

The Very Best Team

Binet, Cole, Kent and Pete

Core Message

What is driving the residential real estate market?

Resources Utilized

The logo for Quandl, featuring the word "Quandl" in a stylized orange font on a dark blue background.The Wells Fargo logo, consisting of the words "WELLS" and "FARGO" in yellow, stacked vertically, on a red background.The Freddie Mac logo, featuring a stylized house icon with a blue roof and green base, followed by the text "Freddie Mac" in black.The Zillow logo, featuring a blue house icon with a white "Z" inside, followed by the word "Zillow" in blue.

Approach & Methodology

Describe the **QUESTIONS ASKED**, and ***WHY*** you asked them

- **WHY?** ... We shared a general curiosity about what factors drive real estate prices up or down,
 - We were limited in the data we could retrieve using free api's in a short period of time. Zillow led us to focus on on **QUESTIONS** about home value, sales price, inventory and days-to-sell:
 - ZALL - Zillow Home Volume Index for all homes
 - ISAM - For Sale Inventory (number of listings on the market)
 - NSAM - Median days to pending (median number of days a listing is on the market)
 - CSAM - Number of listings with price cuts
 - SSAM - Median sale price (median value for listings that sold)
- We incorporated **unemployment rate** because we **hypothesized** that with fewer people getting paychecks there would be less money being invested into real estate.
- We incorporated **mortgage rates** because we **hypothesized** that the cost of borrowing has a significant impact on what people can afford to purchase
- We incorporated **population** volume because we **hypothesized** that Colorado population growth would increase demand for housing and subsequently impact home values and sales indicators

Describe whether you were able to answer these questions to your satisfaction, and briefly summarize your findings

- We were definitely able to find satisfying answers to these questions
 - And, we agreed a lot more research could be done to push our findings even further.

Briefly **SUMMARIZE 'TOP 3' FINDINGS** with largest correlation and examples of stats that were not correlated.

- Population is highly correlated to sales pricing and home values
- Unemployment is mildly correlated to sales pricing
- Surprisingly mortgage rates are uncorrelated to sales pricing, home values, time-to-sell and for-sale inventory

Data Clean-Up and Exploration

- Learning to work with the API
 - Making calls
 - Filtering
 - Parameters
- Understanding the data (Labels, Values, what they mean)
- Getting Data
 - Mortgage Rates (Binet)
 - Population (Binet)
 - Unemployment (Pete)
 - Housing Info (Cole)
- Merge / Clean (Whole Team)
 - Deduplicating, getting Date/Times formatted correctly, etc.
 - Switching everything over to monthly values / making sure they're monthly values
 - Had to look at monthly

```
In [19]: > zillow_indicators = quandl.get_table('ZILLOW/INDICATORS', api_key=api_key, paginate=True)
zillow_indicators = zillow_indicators[["indicator_id", "indicator"]]
zillow_indicators
```

Out[19]:

	indicator_id	indicator
None		
0	ZSFH	ZHVI Single-Family Homes Time Series (\$)
1	ZCON	ZHVI Condo/Co-op Time Series (\$)
2	ZATT	ZHVI All Homes- Top Tier Time Series (\$)
3	ZALL	ZHVI All Homes (SFR, Condo/Co-op) Time Series (\$)
4	ZABT	ZHVI All Homes- Bottom Tier Time Series (\$)
5	Z5BR	ZHVI 5+ Bedroom Time Series (\$)
6	Z4BR	ZHVI 4-Bedroom Time Series (\$)
7	Z3BR	ZHVI 3-Bedroom Time Series (\$)
8	Z2BR	ZHVI 2-Bedroom Time Series (\$)

```
In [11]: > ssam_combined_df = pd.merge(ssam_df, rates_df, on="Year-Month" )
```

```
In [12]: > ssam_combined_df.head()
```

Out[12]:

	Year	Month	Year-Month	Indicator ID	Indicator Name	Region ID	Region Type	Region Name	Value	County Population	Unemployment Rate	Freddie Mac Rates	Wells Fargo Rates
0	2016	Apr	2016-04-01	SSAM	Median Sale Price (Smooth, All Homes, Monthly)	394484	metro	Colorado Springs, CO	246388.0	689481.0	3.5	3.66	3.880
1	2016	Apr	2016-04-01	SSAM	Median Sale Price (Smooth, All Homes, Monthly)	394530	metro	Denver, CO	340150.0	697744.0	3.5	3.66	3.880
2	2016	Apr	2016-04-01	SSAM	Median Sale Price (Smooth, All Homes, Monthly)	394645	metro	Greeley, CO	272475.0	294867.0	3.5	3.66	3.880
3	2016	May	2016-05-01	SSAM	Median Sale Price (Smooth, All Homes, Monthly)	394484	metro	Colorado Springs, CO	251325.0	689481.0	3.4	3.60	3.815
4	2016	May	2016-05-01	SSAM	Median Sale Price (Smooth, All Homes, Monthly)	394484	metro	Colorado Springs, CO	251325.0	689481.0	3.4	3.60	3.815

```
In [13]: > ssam_combined_df = ssam_combined_df.drop_duplicates(subset=["Year-Month", "Region Name", "Value"])
```

```
In [14]: > ssam_combined_df = ssam_combined_df.sort_values("Year-Month")
```

Median Sale Price - Analysis DF

Question: Do unemployment rates, population, and bank rates affect median sale price?

- Used a combined DF
- Grouped by Year-Month
- Last value or Averages based on metric

Analysis DF

```
unemployment_rate = ssam_combined_df.groupby("Year-Month")["Unemployment Rate"].last()
freddie_mac_rate = ssam_combined_df.groupby("Year-Month")["Freddie Mac Rates"].last()
wells_rates = ssam_combined_df.groupby("Year-Month")["Wells Fargo Rates"].last()
median_sale_price = round(ssam_combined_df.groupby("Year-Month")["Value"].mean(), 2)
population = round(ssam_combined_df.groupby("Year-Month")["County Population"].mean())
```

```
analysis_df = pd.DataFrame({"Unemployment Rate": unemployment_rate,
                           "Freddie Mac Rate": freddie_mac_rate,
                           "Mean Population": population,
                           "Avg Median Sale Price": median_sale_price})
```

```
analysis_df.head()
```

.8]:

	Unemployment Rate	Freddie Mac Rate	Mean Population	Avg Median Sale Price
Year-Month				
2016-04-01	3.5	3.66	560697.0	286337.67
2016-05-01	3.4	3.64	560697.0	292826.33
2016-06-01	3.4	3.48	560697.0	297963.33
2016-07-01	3.3	3.48	560697.0	301830.00
2016-08-01	3.3	3.43	560697.0	304584.00

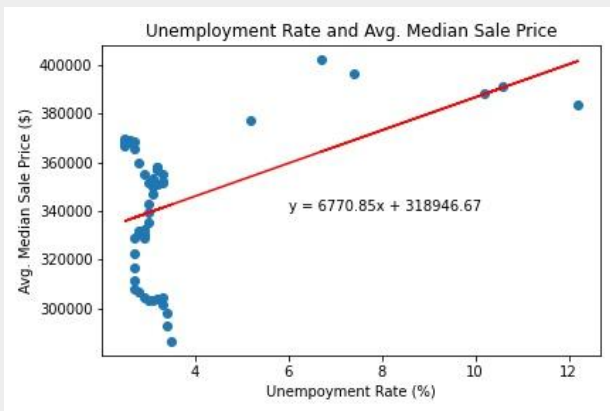
```
analysis_unemp = analysis_df.iloc[:, 0]
analysis_fmr = analysis_df.iloc[:, 1]
analysis_pop = analysis_df.iloc[:, 2]
analysis_msp = analysis_df.iloc[:, 3]
```

Median Sale Price

Question: Do unemployment rates, population, and bank rates affect avg. median sale price?

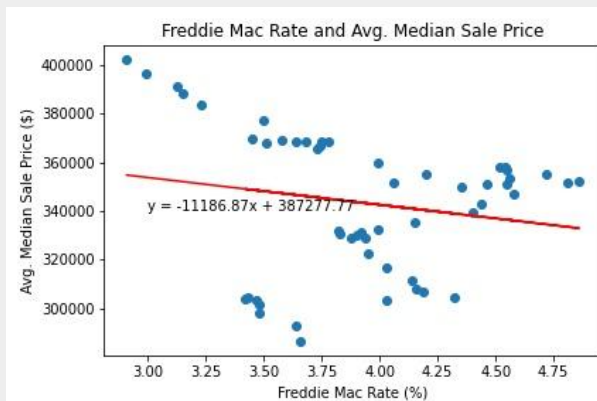
Unemployment

Correlation: 0.4756632334532904



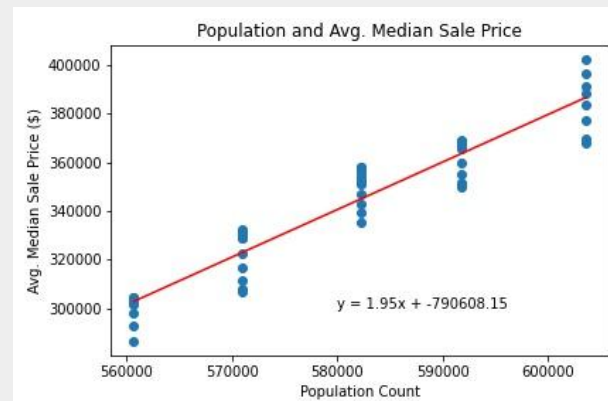
Freddie Mac

Correlation: -0.18462013286947532



Population

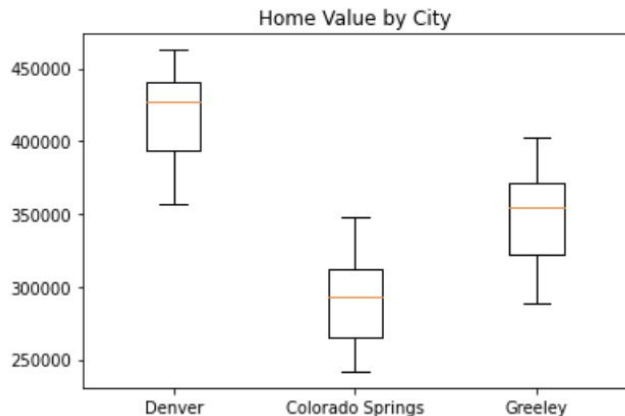
Correlation: 0.9502769873562387



Home Value Data

```
#Home Values By City
Denver_HV = zall_df.loc[(zall_df["Region Name"] == "Denver, CO"),:]
COS_HV = zall_df.loc[(zall_df["Region Name"] == "Colorado Springs, CO"),:]
Greeley_HV = zall_df.loc[(zall_df["Region Name"] == "Greeley, CO"),:]
Final_Denver = Denver_HV["Value"]
Final_COS = COS_HV["Value"]
Final_Greeley = Greeley_HV["Value"]
fig1, ax1 = plt.subplots()
ax1.boxplot([Final_Denver, Final_COS, Final_Greeley])
plt.xticks([1,2,3],['Denver', 'Colorado Springs', 'Greeley'])
ax1.set_title("Home Value by City")
```

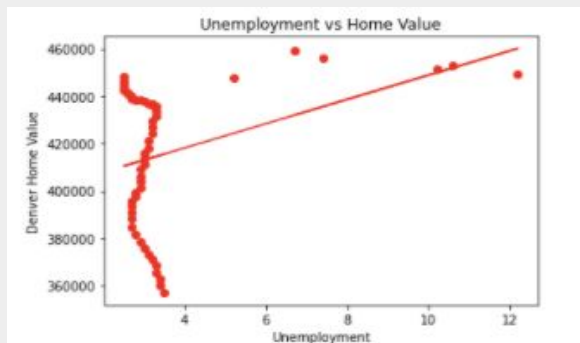
```
Text(0.5, 1.0, 'Home Value by City')
```



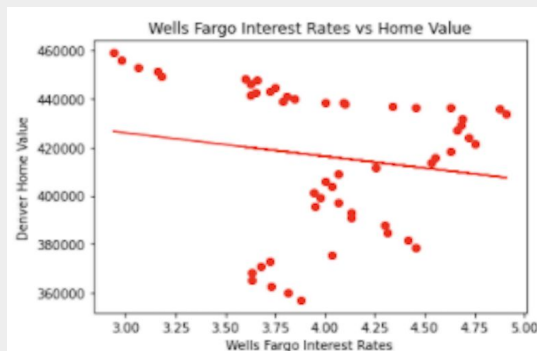
- Using .loc to sort our data
- Denver>Greeley>Colorado Springs

Home Value

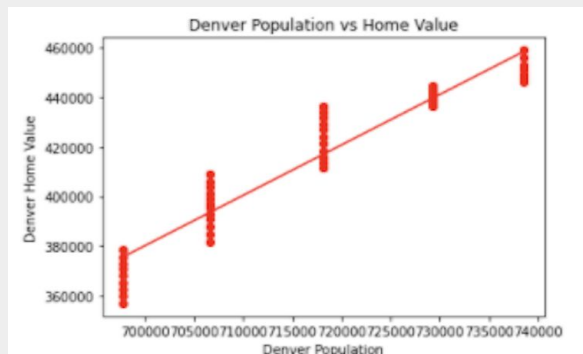
Question: Do unemployment rates, population, and bank rates affect home value?



The correlation coefficient is 0.35
The line equation is: $y = 5126.53x + 397638.26$



The correlation coefficient is -0.16
The line equation is: $y = -9683.59x + 455045.88$



The correlation coefficient is 0.96
The line equation is: $y = 2.03x + -1039561.18$

```
x_axis = Denver_df["Wells Fargo Rate"]
y_axis = Denver_df["Value"]
plt.title("Wells Fargo Interest Rates vs Home Value")
plt.xlabel("Wells Fargo Interest Rates")
plt.ylabel("Denver Home Value")
plt.scatter(x_axis, y_axis, marker="o", color="red")
(slope, intercept, rvalue, stderr) = linregress(x_axis, y_axis)
regress_values = x_axis * slope + intercept
line_eq = "y = " + str(round(slope, 2)) + "x + " + str(round(intercept, 2))
plt.plot(x_axis, regress_values, "r-")
plt.show()
print("The correlation coefficient is " + str(round(st.pearsonr(x_axis, y_axis)[0], 2)))
print("The line equation is: " + str(line_eq))
```

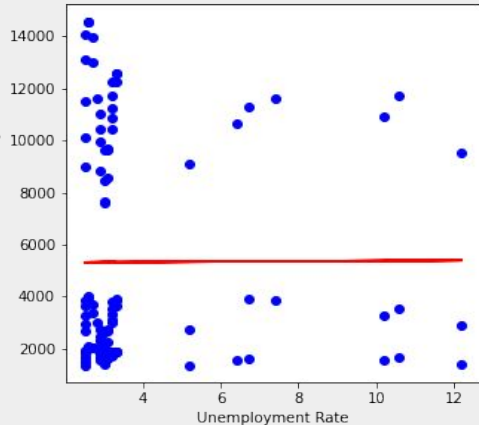
Sales Inventory

Question: Do unemployment rates, population, and bank rates affect the amount of homes available for sale on the market?

Unemployment

Correlation: 0.00

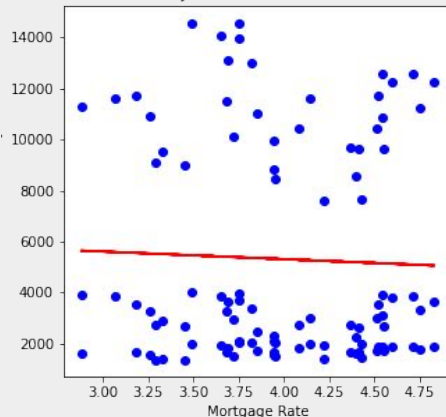
For-Sale Inventory vs. Unemployment Rate (2017-2020)



Freddie Mac

Correlation: -0.04

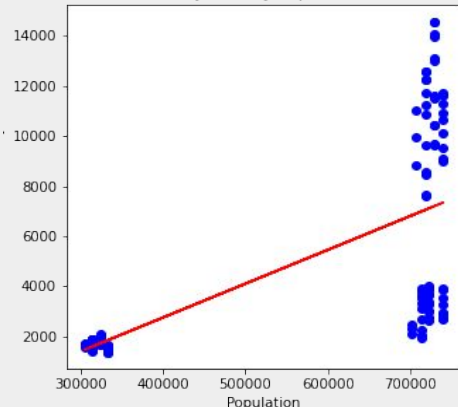
For-Sale Inventory vs. Freddie Mac Rate (2017-2020)



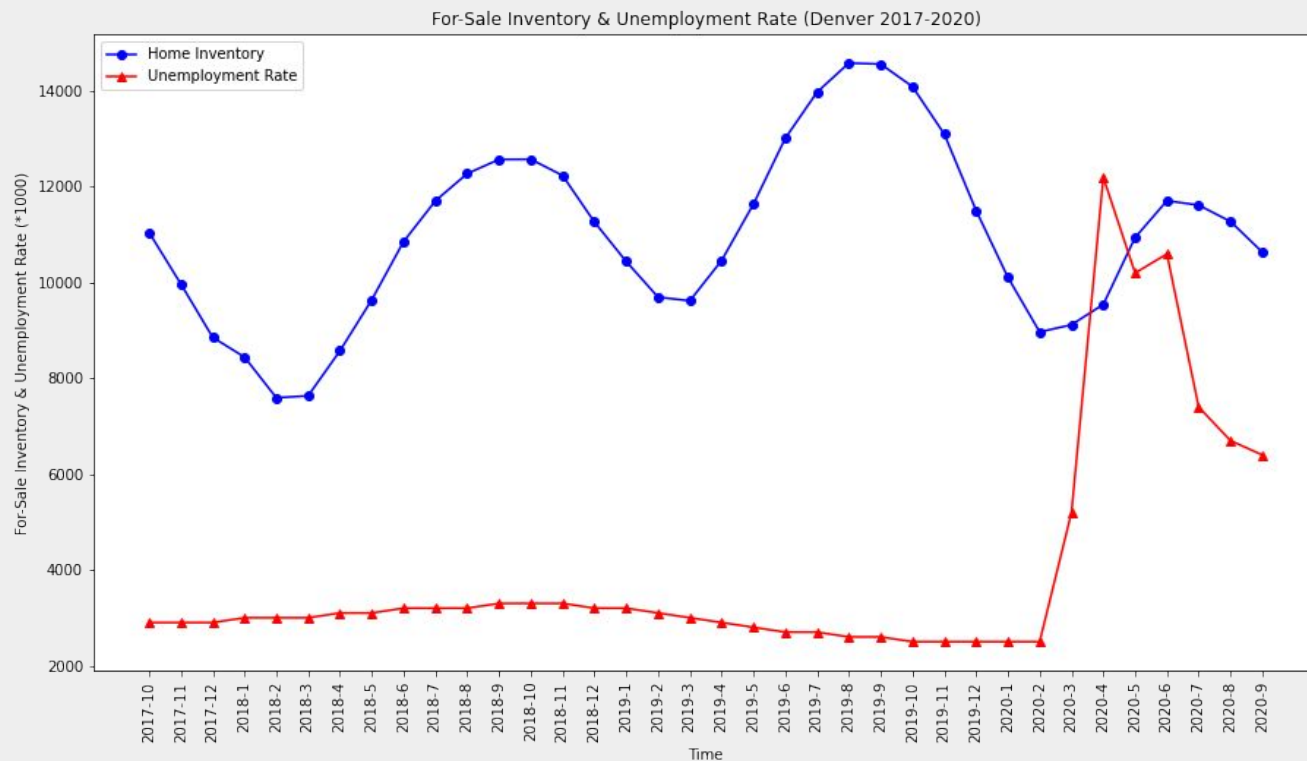
Population

Correlation: 0.6

For-Sale Inventory vs. City Population (2017-2020)



Not enough data to show anything significant

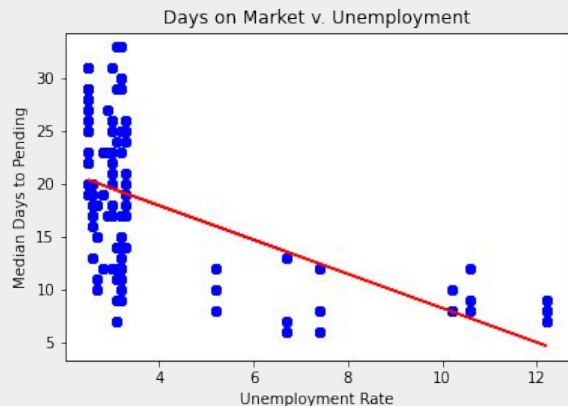


Median Days-to-Pending (Time-to-Sell)

Question: Do unemployment rates, population, and bank rates affect the time it takes to sell property?

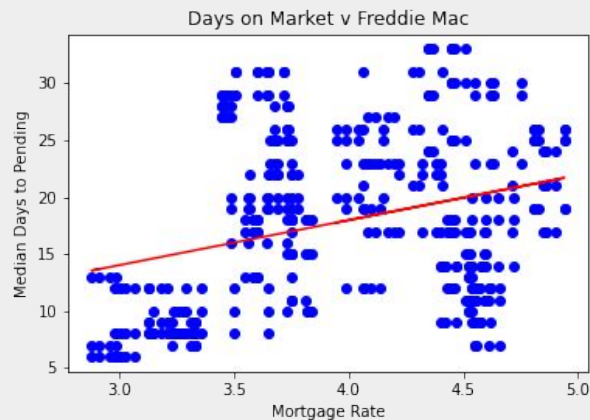
Unemployment

Correlation: -0.56



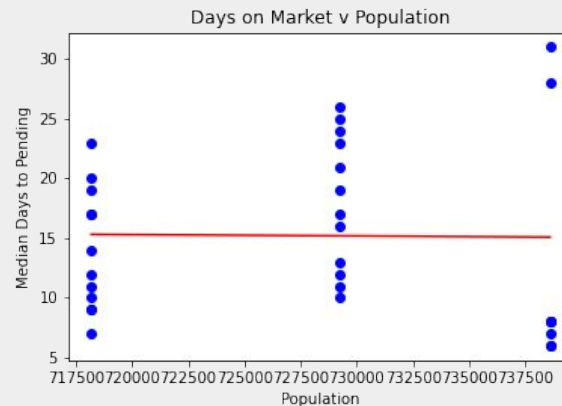
Freddie Mac

Correlation: 0.30



Population

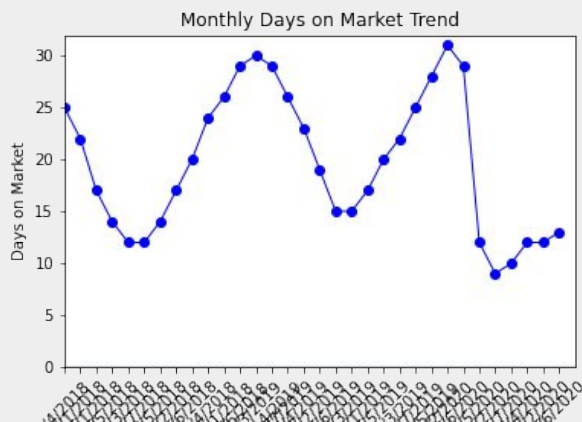
Correlation: -0.13



Additional Insight ‘Nuggets’ ...

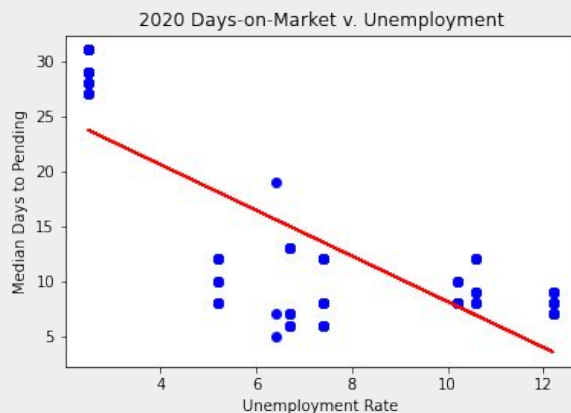
Days-to-Pending (Time-to Sell)

Seasonal, i.e. homes sell faster in the summer months



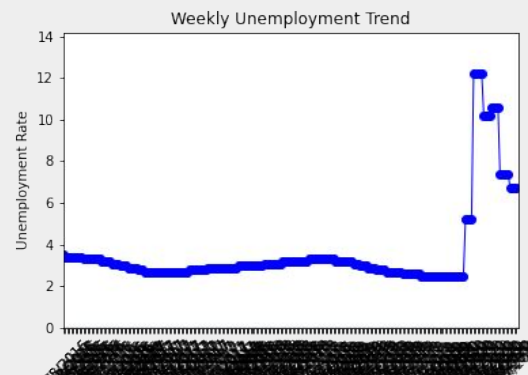
Days-to-Pending (Time-to Sell)

Counterintuitive correlation coefficient of -0.78 in 2020 v. pre-2020 (COVID?)



Colorado Unemployment

It has been a wild few months in 2020



Lessons Learned & Growth Opportunities

Data literacy and manipulation took much longer than expected

- Assessment was much easier once we had a clean and well organized master data set (data frame)
- Aligning our data with dates proved by challenging and important in our assessment. We had different values at different times, e.g. weekly mortgage rate data v. monthly unemployment and real estate data

If we had additional weeks:

- Research inflation rates with these values
- Expand from Zillow-only data
- Explore impact of government programs, e.g. stimulus during COVID
- Examine other factors impacting the housing market such as cost of building, goods/labor)



Image links

- https://upload.wikimedia.org/wikipedia/commons/thumb/3/3b/Zillow_logo19.png/250px-Zillow_logo19.png
- <https://www.quandl.com/images/head/quandl-logo-social.png>
- https://en.wikipedia.org/wiki/Bureau_of_Labor_Statistics
- http://logosvg.com/wp-content/uploads/Freddie_Mac_logo.svg
- https://logos-download.com/wp-content/uploads/2016/04/Wells_Fargo_Bank_logotype_logo_3D-700x700.jpg

Describe the questions you and your group found interesting, and what motivated you to answer them

- Question: How do unemployment rates, population, and mortgage rates affect
 - Home value?
 - Median sale price?
 - Median days to pending?
 - For-sale inventory volume?
- Motivation:
 - Getting an “A”
 - But also, examine and better understand the Colorado real estate market using third party data secured through api's