

BGO IAA Challenge Problem

Deadline: 02/27/2023 (Monday) 8:59:59 EST

As data scientists at BGO and investors in apartment buildings, we're deeply interested in local housing markets and dynamics. Single-family housing tends to compete with apartments in the rental market and is a good proxy for the overall housing market. We'd like you to analyze several public datasets (Zillow and Census data, described below) to help shine a light on this critical business question. Python is required for this assignment.

Primary Research Questions:

1. What are the factors driving growth in housing prices over time when looking through the lens of statistical and/or machine learning?
2. What features can you engineer to improve model performance?
3. Given your findings and predictions, what markets would you suggest BGO investigate for investment over the next 5-10 years? Conversely, what markets would you suggest BGO stay clear of?
4. What other datasets would you like to have to analyze this challenge? What would be your next steps?

Secondary (bonus) Research Questions:

1. What groups (clusters) of markets tend to move together?
2. Are there significant outliers in your analysis that warrant further investigation?
3. How do your forecasts compare with Zillow's forecast? (the second Zillow dataset defined below).
4. Observe and comment on the relationship between home values and rental rates (the third Zillow dataset defined below).
5. Is zip code the appropriate level of analysis, or would you suggest modeling markets as larger clusters and/or geographic entities (e.g., MSA & CBSA)?
6. Is it possible to better understand geospatial proximity & adjacency and incorporate that into your feature space? (hint: analyze using a shapefile and geopandas <https://catalog.data.gov/dataset/tiger-line-shapefile-2019-2010-nation-u-s-2010-census-5-digit-zip-code-tabulation-area-zcta5-na>). This could help answer the previous question (super bonus).

Artifacts:

- Please submit your code in notebook form with comments.
- Please submit a PowerPoint and/or word file with your business deliverable.

Datasets:

- Zillow Home Value Index (zip code level)
 - **Zillow Home Value Index (ZHVI) measures** the typical home value and market changes across a given region and housing type. It reflects the typical value for homes in the 35th to 65th percentile range. We provide 3 Zillow datasets for the challenge. You will need to use the housing price dataset to solve the main problems and the rest two datasets for secondary problems. The detail of the dataset can be found at <https://www.zillow.com/research/data/>
 - Zillow Datasets:
 - Housing price: Zip zhvi uc sfrcondo tier 0.33 0.67 sm sa month.csv
 - Housing price forecast: Zip zhvf growth uc sfrcondo tier 0.33 0.67 sm sa month.csv
 - Rent price: Zip zori sm month.csv
- Census Dataset:
 - Select demographics: BGO_challenge_census_data.csv

You can find these datasets at <https://bgo.egnyte.com/fl/cQk4ee6dBw> and access using the password Z8ELdben6H.

Submission Steps: Please zip all your work in one .zip file (BGO_CHALLENGE_FIRSTNAME_LASTNAME.zip) and submit it directly to both emails below.

If you have any questions, please contact Chris Liedtke (Chris.Liedtke@bentallgreenoak.com) and Zhongshuai Cao (Zhongshuai.Cao@bentallgreenoak.com).