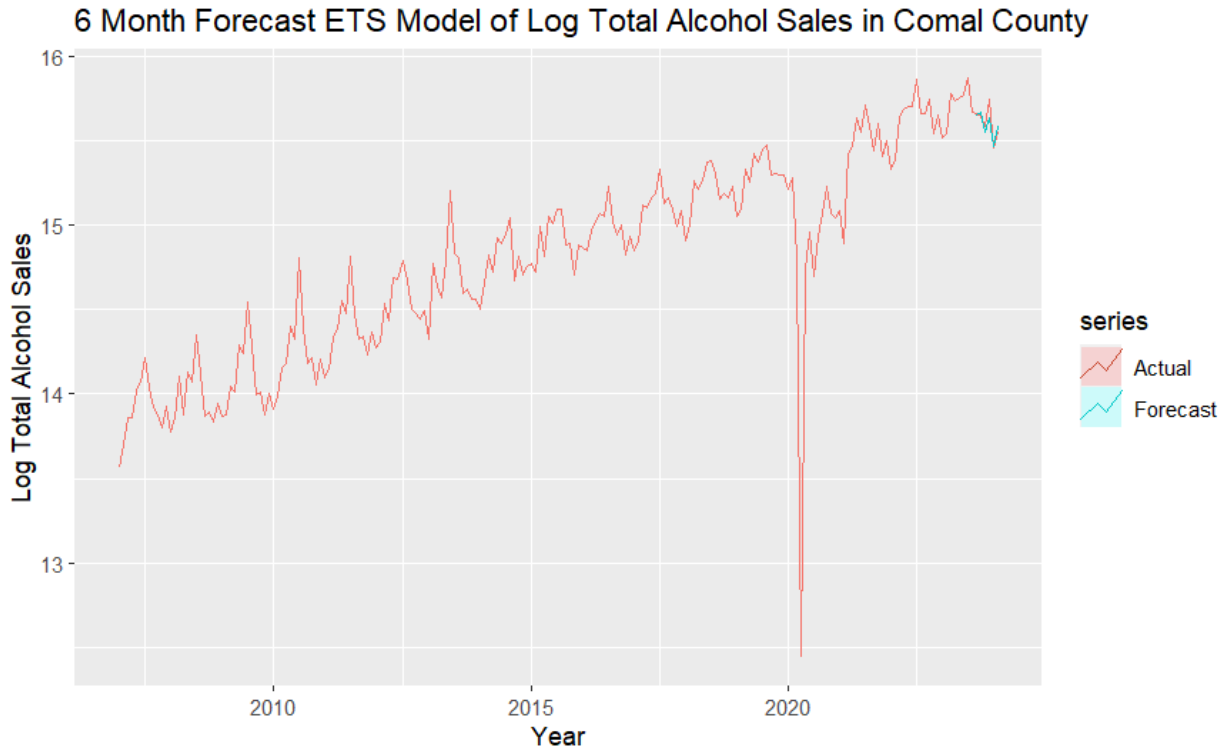


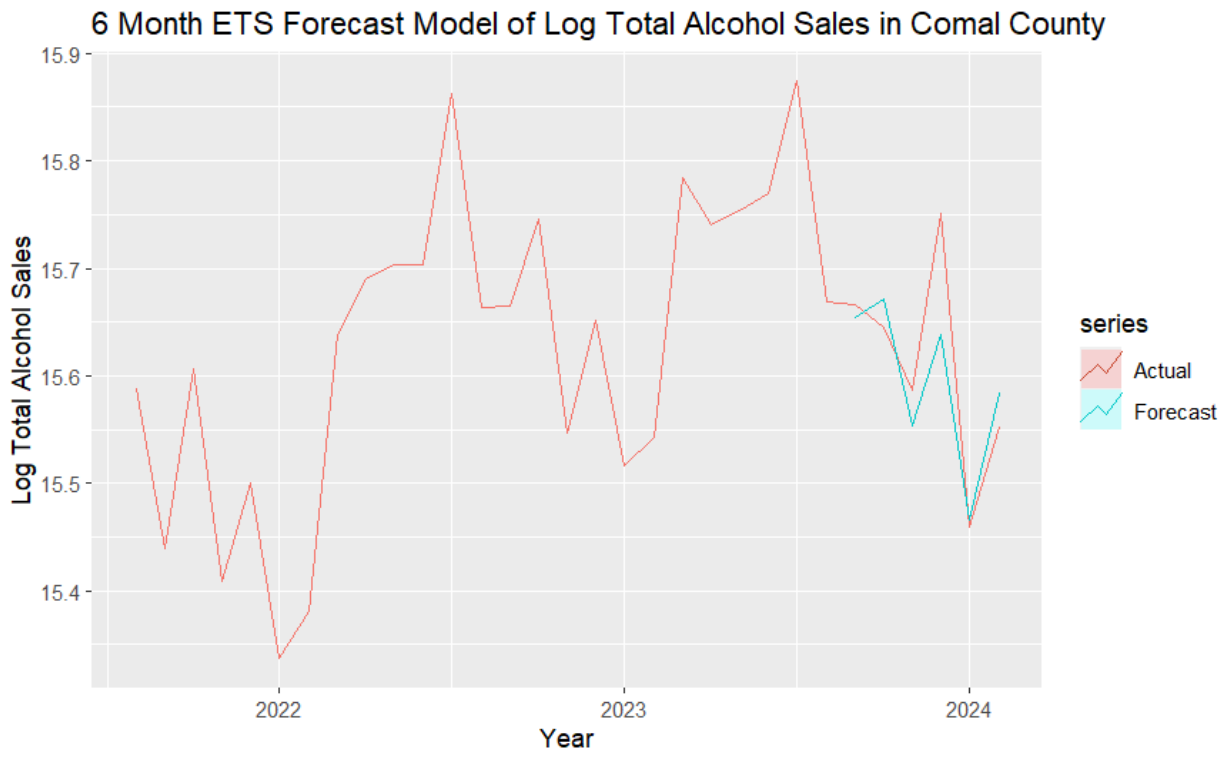
When interpreting our residual diagnostics without a log transformation we can see that the range of the yearly seasonal peaks and troughs are growing over time. This is usually a sign that log transformation suits a piece of data. That being said, we can see that with a log transformation, the range of residual seasonal peaks and troughs remains relatively constant over time. For those reasons, we have determined that a log transformation best suits our data.

To complete the forecasts for the training set we need to identify the necessary ETS and ARIMA models that will be used. To find the specific ETS and ARIMA model we used ets("ZZZ") and auto.arima in R to determine what fit our data best. R determined that an ETS(M,A,A) and an ARIMA (1,1,1)(2,0,0) model fit our data best. The ETS(M,A,A) model will be used for both the standard and the One-Step ahead ETS forecasts. The ARIMA(1,1,1)(2,0,0) model will be used for both the standard and One-Step ahead ARIMA forecasts.

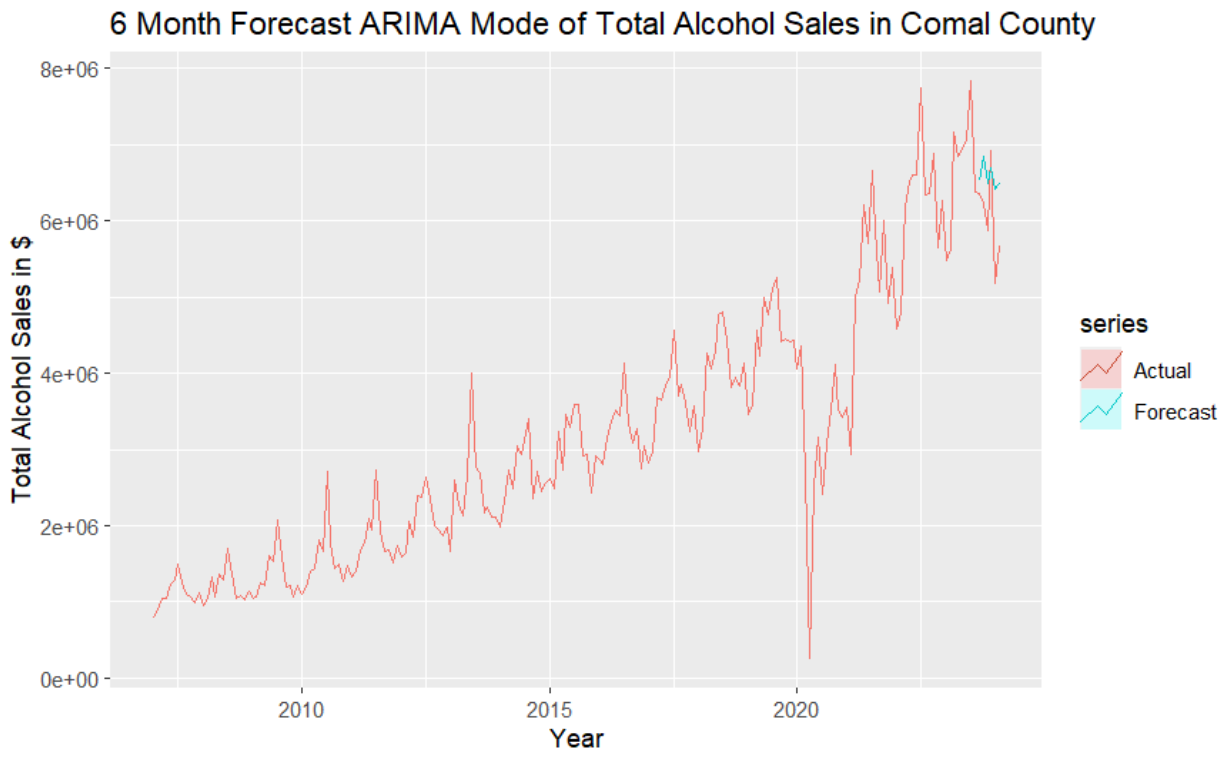
Part II: Six-Month Ahead Forecast and Comparison Discussion

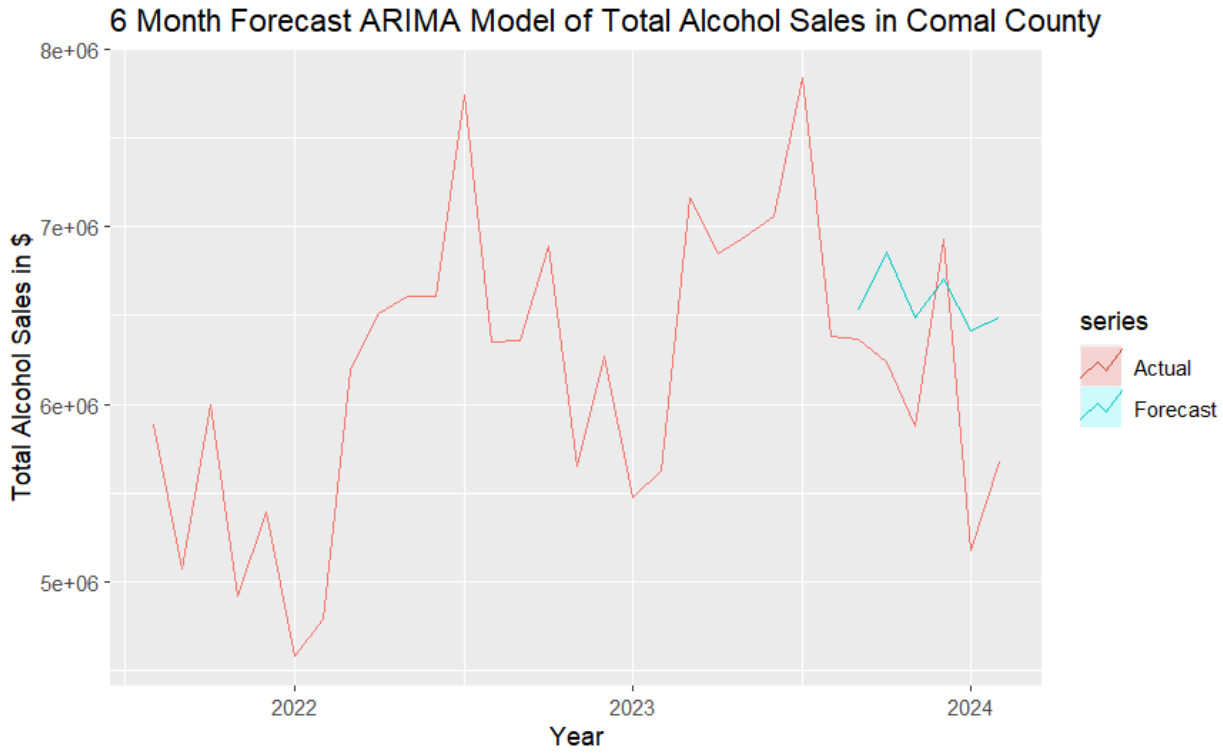
We will begin doing two six-month-ahead forecasts with our training set. We will be doing an ETS forecast and an ARIMA forecast. Using the MAA ETS model we determined was the best fit for this data we can create an ETS forecast six months ahead in R., the following is the ETS forecast layered onto all of the data and the ETS forecast layered onto a two-month window of the data.





Similar to the ETS forecast we can do the same with the ARIMA (1,1,1)(2,0,0) model to forecast six months to test against our training set. The following is the ARIMA forecast on all of the data and the ARIMA forecast on a two-year window. (Note: the ETS and ARIMA forecasts are both using data that is subjected to a log transformation, the ARIMA model undoes the log transformation so they do not share the same scale.)

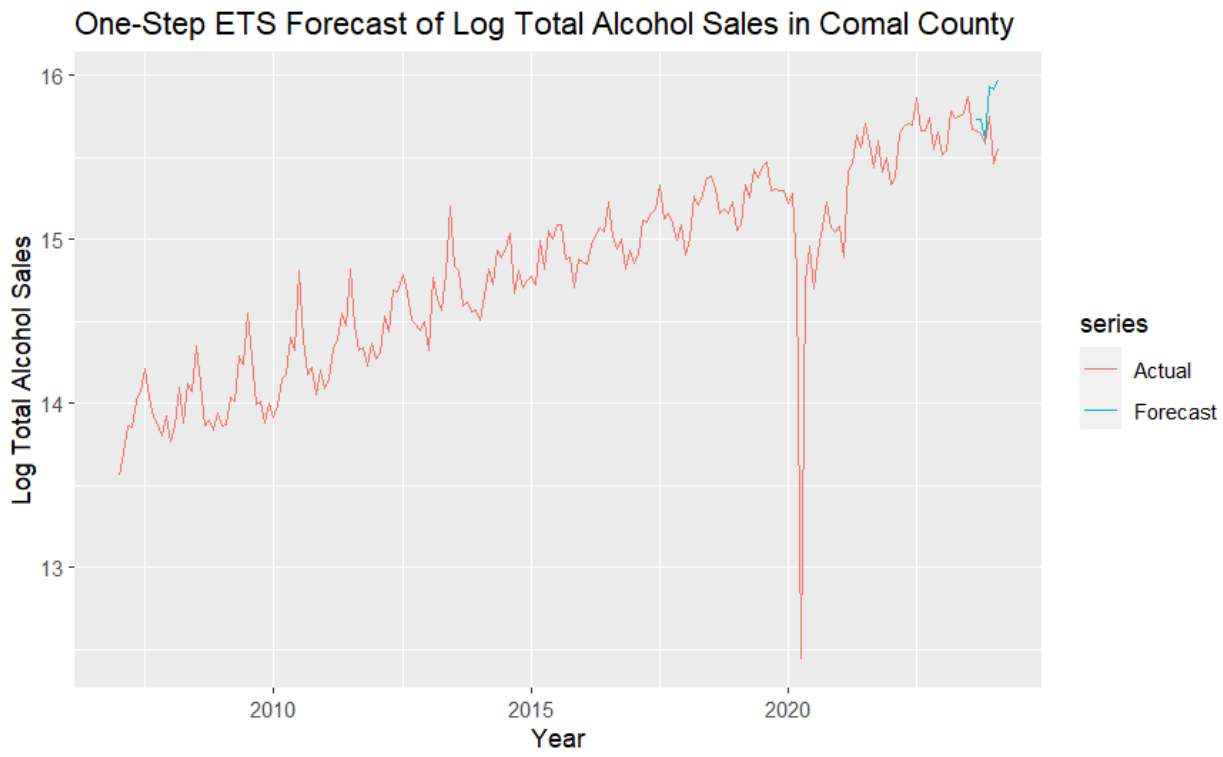

6 Month Forecast ARIMA Mode of Total Alcohol Sales in Comal County
Total Alcohol Sales in $
8e+06
6e+06
4e+06
2e+06
0e+00
2010
2015
2020
Year
series
Actual
Forecast

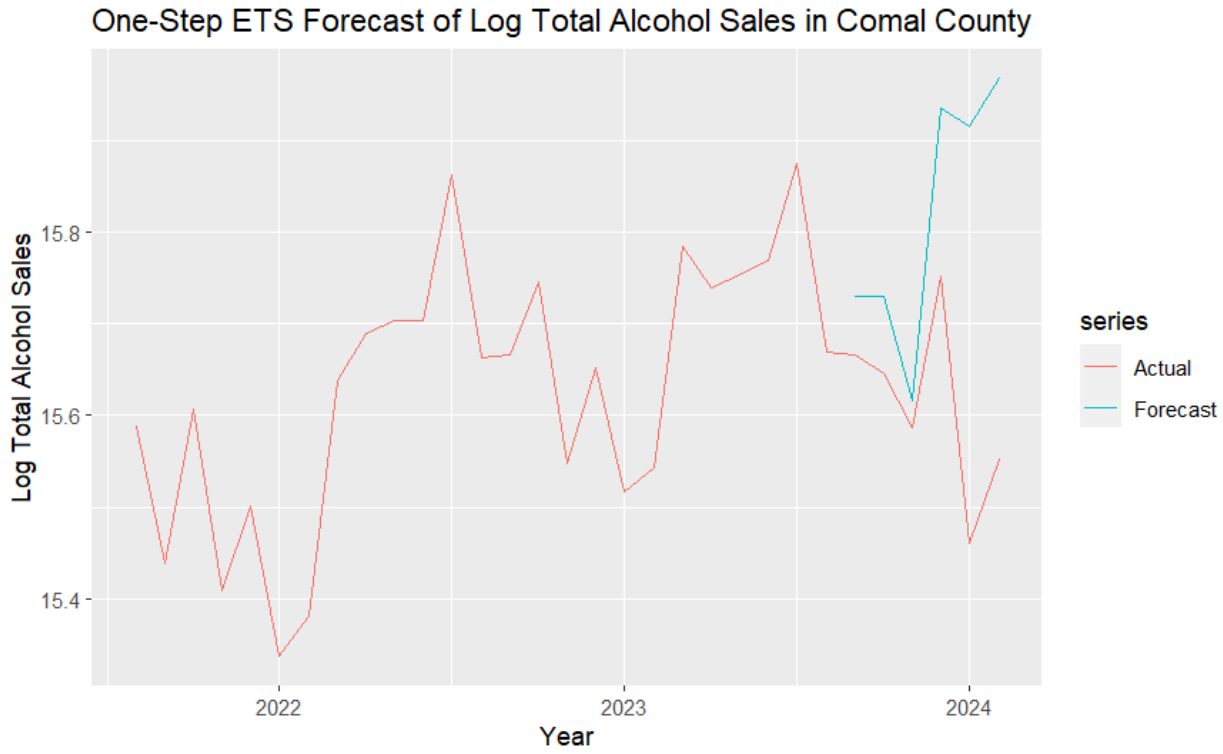

6 Month Forecast ARIMA Model of Total Alcohol Sales in Comal County
Total Alcohol Sales in $
8e+06
7e+06
6e+06
5e+06
2022
2023
2024
Year
series
Actual
Forecast

Immediately it appears evident that the ETS forecast fits the data better when compared to the ARIMA forecast. The ARIMA forecast typically overpredicts total sales and despite accounting for the seasonality it seems to not capture it very well in the forecast. On the other hand, the ETS forecast seems to fit the date better. It forecasts the data well but seems to underpredict upward spikes. That being said, while the two forecasting models do not predict perfectly, the ETS forecasting model appears to be the better fit between the two.

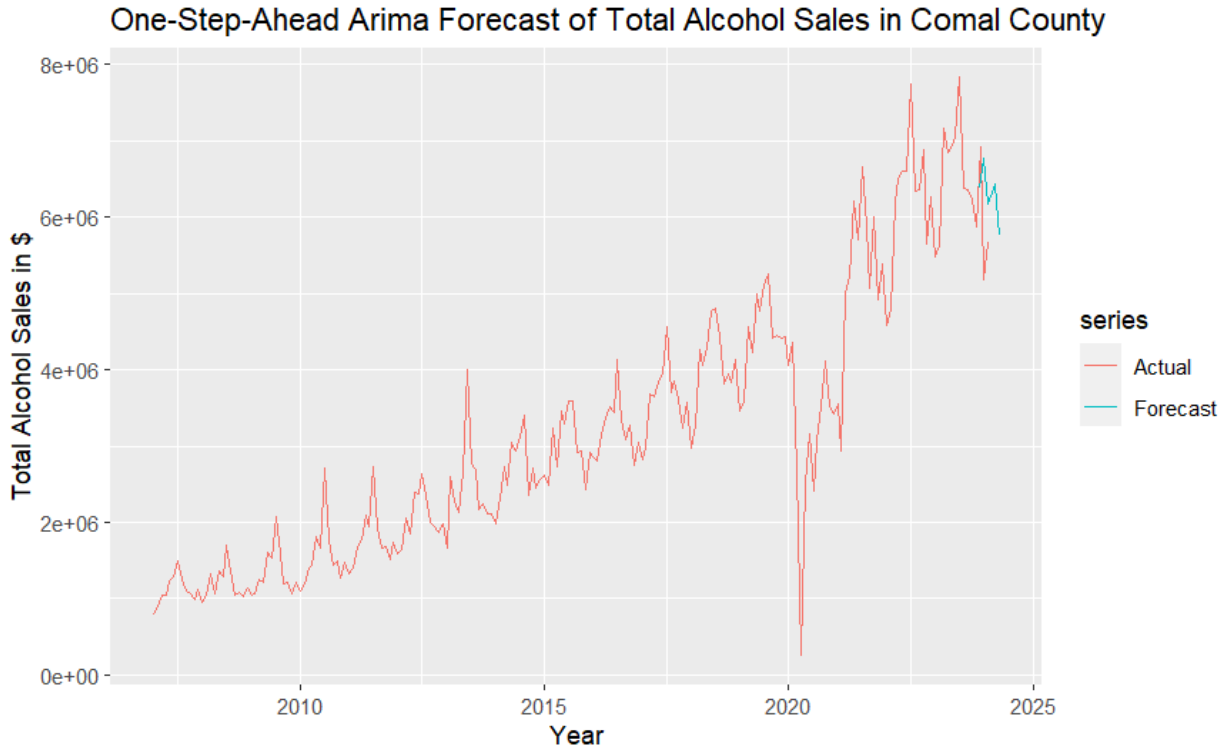
Part III: Six One-Step-Ahead Forecasts and Comparison Discussion

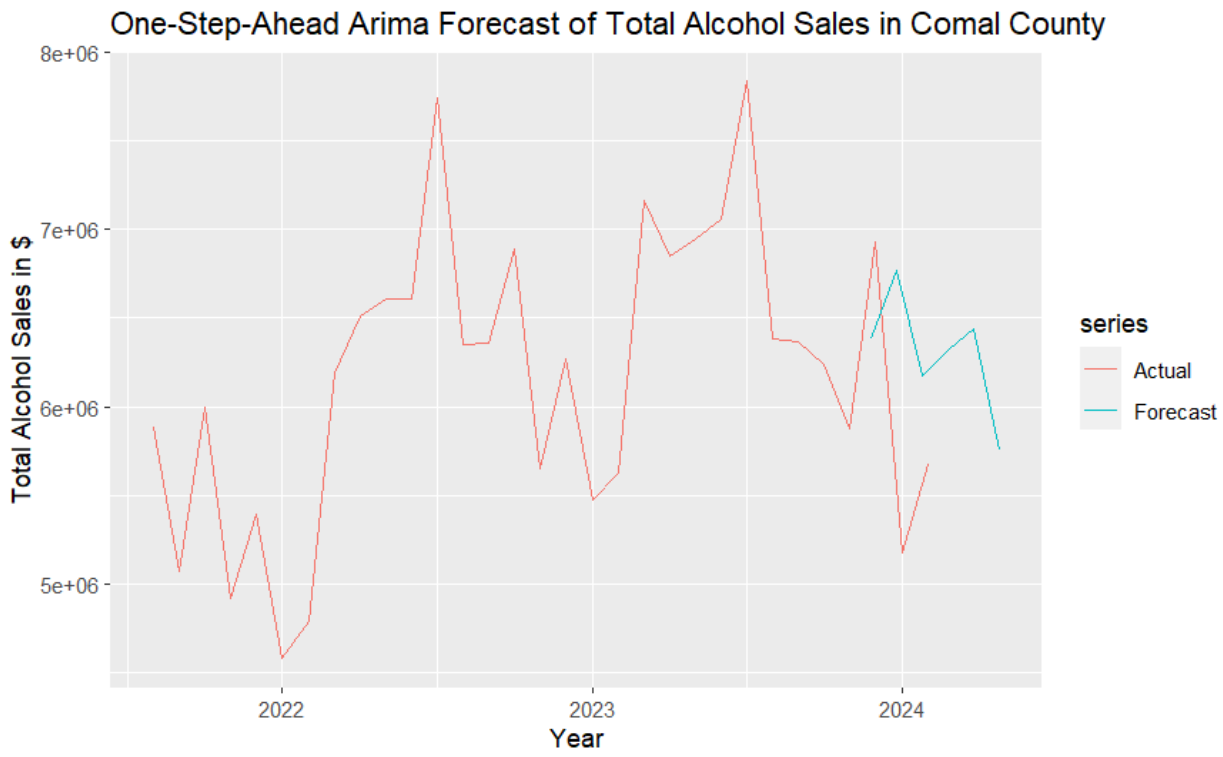
A One-step Ahead ETS and Arima forecast will also be done with the training set to determine if one of those particular forecasting models fits our data best. To do this we create a loop that does a new forecast taking into account the previous periods forecasted. For these models, we will use the same specifications from the six-month forecasts. So our ETS model will be MAA and our ARIMA model will be (1,1,1)(2,0,0). The following two graphs are the One-Step ahead ETS forecast layered onto the real data and a two-month window of that graph.


One-Step ETS Forecast of Log Total Alcohol Sales in Comal County
Log Total Alcohol Sales
16
15
14
13
2010
2015
2020
Year
series
Actual
Forecast


One-Step ETS Forecast of Log Total Alcohol Sales in Comal County
Log Total Alcohol Sales
15.8
15.6
15.4
2022
2023
2024
Year
series
Actual
Forecast

We also will create a six One-Step Ahead ARIMA forecast similarly by using a loop in R that forecasts after the period taking into account the last period's forecast. The two following graphs are the One-Step ARIMA forecast layered on all of the data and a two-year window with the One-Step ARIMA forecast layered onto it.





The One-Step ETS and ARIMA forecasts both perform in similar ways. In the first few periods, both models forecast pretty well staying within a fair margin of the actual data. However, both models following the first couple of periods start overestimating the forecasts pretty significantly.

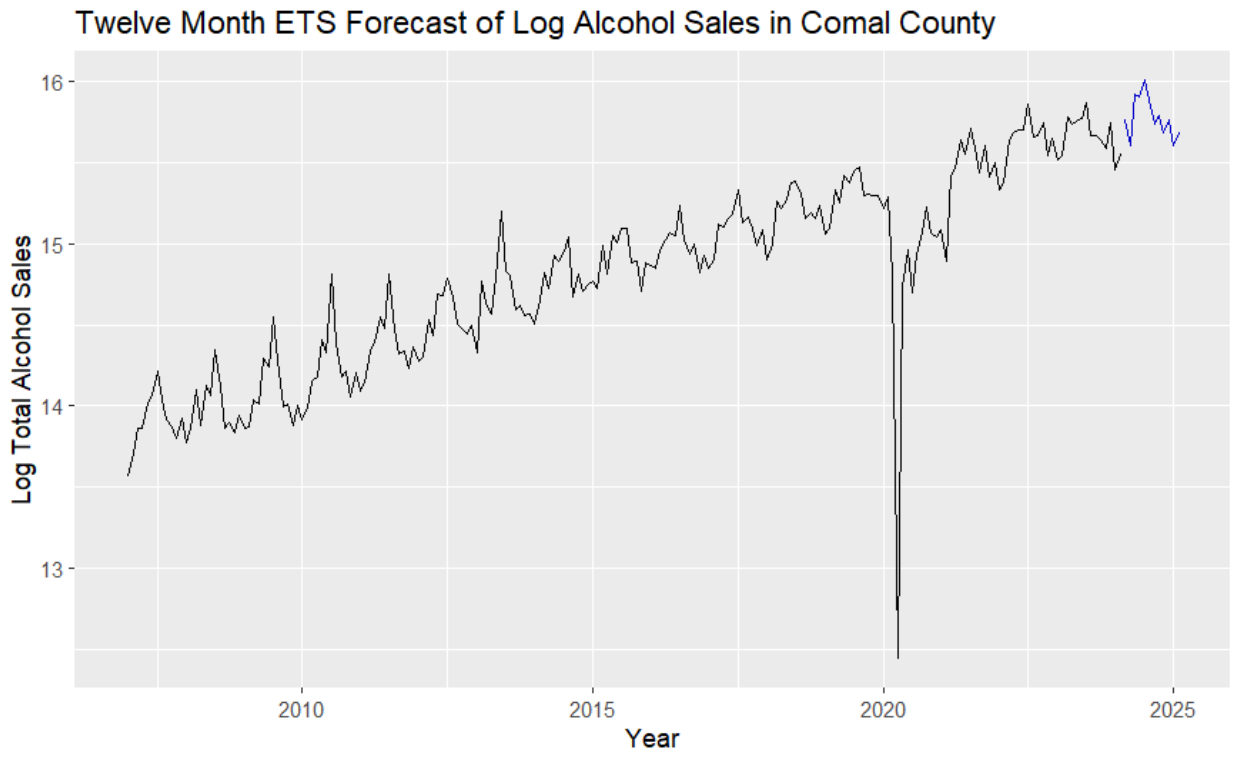
Part IV: Discussion of Preferred Forecasting Model

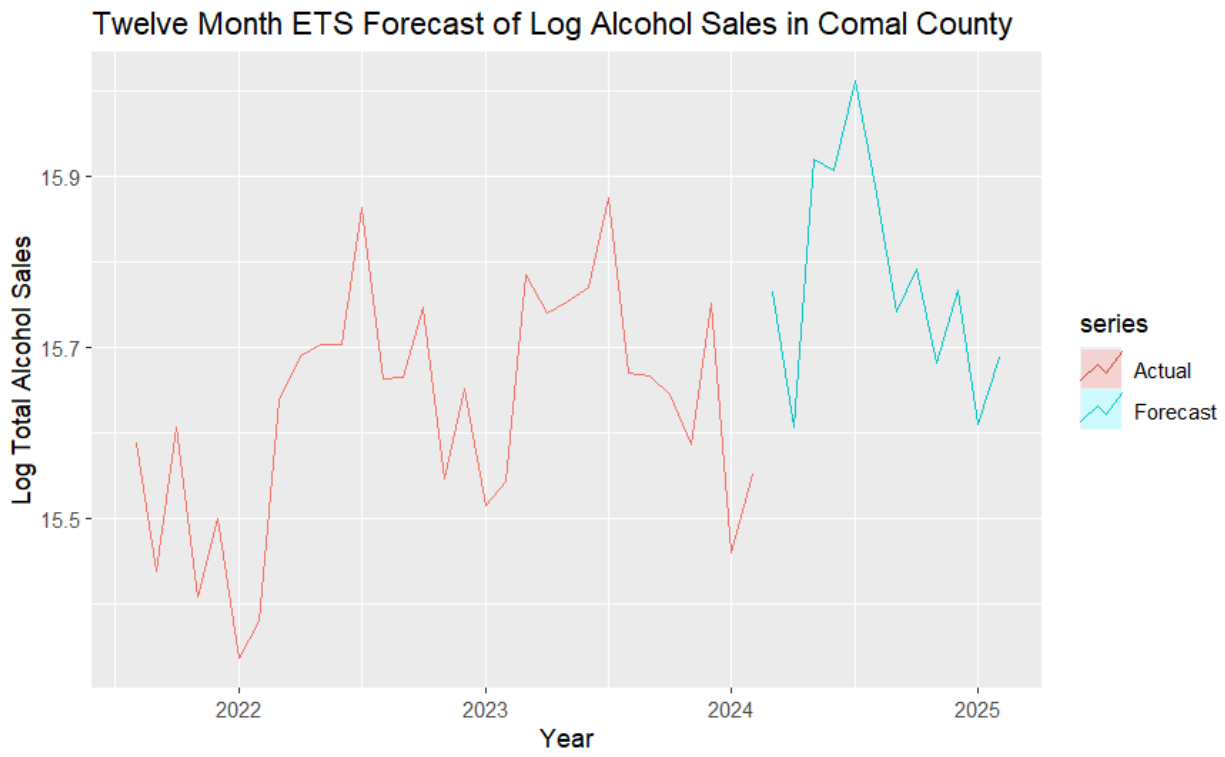
When considering which forecasting model we plan to use for our twelve-month forecasting model we need to compare all of our training set models. First, I believe that we can rule out using the One-Step ETS and ARIMA models as both models severely overestimate a couple of periods into the future. It is very unlikely that we would receive an accurate forecast if either of those models were used. Therefore, we are left to choose between the standard ETS and standard ARIMA forecasts. Of those two models we believe that the ETS model fits best. It does

not significantly overestimate or underestimate when compared to the actual data and it captures the seasonality of our data very well. The ARIMA model in comparison does not overestimate by a significant margin, but it does not seem to capture the seasonality very well, despite being a seasonal ARIMA model. Finally, with all of that being said we will be using a MAA ETS forecasting model to do our 12-month ahead forecast.

Section B: 12-Month Ahead Path Forecast

The following is the 12-month ahead forecasting Total Alcohol Sales in Comal County, Texas using the model we identified that fit our data best, the ETS MAA forecasting model.





Multivariate Forecasting Model

Section A: VAR Model-Based Multivariate Forecasting Model

The statistical outputs of the Vector Autoregression (VAR) model concerning Employed Persons, Construction Employees, and Total Sales reveal several key insights. The model, which incorporates 12 lags for each variable, demonstrates the interdependencies and dynamic relationships among these key economic indicators over a sample size of 193 observations. The VAR model's stability is confirmed by the roots of the characteristic polynomial, all of which are less than one, indicating a stable system over time.

The estimation results for Employed Persons show significant effects from its own past values as well as from past values of Construction Employees, highlighting the interconnected nature of employment sectors. Notably, the second lag of Employed Persons (coefficient = 0.414, p = 0.000367) and the first lag of Construction Employees (coefficient = 0.181, p = 0.000654) are particularly significant, suggesting that previous levels of construction activity and overall employment have a marked influence on current employment levels.

In the Construction Employees equation, the impact of past employment levels across the sectors is evident, with significant contributions from lags of both Employed Persons and Construction Employees. This indicates a strong autoregressive component within the sector, where past construction employment levels are predictive of future levels, underlined by the third lag of Construction Employees being especially significant (coefficient = -0.379, p = 0.000208).

For Total Sales, the model reveals a strong autoregressive nature, where past sales significantly predict future sales. Additionally, the model highlights the influence of the broader economic activity on sales, as evidenced by significant coefficients for past values of Employed Persons and Construction Employees at various lags. For instance, the first lag of Total Sales (coefficient = -0.436, p = 0.000351) and the eighth lag of Total Sales (coefficient = -0.536, p = 0.000035) show a strong predictive power, underscoring the dependency of sales volume on employment levels and sectoral activities.

The statistical significance of the model is robust, with p-values indicating significant predictive power of many lags across all equations. This is further supported by the F-statistic (F = 3.946) and its associated p-value (p = 1.32e-09), which confirm the overall significance of the model in capturing the dynamics among the variables.

The diagnostics of the model, including residual standard errors and Adjusted R-squared values (0.3558 for Total Sales), suggest varying degrees of fit across the different equations, with Total Sales showing a relatively better fit. This indicates the model's adequacy in capturing the underlying patterns in the data, although further checks for autocorrelation, heteroscedasticity, and normality of residuals would be necessary to validate these findings fully.

Overall, the VAR model provides a comprehensive view of the interplay between key economic indicators, offering valuable insights for forecasting and economic analysis. The significant lags identified through the model form a basis for understanding how past economic conditions influence future trends, serving as a crucial tool for economic planning and decision-making.

Section B: Discussion of Variable Selection

In the context of forecasting total alcohol sales in Comal County, the selection of relevant economic variables is pivotal to the accuracy and relevance of the VAR model. For this analysis, two particular series have been chosen to accompany the log-transformed total sales data: the log of Employed Persons in Comal County, TX, and the log of All Employees in the Construction sector in Texas. These selections are grounded in their potential impact and correlation with alcohol sales in the region.

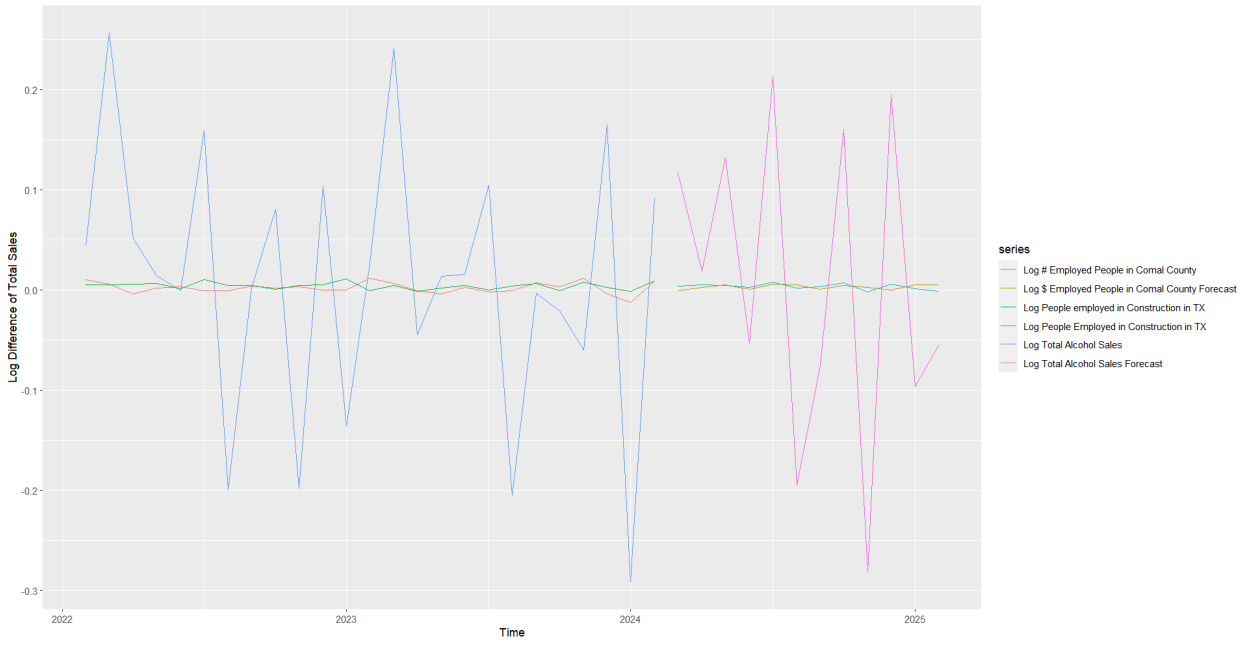
The number of employed persons in Comal County serves as a direct gauge of the local economic environment. Employment levels are closely tied to economic vitality, influencing consumer confidence and disposable income. Higher employment rates often lead to increased disposable income, which can enhance consumer spending on non-essential goods, including

alcohol. Given the specific focus on Comal County, this variable directly reflects the economic conditions that the residents, and consequently the alcohol sales market, are experiencing.

Additionally, the construction sector is a significant component of the state's economy and is often a bellwether for broader economic trends. Activity in this sector can indicate overall economic health through its ties to infrastructure development, investment, and employment creation. By incorporating the employment statistics from this sector, the model can capture the ripple effects of state-wide economic shifts on local markets like Comal County. Although this metric is on a state-wide level, its fluctuations are indicative of economic conditions that can influence consumer behavior in Comal County as well.

Incorporating these variables into the VAR model aims to harness their explanatory power regarding fluctuations in alcohol sales. The county's employment status directly affects the purchasing power of its residents, while the health of the construction industry reflects broader economic conditions that can also influence consumer spending patterns. Together, these indicators provide a robust framework for anticipating changes in alcohol sales, making them essential components of our multivariate forecasting model.

Section C: 12-Month Ahead Path Forecast of Total Sales



Because the two forecasts have different scales, we cannot directly compare the outputs received from the forecasting equations. The two models forecast well given their scales. They both capture the seasonal elements of the dataset. Both capture the peaks and troughs of the seasons and appear to show increased sales in alcohol. The multivariate model forecast appears more stable with less positive growth.

References

Association, T. S. H. (n.d.). Comal County. Texas State Historical Association. Retrieved April 11, 2024, from https://www.tshaonline.org/handbook/entries/comal-county

Association (TSHA), T. S. H. (n.d.). Schertz. Texas Almanac. Retrieved April 11, 2024, from https://www.texasalmanac.com/places/schertz

Comal County, TX | Data USA. (n.d.). Datausa.io. Retrieved April 11, 2024, from https://datausa.io/profile/geo/comal-county-tx#:~:text=From%202020%20to%202021%2C%20employment

Comal County, Texas Population 2023. (n.d.). Worldpopulationreview.com. https://worldpopulationreview.com/us-counties/tx/comal-county-population

Comal County, TX population by year, race, & more. (2024, April 5). USAFacts. https://usafacts.org/data/topics/people-society/population-and-demographics/our-changing-population/state/texas/county/comal-county/

Federal Reserve Bank of St. Louis. (2024, April 30). Retrieved from https://fred.stlouisfed.org/

The Demographic Statistical Atlas of the United States - Statistical Atlas. (n.d.). Statisticalatlas.com. Retrieved April 11, 2024, from https://statisticalatlas.com/county/Texas/Comal-County/Population

Who are the top employers in New Braunfels? | Veramendi. (n.d.). Veramenditx.com. Retrieved April 11, 2024, from https://veramenditx.com/news/who-are-the-top-employers-in-new-braunfels/