# Interim Proposal

**Group P3 members:** Owen Kwong, Eugene Ng, Marc Shi, Amanda Zhang


**Data Source:** This dataset contains information collected by the U.S Census Service in 1970 concerning housing in the area of Boston Mass. It was obtained from the StatLib archive  https://lib.stat.cmu.edu/datasets/boston and originally published by Harrison and Rubinfeld (1978) to examine the relationship between pollution and housing prices. All variables were collected in 1970 in this dataset.

**Variables Description:**
Each observation in the dataset represents a neighborhood in Boston. The dataset includes 14 potential explanatory variables with the last one MEDV being considered the response variable in a regression (see reason next section).
1. CRIM - number of crimes per person in the town, data collected from the FBI
2. ZN - percentage of land in each town that is set aside for residential lots over 25,000 square feet, collected from the Metropolitan Area Planning Commission
3. INDUS - percentage of land used for businesses that are not shops or stores in each town, collected from the Department of Commerce and Development
4. CHAS - dummy variable for if land touches Charles River (1 if yes; 0 otherwise) collected by US Census
5. NOX - amount of nitric oxide pollution in the air (measured in parts per 10 million) collected from Department of Commerce and Development
6. RM - average number of rooms per home in a town, collected from US Census
7. AGE - proportion of owner-occupied units built prior to 1940, collected from US Census
8. DIS - average distance from five Boston employment centers (miles), collected by Harvard University
9. RAD - how easy to access highways from the neighborhood (standardized 1 to 24), collected from MIT
10. TAX - full-value property-tax rate per $10,000, collected from Massachusetts Taxpayers Foundation
11. PTRATIO - average number of students per teacher in each town, from Massachusetts Department of Education
12. B - measure related to proportion of Black residents in town: $1000(Bk - 0.63)^2$ where Bk is proportion, collected from US Census
13. LSTAT - percentage of population in each neighborhood considered lower income, collected from US Census
14. MEDV - Median price of owner-occupied homes in $1000's, collected from US Census

**Research Question and Background:**
   Our research question asks what is the relationship between some of the factors (e.g., crime rate, nitric oxide pollution) and the median value of owner-occupied homes in Boston neighborhoods ?
   In the literature, Ceccato and Wilhelmsson's study (2018) emphasizes that neighborhoods with high crime rates are associated with lower housing prices. In contrast, neighborhoods with higher levels of accessibility to city centers are associated with higher housing prices. Moreover, Jones et al. (2023) show that proximity to greenspaces is also positively related to housing prices.
   Therefore, crime rates (CRIM), distance to employment centers (DIV), and proximity to Charles River (CHAS) are particularly interesting to us to look at their relationship with the median price of homes (MEDV).

**Member Resposibilities:**
Introduction: Amanda Zhang
Explorative analysis and findings: Marc Shi
Analysis explanation and methodology: Eugene Ng
Conclusion and Discussion: Owen Kwong

**References:**

Ceccato, V., & Wilhelmsson, M. (2018). Does crime impact real estate prices? An assessment of accessibility and location. In G. J. N. Bruinsma & S. D. Johnson (Eds.), Oxford Handbook on Environmental Criminology (pp. 518-544). Oxford University Press.

Jones, C. A., Dunse, N. A., Li, E., & Liu, Y. (2023). Housing prices and the characteristics of nearby green space: Does landscape pattern index matter? *Land, 12*(2), 496. https://doi.org/10.3390/land12020496

Harrison, D., & Rubinfeld, D. L. (1978). Hedonic prices and the demand for clean air. Journal of Environmental Economics and Management, 5(1), 81-102. https://doi.org/10.1016/0095-0696(78)90006-2