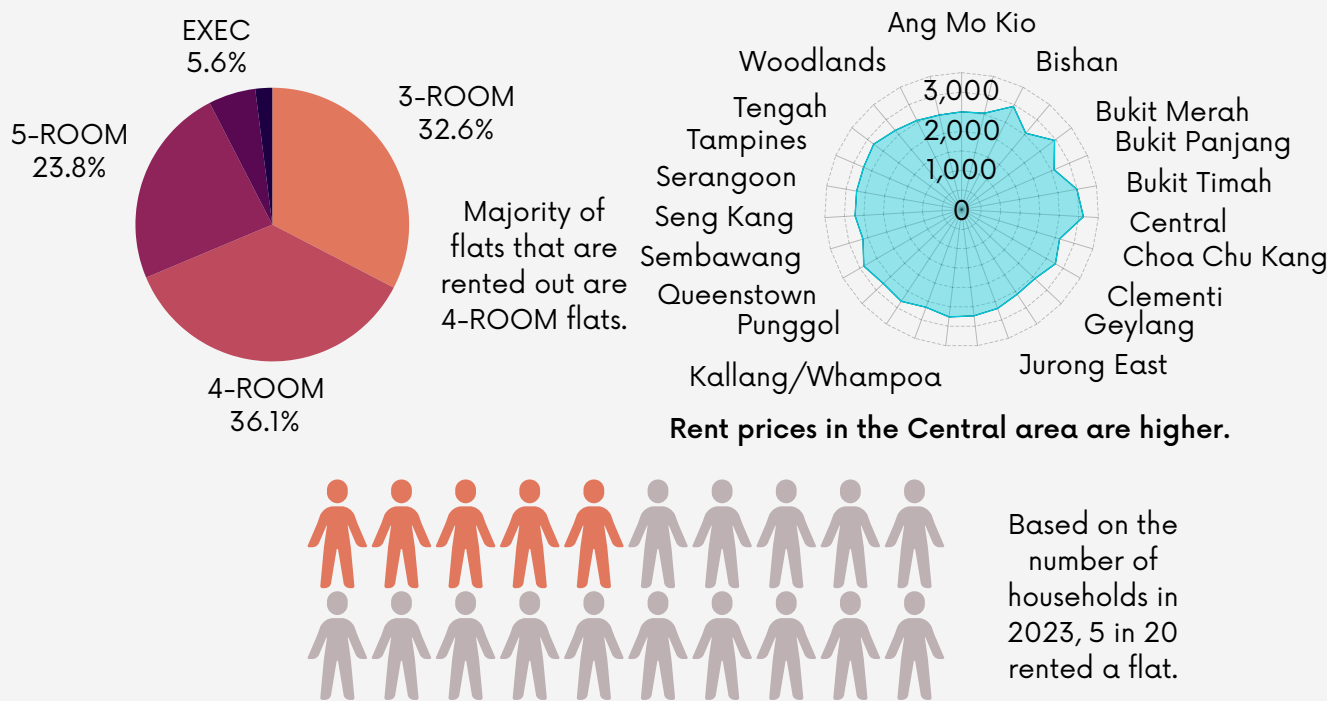# RENT PRICES IN SINGAPORE

The primary goal is to transform raw data into a clean, structured, and enriched dataset for robust data modeling. This process involves data cleaning, integration (fusion with other datasets), exploratory analysis, preparation for modeling as well as modelling and validation.

STEP 1: DATA CLEANING

- Data inspection is done on the rent prices dataset to identify missing data, duplicates and outliers.
  - .info() and .summary() is being used to assess data types, formats, and overall structure.
  - .unique() is used for detecting any data inconsistencies.
- Data manipulation is also carried out to prepare the dataset for fusion with 2 other datasets e.g. extracting mean rent prices, introducing rent_approval_year

STEP 2: EXPLORATORY DATA ANALYSIS

EXEC 5.6%

3-ROOM 32.6%

5-ROOM 23.8%

Majority of flats that are rented out are 4-ROOM flats.

4-ROOM 36.1%

Radar chart labels: Ang Mo Kio, Bishan, Bukit Merah, Bukit Panjang, Bukit Timah, Central, Choa Chu Kang, Clementi, Geylang, Jurong East, Kallang/Whampoa, Punggol, Queenstown, Sembawang, Seng Kang, Serangoon, Tampines, Tengah, Woodlands

3,000 2,000 1,000 0

Rent prices in the Central area are higher.

Based on the number of households in 2023, 5 in 20 rented a flat.

STEP 3: DATA FUSION

Fusion with other datasets - resale prices and HDB information datasets

STEP 4: DATA PREP

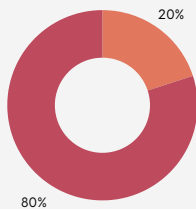Data cleaning and preparation for data modeling.
- One-hot encoding for categorical variables.
- Splitting of dataset into train, test, validation

STEP 5: MODELING AND VALIDATION

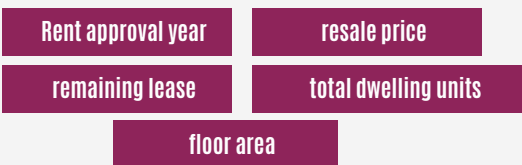Train and evaluate models for prediction or classification.

Model 1: Principal Component Analysis (PCA)

80% of the variance can be explained by 25 out of 54 variables

20%

80%

Model 2: Regression

R2 score of 0.46. High MAE and MSE of 436.6 and 332936.3 respectively. Top 5 feature importance:

Rent approval year	Town_Bukit Timah
Town_Clementi	Multi-storey carpark
Town_Bishan	

Model 3: Random Forest

R2 score of 0.43. High MAE and MSE of 429.6 and 351809.8 respectively. Top 5 feature importance:

Rent approval year	resale price
remaining lease	total dwelling units
floor area	

Model 4: XGBoost

R2 score of 0.54. Lower MAE and MSE of 404.0 and 285363.2 respectively. Top 5 feature importance:

Rent approval year	Town_Jurong East
Resale Price	Town_Bukit Timah
Town_Tampines	