Analysis of Population Age and Residential Rental Prices

Dustin M. Kraft

Western Governors University

**Table of Contents**

[A. Project Overview 3](#_Toc58328659)

[A1. Research Question or Organizational Need 3](#_Toc58328660)

[A2. Context and Background 3](#_Toc58328661)

[A3. Summary of Published Works 3](#_Toc58328662)

[A3a. Relation of Published Works to Project 4](#_Toc58328663)

[A4. Summary of Data Analytics Solution 4](#_Toc58328664)

[A5. Benefit to Organization and Decision-Making Process 5](#_Toc58328665)

[B. Data Analytics Plan 5](#_Toc58328666)

[B1. Goals, Objectives, and Deliverables 5](#_Toc58328667)

[B2. Scope of Project 5](#_Toc58328668)

[B3. Standard Methodology 6](#_Toc58328669)

[B4. Timeline and Milestones 7](#_Toc58328670)

[B5. Resources and Costs 7](#_Toc58328671)

[B6. Criteria for Success 7](#_Toc58328672)

[C. Design of Data Analytics Solution 8](#_Toc58328673)

[C1. Hypothesis 8](#_Toc58328674)

[C2. Analytical Method 8](#_Toc58328675)

[C2a. Justification of Analytical Method 8](#_Toc58328676)

[C3. Tools and Environments of Solution 8](#_Toc58328677)

[C4. Methods and Metrics to Evaluate Statistical Significance 8](#_Toc58328678)

[C4a. Justification Of Methods and Metrics 9](#_Toc58328679)

[C5. Practical Significance 9](#_Toc58328680)

[C6. Visual Communication 9](#_Toc58328681)

[D. Description of Datasets 10](#_Toc58328682)

[D1. Source of Data 10](#_Toc58328683)

[D2. Appropriateness of Dataset 10](#_Toc58328684)

[D3. Data Collection Methods 11](#_Toc58328685)

[D4. Data Quality 11](#_Toc58328686)

[D5. Data Governance, Privacy and Security, Ethical, Legal, and Regulatory Compliance 11](#_Toc58328687)

[D5a. Precautions 11](#_Toc58328688)

[E. Sources 11](#_Toc58328689)

# Project Overview

# A1. Research Question or Organizational Need

# This project will analyze the relevance of population age in market assessments for the real estate investment industry.

# A2. Context and Background

Real estate investors use various metrics to identify geographic areas in which to purchase free-cash-flowing rental properties. If available demographic data can be used to pinpoint an additional metric to improve odds of profitability, it may provide a competitive edge in the rental market.

# A3. Summary of Published Works

Over the past two years, investors with large gains in the stock market, are turning to real estate assets to diversify portfolios. Both large firms and individual investors have used record low interest rates to purchase properties in desirable rental markets across the country. According to an analysis by Redfin, “Real estate investors bought a record 18.4% of the homes that were sold in the U.S. during the fourth quarter of 2021, up from 12.6% a year earlier and a revised rate of 17.4% in the third quarter (Anderson, 2022).” Because of this uptick, the real estate investment market has become increasingly competitive, causing investors to look outside of their own city or state to find new investments.

To increase the likelihood of a profitable investment, potential rent needs to be high enough provide cash flow after covering expenses. A common industry practice is to perform a rental market analysis of new markets to judge the potential for a successful investment. The investor needs to examine various metrics for evidence that they will be able to charge more for rent than the total expenses to own and maintain a property within a new market. In an article for Investopedia, Michele Lerner discusses one such metric known as the “price-to-rent ratio”:

“The price-to-rent ratio is a calculation that compares median home prices and median rents in a particular market. Simply divide the median house price by the median annual rent to generate a ratio. As a general rule of thumb, consumers should consider buying when the ratio is under 15 and rent when it is above 20. Markets with a high price/rent ratio usually do not offer as good an investment opportunity (Lerner, 2022).”

The high ratio situation that Lerner warns investors about happens when median rent is low, thereby creating a ratio with a smaller denominator by which to divide the median home price. Said another way, this is a situation where investors will likely not receive enough rent to cover costs and earn a return on their investment.

High demand is required to drive up rents. Areas with a lower price-to-rent ratio have a higher demand for rental units and a lower demand for home ownership. Because younger residents are more likely to rent, this population may be an indicator of this demand. A study by Pew Research shows that residents of age thirty-five and below are far more likely to rent than own their home. Pew’s discussion of the findings includes the following:

“Younger people – those below the age of 35 – are far more likely to rent than are other age groups: About two-thirds (65.9%) of this age group lives in rentals. This compares with, for example, 42% of those ages 35 to 44, and less than a third (31.5%) of 45- to 54-year-olds (DeSilver, 2021).”

# A3a. Relation of Published Works to Project

In response to the increase in real estate investors found by Redfin (Anderson, 2022), the company has a need for a competitive advantage, this analysis examines whether more residents in the group that is more likely to rent, according to Pew Research (DeSilver, 2021), correlates with increased median rents, creating a market with a better rent-to-purchase ratio. In her article for Investopedia, Michele Lerner explains that this is an indicator for greater investment potential ().

# A4. Summary of Data Analytics Solution

The proposed solution is to perform a correlation and regression analysis on data regarding population demographics and rent by Zip Code Tabulation Area (ZCTA) to determine if there is a positive correlation between the size of the target age group and higher median price-to-rent ratio. Additionally, the strength of that relationship will be measured to determine if it is useful to predict higher rents within a real estate market.

# A5. Benefit to Organization and Decision-Making Process

Under the direction of the Regional Development Manager as key stakeholder, the proposed project will determine if market analysts may incorporate this research and associated dataset into analysis of future markets to increase odds of profitability or if the studied relationship and dataset can be ruled out as unbeneficial.

# Data Analytics Plan

# B1. Goals, Objectives, and Deliverables

* Goal: perform an analysis of Census data regarding population and contract rents.
  + Objective: Collect, clean and aggregate national census data on population demographics and rental real estate.
    - Deliverable: Dataset on US data compiled by Census zip code tabulation area.
  + Objective: Determine if there is a positive relationship between percent of residents in the 20 to 34 age group and residential rent rates.
    - Deliverable: Hypothesis test performed on linear regression model
  + Objective: Determine if percent of residents between the ages of 20 and 34 can be used to predict residential rent rates.
    - Deliverable: Statistical evaluation of the strength of the relationship between the target demographic and residential rent rates.

# B2. Scope of Project

The scope of this project includes the determination of whether or not the gathered dataset suggests a relationship between the target variables can be used in an initial assessment on a real estate market. This is only one element of a full-scale market analysis and, even if a correlation is found, cannot fully determine profitability. Additionally, this analysis will not examine individual properties or types of properties available in a ZCTA.

# B3. Standard Methodology

The project management methodology to be used is Knowledge Discovery in Database (KDD), with the following steps:

1. Goal-Setting and Application Understanding
   * Identify a question and its importance to the company. Form a hypothesis to be tested.
2. Data Selection and Integration
   * Use data.census.gov to select and download tables containing population age, median rent, and median income broken down by ZTCAs.
3. Data Cleaning and Preprocessing.
   * Remove unnecessary columns and add Null values to fields with missing or bad data.
4. Data Transformation
   * Combine tables into one dataframe, remove empty records, limit the dataset to records below the threshold for margin of error and calculate the values for percentage of population in the target age group and median price-to-rent ratio.
5. Data Mining
   * Perform regression analysis on the explanatory and response variables.
6. Pattern Evaluation/Interpretation
   * Create scatterplot and plot of regression line.
   * Calculate relevant statistics for hypothesis test.
7. Knowledge Discovery and Use
   * Evaluate target metrics and complete test of the hypothesis.

# B4. Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone** | **Projected Start Date** | **Projected End Date** | **Duration (days/hours)** |
| Establish requirements | Complete, see section B. Data Analytics Plan | | 3 hours |
| Locate and download appropriate data tables | 9/15/2022 | 9/15/2022 | 2 hours |
| Clean data | 9/15/2022 | 9/15/2022 | 2 hours |
| Perform analysis | 9/16/2022 | 9/16/2022 | 2 hours |
| Generate analysis results document | 9/16/2022 | 9/17/2022 | 3 hours |

# B5. Resources and Costs

|  |  |
| --- | --- |
| **Resource (Personnel, Technology, or Infrastructure)** | **Associated Cost** |
| American Community Survey data | $0.00 |
| Python environment, Excel | $0.00 |
| Computer hardware | $0.00 |
| Labor (estimated 12 hours @ $110/hr) | $1,320.00 |
| **Total estimated project cost:** | **$1,320.00** |

The data to be collected is free to use without attribution. Existing assets, such as the software/development environment and needed hardware, will be used. No additional purchases are required. Billable hours will be invoiced at a data analyst rate of $110 per hour.

# B6. Criteria for Success

Python will be used to combine the data into a usable dataset and build a linear regression model to calculate the associated slope of the regression line, p-value, and r-value. The slope of the regression line (regression coefficient) will determine if there is a positive relationship. The r-value (correlation coefficient) will determine the strength of that relationship and the p-value will be used as the measure of significance compared to a 95% confidence interval.

|  |  |  |
| --- | --- | --- |
| **Criterion/Metric** | **Required Data** | **Cut Score for Success** |
| Regression Coefficient | Slope of the regression line | >0 |
| Measure of Significance | p-value | <0.05 |
| Correlation Coefficient | r-value | >0.5 |

# Design of Data Analytics Solution

# C1. Hypothesis

# A higher percentage of adult residents between the ages of 20 and 34 will correlate with higher price-to-rent ratio in that area.

# C2. Analytical Method

I will use descriptive statistical methods to explore and explain the dataset. I will then perform a hypothesis test on a linear regression analysis to determine if there is a relationship between the chosen variables.

# C2a. Justification of Analytical Method

I will be using linear regression because both the explanatory and response variables are continuous and the question is relational in nature.

# C3. Tools and Environments of Solution

The raw data will be downloaded in .csv format using the U.S. Census Bureau’s robust data platform at data.census.gov. The Census Bureau conducts the American Community Survey on an ongoing basis and is the most complete source of data on U.S. communities available. Because of the manageable size of the .csv data, it will be most efficient to clean in Excel before importing and combining in Python.

# C4. Methods and Metrics to Evaluate Statistical Significance

The explanatory variable (x) will be the percentage of the population between 20 and 34. The response variable (y) will be the median price-to-rent ratio. I will pose a hypothesis stating that a higher percentage of adult residents between the ages of 20 and 34 shows no statistically significant relationship with a higher price-to-rent ratio in the same area. I will then attempt to disprove that hypothesis, thereby proving that a higher percentage of residents in the target demographic shows a statistically significant relationship with higher price-to-rent ratios.

The values that will be used in this test will be the slope of the regression line, the p-value, and a 95% confidence interval. I will also examine the correlation coefficient of the regression to determine if this model can be used to predict rent rates within a new market.

# C4a. Justification of Methods and Metrics

The percentage of the population in the target age group will be used, rather than the estimated total, to equalize between varying population sizes. A ratio of median rent to median home value within the same ZCTA will be used because incomes and housing costs can vary greatly even within a few miles of each other. The ratio will give an indication of the demand for rental housing versus home ownership. A linear regression model will determine if there is a positive relationship between the two variables that is significant beyond a commonly accepted 95% confidence interval.

# C5. Practical Significance

Finding investment properties with high rent potential is crucial to a successful investment. There are over 40,000 zip codes in the United States requiring many man-hours to evaluate individually. If a strong correlation is found in this analysis, an analyst could use a free dataset to target only markets with a percentage of young residents above a certain threshold, before performing further analysis. This strategy can reduce labor costs while improving the chances of profitability.

# C6. Visual Communication

A scatterplot of the datapoints, with the regression line overlayed, will be used as it is the standard method of representing a linear regression model and will best illustrate the relationship as well as the distribution of the data. Other visuals may be incorporated as needed.

# Description of Datasets

# D1. Source of Data

The source of the data is the American Community Survey, performed by the U.S. Census Bureau and available for download through data.census.gov.

Bureau, U. S. C. (2020). *American Community Survey B25058 MEDIAN CONTRACT RENT (DOLLARS)*. Explore census data. Retrieved June 28, 2022, from https://data.census.gov/cedsci/table?q=B25058%3A+MEDIAN+CONTRACT+RENT+%28DOLLARS%29&g=0100000US%248600000&tid=ACSDT5Y2020.B25058

Bureau, U. S. C. (2020). *American Community Survey S1903 MEDIAN INCOME IN THE PAST 12 MONTHS (IN 2020 INFLATION-ADJUSTED DOLLARS)*. Explore census data. Retrieved June 28, 2022, from https://data.census.gov/cedsci/table?q=median+income&g=0100000US%248600000&tid=ACSST5Y2020.S1903

Bureau, U. S. C. (2020). *American Community Survey S2502 DEMOGRAPHIC CHARACTERISTICS FOR OCCUPIED HOUSING UNITS*. Explore census data. Retrieved June 28, 2022, from https://data.census.gov/cedsci/table?t=Owner%2FRenter+%28Householder%29+Characteristics&g=0100000US%248600000&tid=ACSST5Y2020.S2502

# D2. Appropriateness of Dataset

The American Community Survey is compiled on an ongoing basis and is the most complete source of data on United States communities available. It is conducted by the U.S. Census Bureau and is used to assist local policymakers, business leaders, and others to make informed decisions in their location (Bureau, acs\_general\_handbook\_2020\_ch01.pdf 2020). The features chosen from this dataset are the most relevant to answering the identified question, while accounting for differences in population size and cost of living.

# D3. Data Collection Methods

The data collected is part of the American Communities Survey, which has an annual sample size of about 3.5 million homes. The survey contains estimates of data that have been collected over a period of time rather than for a single point. Multiyear estimates are generated for each area by combining data collected over 5-year periods. The most recent dataset available is from the period between 2015 and 2020.

# D4. Data Quality

Because ACS samples are collected over a period of time, they include an associated sampling error for each estimate. This error will have to be taken into account during the analysis as estimates with a large sampling error can cause distortions. Additionally, areas with very few residents will not provide good insight and should be suppressed from the analysis.

# D5. Data Governance, Privacy and Security, Ethical, Legal, and Regulatory Compliance

The data from the American Community Survey are publicly available and free for public use.

# D5a. Precautions

The datasets used contain no personally identifiable or proprietary information. Future use should refer back to the portal at data.census.gov for the most up-to-date data.

# Sources

Anderson, D. (2022, April 6). Real estate investors are buying a record share of U.S. homes. Redfin Real Estate News. Retrieved July 14, 2022, from https://www.redfin.com/news/investor-home-purchases-q4-2021/

Bureau, U. S. C. (2020). acs\_general\_handbook\_2020\_ch01.pdf. Washington D.C.; United States Census Bureau.

Bureau, U. S. C. (2020). American Community Survey B25058 MEDIAN CONTRACT RENT (DOLLARS). Explore census data. Retrieved June 28, 2022, from https://data.census.gov/cedsci/table?q=B25058%3A+MEDIAN+CONTRACT+RENT+%28DOLLARS%29&amp;g=0100000US%248600000&amp;tid=ACSDT5Y2020.B25058

Bureau, U. S. C. (2020). American Community Survey S1903 MEDIAN INCOME IN THE PAST 12 MONTHS (IN 2020 INFLATION-ADJUSTED DOLLARS). Explore census data. Retrieved June 28, 2022, from https://data.census.gov/cedsci/table?q=median+income&amp;g=0100000US%248600000&amp;tid=ACSST5Y2020.S1903

Bureau, U. S. C. (2020). American Community Survey S2502 DEMOGRAPHIC CHARACTERISTICS FOR OCCUPIED HOUSING UNITS. Explore census data. Retrieved June 28, 2022, from https://data.census.gov/cedsci/table?t=Owner%2FRenter+%28Householder%29+Characteristics&amp;g=0100000US%248600000&amp;tid=ACSST5Y2020.S2502

Bureau, U. S. C. (2022, June 2). About the American Community Survey. Census.gov. Retrieved June 20, 2022, from https://www.census.gov/programs-surveys/acs/about.html

DeSilver, D. (2021, August 3). As national eviction ban expires, a look at who rents and who owns in the U.S. Pew Research Center. Retrieved July 14, 2022, from https://www.pewresearch.org/fact-tank/2021/08/02/as-national-eviction-ban-expires-a-look-at-who-rents-and-who-owns-in-the-u-s/

Lerner, M. (2022, July 13). *8 must-have numbers for evaluating a real estate investment*. Investopedia. Retrieved September 6, 2022, from https://www.investopedia.com/financial-edge/0511/8-must-have-numbers-for-evaluating-a-real-estate-investment.aspx