

TBD*

TBD

Ziheng Zhong

April 1, 2024

TBD

Table of contents

1	Introduction	2
2	Data	2
2.1	Source	4
2.2	Method	4
3	Results	4
3.1	Data Trend	4
3.2	Modeling	4
4	Discussion	5
4.1	Demographic Shifts	5
4.2	Health-related Behaviors	5
4.3	Government Policies	5
4.4	Environmental Changes	5
4.5	Possible Improvements	5
5	Conclusion	5
	References	6

*Code and data are available at: https://github.com/iJustinn/House_Price.git

Table 1: Summary statistics of the California housing dataset

longitude	latitude	housing_median_age
Min. :-124.3	Min. :32.54	Min. : 1.00
1st Qu.: -121.8	1st Qu.:33.93	1st Qu.:18.00
Median :-118.5	Median :34.26	Median :29.00
Mean :-119.6	Mean :35.63	Mean :28.64
3rd Qu.: -118.0	3rd Qu.:37.71	3rd Qu.:37.00
Max. :-114.3	Max. :41.95	Max. :52.00
total_rooms	total_bedrooms	population
Min. : 2	Min. : 1.0	Min. : 3
1st Qu.: 1448	1st Qu.: 296.0	1st Qu.: 787
Median : 2127	Median : 435.0	Median : 1166
Mean : 2636	Mean : 537.9	Mean : 1425
3rd Qu.: 3148	3rd Qu.: 647.0	3rd Qu.: 1725
Max. :39320	Max. :6445.0	Max. :35682
households	median_income	median_house_value
Min. : 1.0	Min. : 0.4999	Min. : 14999
1st Qu.: 280.0	1st Qu.: 2.5634	1st Qu.:119600
Median : 409.0	Median : 3.5348	Median :179700
Mean : 499.5	Mean : 3.8707	Mean :206856
3rd Qu.: 605.0	3rd Qu.: 4.7432	3rd Qu.:264725
Max. :6082.0	Max. :15.0001	Max. :500001

1 Introduction

2 Data

Data used in this paper was cleaned, processed and tested with the programming language R (R Core Team 2022). Also with support of additional packages in R: `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `janitor` (Firke 2023), `readr` (Wickham, Hester, and Bryan 2023), `knitr` (Xie 2014), `rstanarm` (Goodrich et al. 2023), `modelsummary` (Arel-Bundock 2023), `tidybayes` (Kay 2023), `loo` (Vehtari et al. 2023), `testthat` (Wickham Year of publication), `KableExtra` (Zhu 2023).

Table 2: Count of missing values for each variable

Variables	MissingValues
longitude	0
latitude	0
housing_median_age	0
total_rooms	0
total_bedrooms	207
population	0
households	0
median_income	0
median_house_value	0
ocean_proximity	0

Table 3: Count of missing values for each variable after cleaning

Variables	MissingValues
longitude	0
latitude	0
housing_median_age	0
total_rooms	0
total_bedrooms	0
population	0
households	0
median_income	0
median_house_value	0
ocean_proximity	0

Table 4: summary of modeling

2.1 Source

2.2 Method

3 Results

3.1 Data Trend

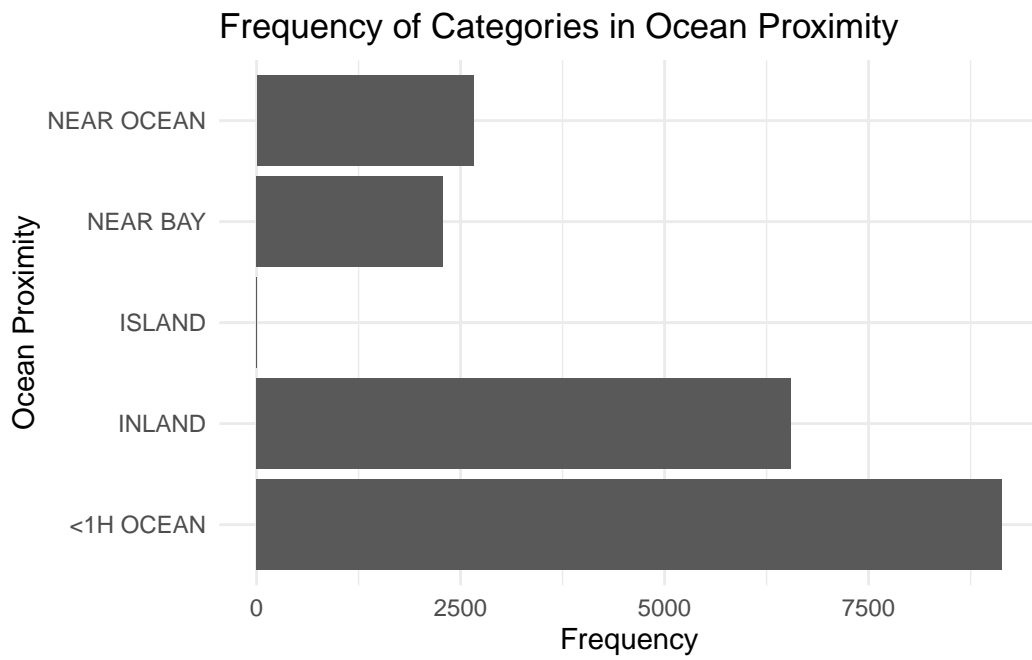


Figure 1: Frequency of Ocean Proximity

3.2 Modeling

Model Results

Table 5: cross validation of modeling

4 Discussion

4.1 Demographic Shifts

4.2 Health-related Behaviors

4.3 Government Policies

4.4 Environmental Changes

4.5 Possible Improvements

5 Conclusion

References

- Arel-Bundock, Vincent. 2023. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. <https://vincentarelbundock.github.io/modelsummary/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, Sam Brilleman, and and others. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. <https://mc-stan.org/rstanarm>.
- Kay, Matthew. 2023. *Tidybayes: Tidy Data and Geoms for Bayesian Models*. <https://mjskay.github.io/tidybayes/>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Vehtari, Aki, Jonah Gabry, Yuling Yao, and Andrew Gelman. 2023. *Loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models*. <https://mc-stan.org/loo>.
- Wickham, Hadley. Year of publication. *Testthat: Get Started with Testing*. <https://CRAN.R-project.org/package=testthat>.
- . 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Xie, Yihui. 2014. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.
- Zhu, Hao. 2023. *kableExtra: Construct Complex Table with ‘Kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.