TBD*

TBD

Ziheng Zhong

April 8, 2024

TBD

Table of contents

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

Table 1: Summary statistics of the California housing dataset

Table 2: Count of missing values for each variable

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 3: Count of missing values for each variable after cleaning

Table 4: TBD

Table 5: TBD

2.1 Source

2.2 Method

3 Results

3.1 Data Trend

3.2 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

A Appendix

A.1 Datasheet

Motivation

Composition

Collection process

Preprocessing/cleaning/labeling

Uses

Distribution

Maintenance

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 Grolemund, 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.