

Title: 10 Real Estate Data Projects

Introductory Hook: Interested in a data-driven, real estate career? Want to discover patterns in the housing market? Here are 10 wonderful projects to get you started.

Body 1: Housing Price Visualization

Description: Housing prices have always been a major point of interest when gauging the state of the economy. Considering that an uncontrolled ramp-up in housing prices can lead to significant economic ramifications, being able to analyze housing price data would help you to anticipate a subsequent economic downturn.

For this project, consider making a visualization that displays median US housing prices over time. You could localize this by charting the median home price in a particular region or the US as a whole. If you'd prefer more of a challenge, consider making your visualization interactive. Studies have shown that interactive visualizations are more effective at attracting and engaging users; additionally, users tend to retain information better when presented via interactive mediums as opposed to static ones.

Reliable data on median US home values can be found at [census.gov](https://www.census.gov). Popular tools for creating interactive visualizations include D3.js, Vega, Matplotlib, and Tableau. If you're naturally an artist, you could also hand-draw your visualizations so long as the information being represented is truthful.

Body 2: Demographics Map

Description: To understand a society, it is important to have knowledge about the different characteristics of a population. This project gives you a chance to learn more about a given population as well as how to present data on a map.

D3.js is a popular library for creating dynamic maps especially when considering that there already exists an abundance of open-source map templates for you to choose from. While D3 does have a pretty steep learning curve, if you want a library that can create some truly stunning visualizations, the payoff is well worth it.

To start, find a template that displays the area whose demographic you're trying to convey. Then, find data that describes the population of that area and connect it to your template. If you're in need of ideas, consider data such as age, race, education level, occupation, criminal history, etc.

Body 3: Quality of Life Analyzer

Description: Sociologists have found that location has a strong impact on a person's quality of life and anecdotally, many people agree. Factors that are inherently tied to one's location such as school quality, water quality, access to food and healthcare, drug usage, crime rates, and job

availability are so effective at predicting a person's life trajectory that they are even utilized by think tanks and law enforcement agencies.

This project will involve aggregating a large range of data and outputting a quantitative measurement that rates an individual's quality of life. In searching and obtaining this data, you will have the opportunity to create and manage a large data lake which has applications in machine learning and on-premise data management.

Body 4: Zillow Scraper

Description: Zillow is the world's leading online real estate and rental marketplace. As such, it has a host of data that can be pulled using a web scraper or through its very own API. For those who have a penchant for backend and data engineering, this project is sure to stir your interest.

Python would be the recommended language of choice for this project due the powerful libraries it has access to such as Pandas and BeautifulSoup, both of which have excellent web scraping capabilities.

Once you've scraped and parsed your desired data, you can use SQLAlchemy, a Python database toolkit, to input that data into a relational database and perform various CRUD operations on it which will allow you to discover new insights in addition to improving your SQL skills.

Body 5: Future Home Value Estimator

Description: While people often erroneously assume that their homes will appreciate in value over time, the truth is that it is only the land on which the home is built that appreciates in value, not the home itself. Regardless, let us remain in accordance with this synecdoche and build a home value estimator that will estimate the future value of a specific home.

Before making any future projections, it is imperative that the present value of the home has already been determined. One way to obtain this is to simply use a valuation that has been provided by an inspector or a third-party source such as Zillow or Redfin. If you feel that the current estimated value is inaccurate, come up with your own sum-of-the-parts valuation (or whatever valuation method you prefer) and use that as the correct present value. From then, input your assumptions on the annual return that you expect the home to deliver and the number of years you wish to forecast.

Body 6: Automated Home Value Forecaster

Description: Piggybacking off the previous project, instead of letting the user make their own assumptions which may be biased or inaccurate, use statistical methods to project the value of a home over time. This will require a lot more research as data-driven prediction algorithms are fairly mathematically complex. There are numerous online resources covering machine learning

prediction models, particularly ones that pertain to home values, and it is recommended that you take the time to study these resources.

Body 7: Homeowners Insurance Underwriting Tool

Description: Home insurance coverage is critical for protecting your home from enormous financial losses and in the event of a total loss, it can provide you a basis to rebuild on.

Moreover, banks also require it as a condition for your mortgage. In this project, you will gain an insight into the underwriting processes that insurers use to analyze your profile, determine your level of risk, and calculate an appropriate premium.

Begin by building a tool that will assess a person's risk profile based on a variety of data inputs. Consider a list of data such as but not limited to: age, race, marital status, credit history, criminal history, education level, occupation, income, location, etc. From there, you can corroborate this data with additional sources to make deeper inferences about the individual such as the proportion of people of his/her age that have been involved in home accidents, the life expectancy and habits of the individual, whether or not the individual has people who resent them strongly, the age and quality of the home, and so on. After you have finished building the risk profile, calculate an appropriate premium that reconciles the individual's level of risk with the insurer's ability to generate a profit; for reference, take a look at company balance sheets or online home insurance quotes.

Body 8: Hazardous Exposure Map

Description: Hazardous materials, if exposed to the public, can have disastrous consequences on public health and safety. Thus, it is crucial that there exist effective warning systems that inform people of whether or not their home or a home that they're looking to purchase/rent has been exposed to hazards.

From a high level, this project consists of a color-coded map that displays areas where hazardous materials have been found and the range that the hazards can affect. The colors are used to represent a variation in data i.e. different colors can represent different hazards (e.g. asbestos, toxic fumes, radiation) or the level of exposure that an area has received (e.g. dangerous - evacuate ASAP, unsafe - stay indoors, mild irritant - no need to panic, just wear a mask and wash your hands). Data can be sourced from local news, government websites, text alerts, word of mouth, and even observations from daily life (e.g. flyers on utility poles, bright orange cones, "do not cross" signs, etc.)

Body 9: Home Renovation Scouter

Description: In recent years, the increase in popularity of property investing has been beneficial for all parties involved. Investors and sellers have a chance to make tremendous financial gains while data enthusiasts are provided with plenty of data to collect and analyze. This project aims

to use housing data to reveal undervalued, neglected, and/or beaten-down homes that the marketplace has either ignored or passed up on and is intended for entrepreneurs who are looking to purchase these properties as fixer-uppers.

One approach would be to once again utilize a map but a vertical display that displays each house as a row entry works too. For each home, display its current value as well as its potential value; the larger the disparity between the home's current value and potential value, the larger the return and the more enticing it is for an investor.

Body 10: Natural Disaster Vulnerability Report

Description: Not only are natural disasters detrimental to human life, they also displace and disrupt large swarms of people which subsequently impacts the value and habitability of real estate. For this project, you will build a calculator that uses data to calculate a home's vulnerability to natural disasters.

Use data inputs such as weather patterns, altitude, nearby environmental risks (e.g. volcanoes, oceans), fault lines, and subsidence to build a risk profile for your desired home/area. You can add complexity to this model by assigning weights to different factors depending on their lethality and likelihood. Information can be sourced from GIS applications, government agency websites, and informational documents from city planners, civil engineers, and architects. Once your profile is complete, implement a calculator that can output a score which represents how vulnerable a house is to natural disasters; if you wish, you could customize this calculator by allowing the user to filter out different natural disaster options.