5-2022

SOCIOECONOMIC LANDSCAPE RESTRUCTURING CAUSED BY PROPERTY APPRECIATION RATES

Abner Garcia

SOCIOECONOMIC LANDSCAPE RESTRUCTURING CAUSED BY PROPERTY APPRECIATION RATES

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Information Systems and Technology:

Business Intelligence and Analytics

by

Abner Garcia

May 2022

SOCIOECONOMIC LANDSCAPE RESTRUCTURING CAUSED BY PROPERTY APPRECIATION RATES

A Project

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Abner Garcia

May 2022

Approved by:

Dr. Nasrin Mohabbati, Committee Chair

Dr. Conrad Shayo, Committee Member

Dr. Jay Varzandeh, Chair, IDS Department

ABSTRACT

Real estate property has historically been considered the bedrock of the American Dream and a primary method used to build wealth. However, homeownership has not been attainable for many people because property appreciation rates have consistently exceeded income growth. Available research indicates that rapid appreciation rates pressure low to average income earners out of their counties of residence towards more affordable counties' residencies. This creates a problem where the receiving counties have increased demand and prices which starts a cycle of migration for lower income populations. Shifting populations can change the economic and demographic characteristics of counties. Previous research that explored determinants of high property appreciation found that it is significantly affected by population growth, demographic characteristics, and proximity to metropolitan cities. However, research regarding demographic or economic change attributed to appreciation rates is scarce. This project sought to answer if property appreciation rate influence changes in racial diversity, income, and population levels in California counties. To accomplish the project's objective, data was collected from various government agencies for eighteen California counties in a thirty-year period. Then yearly change was calculated for each variable in the data set. Additionally, a racial diversity index was calculated using the Simpson diversity index formula. Once the data was cleaned and prepared, linear regression models were used to determine significance of relationships and the effects of characteristics between appreciation rate and other economic and demographic factors. K-means clustering algorithms were also used to determine if the characteristics of counties had an impact on the relationships. The results of the analysis showed that appreciation rate did not impact population change. The relationship between appreciation rate and income levels had some significance in counties with high property values, indicating that those areas attract only high-income buyers. Additionally, the analysis demonstrated that appreciation rates can be considered a determinant for changes in demographic characteristics within some studied California counties. The significance of the relationships was strongly influenced by the underlying characteristics of each county. Counties with large Asian proportions saw a decline in Black and other racial minority proportions as appreciation rates rose. Racial diversity in counties with lower income levels was significantly impacted by the combination of property appreciation and changes in employment level. The research from this project suggests that future studies should be conducted to determine and understand how characteristics of each county can say about the future state of their economies.

ACKNOWLEDGEMENTS

I was always taught that accomplishments were never a single person's effort, they were always achieved with the support of others. Completing this project was no different, I could not have completed it without my support system. My faith, which carried me. My family, who motivated me. My friends who supported me. Professors like Dr. Nasrin Mohabbati, who taught and guided me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE:

INTRODUCTION

> "Next to food and clothing, the housing of a nation is its most vital problem"
>
> – Herbert Hoover 12/14/1931

Homeownership has been the bedrock of the "American Dream," and throughout American history, the federal government has promoted homeownership. Different Presidents such as Bill Clinton and George W. Bush prepared policies that increased homeownership (Goodman & Mayer, 2018). After the Great Recession, President Obama praised the people of Phoenix, AZ for the growth in home prices and sales (Office of the Press Secretary, 2013). Reasons for promoting homeownership include the creation of wealth, reduction of crime, increase in education levels (Beracha et al, 2012) and the reduction of racial inequality (Goodman & Mayer, 2018).

However, there is a disconnect between the federal and local governments. Cities usually receive higher revenues from commercial property and business development than housing and property taxes (Quigley & Raphael, 2005). Given their freedom to set their own rules, there are many cities which regulate favoring retail and commercial development over residential housing (Quigley & Raphael, 2005). This affects the local housing markets by reducing supply, increasing development costs, and ultimately making housing more expensive.

Pricing volatility is higher for properties at lower price points having a greater effect on households of lower income levels (Goodman & Mayer, 2018). Additionally, appreciation rates have increased at a higher pace compared to income growth (Goodman & Mayer, 2018). Not to mention the disconnect between areas with higher levels of diversity, which tend to appreciate at lower rates, reducing the benefits which can be received from homeownership (Anacker, 2010). Black and Hispanic households also tend to have a greater proportion of their wealth composed of home equity (Anacker, 2010).

When the cost of house rent rises beyond residents means, it will cause emigration to counties that are more affordable (Gunderson & Sorenson, 2010). Migrants displaced for this reason will seek neighboring cities or counties to maintain similar lifestyles while reducing their expenses. People migrating for these reasons can have higher income levels than residents of the receiving counties and are willing to pay higher prices in hopes of keeping the amenities they are accustomed to (Gunderson & Sorenson, 2010). In this situation, markets in receiving areas become more competitive increasing property appreciation rates.

This project explored a variety of counties in California, to determine if appreciation rates can cause migration, changes to diversity levels and income growth rates. California was chosen because it has experienced high levels of property appreciation and population size decreases over the last few years (Gunderson & Sorenson, 2010). From 2019 to 2020 median property values appreciated at a rate of 17% (California Association of Realtors, 2022a), during the same period personal income grew at a rate of 8.8% (Bureau of Labor Statistics, 2022). Studying the largest state of the country, in terms of population and economic activity, will provide additional insight into what is going on in other regions of the country.

Problem Statement

Research by Gunderson and Sorenson (2010) highlighted a problem caused by the redistribution of California populations. The research found that most people leaving their counties of residence left due to high property values and were in search of more affordable counties. The problem is their migration created new demand in the receiving counties and increased property values, which created a cycle of migration. This correlation between population growth and property values was also supported by Mulder's (2009) research which indicated domestic migration positively affected appreciation rates. Appreciation rates are also impacted by regulation and policies used by local governments to attract commercial development (Quigley & Raphael, 2005).

Additionally, research by Moye (2014) found that Philadelphia neighborhoods with small proportions of Black and other racial minorities experienced higher appreciation rates than neighborhoods with large proportions of those racial groups. Moye adds, if these neighborhoods are located adjacent to neighborhoods with large proportions of White or Hispanic racial groups, the proportion of minority racial groups declines. Research by Anacker (2010) corroborates the results, adding that there are gaps between the benefits received from property appreciation when comparing Black, Hispanic, and White racial groups. Anacker also expressed how income and employment have played a role in Hispanics and Blacks ability to maintain homeownership. Moye, suggested monitoring and understanding these effects are important for the future.

Though research by Gallin (2003) has shown that income does not have substantive relationship with property appreciation, it is a topic of interest when considering the rising property values. As property values rise, the income requirements to purchase will also rise. The minimum qualifying income required to purchase a home in California during 2020 was $98,400 (California Association of Realtors, 2022b). For the same year, the average income was $70,192 in 2020 (Bureau of Labor Statistics, 2022). The disparity between income and property values was caused by appreciation rates that consistently outpaced income growth.

Most of the research previously performed and described above studied and discovered problems related to property appreciation and affordability. They explore how appreciation can be affected by economic and demographic factors. However, the inverse is not thoroughly explored where property values can be a causing factor to the changes in racial proportions, population growth, and income levels. This leaves gaps in the understanding about the relationship of appreciation rate with economic and demographic factors. This project seeks to provide additional insight into the problems highlighted by Gunderson (2010) and Moye (2014) with a focus on California counties. In some instances, it will replicate analysis models used by previous research to confirm if the results differ in counties studied in this project.

Objective

The objective of this project is to determine if there is any causal relationship between property appreciation rates and changes in the economic and demographic composition of the studied counties in the state of California.

Research Questions

1. Is property appreciation rate impacting changes in income and population levels with focus on the counties in California?
2. Have appreciation rates increased or reduced diversity levels in California Counties over years?

Methodology

To answer these questions, this project relied on the analysis of various data sets and will consist of four stages. The first stage of the project involved the collection of data from various sources. The second stage consisted of data cleaning and standardization because the data came from various sources and it contained different units of measurement and variables which will not be proper for this project. Therefore, there is a need for pre-processing of the data before analysis. The third stage will consist of analyzing the transformed data sets through clustering and regression analysis methods. The fourth stage will review the results of the analysis and provide an interpretation of the results.

The data collected will consist of real estate property data from the California Association of Realtors, income and population data from the Bureau of Labor Statistics, demographic data from the United States Census Bureau, and employment data from the State of California Employment Development Department. To maintain the project within a manageable timeline, only a specific region of California will be selected and analyzed. As part of the study, scholarly research articles will be collected from the library's OneSearch tool and will reviewed to determine if there are other variables which may affect the study. Upon completion of the research, a project manuscript and dashboard will be created to provide visualization of the results.

Organization of the Study

This project is organized as follows: Chapter 1 provides an introduction for the project; Chapter 2 performs a literature review about related topics to this project; Chapter 3 provides a description of the collected data and the analysis methods performed on the data; Chapter 4 examines the results of the regression and cluster analysis; Chapter 5 provides the conclusion remarks, future research directions and recommendations.

CHAPTER TWO:

LITERATURE REVIEW

In this chapter, there will be a brief review of scholarly articles which have some intersection with this project’s topic of study. To contain articles within the range of available data, search results were only explored within the years of 1990 to 2020.The articles reviewed are grouped below in the following categories: articles closely relating to real estate appreciation; articles closely relating to the economic or demographic changes. Table 1 shows the categorization of reviewed articles based on keywords used in the database search.

Table 1: Research

<table>
<tr><th>Database</th><th>Keywords</th><th>Number of Hits</th><th>Relevant Articles</th></tr>
<tr><td>Pfau OneSearch</td><td>Income
Property
Appreciation</td><td>55</td><td>4</td></tr>
<tr><td>Title/Authors:</td><td colspan="3">• “Still Paying the Race Tax? Analyzing Property Values in Homogeneous and Mixed-Race Suburbs” – Anacker. (2010)
• “Homeownership and the American Dream” – Goodman & Mayer. (2018)
• “The Rent versus Buy Decision: Investigating the Needed Property Appreciation Rates to be Indifferent between Renting and Buying Property” – Beracha et al. (2012)
• “Neighborhood racial–ethnic change and home value appreciation: evidence from Philadelphia” – Moye. (2014)</td></tr>
</table>

Pfau OneSearch	Affordable Property Appreciation	84	4
Title/Authors:	• “The impact of land use regulation across the conditional distribution of home prices: an application of quantile regression for group-level treatments” – Leonard et al. (2020) • “Regulation and the High Cost of Housing in California” - Quiqley & Raphael. (2005) • “Silver Bullet or Trojan Horse? The Effects of Inclusionary Zoning on Local Housing Markets in the United States” - Schuetz et al. (2011) • “An Examination of Domestic Migration from California Counties” – Gunderson & Sorenson. (2010)		
Pfau OneSearch	Employment Property Appreciation	122	2
Title/Authors:	• “Financial frictions, the housing market, and unemployment” – Branch et al. (2016) • “Unaffordable housing and local employment growth: Evidence from California municipalities” – Chakrabart & Zhang. (2015)		
Recommendation		N/A	5
Title/Authors:	• “Using Simpson's diversity index to examine multidimensional models of diversity in health professions education” – McLauglin et al. (2016) • “An efficient k′-means clustering algorithm” – Zalik. (2008) • “The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets” – Gallin. (2003) • “The relationship between population and housing” – Mulder. (2009) • “Race and Home Price Appreciation in the United States: 1992–2012” – Hipsman. (2018)		

Real Estate Appreciation Articles

The study by Quiqley and Raphael (2005) researched housing regulation practices of cities and their impact on housing costs. The study covers the decade of 1990 to 2000 and determined a significant relationship between the supply of housing and the number of regulations in a city. The study found cities with two or more housing regulations would cause greater appreciation rates. This study is significant to this project, and highlights a cause of high appreciation rates, however, it does not set appreciation rates as the determinant for the research, instead, it is the dependent variable.

Research by Beracha et al. (2012) studied the indifference residents have between renting a home or buying a property. In their study, they determined that residents weigh the economic value of preparing a down payment for a purchase and compared it to gains if it was invested in different ways. They found that residents favored purchasing over renting, which is caused higher demand levels. This research has relevance but does not deeply explore the housing market or specific regions, instead it focused providing insight of when buyers may change their attitude towards purchasing real estate property.

Jenny Shuetz, et al. (2011) studied the effects of Inclusionary Zoning (IZ) policies, which incentivize developers to produce a prescribed proportion of units within a certain price point when building in IZ. Their research highlights how inclusionary zoning can increase the supply of affordable housing, but it has negative effects during periods of high appreciation. These policies contributed to higher prices and lower rates of production. They conclude that IZ is not an effective policy for creating affordable housing when compared to other policies such as Low-Income Housing Tax Credits. This research has significance to this project as it studies policies aimed to create housing options at lower price points, which are necessary to avoid lower income residents from being pushed out of cities.

The research by Goodman and Mayer (2018) reviewed homeownership levels across various years and counties. In their research, they identified the differences between appreciation rates of the areas with predominantly White residents and African American and other race residents. Additionally, they discuss the reduction in homeownership levels over the last 30 years, partially due to gaps in property appreciation rate and income growth. Their research has relevance to this project; however, it did not address changes caused by appreciation rates rather it only quantifies and mentions inconsistencies in appreciation benefits received by different racial groups.

Mulder (2009) researched the two-sided relationship between appreciation rate and population growth. Mulder explains how population growth drives demand for housing and therefore driving appreciation rates higher. As populations decline the inverse of the relationship can appear but with delayed effects. This research shows significance to the topic of this project as indicator for property appreciation. However, it studies the appreciation because of population growth.

Economic and Social Changes

Katrin Anacker’s (2016) research focused on wealth generating capacity of homeownership for minority groups. The study found, similarly to Goodman and Mayer, that minority households do not benefit as much as their white counterparts, minorities also had a greater proportion of their income and wealth invested in homeownership making them more vulnerable in recessions. Their study covered more depth than Goodman and Mayer in regard to racial differences, and has significance for this project’s topic, but it does not track change in the studied regions, leaving gaps for further study.

Hipsman (2018) researched differences between property appreciation rates within neighborhoods with racial and ethnic characteristics. This research resembles Anacker’s (2016) and validates the results, indicating disparities between property values in predominately white and black neighborhoods exist. Hipsman quantifies the difference in growth in percentage points, indication predominately white neighborhoods saw faster growth than other racial groups. This research is significant because it shows that appreciation can be affected by racial and ethnic characteristics, however it does not show the inverse relationship of appreciation affecting sizes or proportions of racial diversity.

Gunther and Sorenson (2010) performed research on migration within and outmigration from California. In their study, they analyzed destinations Californians migrated to, and attributed this migration to high costs of living and housing prices. This is meaningful to this project as it does cover some of the changes caused by housing costs, however, it did not specifically explore the rates of social or demographic change.

Chakrabart and Zhang (2015) studied the causal relationship between housing affordability and employment growth. Through their study, they investigated relationships through various empirical methods and concluded higher prices lead to slower employment growth. In their research, they indicated a possible reciprocity effect where employment growth can create high housing prices. The significance of Chakrabart's and Zhang's research to this project is that it can corroborate finds of the research.

Branch et al (2016) studied the effects of household finances on labor and housing markets. They created equilibrium models to analyze relationships and pressures on household finances. A center point of their study was the effects of how home equity loans changed the dynamics of housing prices and unemployment. This study is significant to this project because it studied similar variables, however, its purpose was to model the effects of prices on unemployment only.

Moye (2014) studied the wealth benefits received by different racial groups through home appreciation in the Philadelphia metropolitan area. Moye confirmed that areas with larger proportions of minority races show lower rates of appreciation. Additionally, the study found housing prices declined in areas where greater racial integration had occurred. Moye's study has significance to this project in that explains some changes in wealth and diversity but focused on a single metropolitan area which differs from this research project's scope.

Leonard et al. (2020) discussed the impacts of land use regulations on pricing and resulted in migration which spreads the pricing impacts to neighboring areas. Their study analyzed the types of land use regulations and if diverse use of regulations had a different impact on prices. Their research study had some overlaps with the proposed variables for this project, however, they focused more significantly on the regulatory effects on pricing.

Gallin (2003) indicated a misunderstanding of the relationship between income and appreciation rates. Gallin points out that many researchers claim that property appreciation rates have a significant relationship with income changes, but that the flaw in the research is it does not account for any cointegrated relationships. Gallin's research indicates that income and appreciation rate are not cointegrated and therefore could not be used as key indicators of each other. This research is significant but does not dive into the differences in characteristics of the studied areas.

In this chapter various articles were reviewed to determine if the topics proposed for this project have been thoroughly explored and analyzed. The articles collected for this purpose, considered all the same variables but did not consider appreciation as a key determinant of the changes discovered. Additionally, the geographic focus was spread across various states which differs substantially from this project. The next chapter will present the data collection and cleaning process used for this project.

CHAPTER THREE:

DATA COLLECTION AND RESEARCH METHODS

The United States Census Bureau conducts yearly surveys to collect various pieces of information about the people who live in the United States. From these surveys, the Census Bureau prepares estimates used to manage over “$675 billion in federal and state funds distributed each year” (United States Census Bureau, 2022). This data is published on various federal and state agency websites and organized by state, county, and metropolitan statistical areas which are zones with shared commuting and economic activity. For this project, county level data was collected from the following sources:

- Race and ethnicity data from the United States Census Bureau (United States Census Bureau, 2021).
- Population and income data from the U.S. Bureau of Labor Statistics (Bureau of Labor Statistics, 2022).
- Employment data from the California Employment Development Department (Local area unemployment statistics (LAUS), 2022).
- Property pricing data from the California Association of Realtors (California Association of Realtors, 2022a).

The counties of interest for this project were Fresno, Kern, Kings, Los Angeles, Madera, Merced, Monterey, Orange, Riverside, San Benito, San Bernardino, San Diego, San Luis Obispo, Santa Barbara, Santa Clara, Santa Cruz, Tulare, and Ventura. These counties were chosen because they cover nearly half of the geographic area of the state of California and have both urban and rural characteristics.

Preparation

The collected data was in various formats and structures. To analyze all variables, it was necessary to create a cohesive data structure with overlapping observation keys. The number of observations in some of the data sets was larger than in the others, which needed filtering to only include matching dates. Additionally, some contained coded keys that needed to be converted to match the other data sets. R (and RStudio) was used as the primary modeling and analysis tool (R Core Team, 2021). Excel was used for pre-processing, filtering, and extraction of data. Notepad was used as an auxiliary tool to convert a data set which was provided in a text file format.

The first step in creating a cohesive and standard data structure was to remove coded keys from the race and ethnicity data sets. For the data sets containing the years 2000 through 2020, dates and ages were coded into numerical categories. Dates were coded into thirteen categories and age groups were coded into eighteen categorical groups. A dictionary file was provided by the U.S. Census Bureau which was used to filter observations.

The second step was to extract only yearly observations that had overlaps across all the data sets. Data collected from the U.S. Census Bureau and California Association of Realtors contained observations for additional months, which other data sets did not have. After a review of all the data sets, it was determined that all the data sets contained observations for the month of July. Observations for the month of July were extracted from the U.S Census Bureau and California Association of Realtors median property price data sets.

The median property price data from the California Association of Realtors, was missing values for Tulare County for the years 1990 through 2000. To ensure that all counties had the same number of observations, the average price for that year was included in place of the missing observation. Additionally, the linear regression model and k-means clustering algorithm would produce invalid results if values were missing.

The race and ethnicity data were collected in three separate files, one for each decade of the study. Each of these data sets had differences in the number of categories used. Data for 1990 to 1999 only contained eight racial and ethnic category groups, 2000 to 2009 contained thirty-six categorical groups, and 2010 to 2020 seventy-four categorical groups. For the two latter decades, the data contained gender for each categorical group and additional categories for responses of “two or more races.” Male and female for each racial and ethnic category were combined to create a single summed category. Next, only the categories included in the 1990 to 1999 data were preserved from the data.

To validate the accuracy of the extraction, the sum of all extracted categories was compared to a column provided in the two data sets which contained the sum of all respondents. This provided an exact match, which validated the data was extracted without corruption of the estimates. Additionally, the sum of these was compared to the estimates of the population to ensure the estimates were of similar size, which proved to have only a 2% margin of error calculated by the difference between both the sum and population divided by the population estimate.

The third step was to create a unified format across all files. There were several data sets with yearly observations spread across the columns of the data, while others had years spread across the rows. It was necessary to transform all the data into a long format with years contained in one column, counties in another, and the amounts for other variables in their separate columns. By doing so, this created data sets with more rows and fewer columns, reducing the spread and increasing the height. This transformation created equal heights across all data sets, needed to perform comparative analysis across the various data sets.

The fourth step in the preparation of data was to add calculated columns to each data set. For all variables, a yearly percentage of change was calculated by taking the difference between each observation’s value and the value in the preceding year and dividing it by the preceding year’s value. This calculation required each data set to be in the tallest possible format. To accomplish this, the data sets were transformed in RStudio using the Gather and Mutate functions from the “TidyR” package, sample code shown in Figure 1. This created an additional column in the data set for the calculated changes.

```
#Perform gather function to make a tall format
prices_tall = gather(prices, "County", "Price", 3:20)

#Add a percent change column
prices_tall <- prices_tall %>%
    group_by(County, Month) %>%
    arrange(Year) %>%
    mutate(App.Rate = 100*(Price - lag(Price))/lag(Price))
```

Figure 1: Gather and Mutate Functions Used to Calculate Change

For the race and ethnicity data sets, two additional calculations were performed. Proportions of each racial group were calculated by dividing the size of each group by the sum of all racial groups. Secondly, a diversity index was calculated using the Diverse package in RStudio, which enabled the calculation of various types of diversity indices.

The U.S. Census Bureau has published diversity indices for the years 2010 and 2020. The diversity index used by the U.S. Census Bureau is calculated using eight categorical groups; Hispanic combined, non-Hispanic White, non-Hispanic Black or African American, non-Hispanic American Indian or Alaska Native, non-Hispanic Multiracial, and some other race (Jensen, et al., 2021). Unfortunately, the formula was not provided for the bureau's diversity indices. Additionally, the data sets did not contain categories for what is referenced as "some other race."

In lieu of the bureaus approach, the Simpson diversity index formula was chosen for this project using only the eight racial and ethnic groups available across all three data sets. To validate accuracy, the calculated indices for 2010 and 2020 were compared to the indices published by the census bureau and were found to be within an average two points of difference. It is worth noting, the published indices are not based on the yearly American Community Survey but rather on the decennial survey (United States Census Bureau, 2022). The following shows the Simpson Diversity Index formula:

$$Simpson\ Diversity\ Index\ =\ 1\ -\ \frac{\sum n(n\ -\ 1)}{N(N-1)}$$

The Simpson diversity index is commonly used to measure biodiversity in a habitat. It provides a measure of the probability of randomly choosing observations in the same categorical group (McLaughlin et al., 2016). The formula presented above used n to represent the amount of each racial category in the yearly observation and N to show the sum of all groups in the observation. The resulting index represents the probability of choosing someone of the same racial or ethnic category at random. As the index approaches one, the probability of choosing people from the same racial category reduces.

Once all transformations were completed, the data sets were merged for comparative analysis. The various types of transformation were grouped and contained in separate data frames to analyze in common unit size. Three data frames were created to contain the various transformations. One contained all variables in their original form, the second contained the calculated values for change percentages, and the third data frame was prepared to calculate the change in proportions of racial groups.

Table 2: Variable Descriptions

Variable	Description
Year	Observation year, 1990-2020 (calculated variables begin with 1991)
County	County of observation.
NHWA	Non-Hispanic White alone
NHBA	Non-Hispanic Black alone
NHIA	Non-Hispanic American Indian and Alaska Native alone
NHAA	Non-Hispanic Asian alone
HWA	Hispanic White alone
HBA	Hispanic Black
HIA	Hispanic American Indian and Alaska Native alone
HAA	Hispanic Asian Alone
Diversity	Calculation of Simpson diversity index
Population	Population size, based on ACS
Income	Income per capita, based on ACS
Labor Force	Labor force size
Employment	Employment level
Unemployment	Unemployment level
Price	Median price

Descriptive Statistics

1990 was the first year studied for this project. During that year, the lowest real estate property price for studied counties was $88,780, found in Kern County. The highest-priced county in the same year was Santa Clara with a median price of $273,600. The average price of the studied counties was $180,106. The average income during the same year was $19,098 with the lowest income found in Kings County with $13,678. Santa Clara also had the highest income level during 1990, with a per capita income of $25,640. Los Angeles was the most populous and diverse county in the study, with 8,878,157 persons and a diversity index of 0.68. San Benito County had the smallest population, with 36,835 people, and a diversity index of 0.54 which is average for the observed counties. San Luis Obispo was the least diverse county with an index of 0.32.

The year 2020 was the last year which formed part of this project. The lowest-priced county in the final year was Tulare County with $268,000 and Santa Clara had the highest price of $1,380,000. The average median price during this year was $606,642. The average income was $60,513, the highest income was $123,661 in Santa Clara County, and the lowest per capita income was $41,829 in Kings County. Los Angeles County continued to be the most populous and diverse county, with 9,943,046 people and an index of 0.68. San Benito was the least populous county at 64,055 persons, the least diverse county continued to be San Luis Obispo with a diversity level of 0.45.

Each year, the largest racial and ethnic groups were non-Hispanic whites and Hispanic whites. These formed the largest proportion of all eighteen studied counties. The third largest group was non-Hispanic Asian or non-Hispanic Black depending on the county. Counties with higher income levels tended to have larger Asian communities while lower income counties had larger Black communities. The smallest racial groups were the native groups, with their concentration being in counties of lower income levels. This pattern was consistent across all studied counties.

Compared to the state, in 1990 most of the counties studied for this project fell below California's per capita income of $21,485. Only the counties of Los Angeles, Orange, Santa Barbara, Santa Clara, Santa Cruz, and Ventura had income levels above. The median price of real estate property for California in the same year was $193,088, nearly half of the studied counties had median prices below this amount.

In 2020, California state had a per capita income of $70,192. The studied counties mostly fell below this income level, with only Orange, Santa Clara, and Santa Cruz having per capita income levels above the state level. The 2020 median price in California was $666,320, ten counties in this study were below. The counties of San Luis Obispo and Los Angeles were above the $650,000 median price. Descriptive graphical models are included in Appendix A.

Methods of Analysis

The objective of this project was to determine how the appreciation rate changed the economic or demographic landscapes of California counties. Analysis was performed for the data in its original form of sizes, and in calculated yearly changes. The relationships between variables were explored using scatter plots. Counties were also clustered to determine if county characteristics strengthened or weakened such relationships. Finally, regression analysis was performed to quantify the significance of the relationship.

The scatter plots provide a visual indication of the relationships. These were useful to identify relationships across the different variables. the trends were not easily discovered using scatter plots because the models were oversaturated. Each county had different amounts for each variable which impeded the efficacy of the scatter plots. It was determined that filtering the data would help resolve this issue.

K-means clustering method was used to filter the data efficiently and based on each county's different characteristics. K-means clustering is an algorithm which assigns centroids and then calculates the distance of each observation from those centroids with the goal of grouping similar observations (Žalik, 2008). Observations are clustered to the centroid nearest to them. The number of centroids can be selected using the elbow method. For this project, K-means algorithm was implemented in RStudio.

Clustering was used for all five-hundred and fifty-eight observations in the data. The variables used for clustering calculations were income level, employment and unemployment levels, population sizes, and property prices. These variables conform to the basic economic characteristics of each county. Once clustering was completed, counties were grouped by the cluster where most of their observations were found. A separate data frame was created for each group.

Quartile descriptions were also prepared for every unique variable for the year 2020. Variables which may be factors of others were not used in this method. Quartiles were prepared by using the mean as a center point of division, then the above and below average were further divided in half creating four categories. Quartiles allowed for a basic description and characterization of each county in relation to each other. This information was contained in Table 3, along with the cluster grouping results.

After clustering, a review of variable relationships was conducted using scatter plots for each cluster. Where the relation between price or appreciation rate and changes in other variables was identified, the significance of the relationship was tested using linear regression analysis. Linear regression was performed using RStudio and plotted within scatter plots to illustrate the relationship among variables. The yearly change percentage of each economic or demographic factor was used as the dependent variable and the property appreciation rate as the independent variable in the regression model. The same models were prepared using data for all counties and years, and the data filtered by cluster group. As previously noted, relationships between variables were not clearly visible or significant when considering all counties together.

This chapter has presented the data collected and descriptions of the various analysis methods used for this project. Due to the varying sizes and characteristics of the data and counties studied for this project, the counties were grouped based on the economic characteristics exhibited by each county. The next chapter will present the results of the analysis performed, primarily the regression models used to determine the significance between property appreciation and economic and demographic factors.

CHAPTER FOUR:

ANALYSIS AND RESULTS

This chapter will cover the results of the analysis performed through this project. It will begin with a review of the descriptive analysis which discovered some trends across the counties and years, next it will cover multiple linear regression performed on the data containing all counties, and finally the results of analysis within each cluster group.

Descriptive Analysis

This project began with the visual exploration of the data, which highlighted some of the characteristics used to group counties based on similarities. In 1990, the base year of the project, all counties exhibited similarities within racial proportions with the main drivers for diversity being non-Hispanic Whites, Hispanic Whites, and the third group of Black or Asian depending on the county. The largest group across all counties was the non-Hispanic White shown in Figure 2 below in grey.

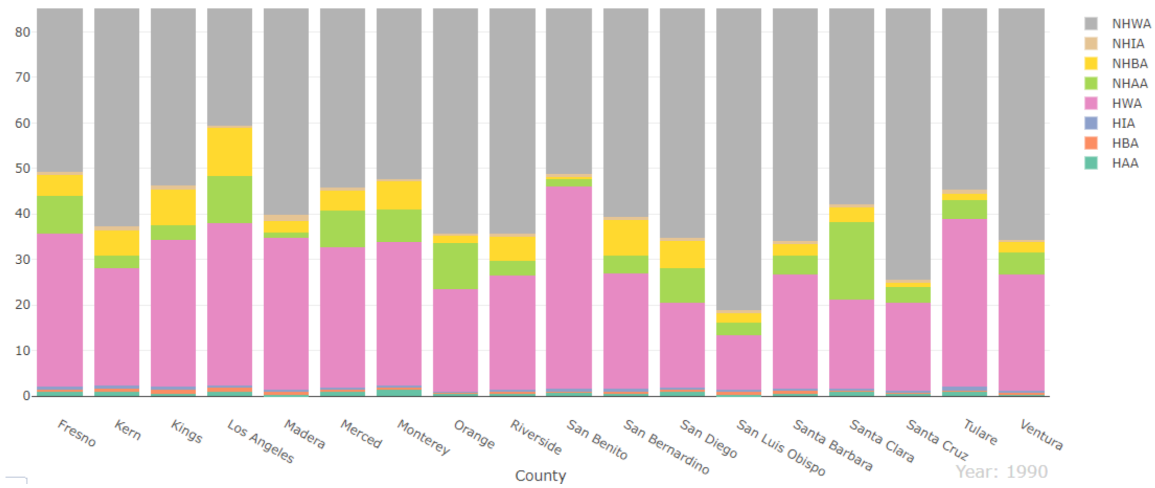

Figure 2: 1990 Racial Proportions

The proportion of non-Hispanic White saw the largest decline over time and Hispanics saw the most growth across all counties. The second largest growth was in the Asian community, but this was not visible in all counties. Counties with existing concentrations of Asian communities saw faster growth compared to others, however, their growth rate was reduced if a county had a high proportion of Black residents. As these communities grew, they reached a level where they were no longer affected by diversity changes. Notably, if the growth of one group is correlated with the decline of another, diversity calculations are negatively impacted. Figure 3 below, shows the racial proportions in 2020, the final year studied through this project.

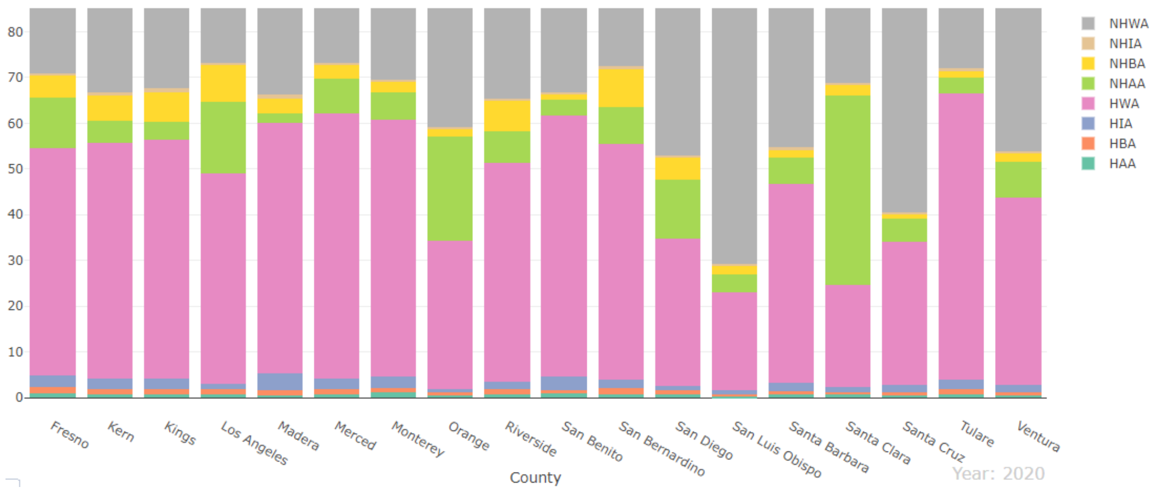

Figure 3: 2020 Racial Proportions

Income was another characteristic that was consistent across all the counties observed in this project. Throughout the periods studied, all the counties maintained a similar rate of growth, with only a couple of exemptions. This is most notable for counties with geographic proximity to each other. All counties experienced consistent growth across all years with only a few periods of yearly decline, which was visible across all counties. These periods of decline were correlated with the periods of property value decline. Income and property values also diverge in the level of volatility experienced within them. Property values experienced high volatility in years of decline whereas income stayed relatively consistent.

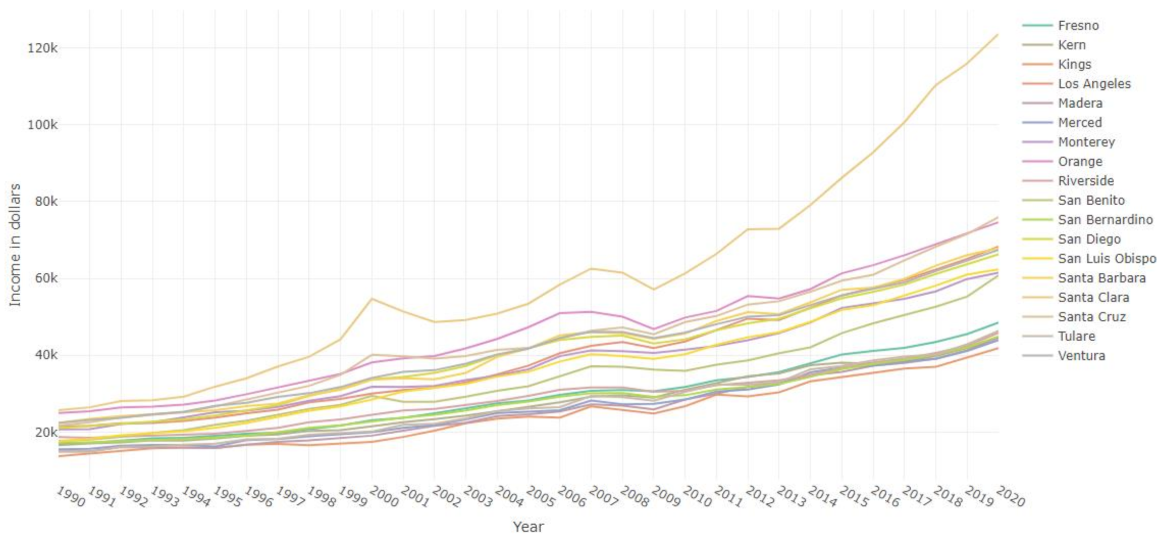

Figure 4: Yearly Income

Multiple Linear Regression

Before performing clustering, linear regression models were prepared for the data set containing all counties and years. Because the data had multiple and repeating counties it was necessary to use dummy variables for years and counties to remove the effects of repetition. A dummy variable assigns a numerical value to categorical variables which allows the model to distinguish between the various categories. This can be done by using the “factor()” function inside the “plm()” function in R.

```{r}
plm(Diversity_Change ~ App.Rate + Pop.chg + Income.chg + Laborforce_Change + factor(County)
    + factor(Year), data = change_merge_naomit)
```

Figure 5: Multiple Linear Regression with Dummy Variables

Data with multiple categorical variables such as year and county is considered panel data. R can be set up to use a function to manage panel data effectively and efficiently, reducing the amount of code needed to create the models. Using the PLM package and “plm()” function, regression models can be created which automatically account for categorical variables by assigning them as an index in the code. The models for this project were created using the PLM package. An example of the code has been shown in the figure below.

```
summary(
  plm(Diversity_Change ~ Pop.chg + Employment_Change + Unemployment_Change,
                      data = filter(change_merge_total, Year > 1990),
                      index = c("County", "Year"),
                      model = "within")
)
```

Figure 6: Multiple Linear Regression Using PLM Function

Once the model structure was chosen, multiple linear regression models were created using only variables which were not factors of others. The backwards elimination was used to create an optimal model, this process removes the variable with the least significance from the model until only variables with high significance are left. The final formulas were as follows:

- Diversity change = b_0 + b_1(Population change) + b_2(Employment change) + b_3(Unemployment change)
- Income change = b_0 + b_1(Appreciation Rate) + b_2(Unemployment change)
- Population change = b_0 + b_1(Appreciation Rate) + b_2(Labor force change) + b_3(Diversity change)

The model for diversity change showed significant effects from changes in population, employment level, and unemployment level with p-values below 0.0001 with an R^2 of 0.14, indicating that the model can explain about 14% of all observances. The model or income changes showed a significant relationship between appreciation rate and unemployment changes. The appreciation rate had a p-value below 0.0001 and unemployment had a p-value of 0.0007, with an R^2 of 0.11, indicating that the model can explain 11% of the observances across all counties. The optimal model for population change showed significance with labor force change and diversity change, resulting in an R^2 of 0.21 and each variable had a p-value well below 0.00001, indicating that this model can explain 21% the observances in the data set. These same models were tested against the clusters created and discussed in the following subsection which discovered a variance in R^2 and p-values depending on the cluster group

Clustering

This subsection reviews the results of the simple linear regression models created for each cluster. Each model assigned a different variable for y and maintained the appreciation rate as x, separate models were created for each variable in the data set. Clustering was completed using K-means clustering which resulted in four cluster groups determined using the elbow method, as shown in Figure 7. The largest group was cluster four containing nine counties. Cluster three was the second largest cluster containing five counties. Cluster two followed in size with three counties included, and cluster one was the smallest group containing only Los Angeles County. Table 3 tabulates the clustering results created using the K-means algorithm and how these counties compared using quartile division of each variable.

```
clust_merge_total_scale <- scale(clust_merge_total[c(11:16)])
fviz_nbclust(clust_merge_total_scale, kmeans, method ="wss") +
  labs(subtitle = "Elbow Method")

k_clust_merge_total <- kmeans(clust_merge_total_scale, centers = 4, nstart = 100)
k_clust_merge_total_scale_clusters <- k_clust_merge_total$cluster
fviz_cluster(list(data=clust_merge_total_scale, cluster = k_clust_merge_total_scale_clusters))
```

Figure 7: K-Means Clustering Code

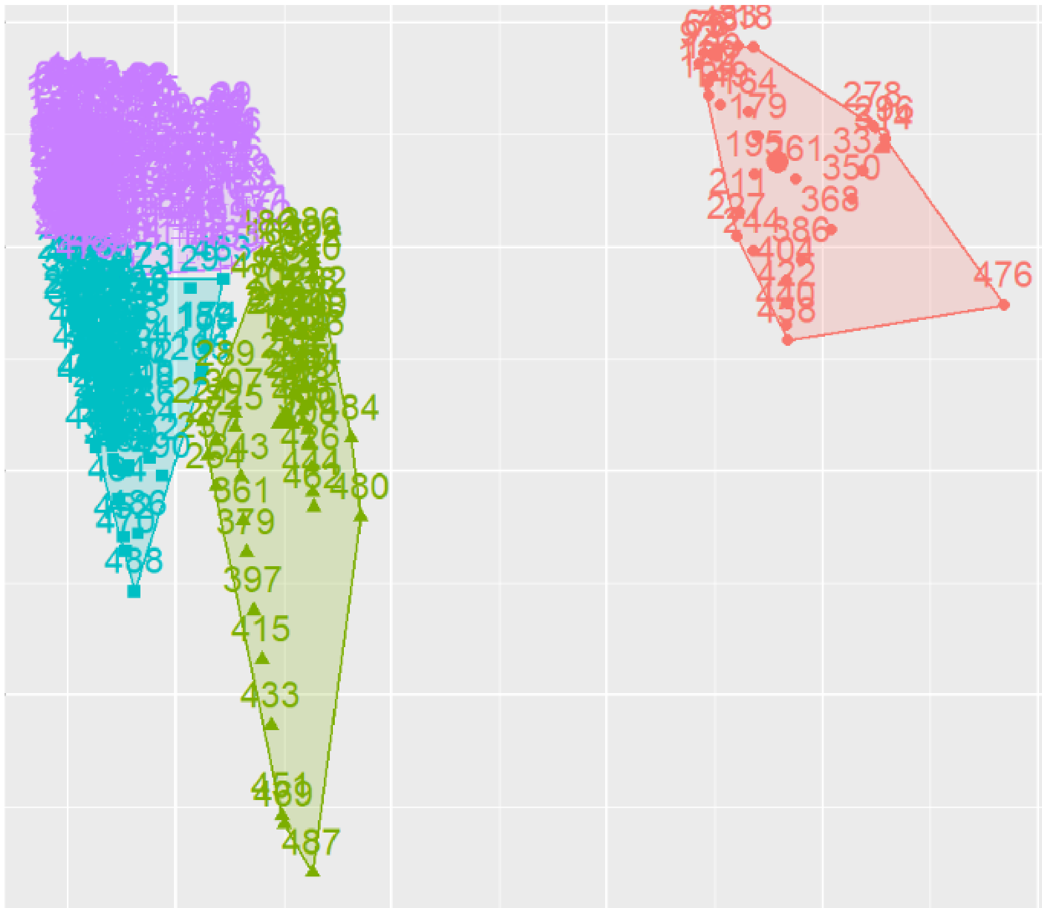

Figure 8: K-Means Clustering

Table 3: Quartiles for Attributes of Each Cluster

Counties	Quartiles					Cluster Group
	Appreciation	Price	Diversity	Income	Population	
Fresno	4	2	4	2	2	4
Kern	2	1	3	1	2	4
Kings	2	1	3	1	1	4
Los Angeles	2	3	4	4	4	1
Madera	1	1	2	1	1	4
Merced	4	1	1	1	1	4
Monterey	4	4	2	3	2	4

Orange	1	4	4	4	4	2
Riverside	2	2	3	2	4	4
San Benito	4	3	1	3	1	3
San Bernardino	3	2	3	1	4	4
San Diego	3	3	4	3	4	2
San Luis Obispo	1	3	1	3	2	3
Santa Barbara	4	4	2	4	2	3
Santa Clara	1	4	4	4	3	2
Santa Cruz	2	4	1	4	1	3
Tulare	2	1	1	2	2	4
Ventura	1	3	3	3	2	3

Cluster 1

All thirty yearly observations of Los Angeles County were clustered into one cluster which was substantially distant from all other cluster groups, red-colored circles in Figure 8. Los Angeles County had the largest population and labor force across all years observed. Property prices were part of the third quartile, and income levels in the fourth quartile indicating that it was higher than average pricing and income. Because of its larger size compared to all other counties, it was an outlier in the model and could not effectively be compared to the other counties.

```{r}
#Simple linear regression for Cluster 1 cosisting of Los Angeles County
plm(Income.chg ~ App.Rate, data = change_regression_1, index = c("County", "Year"))
```

Figure 9: Simple Linear Regression Sample Code

Using simple linear regression, with appreciation as the independent variable, no significant relationship was found between appreciation rate and any other variable. A sample code has been shown in Figure 9. This can be attributed to the large size of the population, which could have seen changes in one sector of the population, but not large enough to cause any noticeable change in the data. The county showed a declining population size from 2016 to 2020, shrinking by 162,662 over that period. The change in population was not correlated with changes in diversity or racial proportions, indicating a balanced decline in all the racial groups. During the same period, the county saw an average appreciation rate of 6.2%, lower than the average of all studied counties.

Cluster 2

Cluster two was comprised of the counties of Orange, San Diego, and Santa Clara. These counties had a similar demographic structure with a substantial non-Hispanic Asian population which was driving the increase in diversity. In all the studied years, these three counties saw significant growth in non-Hispanic Asian proportions. The counties also showed similar income growth and property appreciation pattern, though their scale was different. Santa Clara was the most expensive and highest income level of the three across all thirty years, it also had the smallest population size.

Cluster two showed a significant relationship between appreciation rate and changes in four racial and ethnic groups. Hispanic Asian, Hispanic, and non-Hispanic Black racial groups showed a negative relationship while non-Hispanic Asian showed a positive relationship. The slope presented by the regression model was small for each racial group and contained some outliers within the models. Changes in the non-Hispanic Asian group were most closely aligned to the regression model. Figure 10 illustrates the regression model within the scatter plot of appreciation rate compared to the non-Hispanic Asian racial group.

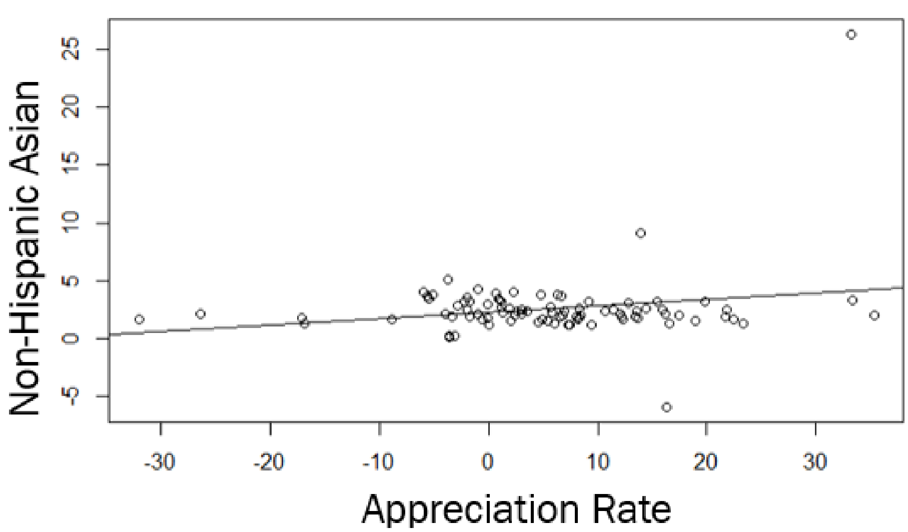

Figure 10: Cluster 2 Appreciation Rate to Non-Hispanic Asian Change

The appreciation rate for counties in cluster two also showed a significant relationship between employment change and income change. Employment change had a similar relationship with some of the racial groups, with many observations that did not closely align with the model. Income change observations did not have as much spread from the linear regression model compared to employment change, illustrated in Figure 11. The higher slope of the model was generally driven by substantial income growth in Santa Clara County. It shows higher-than-average income growth from 2015 to 2020, the average growth among all counties was 4.5%, and Santa Clara County saw an average growth of 7.7%. Orange and San Diego had average income growth of 4.3% during the same period.

Additionally, cluster 2 showed significance from multiple linear regression with income change as the dependent variable, appreciation rate and unemployment change as the independent variables. This same model was used on the data contain all counties and produced R^2 0.11. When the model was replicated for counties in cluster two, R^2 rose to 0.26, and p-values maintained below 0.05. This indicates that appreciation and lower changes in unemployment level can be causal factors for changes in income level for counties in this cluster.

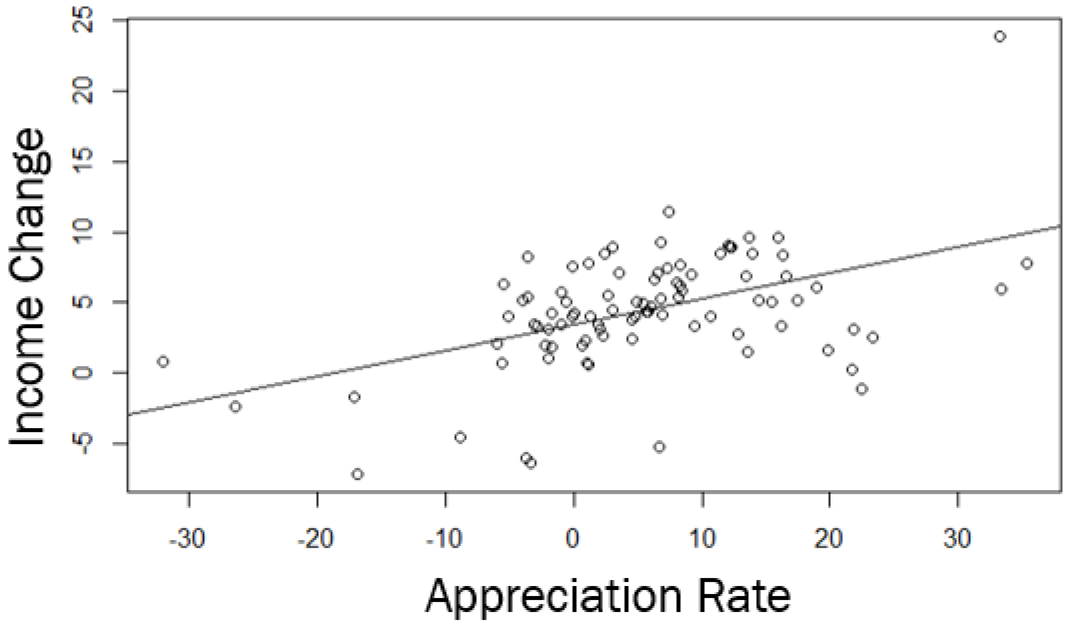

Figure 11: Cluster 2 Income Change to Appreciation Rate

Cluster 3

San Luis Obispo, San Benito, Santa Barbara, Santa Cruz, and Ventura were grouped in cluster three. These counties had lower diversity levels than all other counties studied. The thirty-year average diversity index of all the counties was 0.59, and cluster three averaged 0.51. San Luis Obispo, Santa Barbara, and Santa Cruz had and maintained a predominately non-Hispanic White population, with Hispanic White being the second largest group. San Benito and Ventura had the opposite characteristics. All counties in cluster three had above average income and property prices, with Santa Barbara and Santa Cruz being in the fourth quartile.

Using simple linear regression, appreciation rate had a significant relationship with three other variables, Hispanic White, income change, and employment change. Both income and employment change had a positive relationship with appreciation rate and higher significance, with p-values less than 0.001 and 0.016, compared to .015 for Hispanic White. This indicates that appreciation rate can be a stronger predictor of income changes and employment level changes, as shown in Figure 12.

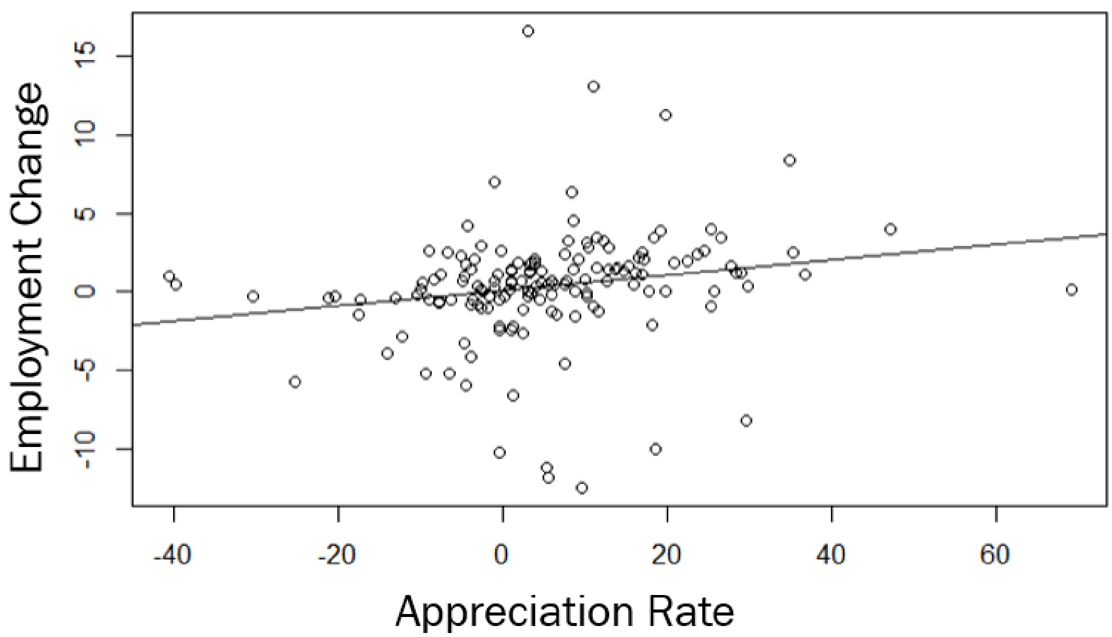

Figure 12: Cluster 3 Appreciation Rate to Employment Change

Cluster 4

Cluster four was comprised of Fresno, Kern, Kings, Madera, Merced, Monterey, Riverside, San Bernardino, and Tulare. All these counties had income levels below the average except for Monterey. These counties mostly had smaller than average populations with the exceptions of Riverside and San Bernardino which had population sizes in the fourth quartile. These counties also contained higher proportions of Black residents compared to all other counties in the study. Cluster four had a split in appreciation rate, with 44% of the counties having higher than average in 2020.

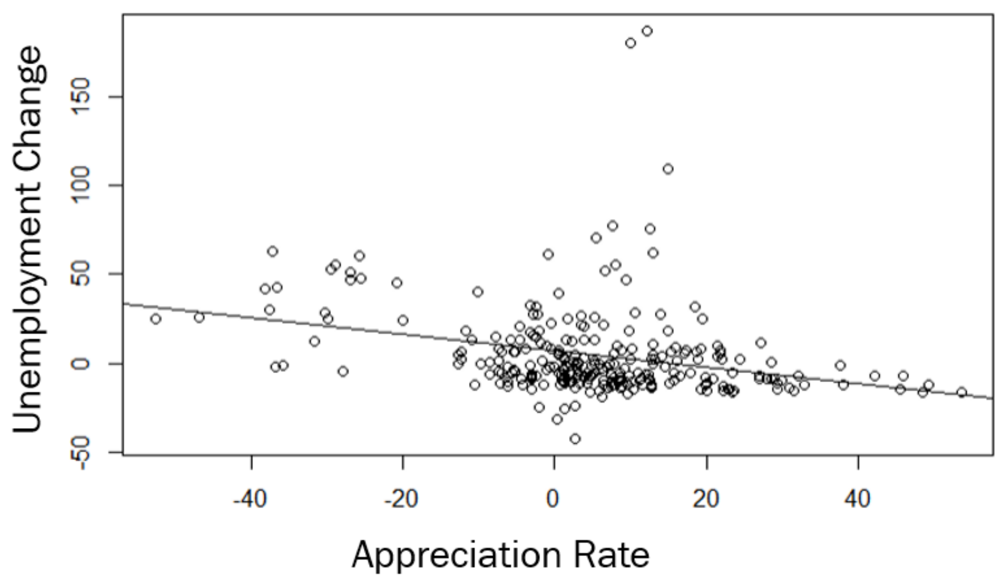

Figure 13: Cluster 4 Appreciation Rate to Unemployment Change

Regression analysis results showed that for the counties in cluster four there is a significant relationship between the property appreciation rate and three variables: A positive relationship was found with income and employment change with p-values below 0.01. Visually, both variables showed a significant spread among the observations indicating some weakness in the model. Change in unemployment level was found to have a negative relationship with the appreciation rate, having a p-value of less than 0.001, shown in Figure 13. Table 4 below, shows p-values resulted from the simple regression analysis using appreciation as the independent variable and each explored variable as the dependent variable. Additionally, multiple linear regression showed that diversity levels in cluster four counties were impacted more by population, employment, and unemployment changes. When the multiple linear regression model replicated for counties of cluster four R^2 rose from 0.14 to 0.20.

Table 4: Simple Linear Regression P-values for Different Cluster Group

Explored Variables	Cluster_1	Cluster_2	Cluster_3	Cluster_4
Non-Hispanic White	0.296	0.211	0.032	0.592
Non-Hispanic Black	0.474	0.049	0.086	0.488
Non-Hispanic Native	0.739	0.084	0.487	0.663
Non-Hispanic Asian	0.705	0.053	0.299	0.361
Hispanic White	0.424	0.221	0.015	0.888
Hispanic Black	0.648	0.005	0.992	0.889

Hispanic Native	0.762	0.070	0.459	0.318
Hispanic Asian	0.785	0.009	0.099	0.746
Population	0.892	0.651	0.604	0.114
Income	0.057	0.000	0.000	0.000
Labor Force Level	0.494	0.243	.246	.226
Unemployment Level	0.391	0.120	0.058	0.000
Employment Level	0.239	0.031	0.016	0.006
Diversity	0.481	0.494	0.078	0.111

This chapter presented the results of the analysis performed throughout this project. Simple linear regression indicated significant relationships exist between appreciation rates and different racial groups. Using k-means clustering, it was determined that the counties studied in this project could be categorized in four clusters. It was also discovered that the characteristics of each cluster can impact the significance of relationships between the different variables explored for this project. The next chapter will provide a discussion of the how this projects results can be interpreted and provide recommendations for mitigation for the research problem. Finally, it will provide a brief overview of topics for future study that were discovered through this project.

CHAPTER FIVE:

DISCUSSION AND RECOMMENDATIONS

This project's objective was to determine how property appreciation rate is related to economic and demographic changes. To accomplish this objective, data was collected from various government agencies and compiled together. Each variable in the data was transformed to include a calculated yearly change. Linear regression models have been developed based on all collected data to explore any existing relationships. Next, the data was clustered by county characteristics to determine if relationships were more dependent on characteristics within different clusters. Finally, regression analysis was performed using appreciation rate as the independent variable in each cluster. The analysis determined there is a significant relationship between property appreciation rates and some economic and demographic changes, however, the relationships varied between counties. This variation can be attributed to each county's economic, demographic, and geographical characteristics. This chapter will provide a brief discussion of how the results of the project can be interpreted and where further study can deepen its findings.

Discussion

The significance of the relationship between appreciation and other variables was substantially affected by the underlying characteristics of each county. Counties with changing demographic structures saw the most significant relationships between appreciation and the racial and ethnic groups. If a county had a racial group with a high level of growth, appreciation would have higher significance with other racial groups. This was most noticeable in cluster two which has had a growing Asian community over the thirty years studied in this project. The growth of this racial group drove property appreciation but also reduced the proportions of other racial and ethnic groups.

Cluster two saw the decline of already small proportions of Black and native racial groups, with statistical significance with Black communities. This created the false flag that diversity would have a significant decline, but because the Asian community was growing, diversity did not see a substantial change. From 2010 to 2020 diversity in Santa Clara declined from 0.69 to 0.68. This held true in most counties where racial proportions decreased. Figure 14 shows the changes in diversity for Santa Clara County.

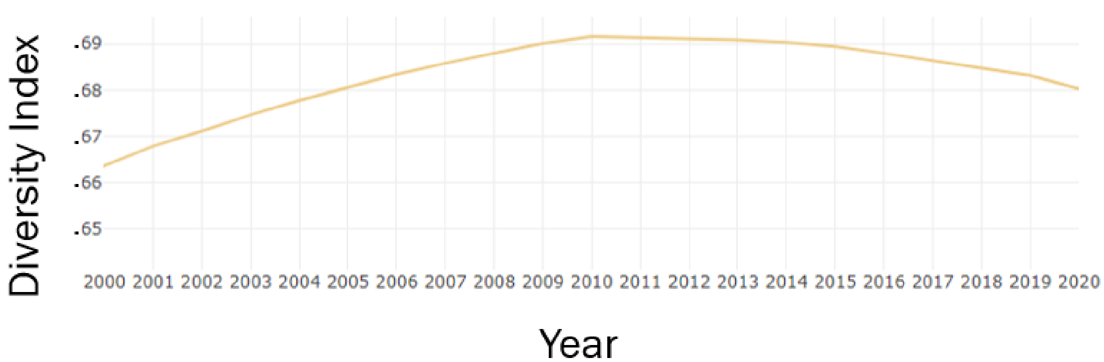

Figure 14: Diversity in Santa Clara County

Counties with large populations did not experience significant effects from the appreciation rate. This project had only one highly populous county, which made it difficult to comparatively analyze it against others of its size. Counties like Los Angeles can be insulated from the effects related to the appreciation rate. Large and populous counties usually are economic hubs full of industrial zones and opportunities which attract affluent residents (Gunderson & Sorenson, 2010). This maintains a consistent level of demand regardless of overall economic seasons.

In summary, appreciation rates have a significant relationship with some changes in counties, however, these are not consistent to consider a causal relationship. Additionally, the relationship between these changes may be caused by the underlying characteristics of each county. Geographic proximity to a large metropolitan area will augment the significance of relationships between variables. Metropolitan areas can drive appreciation, but if a neighboring county had a similar demographic structure the relationships are weakened.

Recommendations

Affordable housing is a significant problem for residents and governments in California. Governments understand that it is important to close the gap between property prices and what people can afford but seem to favor do not implement policies to reduce costs (Quigley & Raphael, 2005). Research by Shuetz (2011) shows quantifies the performance of various cost reduction strategies and policies governments can take to provide some relief to this problem. Additionally, governments need to be aware of how their constituent populations may change, economically or demographically, to better prepare for the future. Using the research strategies conducted by this project, governments can better assess what segments of their population may need attention and promote policies to support them.

Project Retrospect

This project faced challenges due to the wide scope of the variables explored and the varying characteristics of the counties. This is not to say it did not render valuable insight but could have had higher yields if the scope was reduced. This, however, was only discovered after categorizing the studied counties which required a wide exploration and analysis of data. If this project was initiated with this new knowledge, it would have explored only one common area such as the geographic zones containing Los Angeles, Riverside, and San Bernardino counties which were highlighted by Gunderson's (2010) research, indicating people leaving Los Angeles County in favor of Riverside and San Bernardino counties. These could provide a deeper understanding of the type of residents being displaced by the change in characteristics of the receiving counties.

Future Work

This project discovered county characteristics' significance in the changes and effects caused by variables such as property appreciation rate. Future research can dive into a micro-level and categorize cities or neighborhoods based on diversity indices. Creating indices for employment type, income type, and levels then applying predictive analytics to determine if there are any significant causal relationships between various variables.

Conclusion

Demographic and economic changes do not happen organically. These are determined by various quantitative and qualitative variables. One possible determinant is property appreciation rates. As property values rise, demographic characteristics tend to change in counties of high income or low diversity. Counties with lower income are more likely to have economic changes when property appreciation rises, such as rising employment levels. The relationship between appreciation and demographic and economic changes can be influenced by the characteristics of the nearest metropolitan area.

REFERENCES

Anacker, K. B. (2010). Still paying the race tax? Analyzing property values in homogeneous and mixed-race suburbs. *Journal of Urban Affairs*, 55-77. Retrieved from https://web-s-ebscohost-com.libproxy.lib.csusb.edu/ehost/pdfviewer/pdfviewer?vid=0&sid=20cf8895-307c-44d0-bda1-d323dbc08cc7%40redis

Beracha, E., Seiler, M. J., & Johnson, K. H. (2012). The rent versus buy decision: investigating the needed property appreciation rates to be indifferent between renting and buying property. *Journal of Real Estate Practice and Education*, pp. 71-88. Retrieved from https://www-jstor-org.libproxy.lib.csusb.edu/stable/24863202?sid=primo&seq=1

Branch, W. A., Petrosky-Nadeau, N., & Rocheteau, G. (2016, July). Financial frictions, the housing market, and unemployment. *Journal of Economic Theory*, 101-135. Retrieved from https://www-sciencedirect-com.libproxy.lib.csusb.edu/science/article/pii/S0022053115001313

Bureau of Labor Statistics. (2022). *Personal income, population, per capita income (CAINC1)*. Retrieved February 10, 2022, from U.S. Bureau of Labor Statistics: https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=1&acrdn=6

California Association of Realtors. (2022a). *Historical housing data*. Retrieved February 9, 2022, from California Association of Realtors: https://car.sharefile.com/share/view/s0c02663a5c54e23a

California Association of Realtors. (2022b). *Housing affordability index - traditional*. Retrieved April 4, 2022, from California Association of Realtors: https://www.car.org/marketdata/data/haitraditional

Chakrabart, R., & Zhang, J. (2015, May). Unaffordable housing and local employment growth: evidence from California municipalities. *Urban Studies*, 1134-1151. Retrieved from https://www-jstor-org.libproxy.lib.csusb.edu/stable/pdf/26146040.pdf?refreqid=excelsior%3A36c232c2f65cfee47dc52686358fbc80&ab_segments=&origin=

Gallin, J. (2003, April 11). *The long-run relationship between house prices and income: evidence from local housing markets.* Retrieved April 16, 2022, from Federal Reserve Board: https://www.federalreserve.gov/pubs/feds/2003/200317/200317pap.pdf

Goodman, L. S., & Mayer, C. (2018). Homeownership and the American dream. *The Journal of Economic Perspectives*, 31-58. Retrieved from https://web-p-ebscohost-com.libproxy.lib.csusb.edu/ehost/pdfviewer/pdfviewer?vid=0&sid=5decaf6b-129d-4964-9ac3-a073a144ad37%40redis

Gunderson, R. J., & Sorenson, J. D. (2010). An examination of domestic migration from California counties. *Journal of Regional Analysis & Policy*, pp. 34-52. Retrieved from https://www.proquest.com/docview/856829807?accountid=10359&parent

SessionId=IJgo%2Bs0wcbdXgMR4a5F1Cd%2FjcTaUxFOCfvME%2F42iKok%3D&pq-origsite=primo

Hipsman, N. E. (2018, April 17). *Race and home price appreciation in the United States: 1992-2012.* Retrieved April 16, 2022, from Scholars at Harvard: https://scholar.harvard.edu/files/nhipsman/files/housing_race.pdf

Jensen, E., Jones, N., Orozco, K., Medina, L., Bolender, B., & Battle, K. (2021, August 4). *Measuring racial and ethnic diversity for the 2020 Census*. Retrieved March 18, 2022, from United States Census Bureau: https://www.census.gov/newsroom/blogs/random-samplings/2021/08/measuring-racial-ethnic-diversity-2020-census.html

Leonard, T., Yang, X., & Zhang , L. (2020, 11 21). The impact of land use regulation across the conditional distribution of home prices: an application of quantile regression for group-level treatments. *The Annals of Regional Science*, 655-676. Retrieved from https://link-springer-com.libproxy.lib.csusb.edu/article/10.1007/s00168-020-01032-z

Local area unemployment statistics (LAUS). (2022, January 6). Retrieved February 13, 2022, from State of California Employment Development Department: https://data.edd.ca.gov/Labor-Force-and-Unemployment-Rates/Local-Area-Unemployment-Statistics-LAUS-/e6gw-gvii

McLaughlin, J. E., McLaughlin, G., McLaughlin, J., & White, C. (2016, January 3). Using Simpson's diversity index to examine multidimensional models of diversity in health professions education. *International Journal of Medical*

Education, 1-5. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4715903/

Moye, R. (2014, February 17). Neighborhood racial–ethnic change and home value appreciation: evidence from Philadelphia. *Urban Geography*, 236-263. Retrieved from https://www-tandfonline-com.libproxy.lib.csusb.edu/doi/full/10.1080/02723638.2013.856194

Mulder, C. H. (2009, December 01). *The relationship between population and housing.* Retrieved April 16, 2022, from UNECE: https://unece.org/fileadmin/DAM/hlm/archive/Key%20note%20population%20and%20housing.pdf

Office of the Press Secretary. (2013, August 6). *Speeches and remarks*. Retrieved from The White House Archives: https://obamawhitehouse.archives.gov/the-press-office/2013/08/06/remarks-president-responsible-homeownership

Quigley, J. M., & Raphael, S. (2005, May). Regulation and the high cost of housing in California. *American Economic Review*, pp. 323-328. Retrieved from https://web-p-ebscohost-com.libproxy.lib.csusb.edu/ehost/pdfviewer/pdfviewer?vid=0&sid=4d007000-fd08-44e4-9e8a-021b55fafe19%40redis

R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. Retrieved 2022, from https://www.R-project.org/

Schuetz, J., Meltzer, R., & Been, V. (2011, February). Silver bullet or trojan horse? The effects of inclusionary zoning on local housing markets in the United States. *Urban Studies*, 297-329. Retrieved from https://www-jstor-org.libproxy.lib.csusb.edu/stable/43081738?sid=primo&seq=1

United States Census Bureau. (2021). *County population by characteristics*. Retrieved February 14, 2022, from United States Census Bureau: https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html

United States Census Bureau. (2022, March 26). *American community survey (ACS)*. Retrieved April 6, 2022, from United States Census Bureau: https://www.census.gov/programs-surveys/acs

Žalik, K. R. (2008). An efficient k′-means clustering algorithm. *Pattern recognition letters*, 1385-1391. Retrieved from https://www-sciencedirect-com.libproxy.lib.csusb.edu/science/article/pii/S0167865508000706