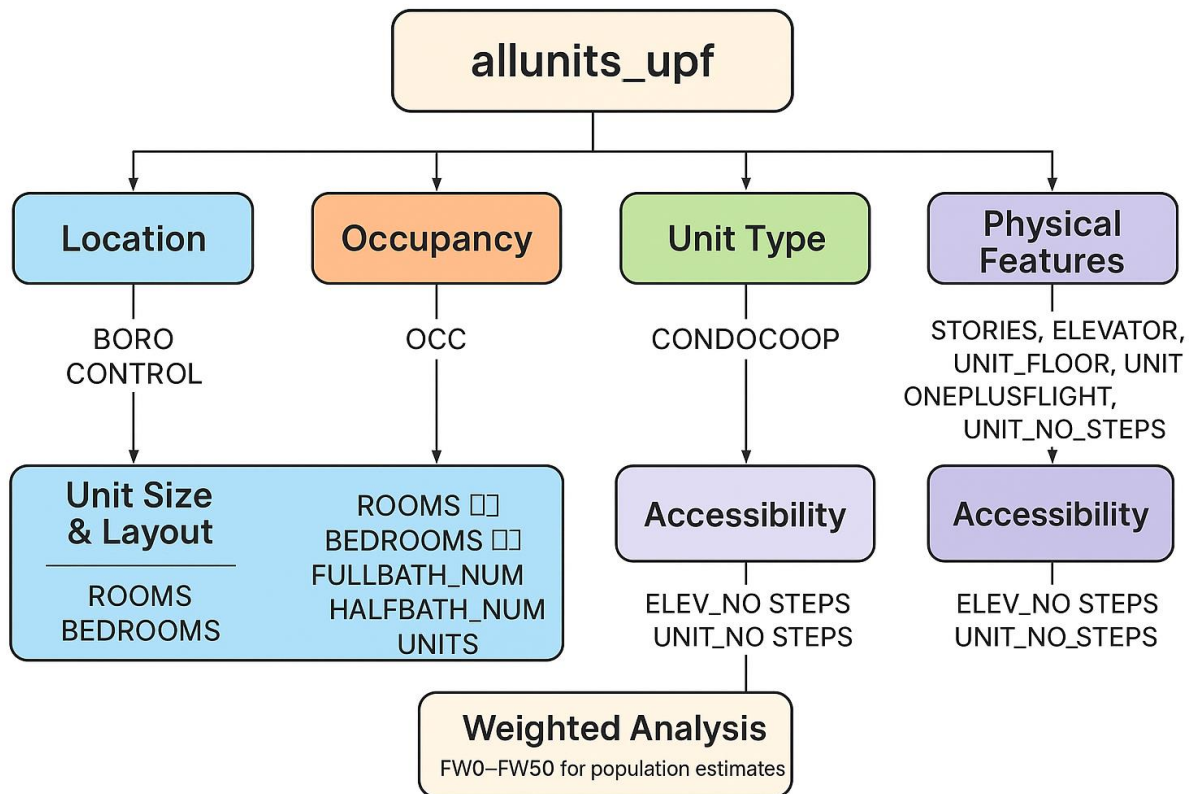


ALL Units puf 2021:



Summarize numeric data

- Columns like **BEDROOMS**, **ROOMS**, **FULLBATH_NUM**, **STORIES** tell about unit size.
- We calculate things like **average number of rooms, bathrooms, stories, etc.**

Summarize categorical data

- Columns like **BORO** (borough), **OCC** (occupied/vacant), **CONDOCOOP**, **ELEVATOR** tell about unit type or location.
- We count **how many units are in each category**.

Make charts

- **Histograms** → Show distribution of bedrooms, rooms, or age of buildings.
- **Bar charts** → Show counts for boroughs, occupancy, condo/co-op, elevator presence, accessibility features.

These charts help **see patterns visually**.

Weighted analysis

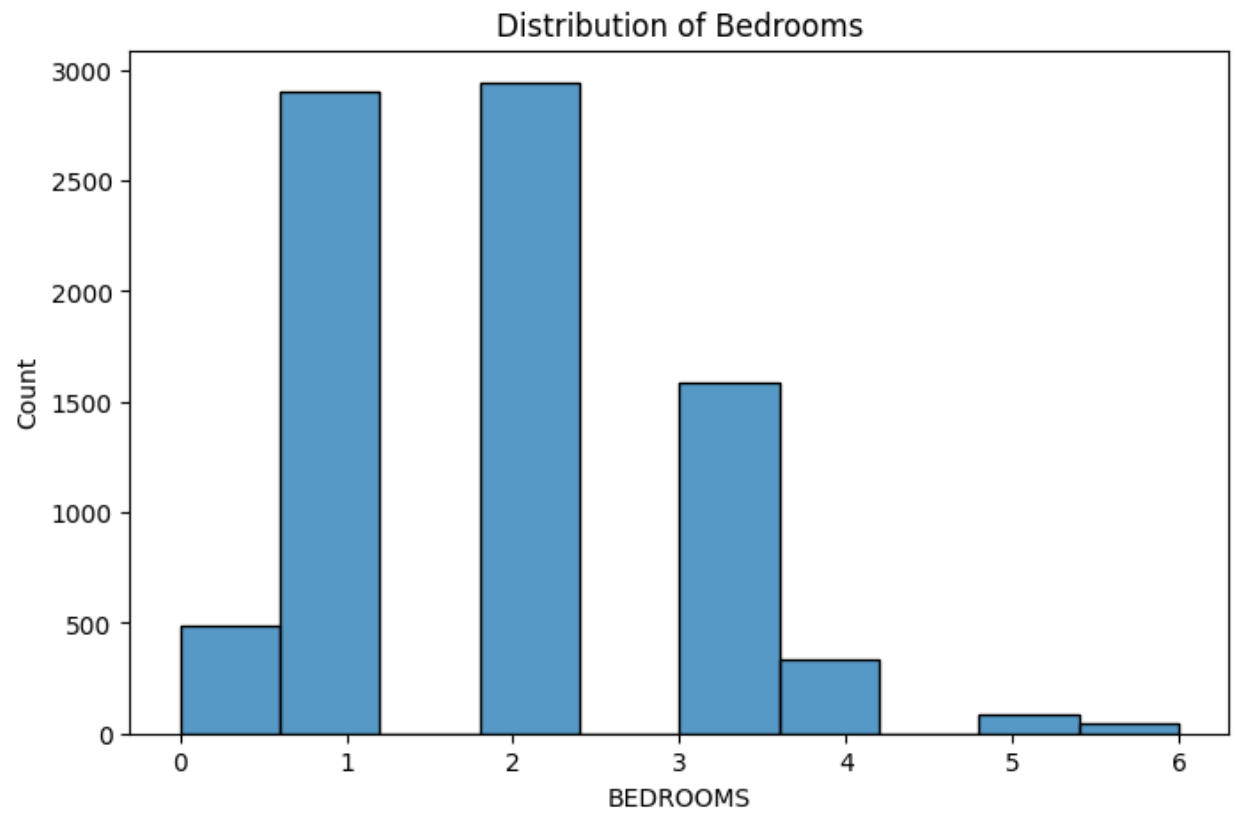
- Columns like `FW0` are **survey weights** that help estimate NYC population totals.
- Example: Weighted average bedrooms per borough tells **how it looks across all NYC housing**, not just your sample.

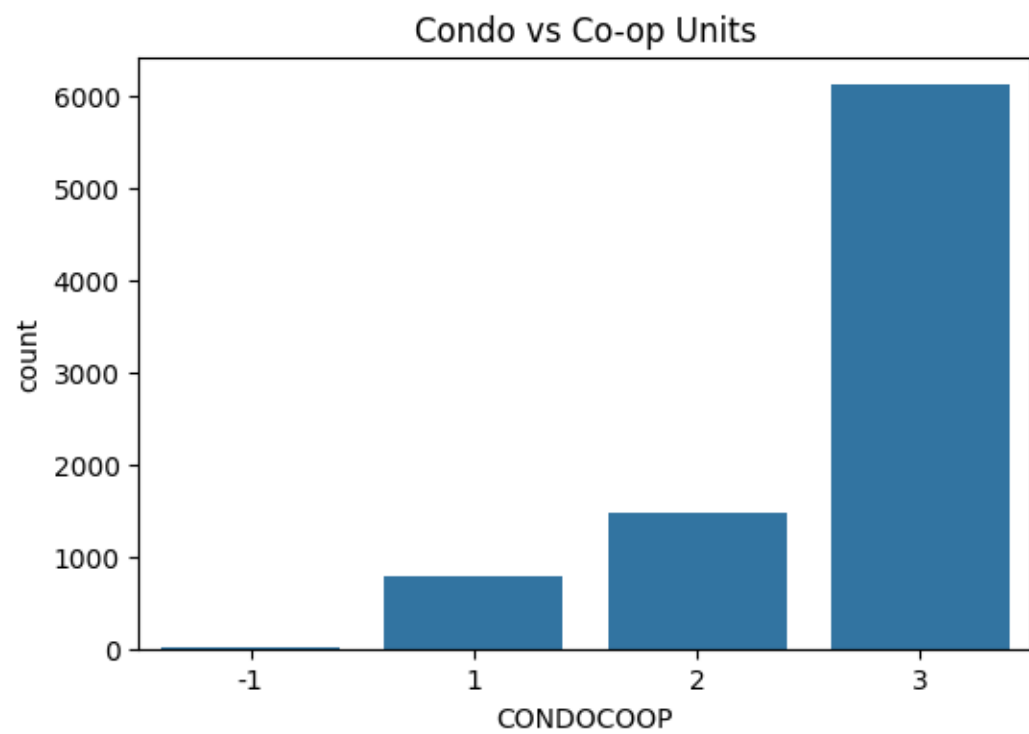
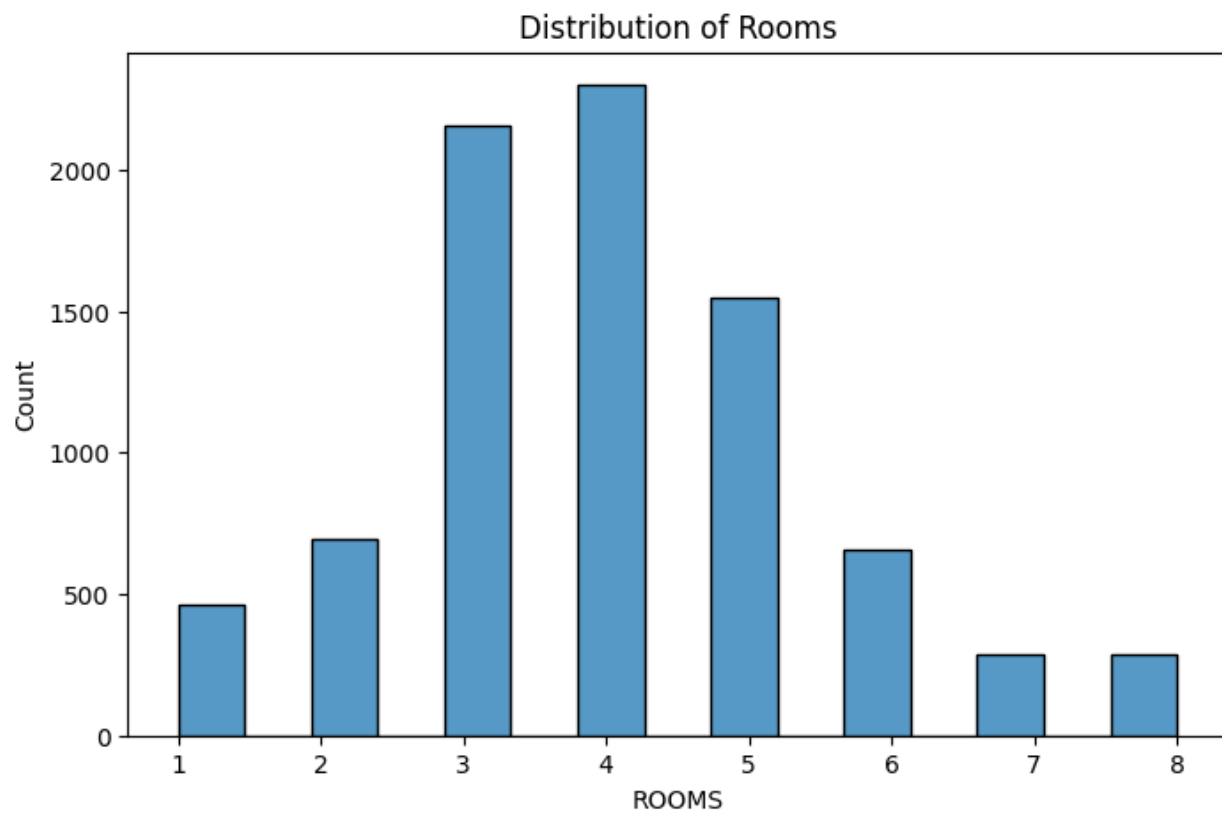
Correlation

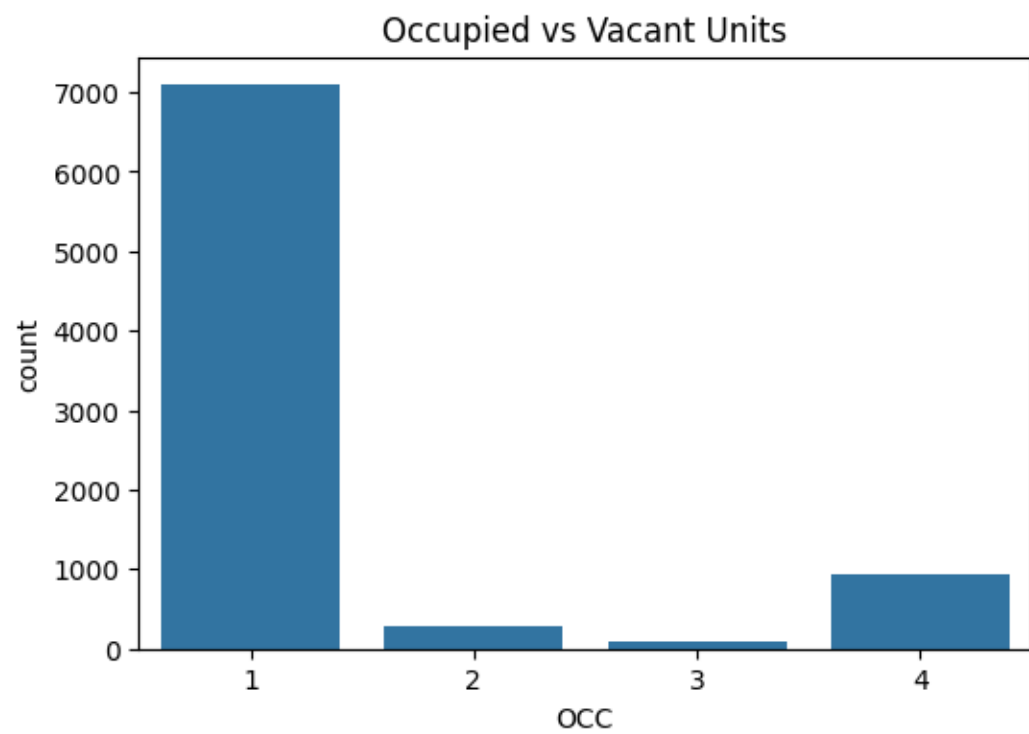
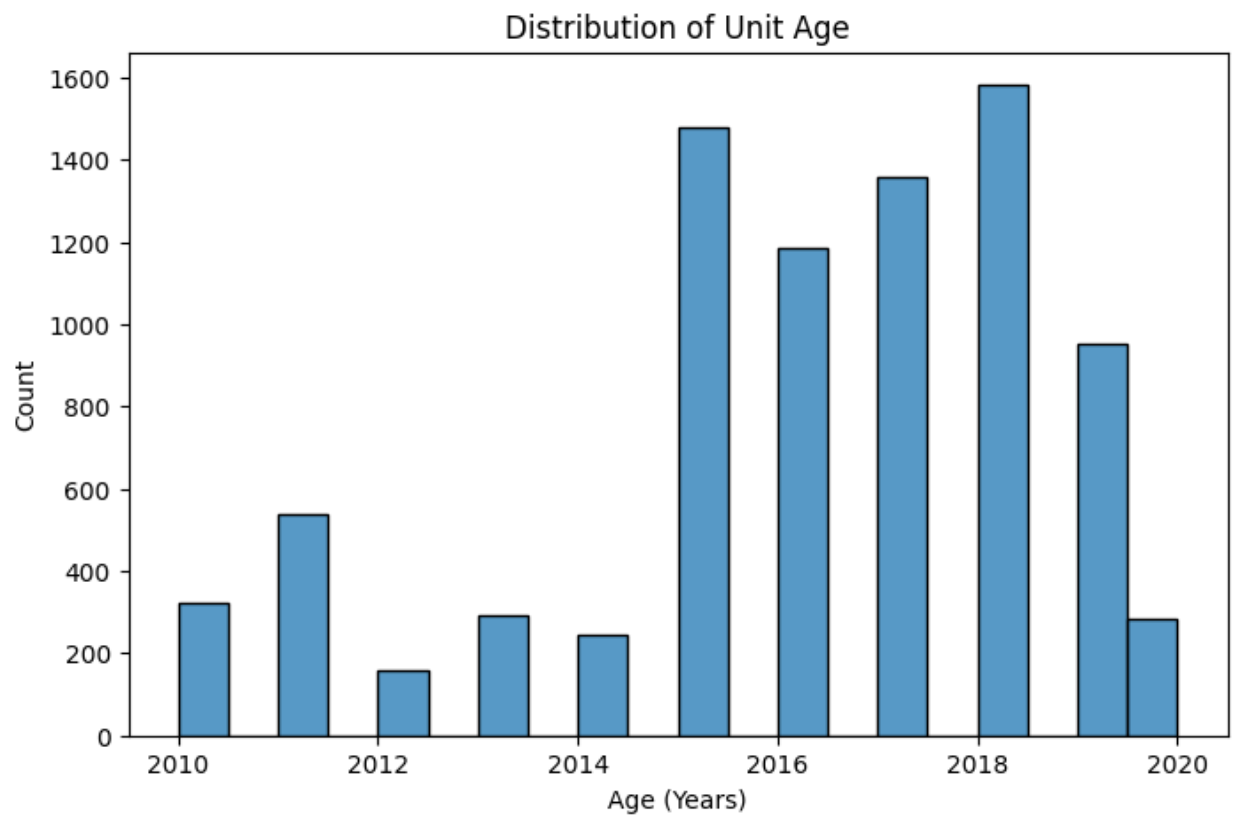
- Check which numeric features are related.
- Example: More rooms → more bedrooms (likely).

Unit age

- Calculate building age: `2021 - YEARBUILT`.
- See how old NYC housing units are. **Summary tables**
- Average rooms, bedrooms, bathrooms **per borough**.
- Occupancy rate per borough → how many units are occupied vs vacant.







Occupied Dataset 2021:

Target variable: You want to analyze **occupied households**, so OCCUPIED or TENURE could be your target depending on your focus.

- **Features:** You have a lot, including:
- **Household composition:** HHSIZE, HH62PLUS, HHUNDER18, HHUNDER6
- **Housing conditions/issues:** NOHEAT, LEAKS, MOLD, ELEVATOR_BROK, RODENTS_UNIT
- **Utilities & costs:** UTIL_ELECTRIC, UTIL_GAS, UTIL_HEAT, UTILCOSTS_SUMMER, UTILCOSTS_WINTER
- **Rent & mortgage info:** RENTASSIST, RENTFEES_AMOUNT, TOTAL_FIRSTMORT, ASSESSMENT_AMT
- **Pets & animals:** ANIMS, SERVICEANIMAL, PETTYPE_*
- **Flags / imputed values:** I_NOHEAT, I_UTIL_ELECTRIC, etc.

Visualization Ideas

- **Household composition** → stacked bar charts
- **Housing issues heatmap** → correlations between issues
- **Utility costs distribution** → box plots for summer/winter
- **Rent & mortgage burden** → violin plot or density plot
- **Clustering result** → scatter plots or cluster maps

