# Brief Summary

Indiana, located in the Midwestern United States, was selected for a comprehensive analysis of its housing and demographic characteristics. The study focused primarily on Indianapolis and surrounding ZIP codes, leveraging multiple datasets to provide an in-depth view of the region's housing landscape.

The datasets included demographic information detailing household compositions, racial distributions, and median income levels for each ZIP code. Housing price data and rental costs were extracted from CSV files, while real-time rental listings were scraped from an online platform using Selenium. Additional housing metrics, such as total housing units and vacancy rates, were retrieved via the U.S. Census API. Policy-related data was also extracted from PDF documents to provide context for housing trends and challenges in Indiana.

The data cleaning process involved standardizing column names and ensuring consistent data types across all datasets. ZIP codes were converted to a unified numeric format to facilitate smooth merging of data. Unnecessary columns were removed, and values were reformatted to align with analysis requirements. A calculated field, the price-to-income ratio, was introduced to assess housing affordability.

Integration steps merged datasets on common ZIP code attributes, creating a comprehensive dataset that links demographic profiles, housing prices, rental costs, and trends. The final integrated dataset was saved for analysis, and intermediate outputs, such as scraped rental data, were preserved for transparency and future use.

This project offers a holistic perspective on Indiana's housing conditions, providing insights into affordability, availability, and demographic influences within the state’s diverse communities. The workflow serves as a robust framework for similar housing studies in other regions.