

Business Scenario & Question

Our analytics team attempts to answer numerous questions in the real estate environment for a hypothetical contracting & construction company. We seek to provide insight into supplyside impacts on housing markets, the availability of loans, cost and profitability of housing, and population shifts and demand for housing. Specifically, our defined business questions are as follows:

Question 1 (Michael) – We are looking to find insights for our client on what the future market environment looks like within borrowing costs for undergoing projects, and if commodity price fluctuations are correlated with those of loans provided by commercial banks. From the Federal Reserve of Economic Data (FRED), I will be using PPI by commodity time series data within lumber, copper, steel, and diesel, as well as construction and land development loans from commercial banks. The models that are most optimal for this analysis are Random Forest Regression for predicting loans, and LSTM to forecast the loans. Our client is also interested in generating rental income by acquiring properties in Connecticut (CT), thus, we will do exploratory analysis on a historical CRE sales dataset from Data.gov for the year 2020 within a multitude of CT cities. This analysis will help the client be more strategic in supply chain management by effectively utilizing resources for undergoing specific projects in CT.

Question 2 (Ivan) – In hopes of identifying viable real estate markets, we are looking to determine the effects of the laws and regulations on short-term rentals, such as Airbnb on the housing market. Is there a causal relationship between airbnb existing in a city before and after legislation and the prices of housing? Is a city that allows short-term rentals more profitable vs a city that bans short-term rentals. In 2016 the city of Austin, Texas voted in a city ordinance requiring all short-term rentals under 30-days to obtain licensing in order to operate and limited the number of short-term rentals in the area. In order to accomplish this we will evaluate a difference in differences model examining the All-Transactions House Price Index for Austin (AHPUS) and additional Austin housing market data from Federal Reserve Economic Data St. Louis Fed and historical Austin Airbnb ('room_type', 'price', 'minimum_nights', 'number_of_reviews', 'license') data from Inside Airbnb.com.

Question 3 (Matt) – Using data from data.ca.gov, Federal Reserve of Economic Data, and USA.com, I will answer the question, Is there a correlation between population density and the choice of housing construction, such as conventional houses or apartments? I will use a random forest predictive method to analyze my question and try other boosting strategies along the way. The independent variables I will use are the Population Density and Population data from USA.com, along with the CNTY_NAME to predict the UNIT_CAT variable. UNIT_CAT is a unit category or what type of building would be constructed, such as “Single-Family Unit,” “2-, 3-, and 4-Plex Unit”, or “5 or More Units Per Structure”. This would be utilized in hope to bring value to a business by predicting where to invest in different construction/housing types.

Question 4 (Alexis) – In particular, the business question I wish to explore is as follows: Over time, how does economic and demographic data, such as loan limits, real interest rates, and real income, impact residential populations in major metropolitan areas within the United States?

Utilizing the United States Census Bureau, the Federal Reserve of Economic Data, and the Federal Housing Finance Agency, data on single household conforming loan limits, real interest rates, the number of households in a region, and real mean and median income has been synthesized into a single dataset for the years 2000-2022 and for the top 50 zip codes in the U.S. Through the use of time series regression and geospatial modeling, we seek to gain insight into housing population fluctuations as a result of the aforementioned features, and ultimately, we will use these methods to predict housing/residential populations (# of households) in future years. It may also be pertinent to perform anomaly detection on this data in order to find unusual patterns, events, or outliers across the past two decades. This knowledge adds firm value by allowing us to predict which zip codes will likely experience population growth or reduction based on given economic conditions, and therefore allow contractors to more effectively and efficiently decide where they should be locating upcoming construction ventures.

Ultimately, our team is trying to determine the influence of the supply chain on housing materials, the types of projects to undertake, and the prime locations for these construction projects.
