Time Series Analysis

Projections and Investor Insights

- Miguel Santana
- Flatiron School
- Data Science | FT Cohort

Introduction

- Methodology
- OSEMN Framework
- Obtain| Scrub
- Exploratory Data Analysis
- Modeling
- Interpret Results | Analyze
- Business Recommendations
- Future Work | Limitations

Methodology

- A real-estate firm is looking to invest in residential property in the state of Colorado. Analyze and provide the top 5 Zip codes respective of the Zillow research dataset.
- Framework: OSEMN
  - Observing Zip Codes with positive trending data to address market volatility.
  - Generating a fair return on investment metric using relevant time periods.
  - Validating model performance and establishing confidence in the results.

Data Science Process

LEAD

OBTAIN | SCRUB | EXPLORE | MODEL | INTERPRET

O Gather data from relevant sources

S Clean data to formats that machine understands

E Find significant patterns and trends using statistical methods

M Construct models to predict and forecast

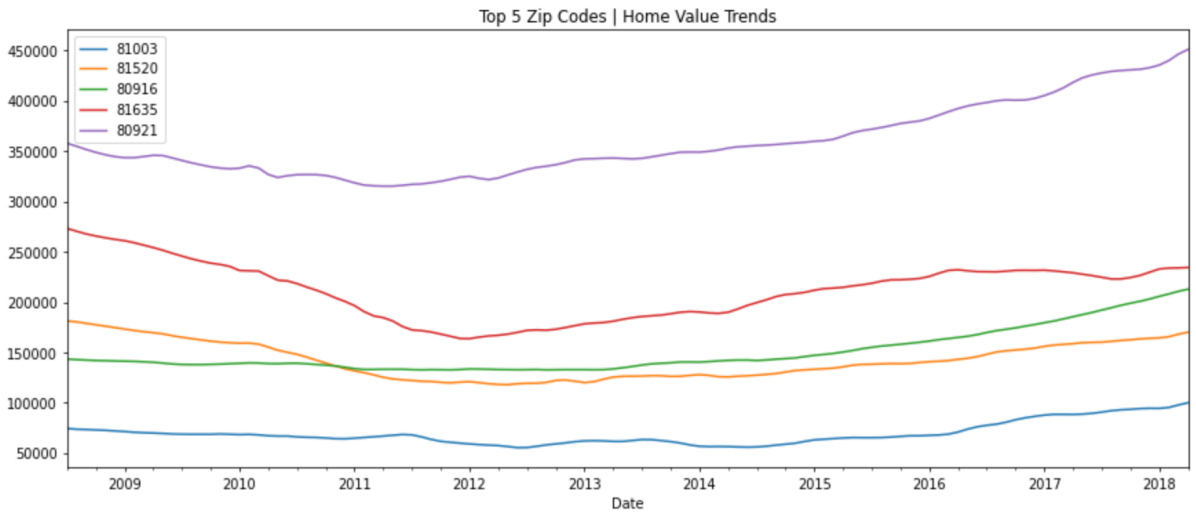
N Put the results into good use

Originally by Hilary Mason and Chris Wiggins

Obtain| Scrub

- The Zillow research dataset included:
  - 14723 Zip Codes (249 in Colorado)
  - Average sales values per Zip Code per month
  - April 1996 – April of 2018 (22 years)
- Scrubbing
  - Narrowing dataset to Colorado.
  - Transforming dataset | wide to long format
  - Dropping monthly data prior to 2009 (Housing Crisis/10 Year relevance).
  - Calculating a 2-year rolling return on investment per Zip Code.
  - Averaging the last 12 rolling ROI values and selecting the 5 largest returns.

Exploratory Data Analysis



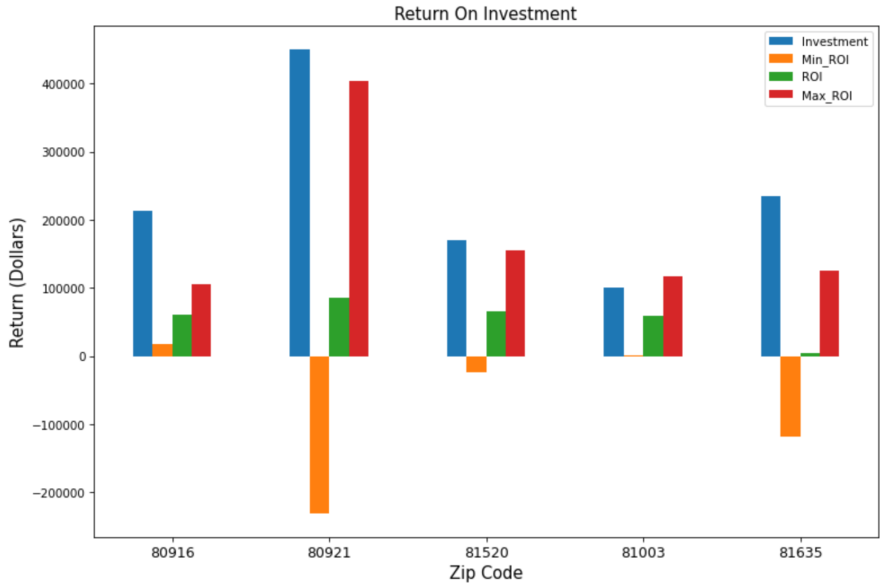
- Goal:
  - Zip Codes showing potential for profit
  - Zip Codes with relatively normal values (to increase predictive results)
- Zip Codes selected:
  - Pueblo, CO – 81003
  - Clifton, CO – 81520
  - Parachute, CO – 81635
  - Colorado Springs, CO – 80916
  - Colorado Springs, CO – 80921

Modeling

- SARIMAX Time Series Models
- Model Parameters: Autoarima grid search
- Model Validation | Error Score (RMSE) within 5-10% of the average Zip Code Value
    - Pueblo, CO 81003 | 10.5%
    - Clifton, CO 81520 | 6.3%
    - Parachute, CO 81635 | 8.9%
    - Colorado Springs, CO 80916 | 3.2%
    - Colorado Springs, CO 80921 | 7.7%

Interpret Results | Analyze

- SARIMAX Time Series Modeling Results
- Zip Code Results:



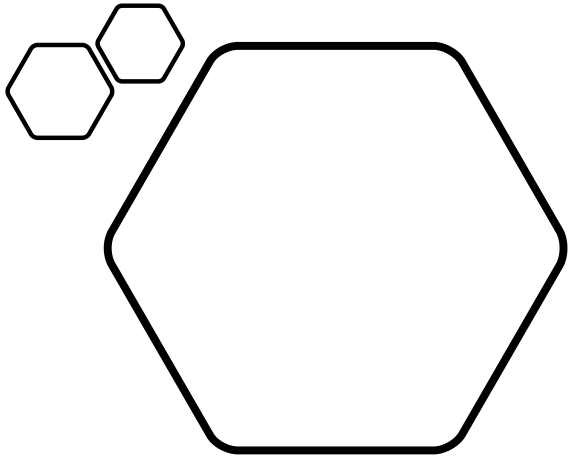
ZipCode	Investment	Min_ROI	ROI	Max_ROI
80916	212800	17888	61387	104887
80921	451000	-231939	85606	403151
81520	170100	-23260	65754	154768
81003	100000	1200	59002	116804
81635	234400	-118066	3810	125686

Business Recommendations

- Zip Codes 80916, 80921, 81520, 81003 and 81635 are the top five Zip Codes for real estate investment Colorado
  - Safest investment (low risk):
    - **Zip Code 81003 & 80916**
    - Both options offer a positive minimum, forecasted and maximum value for return on investment.
  - Most profitable investment (high risk):
    - **Colorado Springs, CO 80921**
    - Minimum: -231,939
    - Projected: 85,606
    - Maximum: 403,151

Future Work | Limitations

- Future Work
  - Future work should incorporate Zillow's additional datasets (census, economic, additional data on research page) for more insight.
- Limitations:
  - Final models were created using single ZVHI values per Zip Code with respect to the last 10 years of data.
  - Awareness of additional factors that may influence prices such as commercial development, federal poverty line data, etc. is needed in order to solidify our business recommendations.



THANK YOU!