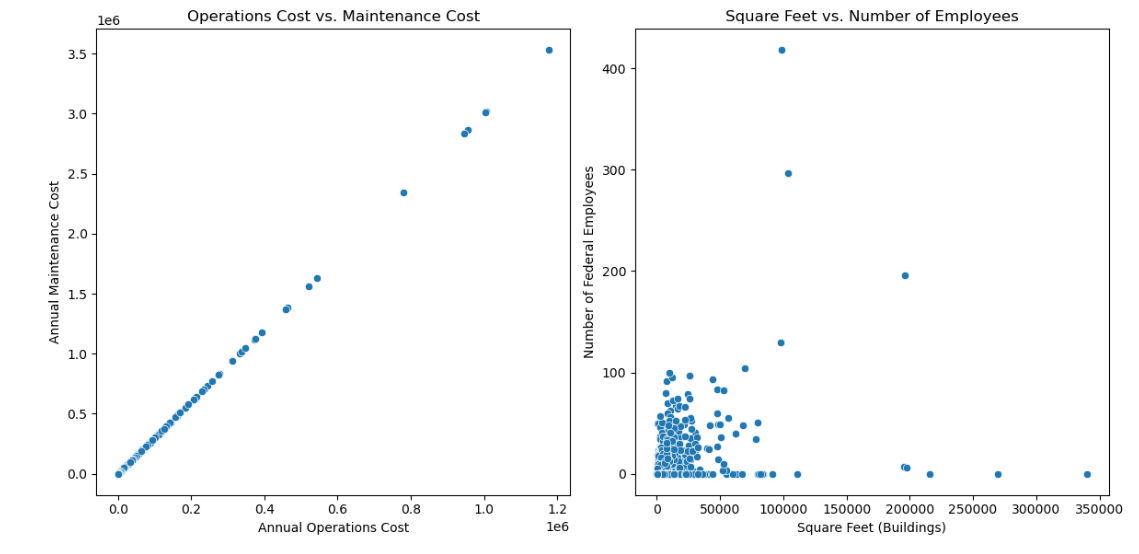
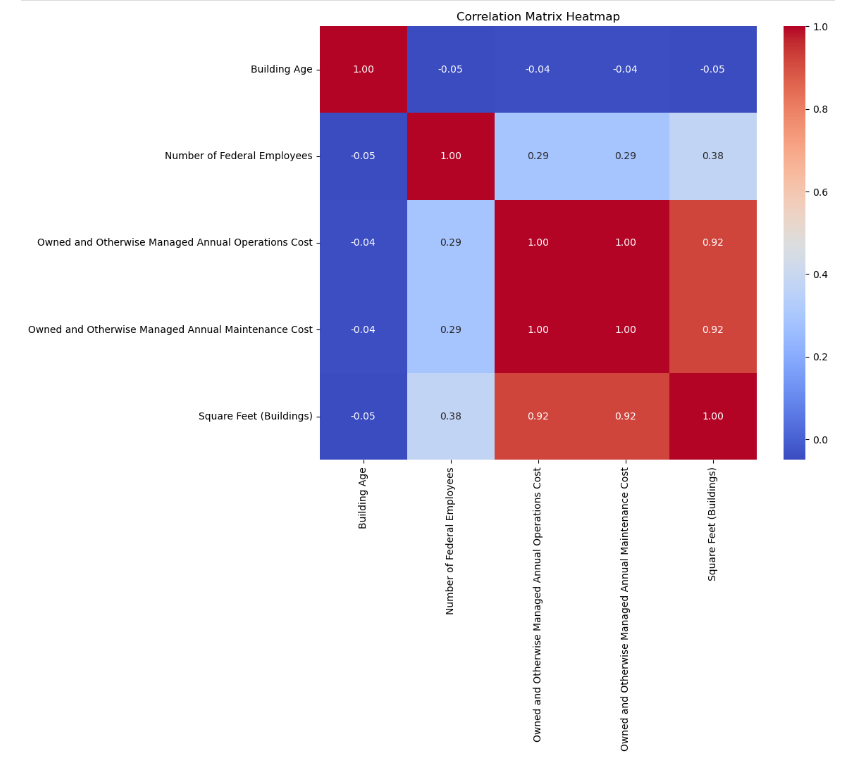
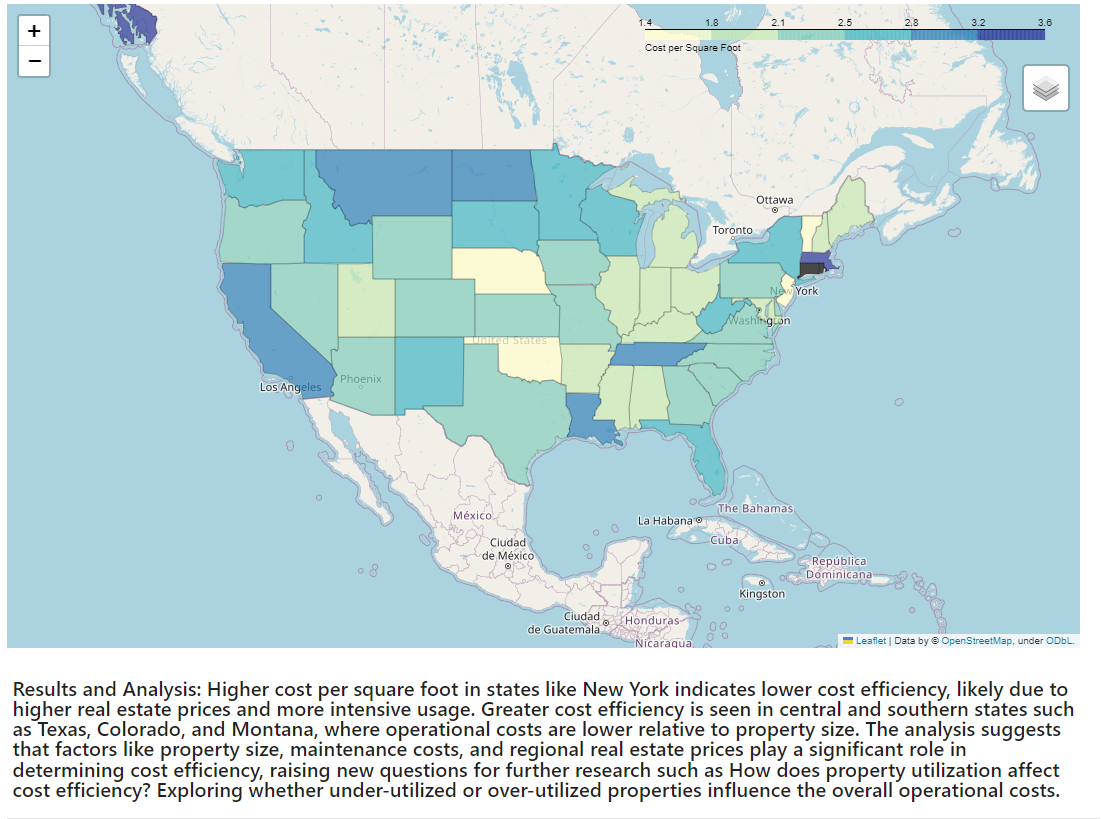
**Python Visuals**

**Task 6.2 Exploring Relationships**

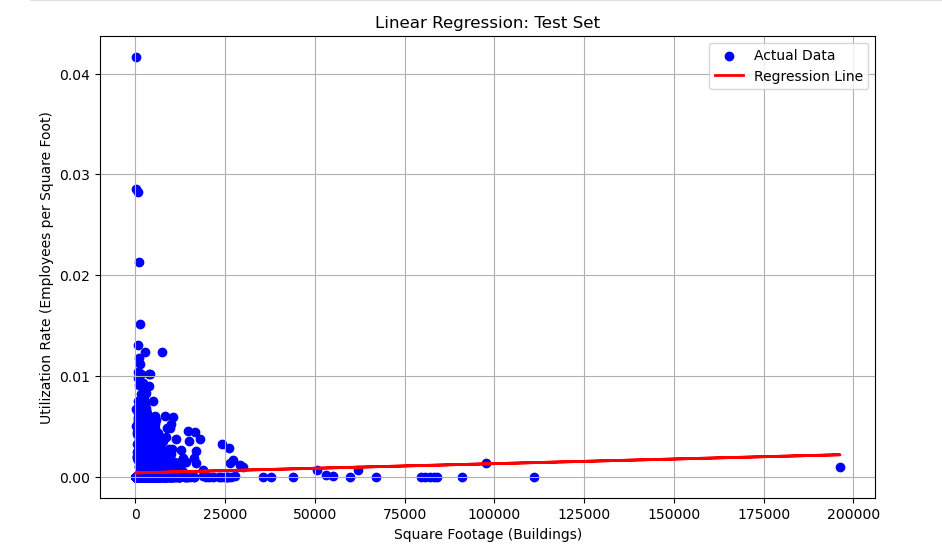




**Task 6.3 Geographical Viz**



**Task 6.4 Supervised ML: Regression**



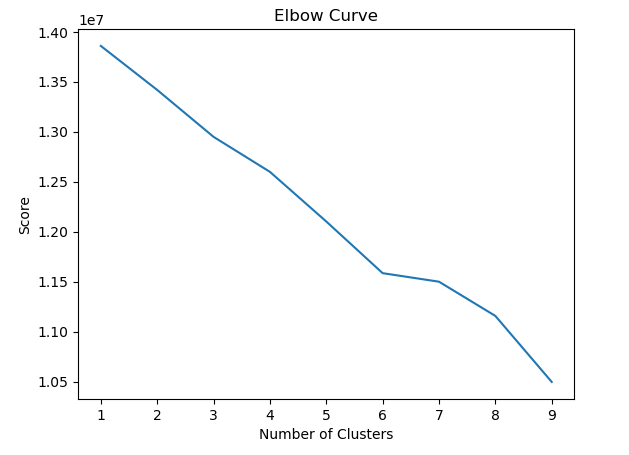
**Observations and reflections**

The model did not perform well on the test set, as shown by the scatter plot where the actual data points (blue dots) are mostly scattered far from the model's predictions (red line). This mismatch, especially for smaller square footages, reflects the low 𝑅^2 score, indicating that the model fails to capture the true variations in utilization rates.

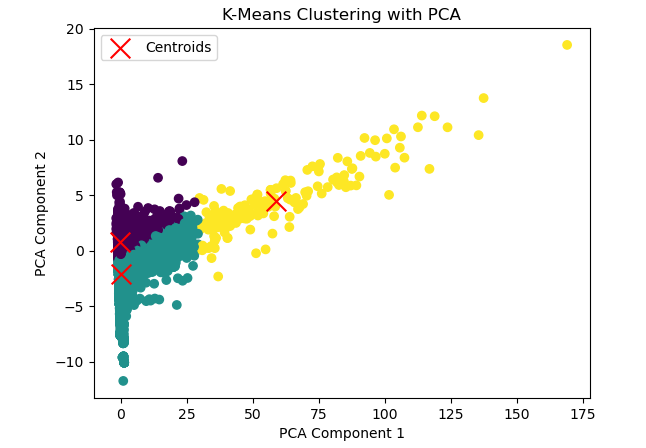
One possible reason for this poor performance is data bias. If the dataset is dominated by buildings with similar square footage or usage patterns, the model may become biased, leading to inaccurate predictions for buildings that don't fit this common pattern. This bias limits the model's ability to generalize and make accurate predictions across different types of buildings.

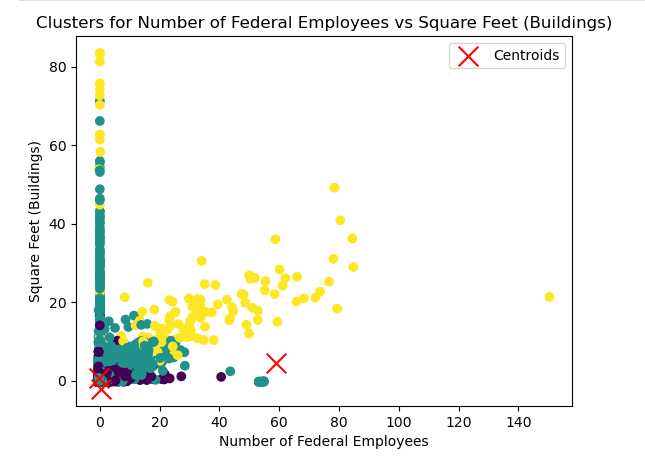
In terms of next steps, it might be worth considering exploring additional features (such as building type, location, or year built) or trying more complex models (e.g., polynomial regression, decision trees) to see if they can better capture the variability in the data. Or perhaps choose another hypothesis to test altogether.

**Task 6.5 Unsupervised ML: Clustering**



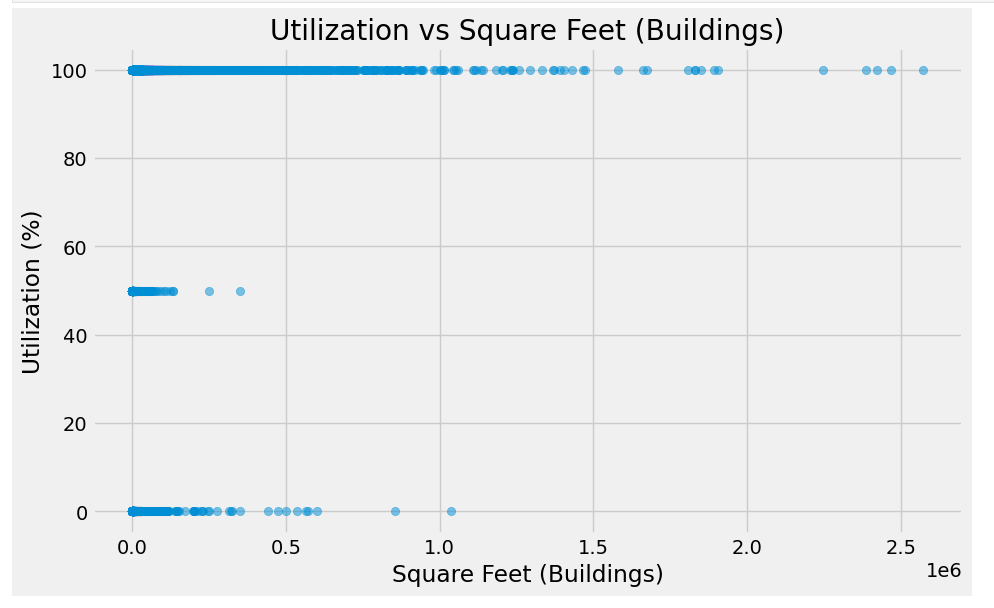
The elbow plot shows that using 3 clusters is the best choice because it's the point where adding more clusters doesn't significantly improve the grouping of the data. After 3 clusters, the improvement slows down, meaning 3 is the most efficient number for capturing the main patterns in the data.



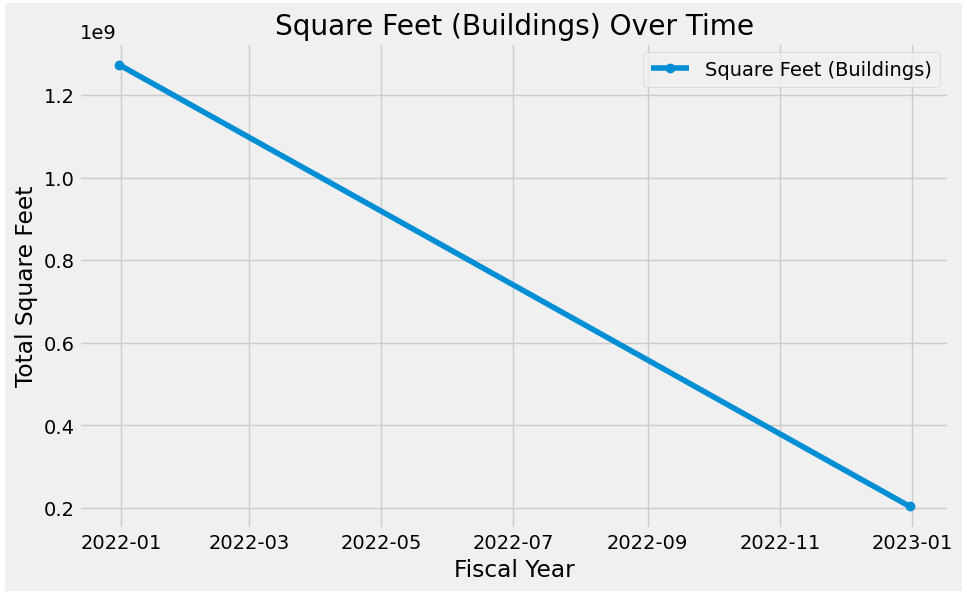


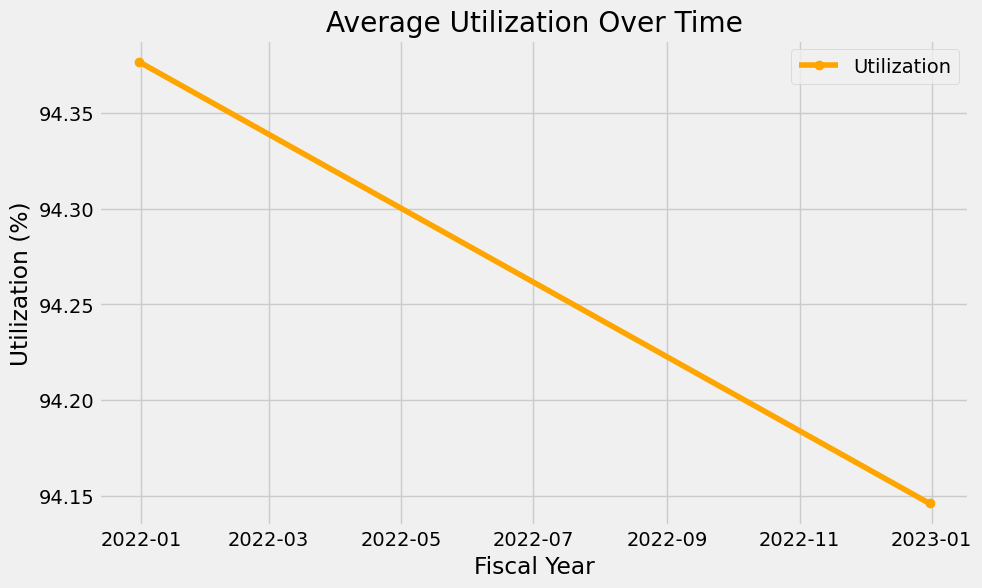
The clusters in the plot show logical groups of buildings based on their size and the number of employees, with each group represented by different colors. It might be helpful to look at more details or use different methods to better understand the patterns.

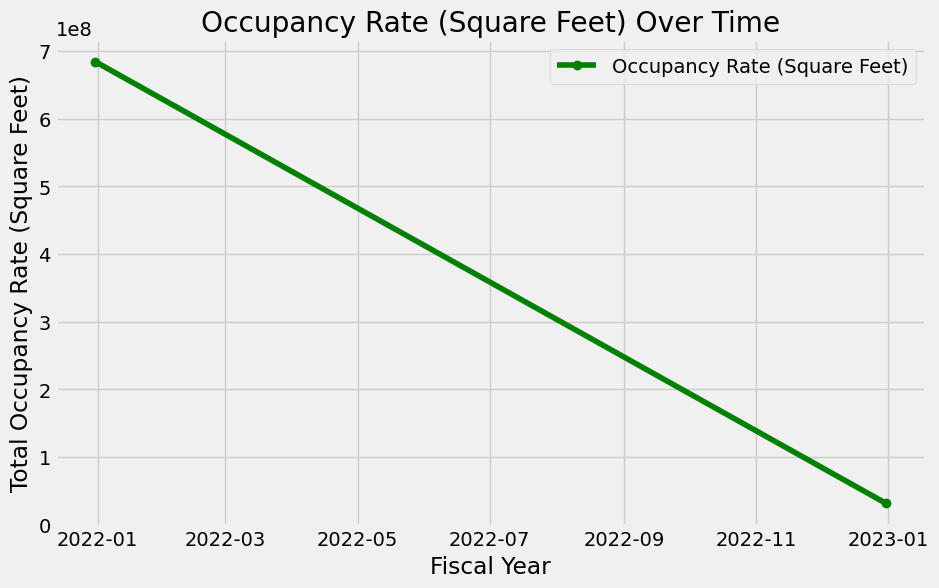
Task 6.6 Time Series



The correlation coefficient between Square Feet (Buildings) and Utilization is approximately 0.012337. This value is very close to zero, indicating that there is virtually no linear relationship between the size of the buildings and their utilization percentage in the dataset. Proceeding to analyze how utilization changes over time for different buildings.

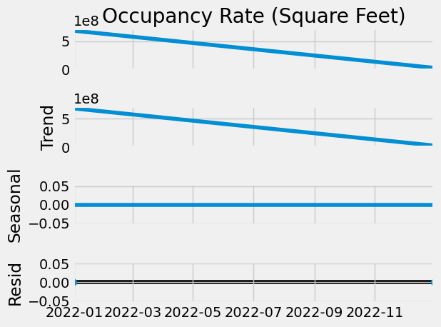
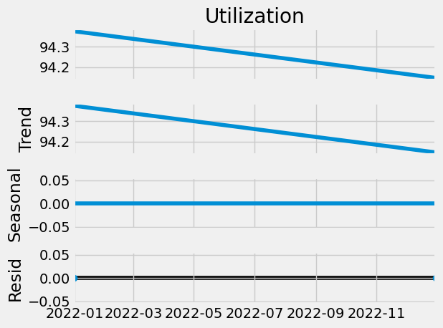
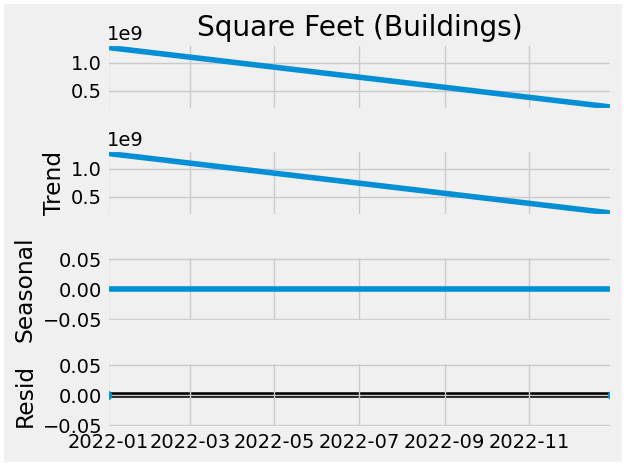






Analysis of line charts: Square Feet (Buildings) Over Time: The total square footage of buildings shows a consistent and significant decline throughout the fiscal year 2022. Average Utilization Over Time: The average utilization percentage experiences a slight but steady decrease over the same period. Square Feet (Buildings) Over Time: The total square footage of buildings shows a consistent and significant decline throughout the fiscal year 2022. Average Utilization Over Time: The average utilization percentage experiences a slight but steady decrease over the same period.

Decomposing the time series



Analysis of the result is overall decline. The primary observation across all metrics is a consistent decline over the fiscal year 2022. The total square footage of buildings has decreased significantly, which is accompanied by a small decline in both utilization percentage and occupancy rate. This could suggest a downsizing of the portfolio, a focus on reducing underutilized space, or other strategic decisions impacting the space available.

Since I am working with annual data and possibly have only two data points (for the years 2022 and 2023), the test does not have enough data points to run the Dickey-Fuller test. Unfortunately, I do not have more granular data to use for the time series.

Given the very limited number of data points, analyzing autocorrelations would not be meaningful in this context.

