## Exploratory Data Analysis

## Objective

The ultimate result of this EDA is to set the grounds on which we can create a dashboard that helps identify, analyze, and compare different properties to invest. Therefore, on this EDA we will:

1. Explore the available data sets.
2. Define which KPIs are relevant for the dashboard.
3. Choose which columns from the dataset we will use to estimate each KPI.
4. Review if the selected columns are in proper condition to be used or if they need any transformation.
5. Define which dimensions we are going to use in the dashboard.
6. Review data quality on selected dimensions.
7. Next steps and proposals

## Datasets Exploration

### property\_dataset\_sample

* Description: The property dataset contains commercial real estate property details across various U.S. states
* Number of columns: 100
* Number of records: 510

This dataset will be crucial for determining most of the metrics and dimensions from the dashboard, therefore, it will be used as the central table of the model.

### demographics\_US

* Description: State-level demographic indicators from the U.S. Census Bureau
* Number of columns: 55
* Number of records: 52

This data set will be used to compare the available properties against state market values.

### states\_codes\_mapping

* Description: Mapping between U.S. state codes and names
* Number of columns: 3
* Number of records: 56

This dataset will be used as an intermediate table to join the other two datasets.

## KPI Definitions

Net Operating Income (NOI):

This financial metric shows the difference between Income (Rent, fees, etc.) and operative expenses (maintenance, utilities, etc.). A higher NOI means a more profitable opportunity.

Cap Rate:

This ratio represents the proportion on which a property yearly net operative income covers the initial investment. This KPI gives an approximation on return on investment. A higher Cap Rate means a faster ROI.

Rent:

This KPI represents the main income when investing in CRE. Rent depends on the total units rented and the price paid for each rented unit. A higher rent KPI means more income, therefore a more profitable property.

Cash On Cash Return:

This KPI ratio shows a more accurate picture on the return on investment term since it not only considers operative expenses but also considers the debt on the property. A higher Cash on Cash Return means a more profitable opportunity.

Loan To Value:

This metric shows the ratio between what is debt for a property and its market value. This KPI is important because the less debt It has, there are lower chances of foreclosure and therefore, it is a less risky investment. A value higher than 1 means that the debt on the property is higher than its market value.

Debt Service Coverage Ratio:

This KPI represents how much of the properties’ monthly debt payment is covered by the current monthly rent income. A value higher than 1 means that the income received by rent exceeds the totality of the monthly debt payments and that.

Cut-off Date Balance Unit:

This measure represents how much a single unit of the property should cost to pay monthly debt. The higher the number is, the higher the minimum rent per unit must be a profitable investment.

Occupancy Rate:

Shows how many units in the property are rented (At a specific moment in time) over the total number of units. A higher rate means more income (at least short term)

Demographic replacement rate:

Represents how demographics will move in the next 10 years. A low replacement rate means that in the future, less people will be in the state, therefore, less chances to rent units.

Rent Affordability Ratio (Median):

### It shows how affordable the monthly rent of a unit of property is against the median income of a household. The higher the ratio is, the less people are going to be able to afford the rent of the property, which directly affects occupancy and ultimately, income. However, a higher value means that when units are rented, there will be more income. To fully analyze this KPI, occupancy rate should also be considered.

Rent Affordability Ratio (Mean):

### It shows how affordable the monthly rent of a unit of property is against the mean income of a household. The higher the ratio is, the less people are going to be able to afford the rent of the property, which directly affects occupancy and ultimately, income. However, a higher value means that when units are rented, there will be more income. To fully analyze this KPI, occupancy rate should also be considered.

## Columns Selection for KPI Estimation

|  |  |  |
| --- | --- | --- |
| **Metric Name** | **Indicator Type** | **Formula** |
| Average NOI (Historic) | Financial | Average (most recent noi+2nd most recent noi+3rd most recent noi) |
| Underwritten NOI | Financial | uw noi |
| CAP Rate | Financial | uw noi / appraised value |
| Rent | Financial | total units\*occupancy %\*monthly rent per unit |
| Cash On Cahs Return | Financial | (uw noi - monthly debt service amount (amortizing)) / appraised value |
| Loan To Value | Risk | cut-off date loan amount / appraised value |
| Debt Service Coverage Ratio | Risk | (total units\*occupancy %\*monthly rent per unit) / monthly debt service amount (amortizing) |
| Cut-off Date Balance Unit | Risk | cut-off date balance unit |
| Occupancy Rate | Occupancy | occupancy % |
| Demographic Replacement Rate | Occupancy | [population\_total\_count - (population\_18yr\_over\_count + population\_25yr\_over\_count + population\_25yr\_over\_with\_bachelor\_degree\_count + population\_18\_34yr\_over\_count + population\_25\_64yr\_over\_count)] / population\_65yr\_over\_count |
| Rent Affordability Ratio (Median) | Occupancy | monthly rent per unit / (household\_income\_median / 12) |
| Rent Affordability Ratio (Mean) | Occupancy | monthly rent per unit / (household\_income\_mean/ 12) |

## KPI Columns data inspection

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Column Name** | **Column Type** | **Null Values** | **Min Value** | **Max Value** | **Mean** | **Ready To Use** | **Notes** |
| total units | Measure | - | 10 | 4,249.00 | 262.56 | Yes |  |
| cut-off date balanceunit | Measure | 6 | 17,989.58 | 638,245.90 | 133,064.70 | No | These six null values can be calculated with total units and monthly debt |
| occupancy % | Measure | - | 0.76 | 1 | 0.94 | Yes |  |
| cut-off date loan amount | Measure | - | 995,172.59 | 947,000,000.00 | 28,352,332.44 | Yes |  |
| monthly debt service amount (amortizing) | Measure | - | 6,957.52 | 4,224,672.22 | 156,634.10 | Yes |  |
| loan term (remaining) | Measure | - | 35 | 120 | 98.33 | Yes |  |
| appraised value | Measure | - | 44,384.00 | 1,843,000,000.00 | 63,050,990.00 | Yes |  |
| uw noi | Measure | - | 126,724.33 | 80,420,149.28 | 2,994,541.00 | Yes |  |
| most recent noi | Measure | - | 81,276.28 | 85,366,589.10 | 2,856,050.00 | Yes |  |
| 2nd most recent noi | Measure | 18 | -473,258.83 | 86,912,883.92 | 2,651,376.00 | Yes | Null values would not affect KPI formula. No treatment needed |
| 3rd most recent noi | Measure | 53 | -333,250.24 | 82,272,789.23 | 2,367,473.00 | Yes | Null values would not affect KPI formula. No treatment needed |
| monthly rent per unit | Measure | - | 309 | 6,998.36 | 1,644.55 | Yes |  |
| uw expenses | Measure | - | 101,772.60 | 47,075,164.22 | 2,134,569.67 | Yes |  |
| population\_total\_count | Measure | - | 579,761 | 39,242,785 | 6,454,662 | Yes |  |
| population\_18yr\_over\_count | Measure | - | 447,402 | 30,513,773 | 5,027,947 | Yes |  |
| population\_65yr\_over\_count | Measure | - | 81,628 | 5,814,110 | 1,057,535 | Yes |  |
| household\_income\_median | Measure | - | 25,096 | 106,287 | 77,173 | Yes |  |

## Dimensions Definition

|  |  |  |
| --- | --- | --- |
| **Dimension** | **Dimension Type** | **Purpose** |
| State | Geographical | Identify the state where the property is located on |
| Property City | Geographical | Identify the city the property is located on |
| Latitude | Geographical | Allows to pin the exact location. Useful to see if location is attractive or not |
| Longitude | Geographical | Allows to pin the exact location. Useful to see if location is attractive or not |
| Appraisal Valuation Date | Date | Shows when the appraisal value was estimated. The oldest the appraisal date, the less trustworthy the appraisal is |
| Occupancy As of Date | Date | Shows when the Occupancy % value was estimated. The oldest the date, the less trustworthy the occupancy % is |
| Year Built | Text | Most recent buildings tend to be more attractive and require less maintenance |
| Year Renovated | Text | Renovated building might be more attractive and require less maintenance |
| Loan Term (Remaining) | Text | When paying a Loan, debt value is important, but also for how long, especially if there are early payment fees |
| Property Type | Text | Describe what type of CRE property is |
| Property Subtype | Text | Describe what subtype of CRE property is |
| Property Name | Text | Uniquely identifies the property (Will serve as the key of the main table) |
| Loan Purpose | Text | Shows what type of loan has the property |
| Additional Financing in Place | Boolean | True or False value that shows if there is more than one loan for the property |

## Dimension Columns data inspection.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dimension** | **Dimension Type** | **Null Values** | **Unique Values** | **Min Value** | **Max Value** | **Ready To Use** | **Notes** |
| State | Geographical |  | 5 |  |  | Yes |  |
| Property City | Geographical |  | 210 |  |  | Yes |  |
| Latitude | Geographical | 20 | 485 |  |  | Yes | This field is not critical, therefore, null records should still be use |
| Longitude | Geographical | 20 | 485 |  |  | Yes | This field is not critical, therefore, null records should still be use |
| Appraisal Valuation Date | Date |  |  | 7/7/2021 | 8/29/2023 | Yes |  |
| Occupancy As of Date | Date |  |  | 13/07/2022 | 11/8/2023 | Yes |  |
| Year Built | Text |  | 67 | 1913 | 2022 | No | There is one record with value = "Various" (Proposal: remove it) |
| Year Renovated | Text | 243 | 17 | 1998 | 2023 | Yes | Null values mean the property has not been renovated |
| Loan Term (Remaining) | Text |  | 61 | 35 | 120 | Yes |  |
| Property Type | Text |  | 1 |  |  | Yes | Multifamily is the only value. However, is included in case a broader dataset is connected to the dashboard |
| Property Subtype | Text |  | 12 |  |  | Yes |  |
| Property Name | Text |  | 510 |  |  | Yes |  |
| Loan Purpose | Text |  | 3 |  |  | Yes |  |
| Additional Financing in Place | Boolean |  | 2 |  |  | Yes |  |

## Next Steps

1. Transform and clean the datasets to create a single data model.
2. Create a report with the metrics described above and slicers with the dimensions.

## Proposals and clarifications

1. An alternative solution to evaluate mid- and long-term risks and profitability could be using simulations combined with AI algorithms.
2. In this EDA we did not consider County, ZIP Code nor Adress because geographical data is on State granularity; however, a more detailed geographical data can be useful for a more informed decision.
3. Dimension “secondary financing in place (existing) (yesno)” was going to be use, but it was empty.