In my Exploratory Data Analysis, I began by examining the dataset's structure and summary statistics using df.info() and df.describe(), which confirmed that I had a reliable dataset with a manageable number of rows, columns, and minimal (if any) missing values. Next, I generated **histograms** and **KDE plots** for three important numeric features—**NumUnderPov** (poverty rate), **MedRentPctHousInc** (median rent as a percentage of household income), and **PctHousOwnOcc** (percentage of owned housing)—to understand their overall shape and dispersion. The **histograms** provided a clear look at the frequency distribution, revealing in some cases a right-skewed pattern (where most data clustered at lower values but a longer tail extended toward higher values). Adding **KDE curves** helped smooth out the distribution, making it easier to spot whether the data were unimodal or hinted at multiple clusters. For instance, **NumUnderPov** typically centered around a moderate percentage, yet there were a few observations with significantly higher poverty rates, indicating possible outliers or naturally high-poverty regions.

The **boxplots** further highlighted these extremes by showing how far certain data points deviated from the bulk of the distribution. In particular, **MedRentPctHousInc** had a handful of areas with rent burdens far above the typical range, which underscores the need to investigate whether those values are genuine high-cost regions or potential data-entry anomalies. Similarly, **PctHousOwnOcc** generally hovered in a middle range but included a few notable low-percentage ownership areas—possibly high-rental markets. Visually, these boxplots are especially valuable because they present the median, quartiles, and any data points that lie far from the central "box." Identifying these outliers or unusually large spreads within specific variables prompted me to consider potential data transformations (such as a log transform on heavily skewed variables) or targeted feature engineering (like capping extreme values or creating ratio features) to ensure the modeling process remains robust. Overall, these visualizations not only confirmed the dataset's viability but also provided key insights about the natural variability and distributional quirks of each feature, ultimately guiding the next steps in the data preparation and modeling workflow.