Cadre Take-home Data Assignment

October 15, 2019

Overview

This take home assignment is designed to get a feel for the candidate's approach to data analysis and modeling, as well as, provide them with an example of the sort of problem Cadre's data science team might work on. It is open ended in nature and candidates should use their own discretion regarding what is feasible given the time and data constraints.

Incomplete but best-effort submissions to this assignment are accepted. If there's a certain area that you feel you can't address in the time allotted, insight into what approaches you tried that weren't successful or would try given more time would be helpful.

Assignment

Using data to price and forecast commercial real estate is a core-focus of the data science team at Cadre. In this assignment, sample pricing series for 10 major metropolitan markets are provided. The central task is to produce price forecasts for these price series and ultimately arrive at a recommendation regarding which market we should seek to invest in.

Along with the price series, several additional open-source data sets have been provided. While the candidate is encouraged to explore and evaluate the usefulness of this data, it is not necessary to include them in any final forecasting model.

Data Overview

Candidate forecast submissions will be evaluated against a holdout set of two years of price data from 4Q 2017 to 3Q 2019, inclusive. All data provided in this assignment was available as of 3Q 2017 and the candidate should restrict themselves to only information available at this point in time when providing their final forecasts for the holdout period.

- **Cadre Price Indexes** (cpi.csv) These are internally developed pricing series for 10 major US markets. The data is produced quarterly and

populated historically through Q2 2001. The price indexes are meant to track overall commercial real estate performance blended across asset type (multifamily, office, retail, etc) and class. Index levels for all markets have been pegged to a value of approximately 100 in 2007.

- **30 Year Fixed Rate Mortgage Average** (30_Year_FRM.csv) Data from Freddie Mac tracking the average 30-year fixed mortgage rate, non-seasonally adjusted.
- **Freddie Mac Housing Price Index** (fmhpi.csv) Data from Freddie Mac that tracks single family house prices. This data is produced monthly and is available at the market-level.
- **Zillow Median Income** (zillow_mi_market.csv) Data released from zillow's research group that tracks the median income level by market. The data is released quarterly with an approximate 1-Quarter lag.
- **Zillow Home Value Index** (zillow_hi_market.csv, zillow_hi_zip.csv) Data released from zillow's research group that tracks the median monthly value of single family homes. The data is available at both the zip code and market level.
- **Zillow Rent Value Index** (zillow_ri_market.csv, zillow_ri_zip.csv) Data released from zillow's research group that tracks the median monthly rental prices of single family residences. This data is provided at both the market and the zip code level.
- **ACS population data** (market_pop.csv, zipcode_pop.csv) Data collected by the US census bureau measuring population at both the market and zip code level. The market data is presented as a set of yearly time series, while the zip code data is a single year snapshot from 2015.
- **zip to cbsa crosswalk** (zip_to_market_corr.csv) Geographic correspondence table between zipcodes and markets.
- **market to name pairings** (market_to_name.csv) Mapping from market codes to string names.

What To Submit

1. A novel or interesting insight into the data with supporting visualization(s). The data science team delivers value to Cadre by helping to inform investment decisions. Part of this involves finding and communicating interesting market "data-narratives" to diverse audiences. Please prepare a small write up and executable code that surfaces one such data-narrative of your choosing.
2. A short write-up detailing the forecasting method employed, the rationale for choosing it, and the approach taken to validating the model.

3. A csv file containing forecasts for all 10 of the price series for each quarter between 4Q 2017 to 3Q 2019, inclusive. Please, ensure that this csv file is formatted similarly to cpi.csv. A measure of mean absolute percentage error will be used to evaluate candidate forecasts against the true, held-out, index values.

The submission format for items 1 and 2 can be stand-alone scripts and a written document, or a well organized notebook.