CALIFORNIA HOUSING MARKET ANALYSIS

Seattle Pacific University
Programming for Data Analytics: Python & Machine Learning
ISM6353

November 21, 2023

By:
Rithika Yalamanchi
Robert Ponce
Sarah Brachvogel
Xiaoyan Zhang

Abstract

This document outlines a comprehensive analysis conducted by a group of real estate agents tasked with assisting a high-profile client in purchasing a new estate in California for their adult son. The analysis focuses on estimating the median house value in California using a dataset. The primary goal is to examine the exploratory data analysis to generate bar charts, line charts, and scatter plots to visualize relationships using the variables and discover patterns in the data. The main analysis employs logistic regression which explores multiple models to predict the median house value. The document includes through data preparation, exploratory analyses, and a primary focus on logistic regression models, unraveling the intricate relationships between variables and focusing on crucial factors influencing housing prices in California which acts as a valuable resource for real estate professionals and decision-makers in the industry.

1. Business Insights

1.1. Business Problem and Motivations

We are a group of real estate agents that were recruited by our boss to help a high-profile client purchase a new estate for their adult son who is moving from Seattle to California. While the client does well financially, they want to purchase a modest house for their son as he is moving out and living alone for the first time. In order to best help determine houses that our client would be interested in, we need to do some preliminary research about the housing market in California to better understand the ever-changing housing market.

The goal is to use the given dataset to estimate the median house value of a house in California. This analysis is extremely important in order for us to help serve our client, and perform our jobs well for our boss. It will help us know how much money a person could expect to spend on a house, how much money will be necessary to plan to save for a house, or help

advise our client to move elsewhere in the event that the median house value is too high for their price range. Especially during this time period of extreme economic growth, the prices of houses have been increasing significantly. This data analysis will be helpful to see how much money the average person spends on a house in California. This information is not only beneficial to us to help better serve our client to tell them what to expect, but it is also beneficial to our real-estate company as a whole in the event that we continue to do business in the California real estate market.

1.2 Stakeholders

In the context of analyzing California housing data, potential stakeholders may include real estate agents, homeowners, homebuyers, property developers, construction companies, and more.

1. **Real-estate Agents:** Professionals in the real estate industry can use the analysis to provide valuable market information to their clients. This helps in setting realistic expectations and making informed buying or selling decisions.
2. **Property Developers:** Developers looking to invest in housing projects can benefit from understanding the trends in median house values. This information guides their decisions on where to develop properties for maximum return on investment.
3. **Prospective Homebuyers**: Individuals looking to purchase a home in California can use the analysis to gain insights into the median house values. This information helps them make informed decisions about their budget and where to search for a home.
4. **Government Agencies:** Public officials and policymakers can use the analysis to inform housing policies, urban planning, and allocation of resources for affordable housing initiatives.

5. **Construction Companies:** This includes the entities engaged in building and construction activities. They might be interested in the demand for new construction and renovation projects.

These stakeholders have a direct interest in the housing market data as it can impact their decision-making processes. Analyzing housing data can provide valuable insights into market trends, property values, demand and supply dynamics, and other factors that influence the real estate industry.

2. Data Analysis

2.1. Data Description

The data contains information for the 1990 California census, and it pertains to the houses found in a given California district and some summary stats about them based on the 1990 census data. In detail, the dataset contains 20,641 rows and 10 columns. It contains one row per census block group. A block group is the smallest geographical unit for which the U.S. Census Bureau publishes sample data (a block group typically has a population of 600 to 3,000 people). As it is shown in Figure 2.1, each column represents a feature such as housing median age, total rooms, total bedrooms, populations, median income, median house value and ocean proximity.

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY
5	-122.25	37.85	52.0	919.0	213.0	413.0	193.0	4.0368	269700.0	NEAR BAY
6	-122.25	37.84	52.0	2535.0	489.0	1094.0	514.0	3.6591	299200.0	NEAR BAY
7	-122.25	37.84	52.0	3104.0	687.0	1157.0	647.0	3.1200	241400.0	NEAR BAY
8	-122.26	37.84	42.0	2555.0	665.0	1206.0	595.0	2.0804	226700.0	NEAR BAY
9	-122.25	37.84	52.0	3549.0	707.0	1551.0	714.0	3.6912	261100.0	NEAR BAY

Figure 2.1 Data sample

This dataset is a modified version of the California Housing dataset available from Luís Torgo's page (University of Porto). Luís Torgo obtained it from the StatLib repository (which is closed now). The dataset may also be downloaded from StatLib mirrors. This dataset appeared in a 1997 paper titled *Sparse Spatial Autoregressions* by Pace, R. Kelly and Ronald Barry, published in the *Statistics and Probability Letters* journal. They built it using the 1990 California census data.

2.2. Data Preparation

Some data issues existed in the original dataset we downloaded from Kraggle. First of all, there are some missing values in the original data, so we deleted all the rows which contain missing values. Secondly, as you can see from *Figure 2.1,* the house locations are represented by longitude and latitude. In order to make further analysis easier , we converted longitude and latitude to cities and counties by Nomiatim which was imported from geopy.geocoders. After the two steps data cleaning process, the data point was reduced to 18,181 rows.

number	city	county	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	Berkeley	Alameda	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
1	Oakland	Alameda	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
2	Oakland	Alameda	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
3	Oakland	Alameda	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
4	Oakland	Alameda	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY
5	Oakland	Alameda	52.0	919.0	213.0	413.0	193.0	4.0368	269700.0	NEAR BAY
6	Oakland	Alameda	52.0	2535.0	489.0	1094.0	514.0	3.6591	299200.0	NEAR BAY
7	Oakland	Alameda	52.0	3104.0	687.0	1157.0	647.0	3.1200	241400.0	NEAR BAY
8	Oakland	Alameda	42.0	2555.0	665.0	1206.0	595.0	2.0804	226700.0	NEAR BAY
9	Oakland	Alameda	52.0	3549.0	707.0	1551.0	714.0	3.6912	261100.0	NEAR BAY

Figure 2.2. Data preparation

2.3. Exploratory Data Analysis (EDA)

For these exploratory analyses, we will focus on creating bar charts and line charts in order to visualize the change in the Median House Value by the other variables.

First, there is a bar chart *Figure 2.3.1* to show the house median value according to ocean proximity. As the chart shows, the houses located on the Island have extremely high median value compared to the other locations. The following is near the ocean which is slightly higher than near the bay. The houses with lowest median value are located inland.

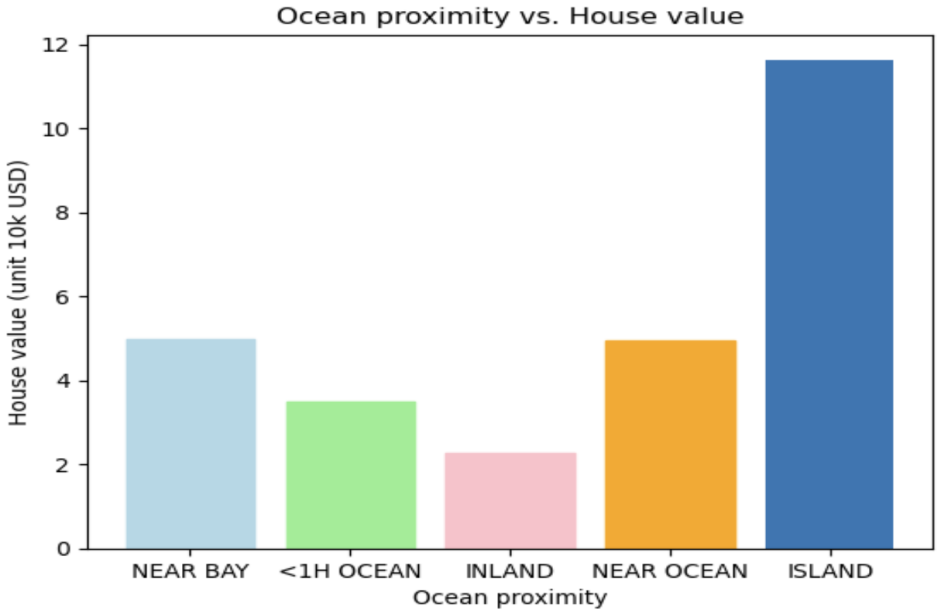

Figure 2.3.1. Ocean proximity vs house value

The following distribution plot *Figure 2.3.2* shows the house value's change by the age of the house. As you can see, the new house which is 0-5 years old has the highest value. However, the value of a house is not always decreasing when it's getting older. The house value hits its lowest point around 6-10 years old due to the expiration of all the fixtures of the house like HVAC, furnace, roof and so on. After the update, the value of the house will increase and keep stable for another decade until the next update time.

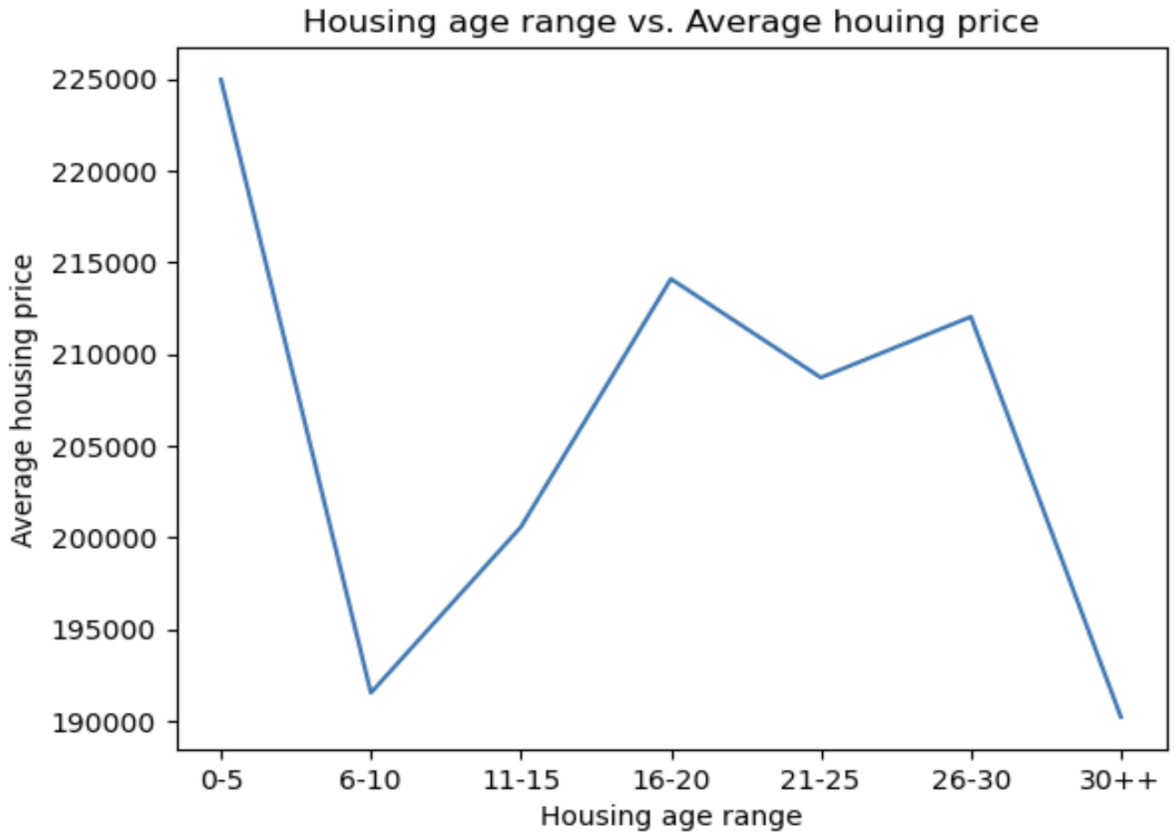

Figure 2.3.2. House age range vs average housing price

The final scatter plot (Figure 2.3.3) shows the positive correlation between housing price and median income. It is easy to understand that a higher income can afford a higher value

house. From this plot we also find out the quantity of low to middle value houses is larger than high value houses.

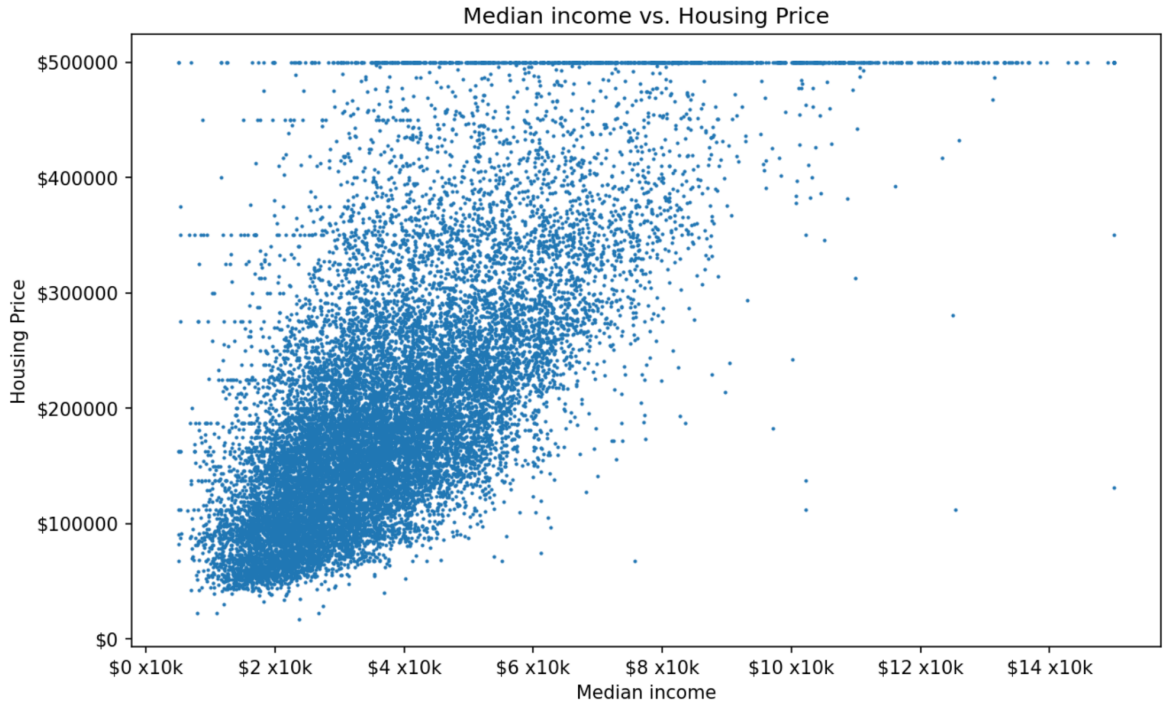

Figure 2.3.3. Median income vs housing price

3. Results and Evaluation

3.1. Overall Analysis

The main analysis used for this project includes logistic regressions with multiple models that have different variables tested for each regression. We ran three different logistic regressions in order to analyze and model our data. As previously mentioned we converted the latitude and longitude columns from the original dataset to city and county in order to better label our data to understand which California cities and counties had the highest median house prices. Overall, we kept all attributes that were included in the original dataset because they provided value to the models we wanted to run to analyze the data. We also included a “number” column in the final

dataset to assign a numerical value to each row of information for easier analysis. Unfortunately, two of the three logistic regressions were not viable because they had extremely high p-values for the counties and cities categories. If a p-value is greater than 0.05, there is no statistical significance between the values, meaning that there is no correlation worth noting in the final model. Both regression Models 1 and 3 were dropped from our final analysis because counties and cities values had statistically insignificant p-values, but regression Model 2 did not follow this pattern, so this was the model we explored to better understand and visualize the data.

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.196e+04	4.53e+04	0.485	0.628	-6.68e+04	1.11e+05
housing_median_age	538.7200	48.108	11.198	0.000	444.423	633.017
total_rooms	0.1429	0.787	0.182	0.856	-1.399	1.685
total_bedrooms	35.4761	7.188	4.936	0.000	21.387	49.565
population	-40.1943	1.055	-38.114	0.000	-42.261	-38.127
households	87.1810	7.765	11.227	0.000	71.960	102.402
median_income	3.613e+04	348.280	103.740	0.000	3.54e+04	3.68e+04
County_Alameda	7036.5505	4.53e+04	0.155	0.877	-8.18e+04	9.59e+04
County_Amador	-2.019e+04	4.78e+04	-0.423	0.673	-1.14e+05	7.34e+04
County_Anaheim	-1840.0957	7.83e+04	-0.023	0.981	-1.55e+05	1.52e+05
County_Angeles	5.433e+04	4.53e+04	1.201	0.230	-3.44e+04	1.43e+05
County_Arvin	-1.44e+04	6.39e+04	-0.225	0.822	-1.4e+05	1.11e+05

Figure 3.1.1. Logistic Regression Model 1: County p-values were too high, making the model unusable.

City_10	-2.884e+04	7.91e+04	-0.365	0.715	-1.84e+05	1.26e+05
City_108	1.617e+04	7.45e+04	0.217	0.828	-1.3e+05	1.62e+05
City_112	-1.157e+04	9.13e+04	-0.127	0.899	-1.9e+05	1.67e+05
City_13	1.152e+04	9.12e+04	0.126	0.900	-1.67e+05	1.9e+05
City_14	-5415.8144	9.12e+04	-0.059	0.953	-1.84e+05	1.73e+05
City_140	-1.92e+04	9.12e+04	-0.210	0.833	-1.98e+05	1.6e+05
City_144	2915.2980	7.91e+04	0.037	0.971	-1.52e+05	1.58e+05
City_152	3937.4950	9.12e+04	0.043	0.966	-1.75e+05	1.83e+05
City_16	-1.521e+04	9.13e+04	-0.167	0.868	-1.94e+05	1.64e+05
City_176	1.146e+04	9.12e+04	0.126	0.900	-1.67e+05	1.9e+05
City_180	-4.959e+04	9.13e+04	-0.543	0.587	-2.28e+05	1.29e+05
City_188	2.386e+04	9.12e+04	0.261	0.794	-1.55e+05	2.03e+05

Figure 3.1.2. Logistic Regression Model 3: City p-values were too high, making the model unusable.

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	3.057e+04	2567.388	11.906	0.000	2.55e+04	3.56e+04
housing_median_age	1035.9396	46.421	22.316	0.000	944.949	1126.930
total_rooms	-6.9150	0.811	-8.522	0.000	-8.506	-5.325
total_bedrooms	75.2449	7.342	10.248	0.000	60.853	89.637
population	-36.3313	1.103	-32.930	0.000	-38.494	-34.169
households	76.1506	7.937	9.594	0.000	60.593	91.708
median_income	4.067e+04	347.785	116.930	0.000	4e+04	4.13e+04
Proximity_INLAND	-6.779e+04	1295.862	-52.314	0.000	-7.03e+04	-6.53e+04
Proximity_ISLAND	1.93e+05	3.4e+04	5.669	0.000	1.26e+05	2.6e+05
Proximity_NEAR BAY	-1.551e+04	1835.488	-8.450	0.000	-1.91e+04	-1.19e+04
Proximity_NEAR OCEAN	1.089e+04	1600.391	6.806	0.000	7754.552	1.4e+04

Omnibus:	4731.398	**Durbin-Watson:**	0.994
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	18858.451
Skew:	1.245	**Prob(JB):**	0.00
Kurtosis:	7.324	**Cond. No.**	2.70e+05

Figure 3.1.3. Logistic Regression Model 2: Usable as p-values were not too high. All other variables were factored in, including the proximity to water variables, which needed dummy variables.

The p value, or probability value, tells how likely it is that the data could have occurred. Therefore, the smaller the p-value, the stronger the probability that the correlation didn't randomly occur. In order for the p-value to be significant, $p<0.05$. In the case of Model 2 all p-values for each variable were 0.000, meaning they were all significant to the model.

In terms of coefficients, a positive coefficient implies that the p-value will increase as the corresponding x-value increases. The coefficients measure the association between x-variables and the probability of an instance belonging to a category.

3.2. Coefficient Analysis

3.2.1. Water Proximity Coefficients

Based on this logic and looking at the logistic regression results, it was more likely for the median house price to increase if the house was closer to the water. Every other variable for distance from water follows this logic. For example, proximity to an island and proximity to an ocean have large positive coefficients (1.93e+05 and 1.089e+04 respectively), meaning that there is an increase in the median house value based on distance from the water, and this increase is extreme. With this, inland proximity has a coefficient of -6.779e+04, meaning that the median house price is significantly lower for a house that is more inland. The only coefficient that does not follow this logic is the proximity to a bay, which had a coefficient of -1.551e+04, meaning that the median house price is significantly lower if the house is closer to a bay. This is interesting because usually houses that have waterfront or are close to the water in some way, especially in sunny California, are more likely to have a higher median price point. Because the

weather in California is so nice, and there are a lot of recreational activities to be had near the water, properties closer to the water are always in higher demand. The analysis based on the near bay proximity variable, however, disproves that point when it comes to proximity to bays, which could be explained by the fact that it could be more challenging to take care of a house that is in a bay because it would be constantly exposed to heavy moisture winds year round. This exposure could be worse than normal waterfront properties because waves and winds are more susceptible to being trapped in a contained area for longer periods of time as opposed to the winds and waves a house may face being placed on a normal waterfront property. Depending on the orientation of the bay, there could also be less exposure to sunlight, and less room for housing property since it would be in a more confined area.

3.2.2. Median Income Coefficient

Additionally, it can be inferred that the median housing price is more likely to increase given a person's median income based on the value of the coefficient: 4.067e+04. Because this value is large and positive, this means that the more income a person or household has, the more likely the median house price is to increase. This makes sense economically because a place such as California that has great wealth coming from various places and industries such as Simi Valley and Hollywood, are more likely to have people that are making a large income. These people, whether famous or just rich, are more likely to have more money to spend on a house in California, therefore contributing to the increase in median housing prices.

3.2.3. Households Coefficient

Following this logic, the median housing price is more likely to increase if there is a bigger household of people (demonstrated by its coefficient: 76.1506). This makes logical sense because the more people there are in a household, the more likely the household is to purchase a

house that fits all of the tenants. Typically, houses that have larger households have more rooms, and houses that have more rooms are typically more expensive. While this coefficient is positive, it is rather small in comparison to the other coefficients previously analyzed, meaning that there isn't an extremely large increase in median housing price due to household size. This could be due to the fact that a lot of larger households are families, and a lot of these families could have children who end up sharing rooms. For example, if a family of six ended up purchasing a house with three bedrooms, sticking two children in each of the two rooms and the parents in the master bedroom, the household size wouldn't effect the house price as much because the family of six is paying for a smaller house than they would be if they were to buy a house with bedrooms for each individual kid. Therefore, household size would only affect the median house price to a certain degree.

3.2.4. Bedrooms Coefficient

In continuation of the above thought process, the median housing price is more likely to increase if there are more bedrooms (demonstrated by its coefficient: 75.2449). This makes sense because the more bedrooms included in a house, the larger the house will be. That means that there are more resources that go into building the house, and the house is overall bigger, meaning it will be more expensive. Once again, the number of bedrooms most likely does not affect the median price of the house to the degree of some other variables based on the knowledge of room sharing that was previously mentioned. In a recent study, it was also reported that about 60% of individuals claimed they shared rooms with their siblings or other family members growing up, showing that it is way more common for households to share rooms in possible search of more affordable housing (Shortsleeve, 2023). Continuing the previous example, if you have a household of six people, there is most likely going to be a necessity to place each person in a

bedroom. So while the number of bedrooms does increase the median housing price, it makes sense that the coefficient between household and number of bedrooms is almost equivalent because typically, those two are highly correlated.

3.2.5. Rooms Coefficient

The one variable that does not directly follow the logic previously discussed is the number of total rooms variable. Because the coefficient was -6.9150, this means that the median house price is more likely to be lower with more total rooms. This could be because there is a difference between having more bedrooms and having more rooms in general. Bedrooms typically are defined as having a door that can be opened and closed, a closet, a window, and 70 square feet of floor space (Vila, 2020). A room, however, doesn't need to have a closet, a door, or even a minimum sqft, it just needs to be separated by walls or some kind of partition (Dictionary.com). While a house is valued based on numerous metrics such as location, condition, and yes, number of rooms or bedrooms, it is more than that. A house is valued mainly by square footage, certainly over simply just the number of rooms. Nowadays most houses with open concept floor plans are in more demand, and therefore houses with a bunch of random rooms not even having to be a certain amount of square feet can allow for choppy and undesirable houses. Therefore, it can be understood that houses with more rooms can lead to a slightly lower median house value because of the demand of open-concept housing, as well as the possible lack of specificity in room square footage.

3.2.6. Population Coefficient

Another variable that is likely to contribute to a lower median housing price is the population coefficient which was valued at -36.3313. Because the coefficient is negative, this means that the greater the population becomes, the more likely there is to have a lower median

housing price. Usually the higher populated areas have a higher housing demand which means there are higher housing prices. This coefficient disproves this logic, but because the coefficient is not extremely large, the population doesn't have that much of an effect on the median housing prices. Nonetheless, this effect could be because of the type of population of people moving to the state of California. Because California is extremely close to the Mexican border, most of the people moving to California are immigrants with little money. From 2021 to 2022 alone, there have already been 126,000 immigrants that have moved to California alone (Perez et al., 2023). It is difficult for these immigrants coming to the U.S. with nothing to afford California housing prices. In an attempt to help the impoverished attain affordable housing, the state of California has already spent $757 million in funding to over 16 communities to create more affordable housing communities and transportation (State of California, 2023). Based on this analysis, simply because the population increase is causing a spike in the housing demand, the type of population of people moving to California and their overall income can have an impact on the median housing price.

3.2.7. Median Housing Age Coefficient

Finally, the last coefficient we looked at was the median age in which people are purchasing houses. The coefficient value came out to be 1035.9396, which means that the older a person is, the higher the median housing value is going to be. This is because, typically, people who are older and more established in their careers have more money to spend on buying a house. People living in California in their 20s or late teens are most likely college students or just starting their careers, and they are most likely renting or in college and cannot afford to purchase a house. Obviously those under the age of 18 are unable to purchase their own house anyways,

and those who are elderly and retiring are more inclined to spend more money on a nice house to live out the rest of their days as a retiree.

Conclusion

Overall it would have been nice to have multiple more accurate logistic regression models to compare, but we were grateful that at least one regression model worked and made sense with the data. With this Model 2 we were able to form the above evaluations of the analyzed data. These evaluations are crucial to the following recommendations that we are making for our client.

4. Recommendations

Based on the results derived from the evaluation of the model analysis, the team came up with a few recommendations which would be helpful for the stakeholders to improve the housing situation in California. These recommendations can help the stakeholders to make informed decisions to navigate the dynamic California housing market effectively.

1. **Focus on waterfront properties**: Upon the analysis, stakeholders are advised to prioritize marketing and promoting waterfront properties due to their positive impact on median house values. Real estate agents should focus on waterfront properties when assisting clients and encourage clients to consider properties closer to water, especially islands or the ocean, for potential higher median house values.
2. **Target high income demographics:** Since there is a positive correlation between median income and median house values, real estate agents are advised to tailor their marketing strategies by highlighting the amenities and features of the locality through marketing campaigns or digital marketing to promote areas with better amenities and services.
3. **Review household size:** From the analysis, it is evident that larger households tend to be associated with higher median house prices. Stakeholders can improvise their marketing and promotional efforts by focusing on families or larger households that include properties with more bedrooms, larger living spaces, or family-friendly layouts.

4. **Population and location impact:** Population dynamics is a critical variable for the stakeholders, and they need to be cautious about this because the negative correlation between population and median house values contradicts typical expectations. So, it is crucial to investigate the demographics of the population, such as income levels and housing preferences as higher populations usually indicate higher demand and prices. Understanding the type of population may provide insights into housing demand in various localities in the country.
5. **Age demographics:** Finally, the team would recommend the stakeholder focus on this variable as there is a positive correlation between median age and median house values. Stakeholders can use these demographics to promote properties by stating the accessibility, amenities of the locality, budget, and requirements of the clients. Often, older individuals may have higher incomes, savings and be more willing to invest in buying homes so it's crucial to understand the age demographics based on the patterns.
6. **Suggestion for client:** The team recommends exploring modest homes within the client's budget, as this aligns with the practical and suitable living space preferred for their adult son, who is moving out and living alone for the first time. Additionally, considering the client's son is an older individual, targeting areas with demographics that align with his age group can provide a sense of community and amenities suitable for independent living.

4.1. Future actions

While this project has provided few valuable insights into the housing analysis in various locations in California, there are certain considerations and improvements that can lead to better decision-making opportunities for the stakeholders interested in this analysis of the project.

1. **Explore alternate models:** To generate more insights about this project, it's essential to explore alternative predictive models for analysis, such as decision trees or ensemble methods as it can provide a more comprehensive understanding of complex relationships within the dataset. However, implementing alternative models may require additional resources, refined dataset with additional variables for evaluation and expertise to perform the analysis.

2. **Enhance digital marketing:** Stakeholders may be required to conduct digital marketing campaigns for waterfront properties, emphasizing unique features and amenities to attract the clients. Over a period, waterfront properties have shown a significant rise in interest and an increase in median house values with increased demand. But the stakeholder needs to be aware of the limited availability of waterfront properties which may restrict the market reach. So, it's important to understand the dynamics of household types and size within a locality that are available in the market for attracting clients.
3. **Monitoring market trends:** It is crucial to regularly monitor market trends and adjust recommendations based on changing dynamics. By leveraging data analytics tools it's easy to track real-time market dynamics and identify patterns or shifts that may impact the housing market. And stakeholders need to be aware of the rapid market fluctuations which may be a challenge in predicting and responding to trends effectively.
4. **Develop personalized approach:** One of the potential actions is to develop a personalized approach for each client by considering a combination of factors such as lifestyle, family size, income, budget, and specific location preferences. It can enhance client satisfaction, build trust, and increase the likelihood of successful property matches but this approach may require additional time and resources to prove a personalized approach.
5. **Provide housing resources:** Stakeholders need to provide clients with a comprehensive understanding of the California housing market through educational resources, such as informational guides or webinars, to provide clients with insights into market trends, influencing factors, and potential investment opportunities. Through regular communication updates on the California housing market ensures that clients are well-informed and empowered to make decisions aligned with their goals. However, this may require ongoing research and communication efforts and some clients may prefer a more hands-off approach, relying on the agent to navigate market complexities.

Overall, by implementing a holistic approach that combines regular market trend monitoring, personalized client interactions, and transparent communication about the California housing market. This decision aligns with the goal of staying responsive to market changes while

ensuring that clients receive tailored support and are well-informed throughout their real estate journey.

4.2. Faith and ethical implications

Proverbs 4:7: "The beginning of wisdom is this: Get wisdom. Though it cost all you have, get understanding"

During the market research, we came across this verse that was relevant to the project. This verse emphasizes the importance of acquiring wisdom and understanding, even if it requires a significant investment. This seemed related to our project as it encourages a thoughtful and intentional pursuit of wisdom and understanding. It was crucial to include ethical considerations in the project as it ensures data privacy, avoiding biases in the analysis, and providing fair and transparent recommendations which aligns with the principles. Also, acquiring wisdom is crucial in the decision-making process, especially in real estate where significant financial transactions are involved. The project involves analyzing housing data to provide valuable insights for decision-making where wisdom guides the path making informed decisions based on a deep understanding of the data and its implications. For the project, this aligns with the need for careful analysis, ethical considerations, and a deep understanding of the data to provide valuable insights for the stakeholders.

4.2.1. Ethical implication of the project

Ethical considerations in this project center on crucial elements like data privacy, transparency, and the conscientious application of analytical discoveries. These aspects guarantee the confidentiality of housing data, upholding the privacy rights of individuals. Our emphasis was on decision-making and recommendations resulting from the analysis, placing a priority on fairness and equity while actively avoiding any trace of discrimination or bias in housing practices. In the context of data analytics ethics, our commitment was to responsible data handling, ensuring data accuracy, and utilizing analytics for positive and constructive purposes.

In the context of business, ethics underscore values of honesty, integrity, and equitable treatment for employees, customers, and competitors where transparent and ethical business practices not only build trust but also contribute positively to the community. It encourages

treating others as one would like to be treated, extending to employer-employee relationships with a focus on fair wages, good working conditions, and mutual respect.

This comprehensive ethical framework not only ensures the responsible execution of this project but also reflects our commitment to values that extend beyond mere analytics, resonating with broader principles of justice and integrity.

4.2.2. Team's personal reflection

From the team's personal experience, the ethical dilemmas and considerations that emerged during the project were about handling sensitive information involving the privacy and confidentiality of individuals represented in the housing data. We have addressed these ethical challenges through careful decision-making, open communication, fair analysis, and a commitment to ethical standards and successfully completed our project.

Ethical considerations in handling sensitive information for making choices about data handling involve respecting the privacy and confidentiality of individuals represented in the housing data, this includes personal details about the homeowner's geographic locations. To address this concern, the team included robust data protection measures for the sensitive information and ensured that data is used only for the intended purposes of the analysis. Also, it was crucial to address ethical implications related to making choices about data handling, analysis methods, and formulating recommendations to ensure that the selected course of action aligns with integrity, transparency, and fairness.

Through this project, the team has observed that ethical challenges often benefit from open communication and being transparent, which ensures that everyone involved is aware of the ethical considerations, potential challenges, and the steps taken to address them. It helped to foster a collaborative approach to ethical decision-making and encouraged a shared commitment to maintaining high ethical standards throughout the project.

5. References

1. Dictionary.com. (n.d.). Room definition & usage examples. Dictionary.com. https://www.dictionary.com/browse/room

2. Perez, C. A., Mejia, M. C., & Johnson, H. (2023, January). Immigrants in California - public policy institute of California. Immigrants in California. https://www.ppic.org/wp-content/uploads/jtf-immigrants-in-california.pdf
3. Shortsleeve, C. (2023, July 26). 6 common questions about room sharing for kids, answered. Sleep Enlightened. https://www.saatva.com/blog/siblings-sharing-a-bedroom/#:~:text=A%20recent%20CBS%20News%20poll,18%2C%20kids%20share%20a%20room
4. State of California. (2023, August 30). California invests $757 million to create affordable housing and Clean Transportation. California Governor. https://www.gov.ca.gov/2023/08/30/california-invests-757-million-to-create-affordable-housing-and-clean-transportation/#:~:text=%E2%80%9CWith%20over%20%24757%20million%20in,families%20with%20children%20can%20prosper.%E2%80%9D
5. Vila, B. (2020, February 11). What makes a room a “bedroom”? Bob Vila. https://www.bobvila.com/articles/406-what-makes-a-room-a-bedroom/
6. *Bible gateway passage: Proverbs 4:7 - king james version*. Bible Gateway. (n.d.). https://www.biblegateway.com/passage/?search=Proverbs+4%3A7&version=KJV