**Topic of interest:**

**Statistical Analysis of the Nebraska Real Estate Market**

Jacob Pickrel-Smith

Department of Data Science, Bellevue University

DSC530-T302: Data Exploration and Analysis

Professor Cary K. Jim

# August 10, 2024

# Summary of Analysis

## Research Question

This analysis aims to explore how key real estate market variables—such as pending sales, months of supply, median sale price, median days on market (DOM), and inventory—interact to influence overall market trends in Nebraska. The study seeks to understand the relationships between supply, demand, and pricing, and how these factors collectively shape the behavior of the real estate market.

## Outcome of the Exploratory Data Analysis (EDA):

**The EDA revealed several key insights:**

1. Distribution and Outliers: Histograms and KDE plots for each variable showed the general distribution patterns, with some variables like median\_sale\_price and pending\_sales being relatively normally distributed, while others, such as months\_of\_supply, showed more variability. Outliers were identified and managed using winsorization to prevent them from skewing the analysis.

2. Descriptive Statistics: The mean, mode, spread, and tails of each variable were calculated. The data indicated that median\_sale\_price had a mean close to the mode, suggesting a somewhat symmetric distribution. Inventory and months\_of\_supply showed more skewness, suggesting they might have a more significant impact on market conditions.

3. Correlations: The correlation matrix highlighted that pending\_sales had a strong positive correlation with inventory, indicating that as more homes are available, more sales are pending. However, the relationship between median\_sale\_price and months\_of\_supply was slightly negative, suggesting that as supply increases relative to demand, prices tend to drop.

**What Was Missed During the Analysis?**

The analysis was comprehensive, but further granularity could have been achieved by segmenting the data by property type (e.g., single-family homes, condos) or by different regions within Nebraska. This would provide a more nuanced understanding of how these factors influence prices across various segments of the market.

**Were There Any Variables That Could Have Helped?**

Additional variables, such as interest rates or economic indicators like employment rates, could have provided more context to the housing market dynamics. These factors often influence buyer behavior and, consequently, market prices.

**Were There Any Assumptions Made That Were Incorrect?**

The assumption that the relationships between variables are linear and can be captured through simple correlation and regression may not fully capture the complexities of the real estate market. Non-linear relationships or interactions between variables could play a significant role and may require more advanced modeling techniques.

**Challenges Faced and Areas of Uncertainty?**

One of the main challenges was handling missing data and ensuring that outliers were appropriately managed without losing critical information. Additionally, interpreting the relationships between variables, especially when they were not strongly correlated, required careful consideration of the underlying market dynamics.

**Summary**

In summary, the analysis provided valuable insights into the housing market in Nebraska, showing that supply and demand factors play crucial roles in determining median sale prices. However, additional data and more sophisticated models that include more of the available features in the utilized dataset could further enhance the understanding of these dynamics.

**References**

*Redfin. (n.d.). Data Center. Redfin. Retrieved [August 10, 2024], from https://www.redfin.com/news/data-center/*