

IBM 6520.P01 (S24-Sp Ses Dgr) Market Forecasting

Texas Home Forecasting

Andrea Aceves, amaceves@cpp.edu

Joshua Arias, jdarias@cpp.edu

Jonathan De La Torre, delatorre1@cpp.edu

Andrew Mao, abmao@cpp.edu

May 12th, 2024

Nature of the Forecasting Problem

Our group chose the data set “txhousing” from ggplot2, which provides comprehensive data about the housing market in Texas from the years 2000 to 2015. This data allows us to analyze trends, patterns, and fluctuations in the real estate market across various regions of Texas. We aim to forecast future housing prices and sales volumes, with a specific focus on identifying potential investment opportunities. Our analysis will primarily focus on the top 5 counties by median price, aiming to forecast trends over the next 3-5 years.

Key Data, Statistics, or Assumptions Used in Identification of Key Factors

We analyzed the median housing prices over the past 10 years in Texas to identify any seasonality or trends. Our findings indicate that there is no significant seasonality in median prices, even when considering the 2010 recession, which was followed by a price recovery. This lack of seasonality is attributed to consistent supply and demand dynamics within the top 5 counties of Texas. The data was sourced from the “txhousing” dataset, and preprocessing

involved filtering and cleaning to ensure accuracy. Additionally, we considered key economic factors such as interest rates and housing policies that could impact the housing market.

Median Price Over 10 Years

```
library(ggplot2)
library(dplyr)
library(lubridate)

data(txhousing)

txhousing$date <- as.Date(paste(txhousing$year, txhousing$month, "01", sep = "-"))

latest_year <- max(txhousing$year)
txhousing_10_years <- txhousing %>%
  filter(year >= (latest_year - 10))

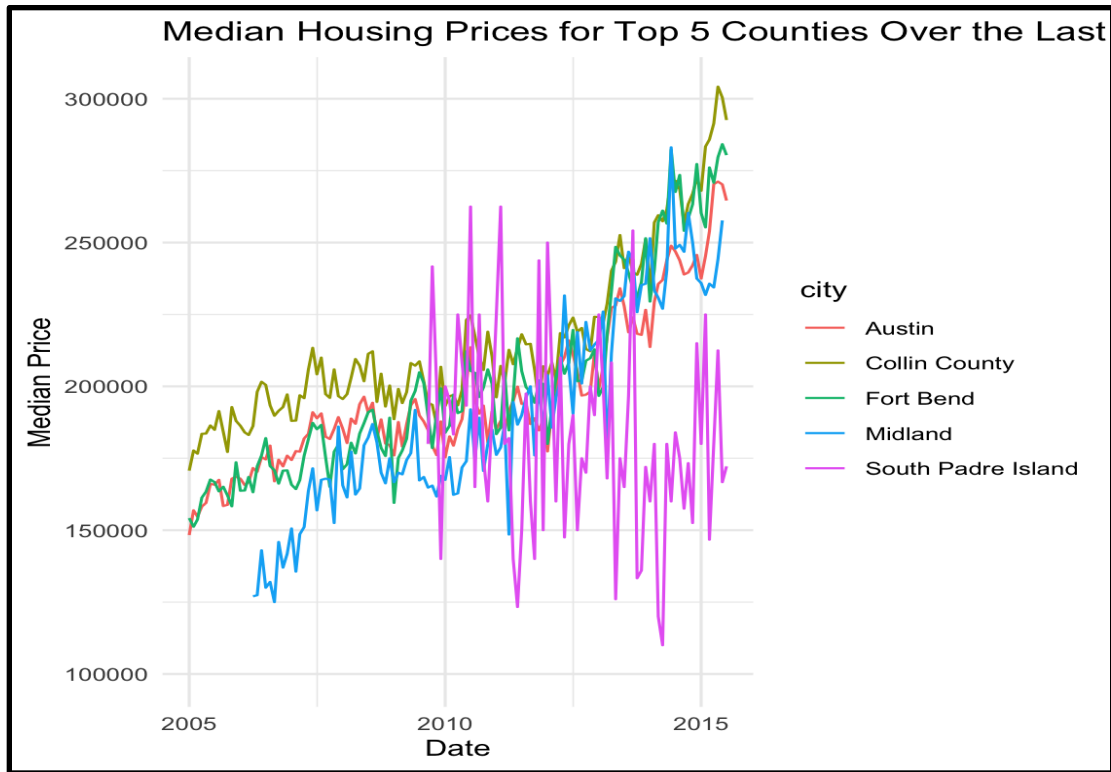
median_prices <- txhousing_10_years %>%
  group_by(city) %>%
  summarise(median_price = median(median, na.rm = TRUE)) %>%
  arrange(desc(median_price))

top_5_counties <- median_prices %>%
  top_n(5)

txhousing_top_5 <- txhousing_10_years %>%
  filter(city %in% top_5_counties$city)

top_5_plot <- ggplot(txhousing_top_5, aes(x = date, y = median, group = city, color = city)) +
  geom_line() +
  labs(title = "Median Housing Prices for Top 5 Counties Over the Last 10 Years", x = "Date", y
= "Median Price") +
  theme_minimal()

print(top_5_plot)
```



When examining the median housing prices from 2005 to 2015, the analysis shows that despite economic fluctuations, prices have generally trended upwards, indicating robust market resilience. For potential investors, this trend suggests that investing in these counties could be

advantageous, as the housing market has demonstrated a capacity to recover from downturns and continue growing.

Sales Over 10 Years:

```
library(ggplot2)
library(dplyr)
library(lubridate)
```

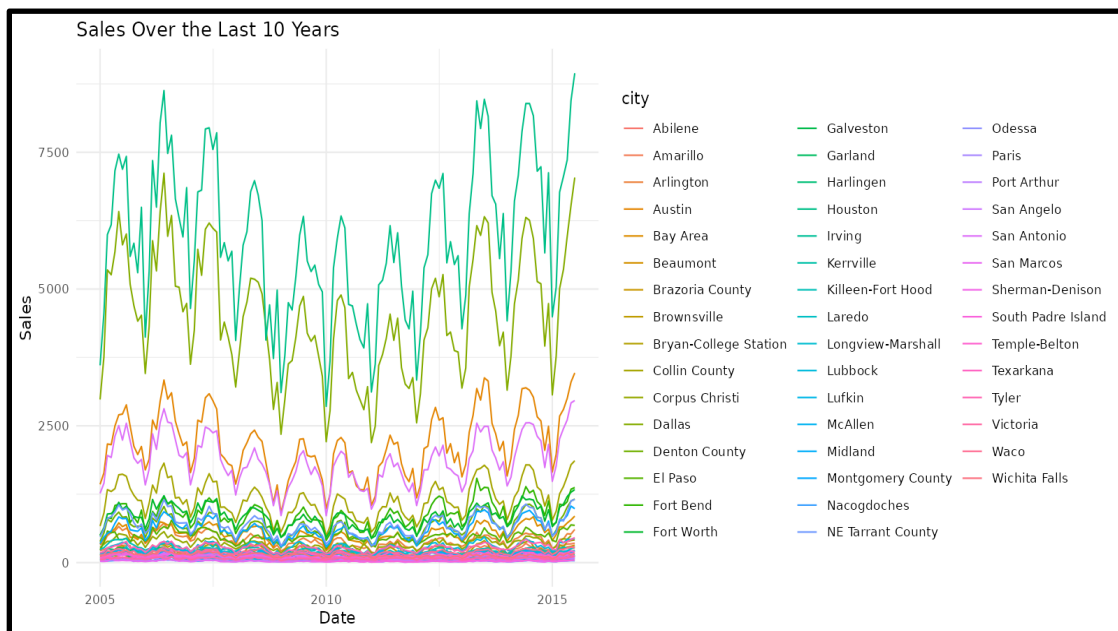
```
data(txhousing)
```

```
txhousing$date <- as.Date(paste(txhousing$year, txhousing$month, "01", sep = "-"))
```

```
latest_year <- max(txhousing$year)
txhousing_10_years <- txhousing %>%
  filter(year >= (latest_year - 10))
```

```
sales_10_years_plot <- ggplot(txhousing_10_years, aes(x = date, y = sales, group = city, color =
city)) +
  geom_line(na.rm = TRUE) + # Remove rows with missing values
  labs(title = "Sales Over the Last 10 Years", x = "Date", y = "Sales") +
  theme_minimal()
```

```
print(sales_10_years_plot)
```

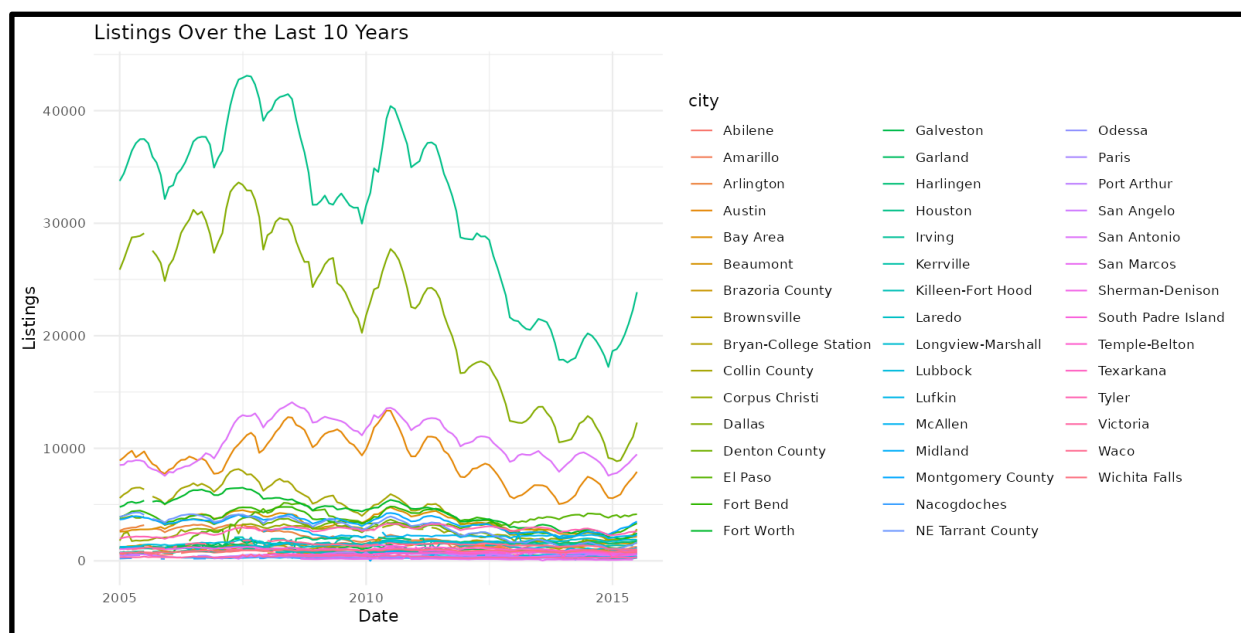


The analysis of sales volumes over these 10 years reveals a pattern of growth consistent with the increase in median prices. This correlation suggests that higher sales volumes are likely driven by increased housing demand, contributing to the upward pressure on prices. The

consistent sales trends indicate a stable and growing housing market, making it an attractive area for investment.

Listings Over 10 Years

```
library(ggplot2)
library(dplyr)
library(lubridate)
data(txhousing)
txhousing$date <- as.Date(paste(txhousing$year, txhousing$month, "01", sep = "-"))
latest_year <- max(txhousing$year)
txhousing_10_years <- txhousing %>%
  filter(year >= (latest_year - 10))
listings_10_years_plot <- ggplot(txhousing_10_years, aes(x = date, y = listings, group = city,
color = city)) +
  geom_line(na.rm = TRUE) + # Remove rows with missing values
  labs(title = "Listings Over the Last 10 Years", x = "Date", y = "Listings") +
  theme_minimal()
print(listings_10_years_plot)
```

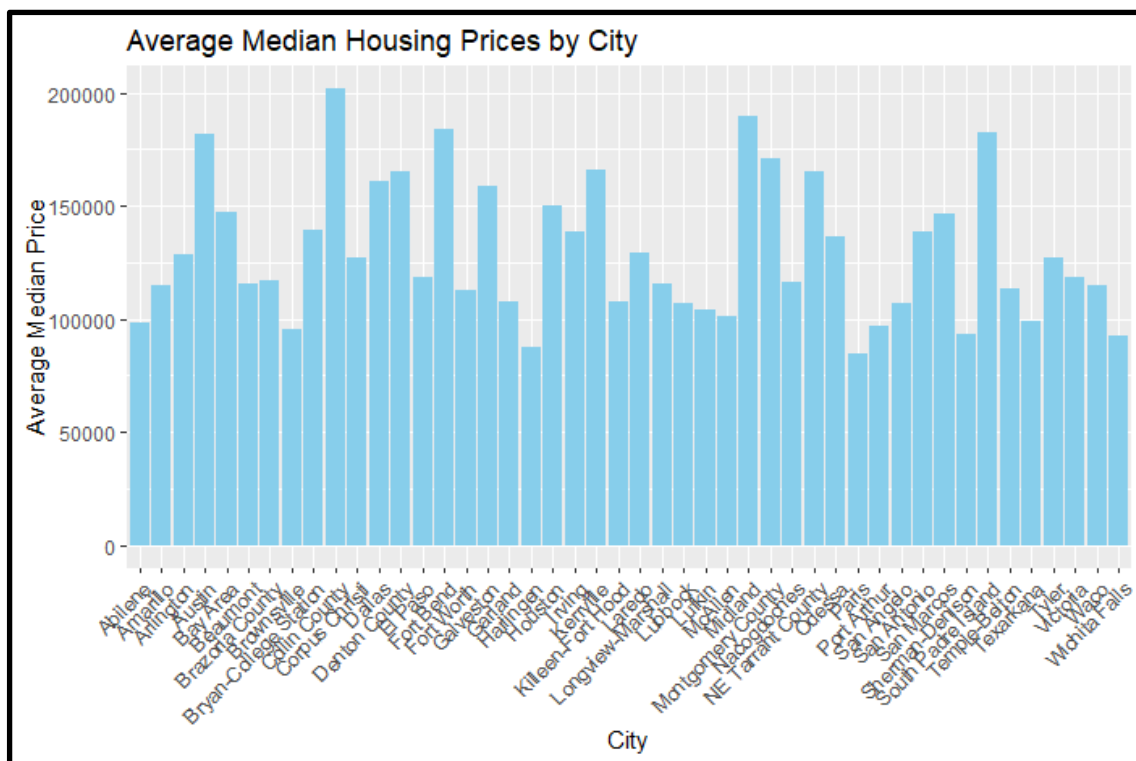


The number of listings over these 10 years has not exhibited the same seasonality as sales, aligning more closely with the rise in median prices. This trend suggests that limited supply, relative to demand, has been a significant factor in driving up prices in the top 5 counties.

Investors should be aware of these supply constraints, as they could lead to continued price increases if demand remains strong.

Average Prices Per Region in Texas:

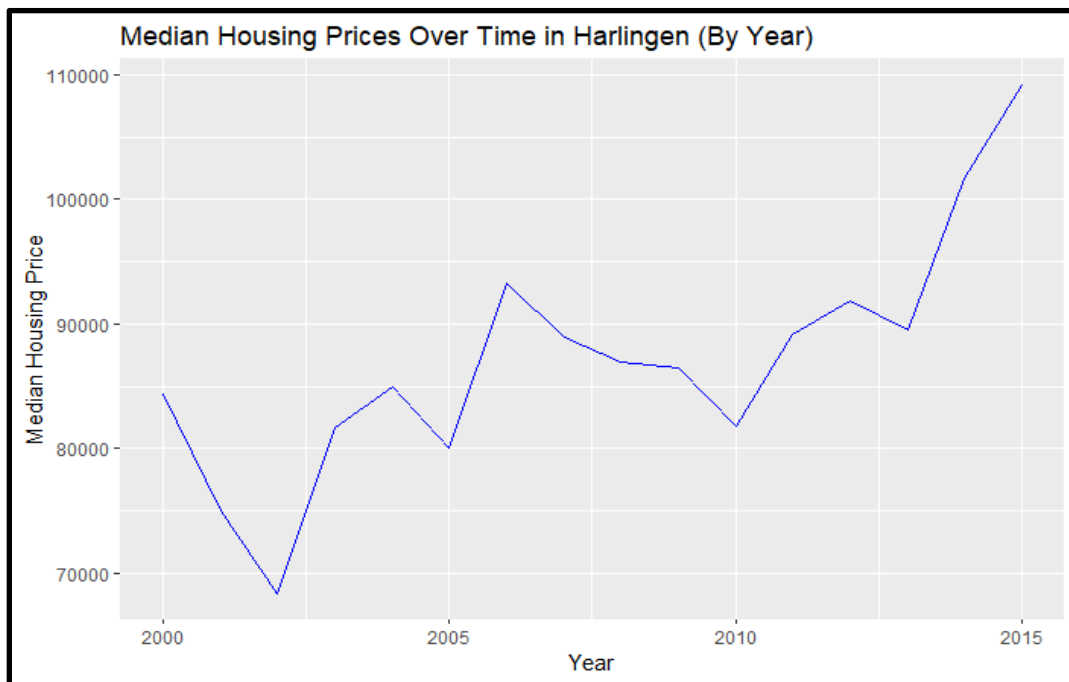
```
library(ggplot2)
library(dplyr)
data(txhousing)
region_prices <- txhousing %>%
  group_by(city) %>%
  summarise(avg_median_price = mean(median, na.rm = TRUE))
ggplot(region_prices, aes(x = city, y = avg_median_price)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Average Median Housing Prices by City",
       x = "City",
       y = "Average Median Price") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



The average median housing prices by city provide insights into regional market variations, with Harlingen having the lowest median prices and Collin County the highest. This data can guide investors in choosing cities that align with their budget and investment goals. For instance, Harlingen may offer more affordable entry points, while Collin County, with its consistent growth, may be more suitable for investors seeking long-term appreciation.

Harlingen Prices Over Time

```
library(ggplot2)
library(dplyr)
harlingen_data <- txhousing %>%
  filter(city == "Harlingen")
harlingen_data$date <- as.Date(paste0(harlingen_data$year, "-", harlingen_data$month, "-01"))
median_prices_by_year <- harlingen_data %>%
  group_by(year) %>%
  summarise(median_price = median(median, na.rm = TRUE))
ggplot(median_prices_by_year, aes(x = year, y = median_price)) +
  geom_line(color = "blue") +
  labs(title = "Median Housing Prices Over Time in Harlingen (By Year)",
       x = "Year",
       y = "Median Housing Price")
```

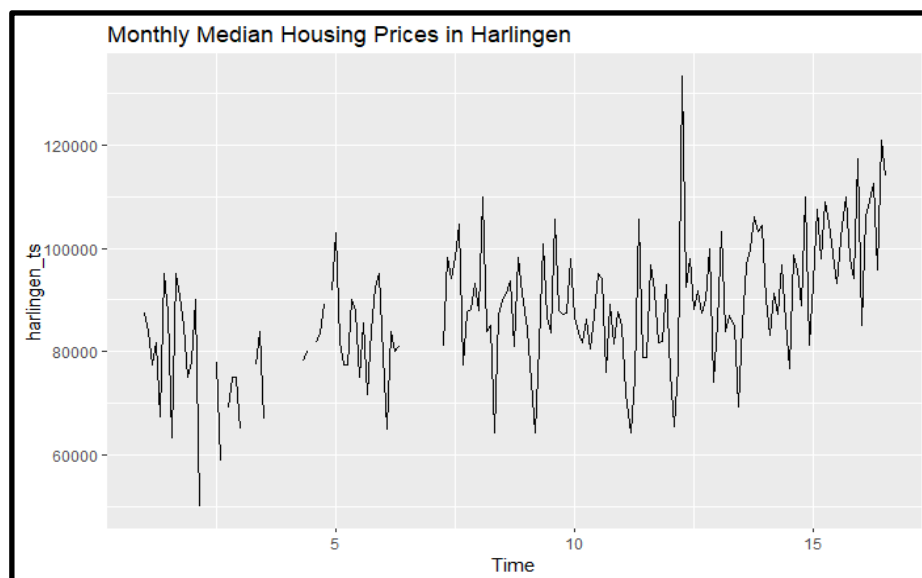


Harlingen's median housing prices have shown a significant increase from 2014 to 2015, despite some fluctuations in previous years. This upward trend suggests a positive outlook for

real estate investment in this area. However, investors should consider the potential for future price dips and be prepared for longer holding periods to achieve desired returns.

ARIMA Forecast for Harlingen

```
library(forecast)
library(ggplot2)
library(dplyr)
harlingen_data <- txhousing %>%
  filter(city == "Harlingen")
harlingen_data$date <- as.Date(paste0(harlingen_data$year, "-", harlingen_data$month, "-01"))
harlingen_ts <- ts(harlingen_data$median, frequency = 12)
autoplot(harlingen_ts) + labs(title = "Monthly Median Housing Prices in Harlingen")
train <- window(harlingen_ts, end = c(2016, 12))
test <- window(harlingen_ts, start = c(2017, 1))
arima_model <- auto.arima(train)
arima_forecast <- forecast(arima_model, h = length(test))
accuracy(arima_forecast, test)
autoplot(arima_forecast) +
  autolayer(test, series = "Actual") +
  labs(title = "ARIMA Forecast for Median Housing Prices in Harlingen")
```

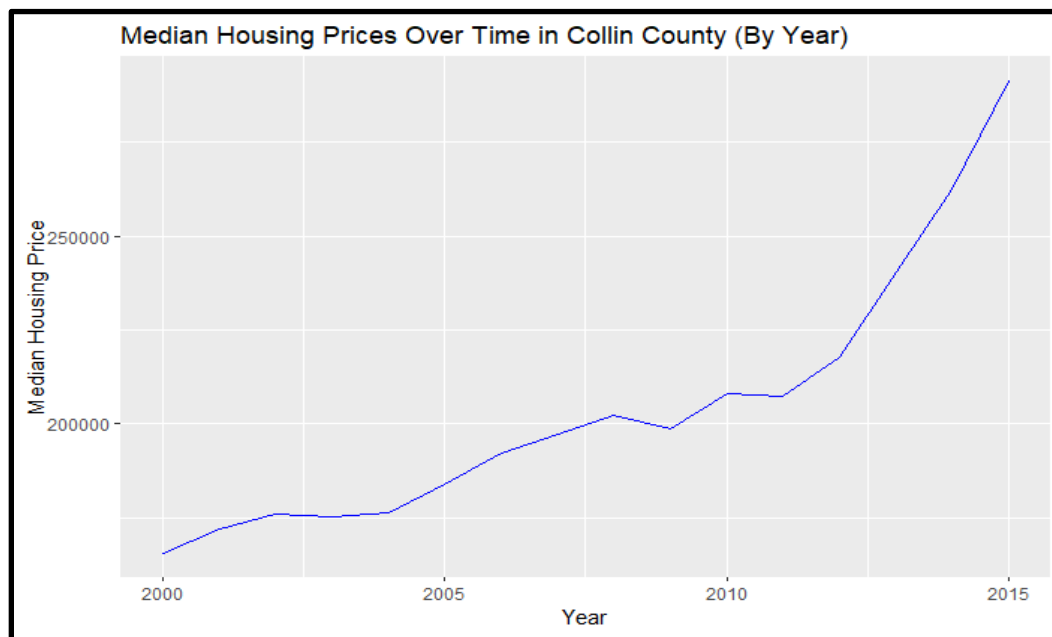


The ARIMA forecast for Harlingen's median housing prices indicates potential market volatility, with forecasted prices showing variability. The accuracy metrics suggest a reasonable fit, but investors should proceed with caution due to the potential for supply and demand

imbalances. This forecast highlights the importance of thorough market analysis before making investment decisions.

Collin Country Price Over Time

```
library(ggplot2)
library(dplyr)
collin_data <- txhousing %>%
  filter(city == "Collin County")
collin_data$date <- as.Date(paste0(collin_data$year, "-", collin_data$month, "-01"))
median_prices_by_year <- collin_data %>%
  group_by(year) %>%
  summarise(median_price = median(median, na.rm = TRUE))
ggplot(median_prices_by_year, aes(x = year, y = median_price)) +
  geom_line(color = "blue") +
  labs(title = "Median Housing Prices Over Time in Collin County (By Year)",
       x = "Year",
       y = "Median Housing Price")
```



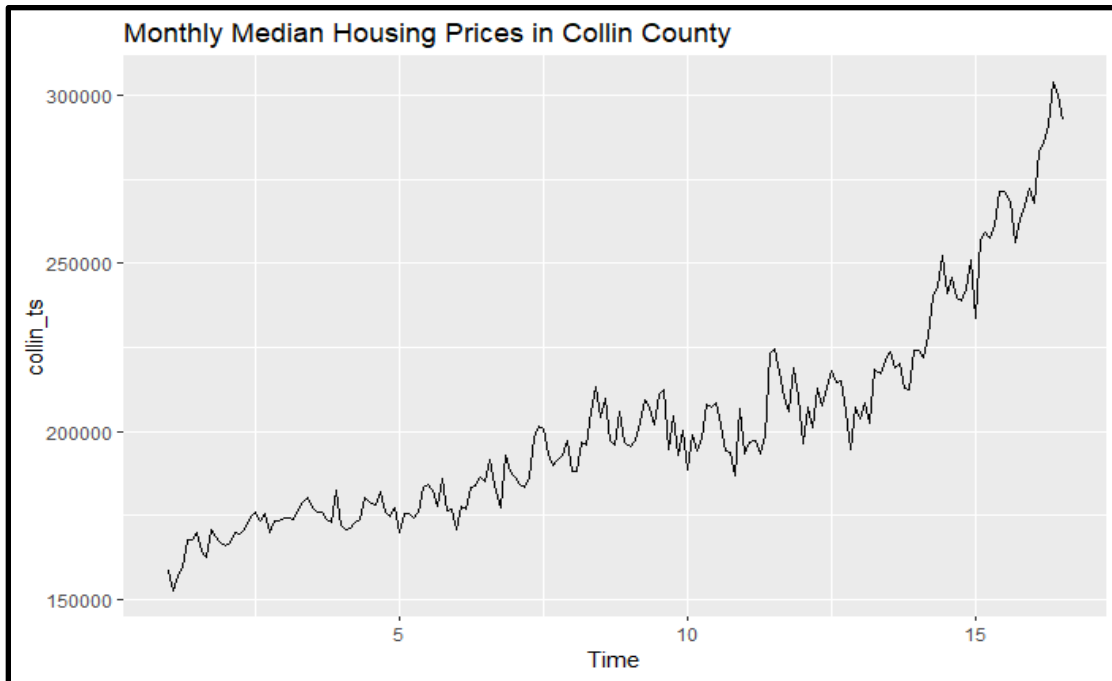
Collin County has experienced a nearly continuous increase in housing prices from 2000 to 2015, indicating strong market demand and limited supply. This consistent growth suggests

that Collin County is a promising area for real estate investment, offering opportunities for

substantial appreciation over time.

ARIMA Forecast for Collin County

```
library(forecast)
library(ggplot2)
library(dplyr)
collin_data <- txhousing %>%
  filter(city == "Collin County")
collin_data$date <- as.Date(paste0(collin_data$year, "-", collin_data$month, "-01"))
collin_ts <- ts(collin_data$median, frequency = 12)
autoplot(collin_ts) + labs(title = "Monthly Median Housing Prices in Collin County")
train <- window(collin_ts, end = c(2016, 12))
test <- window(collin_ts, start = c(2017, 1))
arima_model <- auto.arima(train)
arima_forecast <- forecast(arima_model, h = length(test))
accuracy(arima_forecast, test)
autoplot(arima_forecast) +
  autolayer(test, series = "Actual") +
  labs(title = "ARIMA Forecast for Median Housing Prices in Collin County")
```

The ARIMA forecast for Collin County's median housing prices projects continued growth, supported by consistent historical trends. The forecast accuracy is robust, making it a valuable tool for investors planning their strategies. Given the positive outlook, investors may consider Collin County a favorable area for long-term investments, capitalizing on the expected price increases.

Executive Summary With Actionable Recommendations

The Texas housing market analysis, based on the "txhousing" dataset from ggplot2, covers data from 2000 to 2015, focusing on median housing prices, sales volumes, and listings across various regions. The primary objective was to identify trends, assess market stability, and forecast future outcomes to guide potential investments.

Key Findings:

- **Median Price Trends:** Analysis of the top 5 counties shows a consistent upward trend in median housing prices, indicating a resilient market with strong recovery post-recession.
- **Sales and Listings Dynamics:** Sales volumes and listings exhibit patterns reflecting high

demand and constrained supply, driving prices upward.

- Regional Insights: Significant variations exist across regions, with Collin County showing the highest median prices and Harlingen the lowest.
- Forecast Analysis: ARIMA forecasts for Harlingen and Collin County indicate potential for growth, with Collin County showing stronger and more consistent positive trends.

Actionable Recommendations:

Investment in High-Growth Areas

- Focus on Collin County: Due to its strong historical growth and favorable forecast, Collin County is recommended for long-term investments, promising substantial appreciation.
- Evaluate Harlingen Cautiously: Despite recent growth, potential volatility suggests that investments in Harlingen should be approached with caution, favoring a strategy of long-term holding to mitigate risks.

Monitor Market Conditions

- Track Supply and Demand Trends: Continuous monitoring of listings and sales data is crucial to identify emerging trends and adjust investment strategies accordingly.
- Assess Economic Indicators: Keep an eye on broader economic factors such as interest rates and housing policies that could impact market dynamics.

Diversify Investments:

- Explore Emerging Markets: Beyond the top 5 counties, consider emerging markets within Texas that show signs of growth but may offer more affordable entry points.
- Balance Portfolio: Diversify investments across regions with different risk profiles to balance potential returns with market stability.

Leverage Forecasting Tools:

- Use ARIMA Models: Utilize ARIMA and other forecasting models regularly to update predictions and refine investment decisions based on the latest data trends.

By following these recommendations, investors can capitalize on the robust growth in Texas's housing market while mitigating risks associated with market volatility and economic fluctuations.

Appendendix

Ruhil, A. V. (n.d.). Advanced Data Visualizations. Retrieved from
<https://people.ohio.edu/ruhil/RBookNew/advanced-data-visualizations.html>