# When Does the Market Heat Up?

## Charlottesville Housing Price and Seasonality

Group 8: Grayson Fiveash (Leader),
Amanda Nordwall, and Bridget Holt
DS 4002 Prototyping
3/25/2025

Details | Data Acquisition | Analysis Plan | Tricky Decision | Results

# Project Details

## Research Question

How does seasonality affect Charlottesville housing prices regionally and by neighborhood?

## Hypothesis

Median housing prices increase during the spring months (i.e., April and May) in Charlottesville, VA, on a regional and neighborhood basis.

## Motivation

Charlottesville has the second most expensive real estate market in Virginia behind Northern Virginia, and as residents, we wanted to identify how past trends will influence the future of the housing market.

## Modeling Approach

Creating a multiple linear regression model, measuring results with R-squared and RMSE, and conducting cluster analysis

# Data Acquisition & Explanation

- Downloaded Redfin Charlottesville housing market data from January 2020 to January 2025
- Selected 6 prominent Charlottesville neighborhoods – Venable, Belmont, Jefferson Park, Fry's Spring, Fifeville, and Martha Jefferson
- Collected median housing and houses sold data from the 6 neighborhoods from Redfin
- Cleaned and merged the Charlottesville regional data with the neighborhood data into a single dataset
- Redfin is a public-facing site with accessible historical market data, so there were no licensing or ethical concerns

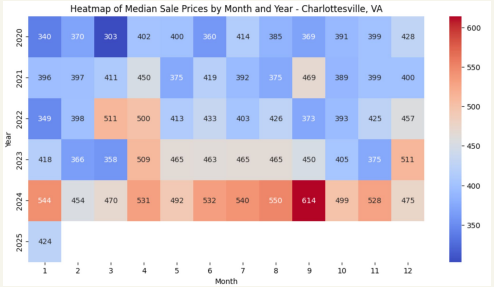
| | |
|---|---|
| **Date** | Month and year |
| **Location** | Area (Charlottesville or by neighborhood) |
| **Median Sale Price** | Median sale price in thousands of dollars |
| **Median Sale Price MoM** | Percent change in median sale price from the previous month |
| **Median Sale Price YoY** | Percent change in median sale price from the previous year |
| **Homes Sold** | Number of homes sold |
| **Homes Sold MoM** | Percent change in homes sold from the previous month |
| **Homes Sold YoY** | Percent change in homes sold from the previous year |

# Analysis Plan + Justification

Gather + Clean Data | EDA | Linear Regression | Clustering

- Cleaned housing market data, on a neighborhood level and regionally, from Redfin
- Removed unnecessary columns, formatted using Datetime, and cutting dataset
- Merged the 4 datasets

- Performed EDA to identify initial visual patterns in the relationship between median housing prices and seasonality

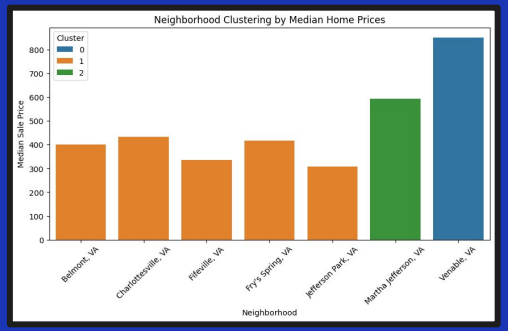


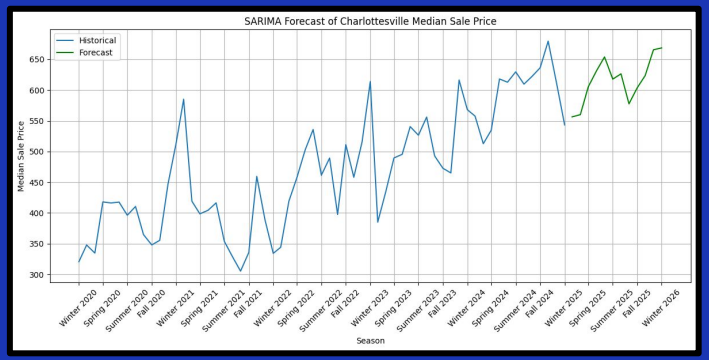

- Create a multiple linear regression model to determine if there are seasonal trends in median housing prices, based in historical data
- Use RMSE and R-squared to determine how much of the variation in housing prices can be explained by our selected features

- Originally aimed to group prices by neighborhood and season
- Question of whether or not it really gives us insight into our hypothesis
- Can help reveal potential hidden structures in the data that aren't immediately obvious through basic summary statistics

# Tricky Analysis Decision

- Initially planned to use K-mean clustering as a secondary analysis tool; however, we concluded that it was ineffective in helping answer our research question, which is more suited for a time series analysis
- Clustering does give us insight into how neighborhoods are priced similarly
- Pivoted to an SARIMA model to capture seasonality
    - Gives us insight into both current and future predicted housing prices over time
- Trial and error process to select p, d, q, and s parameters
- Peak in Spring 2025, but also Winter 2026, so further analysis needed



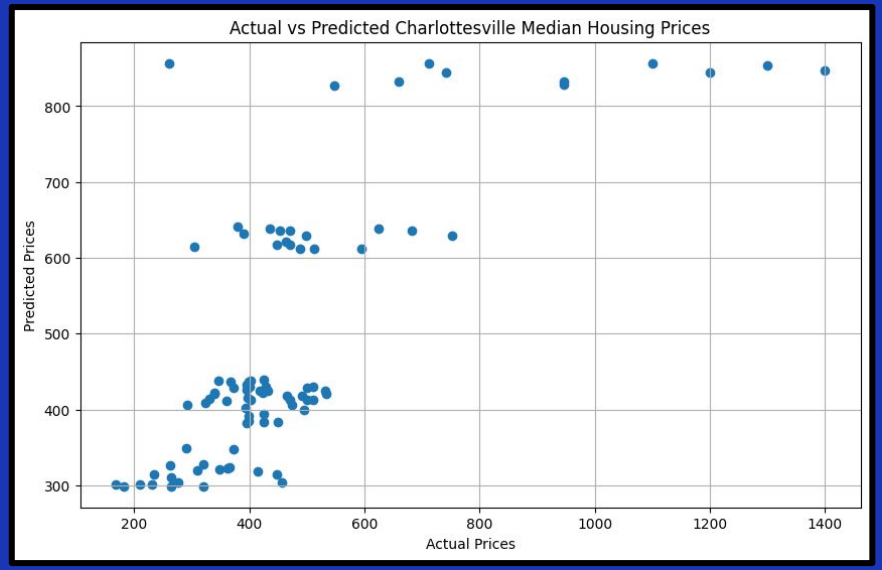



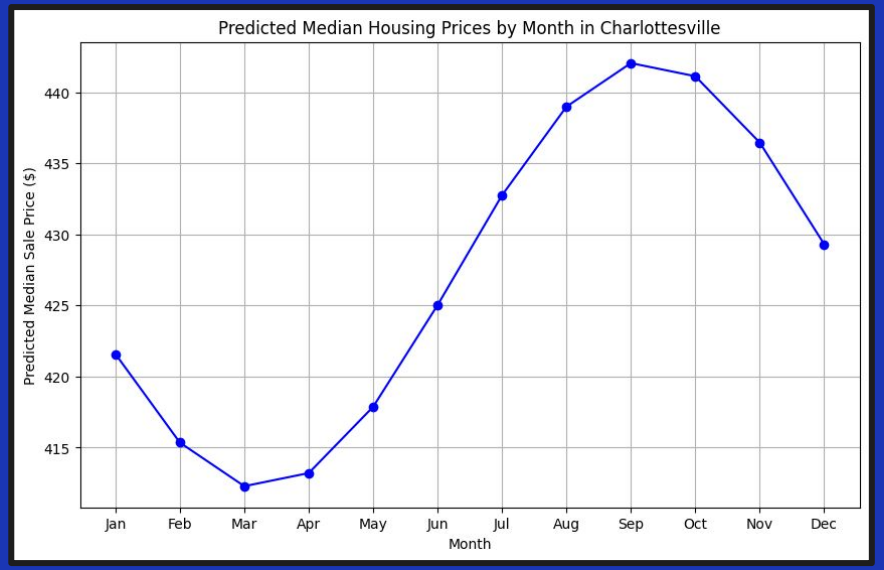
# Bias and Uncertainty

- Not all Charlottesville neighborhoods had monthly housing price data available, so our dataset may not fully capture the entire Charlottesville housing market
- There are many external factor that influence housing market prices, including interest rates, inflation, employment rates, or even COVID-19
- Because the neighborhood housing market data covered only 5 years, we decided to limit the Charlottesville regional data to start in January 2020, so our dataset and analysis excludes longer-term trends in housing prices

# Results and Conclusions

R-squared = 0.55
Root Mean Squared Error = 149.76 (in thousands of dollars)



# Next Steps

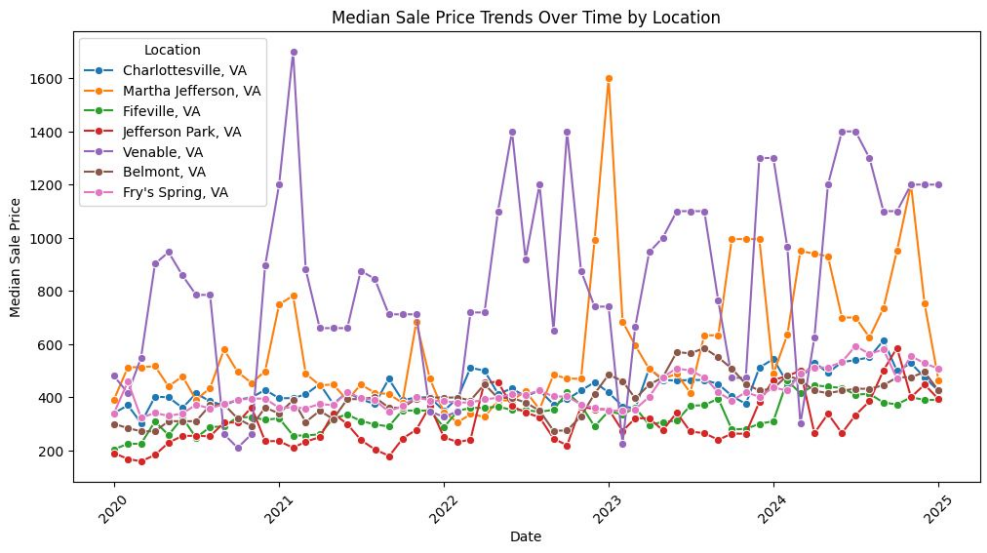
1. Expand data collection
    a. Include additional Charlottesville neighborhoods in our dataset
    b. Draw on additional factors into that influencing housing prices, such as inflation or unemployment
2. Add more features to our multiple linear regression model
    a. Features beyond just location and date impact trends in median housing prices over time
3. Explore non-linear models of analysis, such as Random Forest, which can capture seasonal effects on median housing price that may not be linear

# References

- **Github Repo:**
  - https://github.com/gffiveash/DS4002
- **Dataset:**

  - Redfin “Data Center” *Redfin* https://www.redfin.com/news/data-center/ [Accessed Feb. 27, 2025]
  - Redfin “Charlottesville, VA Housing Market” *Redfin* https://www.redfin.com/city/3867/VA/Charlottesville/housing-market [Accessed Feb. 27, 2025]
  - Redfin “Fifeville, VA Housing Market” *Redfin* https://www.redfin.com/neighborhood/761438/VA/Charlottesville/Fifeville/housing-market [Accessed Mar. 6, 2025]
  - Redfin “Jefferson Park, VA Housing Market” *Redfin* https://www.redfin.com/neighborhood/761434/VA/Charlottesville/Jefferson-Park-Avenue/housing-market [Accessed Mar. 6, 2025]
  - Redfin “Martha Jefferson, VA Housing Market” *Redfin* https://www.redfin.com/neighborhood/761425/VA/Charlottesville/Martha-Jefferson/housing-market [Accessed Mar. 6, 2025]
  - Redfin “Belmont, VA Housing Market” *Redfin* https://www.redfin.com/neighborhood/761439/VA/Charlottesville/Belmont/housing-market [Accessed Mar. 6, 2025]
  - Redfin “Venable, VA Housing Market” *Redfin* https://www.redfin.com/neighborhood/761432/VA/Charlottesville/Venable/housing-market [Accessed Mar. 6, 2025]
  - Redfin “Fry’s Springs, VA Housing Market” *Redfin* https://www.redfin.com/neighborhood/761436/VA/Charlottesville/Fry-s-Spring/housing-market [Accessed Mar. 6, 2025]
  - Geeks4Geeks “SARIMA (Seasonal Autoregressive Integrated Moving Average) *Geeks4Geeks* https://www.geeksforgeeks.org/sarima-seasonal-autoregressive-integrated-moving-average/ [Accessed Mar. 15, 2025]
  - ChatGPT (Version 4). *OpenAI* https://chatgpt.com/

# THANK YOU!


Median Sale Price Trends Over Time by Location
Location
Charlottesville, VA
Martha Jefferson, VA
Fifeville, VA
Jefferson Park, VA
Venable, VA
Belmont, VA
Fry's Spring, VA
Median Sale Price
1600
1400
1200
1000
800
600
400
200
2020
2021
2022
2023
2024
2025
Date