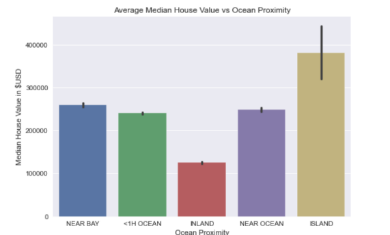
**Callum, Question: The specific research question is: how does ocean proximity affect house value?**

### Data Cleaning

We see that there is only 5 homes in the island data. These are homes which are classified as being on an island.

**Fig 1: Average Median House Value vs Ocean Proximity**



According to the law of small numbers, small random changes have a large apparent effect on the analysis of the data. This is shown by the large error bar of the values of the island data, marked by the black line in the graph. Such a small dataset of five homes, means that if even one home was an outlier, then this whole dataset would be largely affected by that one outlier.

For now on, we will remove the island data, as this data is unreliable with such a small data set.

Now, we need to scan for further outliers with the rest of the data using a box plot. It will mark outliers with a dot (since they lie further than 1.5 multiplied by the interquartile range).

**Fig 2: Median House Value vs Ocean Proximity**

Chart, box and whisker chart

Description automatically generated

It is important to say however, that all the data is shown to be in 'reasonable' values (no median house prices above $10,000,000 for example). Furthermore, excluding homes in the inland, there are not too many outliers present. We can forgive the inland data for having outliers. This is since there were over 6500 homes, which means there will be plenty of individual homes which could be outliers in this large sample. Furthermore, the regions in California considered 'inland' are vast and very different, so the house prices will be very varied. For example, you cannot compare the house prices to the relatively built-up urban Sacramento (the capital of California), to a small country town in the Mohave Desert except they are both classified as 'inland'.

**Analysis of Data**

Looking at Fig 3 We see that the homes located near the bay have a higher median housing value in 1990 compared to the homes in any of the other areas (about $9552 more than homes near the ocean, despite both being near the ocean). Now this makes sense, as any homes which are near the bay are in the San Francisco Bay area.

**Fig 4: Location of homes ‘near the bay’**

Chart, scatter chart

Description automatically generated

We indeed see that the homes near the 'bay' are all in a similar location. That location is the San Francisco Bay area. From 2014 to 2020, San Francisco was ranked the most expensive city to live in the United States (Bote, 2022). Even in 1998 (close to the year the data was collected), a suburb called Atherton in San Francisco was the most expensive zip code in the United States (DePietro, 2018). This is perhaps why San Francisco data was singled out, since its homes would be significantly expensive, compared to perhaps the rest of California, and may be outliers which may skew the data. However, we see that the difference of the median value of homes in San Francisco compared to other homes near the ocean is quite small ($9552), we need to consider that if we adjust for inflation ($1 in 1990 is now $2.21 today ((CPI inflation calculator, n.d.)), we see that the difference in the median house value is $18,750. This does still show that even in 1990, San Francisco had substantially more expensive housing on average than the rest on the state. Factoring how the US also had an interest rate of 8.31% compared to 0.75% today (Macrotrends, 2022), this shows that this different in average price would cost more in interest then compared to today.

**Analysis of General Trend**

**Fig 4: Median House Value vs Ocean Proximity**

**Chart, bar chart

Description automatically generated**

Even in homes not in San Francisco, we see a general trend. As homes get further away from the ocean, we see the average price of those homes being cheaper. We see that inland homes are significantly less expensive than homes near the coast ($103,068 less than the homes that are near the coast, and $100,000 less on average then homes less than an hour from the coast).

### Why are housing prices closer to the ocean more expensive?

The law of supply and demand demonstrated how an increased demand for housing in a particular area, as well as the limited supply of housing combine to drive the price upward (Chappelow, 2019).

### Possible Demand Factors

#### Population

You would expect that with a higher population, there would be a higher demand for homes (with more people needed to live in housing), and as such driving up housing prices (Mulder, 2018). However, is this really the case?

We will investigate this, by calculating a test score for regression line of best fit for median house value over for ocean proximity and population.

𝑦̂ = 𝛽0 + 𝛽1x+ 𝛽2x + 𝛽3x+ 𝛽4x = 223689.99 -100860.63x + 1.086.5640x, 2.652.7500x - 200.0000x

A test score of 0.24, shows us that population has only a weak positive correlation with the median house price based on the knowledge of the home’s ocean proximity. Therefore, it seems that population is not related to why homes near the ocean have a higher house value then homes inland.

#### Median Income

#### An increase in median income theoretically increases demand for housing as well. As income increases, consumers will have more money to perhaps upsize their dwelling (with their increase in income allowing them to be supported with a mortgage to buy more expensive housing) (admin, 2020).

**Fig 5: Average Median Income vs Ocean Proximity**

Chart, bar chart

Description automatically generated

This seems to be much more of a clear indicator of how the median housing value will be affected in each area then the population. We see that this graph looks like Fig 3. We see also that median income decreases as the homes are away from the ocean. This is in line with how housing prices decrease as you move further inland. Calculating the test score for the relationship between median income and median house value to ocean proximity, we see that it is 0.57, a big improvement compared to the regression model showing how population and ocean proximity relate to the home’s median house value.

#### The Climate

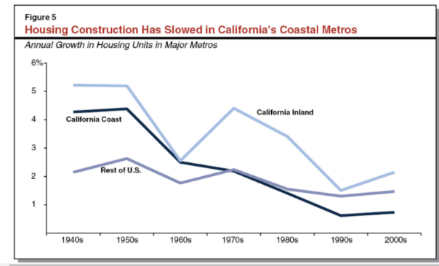
In summer San Francisco (which is near the ocean) has an average temperature of a pleasant 20C (US Department of Commerce [USDOC]). This can be compared to the temperature in the Mojave Desert (the most inland portion of California) in summer, that area is a boiling 36C on average (USDOC, 2015). Clearly, San Francisco and areas near the ocean in general, have a nicer climate then areas further inland. Therefore, there will be higher demand for living in areas near the ocean, which will thus drive-up housing prices.

**Possible Supply Factors**

#### Less homes have been built then has been demanded in California's coastal areas

There have been less homes being built in California's coastal areas then even in the inland areas of California. This is shown by this graph...

**Fig 6: Annual Growth in Housing Units in California vs Rest of US (Legislative Analyst's Office [LAO], 2015)**



Land Prices are Expensive

The cost to buy land to build housing in California's coastal areas is one of the most expensive in the United States (LAO, 2015). Therefore, there is not much housing built because of the huge cost to buy the land to build it on. Land prices in the inland regions of California, however, are at or below the US average (LAO, 2015).

**Reference List:**

CPI inflation calculator. (n.d.). *$1 in 1990 → 2022 | Inflation Calculator*. Www.in2013dollars.com. Retrieved May 21, 2022, from https://www.in2013dollars.com/us/inflation/1990?amount=1#:~:text=Value%20of%20%241%20from%201990

Effectivioology. (2014). *The Law of Small Numbers: Overestimating the Representativeness of Small Samples – Effectiviology*. Effectiviology. https://effectiviology.com/law-of-small-numbers/

Mulder, C. (2018). *Sixty-ninth session of the UNECE Committee on Housing and Land Management Key note presentation The relationship between population and housing*. https://unece.org/fileadmin/DAM/hlm/archive/Key%20note%20population%20and%20housing.pdf

Legislative Analyst's Office. (2015, March 17). *California’s High Housing Costs: Causes and Consequences*. Ca.gov. https://lao.ca.gov/reports/2015/finance/housing-costs/housing-costs.aspx

DePietro, A. (2018, July 31). *Housing 1998-2018: America’s Most Expensive Zip Codes, Then And Now*. Forbes. https://www.forbes.com/sites/andrewdepietro/2018/07/31/housing-1998-2018-most-expensive-zip-codes/?sh=7a49f9e41aea

Chappelow, J. (2019, September 29). *Law of Supply and Demand*. Investopedia. https://www.investopedia.com/terms/l/law-of-supply-demand.asp#:~:text=The%20law%20of%20demand%20says

admin. (2020, November 29). *Factors Affecting Housing Market Supply and Demand In Australia*. SuburbsFinder. https://www.suburbsfinder.com.au/resources/factors-affecting-housing-market-supply-and-demand-in-australia/#:~:text=An%20increase%20in%20income%20means

Bote, J. (2022, March 2). *SF dethroned as most unaffordable housing market in America*. SFGATE. https://www.sfgate.com/realestate/article/San-Francisco-not-most-unaffordable-city-16971811.php

US Department of Commerce, N. (2022, May 21). *Climate*. Www.weather.gov. https://www.weather.gov/wrh/Climate?wfo=mtr

Macrotrends. (2022). *Federal Funds Rate - 62 Year Historical Chart*. Macrotrends.net. https://www.macrotrends.net/2015/fed-funds-rate-historical-chart