

# HPA Data Assessment

## 1 About the Data

The underlying dataset in this assessment shows whether an existing lease on a home was renewed or not. Each row in the data represents a lease on a home, its characteristics, and whether the resident renewed the lease for the next term. Below is the description of each column in the dataset. The data in all columns is binary 0 or 1 (except the lease ID).

Column	Description
lease_id	Unique ID for each lease
no_rent_change	The lease rent was not changed for the next term
rent_change_10	The lease rent was changed by at most 10% for the next term
rent_change_20	The lease rent was changed by more than 10% but at most 20%
lease_length_1	The resident has lived for at most 1 year on the lease
lease_length_2	The resident has lived for more than 1 but at most 2 years on the lease
lease_length_3	The resident has lived for more than 2 but at most 3 years on the lease
age_range	Set of columns indicating average age of residents in the household
NoFinesViolations	Indicates the resident had no fines or violation on the lease
PositiveSurvey	Indicates the resident provided a positive feedback on the survey
LatePayments	Indicates the resident has never been late on making payments
HOA_mandatory	Indicates whether there's a mandatory HOA fee on the lease
Renewed	Indicates whether the resident renewed the lease or not

## 2 Instructions

- Please answer the questions in the section below
- Please include your code and thought process used to navigate the questions below. Code is required for this exercise, and Python scripting language (or Jupyter Notebook) is preferable for solutions. Please include your detailed thought process and analysis

either as a part of the Jupyter Notebook, or as a separate document which duly references the code

- Clear, concise, and thorough answers are appreciated. While you can include technical jargon in your analysis, please include plain text conclusions and insights as well.
- Feel free to include charts or tables, but please provide context around all analysis

### 3 Questions

**Part I:** The business would like to understand what are the drivers that explain whether a resident will renew the lease or not. In other words, to understand what are the renewal promoters and detractors. At this stage, the purpose is to keep the analysis simple, interpretable, and easy to communicate, while also ensuring effectiveness and sound statistical foundation.

1. Please conduct analysis to help the business with the aforementioned objective
2. Please include detailed visualization, charts, tables, graphs, etc. to support your case
3. Please utilize the core probability and statistical theory to provide strong foundation and generate confidence in your results and arguments, and include it in your submission clearly and explicitly
4. Please include a final section at the end that presents your analysis, insights, and conclusions to the business

**Part II:** Assuming the business is interested in predicting resident renewal, use the dataset to create a renewal likelihood prediction model. Please include supporting work to determine and demonstrate how effective is the prediction model.