Real Estate Price Predictor in Egypt

predict you target house prices pt.1

By / Mohamed Waleed EL-masry

Real Estate Price Predictor in Egypt

01. Introduction

02. Data Collection

03. Data Cleaning & Preprocessing

04. Exploratory Data Analysis (EDA)

05. Current Dataset Status

06. Next Steps

01

Introduction

Introduction

01

Objective

Predict prices of ready homes based on property features

02

Data Source

Web scraping from Bayut website

03

Date Set

after cleaning around 46,000 records

02

Data Collection using (python)

Data Collection

Method

Web Scraping (Python)

Source

Bayut Egypt Real Estate Listings

Extracted features

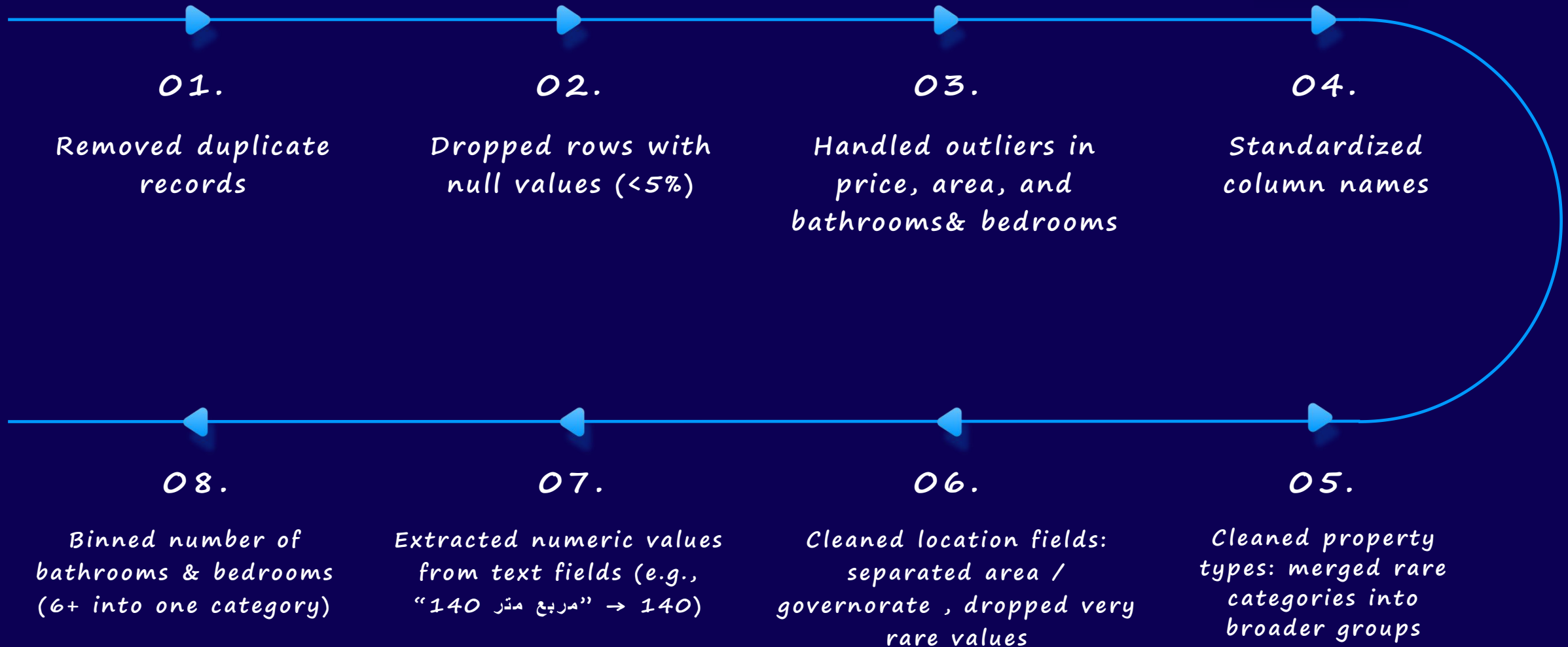property type, price, area, bathrooms, bedrooms, location

Raw dataset size

~50,000 rows

03

Data Cleaning & Preprocessing using (power query)

Data Cleaning & Preprocessing

01.
Removed duplicate records

02.
Dropped rows with null values (<5%)

03.
Handled outliers in price, area, and bathrooms& bedrooms

04.
Standardized column names

08.
Binned number of bathrooms & bedrooms (6+ into one category)

07.
Extracted numeric values from text fields (e.g., "140 متر مربع" → 140)

06.
Cleaned location fields: separated area / governorate , dropped very rare values

05.
Cleaned property types: merged rare categories into broader groups

Average Price of Each Type of Houses

key findings

- Villas have the highest average price at about 30.8M EGP, far exceeding other property types.

- Private Houses rank second with an average price around 17.4M EGP.

- Chalets average about 11.6M EGP, while Apartments have the lowest average price at 7.6M EGP

note :price gap between Villas and the rest of the property types, which might strongly influence model predictions.


Avg Price of Each Type of Hoses

11,613,916.04	7,659,310.43	30,786,224.28	17,388,815.13
شاليه	شقة	فيلا	منازل خاصة
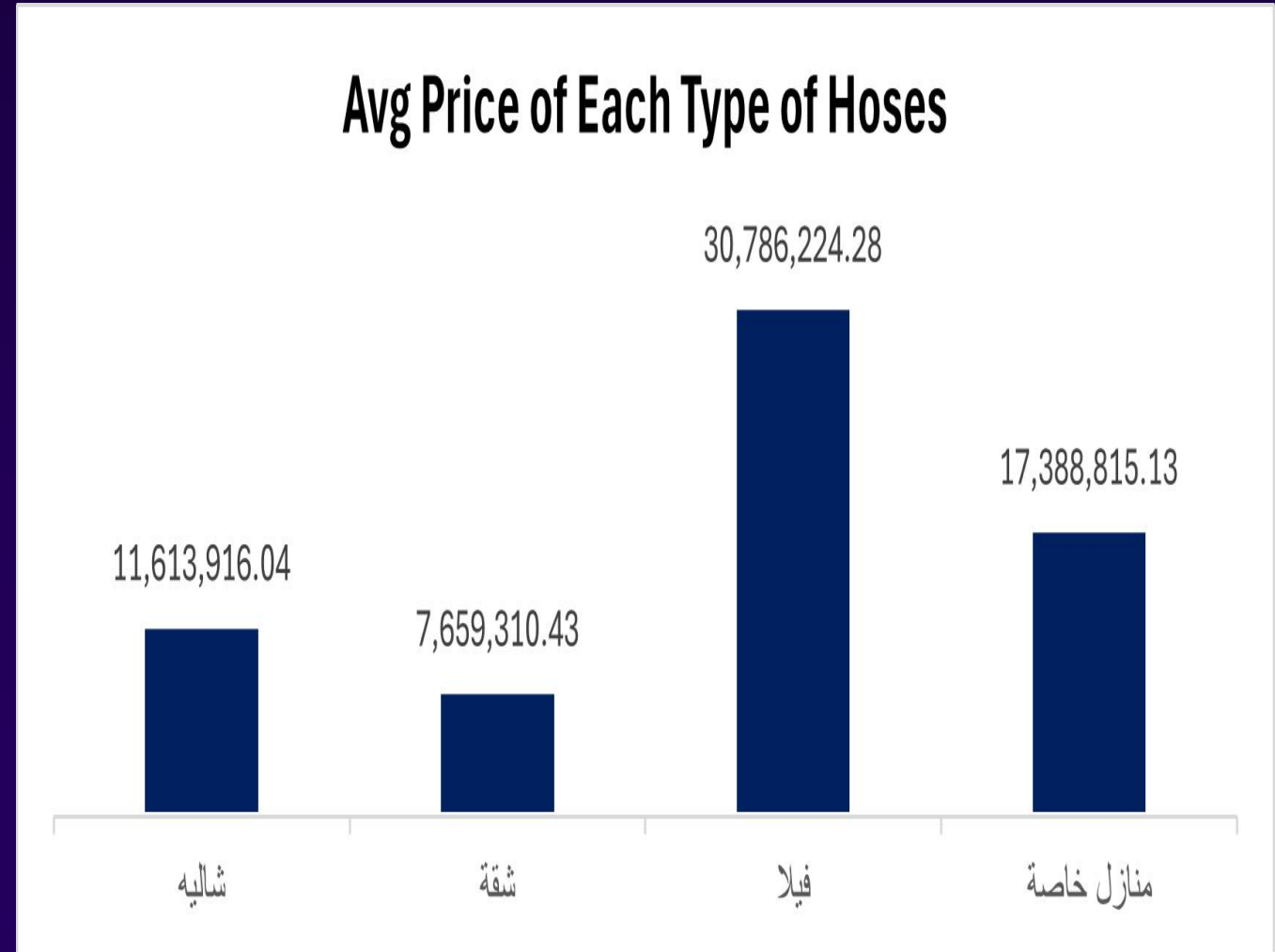
Average Price of Each Type of Property

key findings

– North Coast area records the highest average property prices compared to all other locatio

– Villas have the highest average prices among all property types

– Chalets have the lowest average price in Alexandria compared to other property types.

There's a large gap in pricing between villas and the other property types
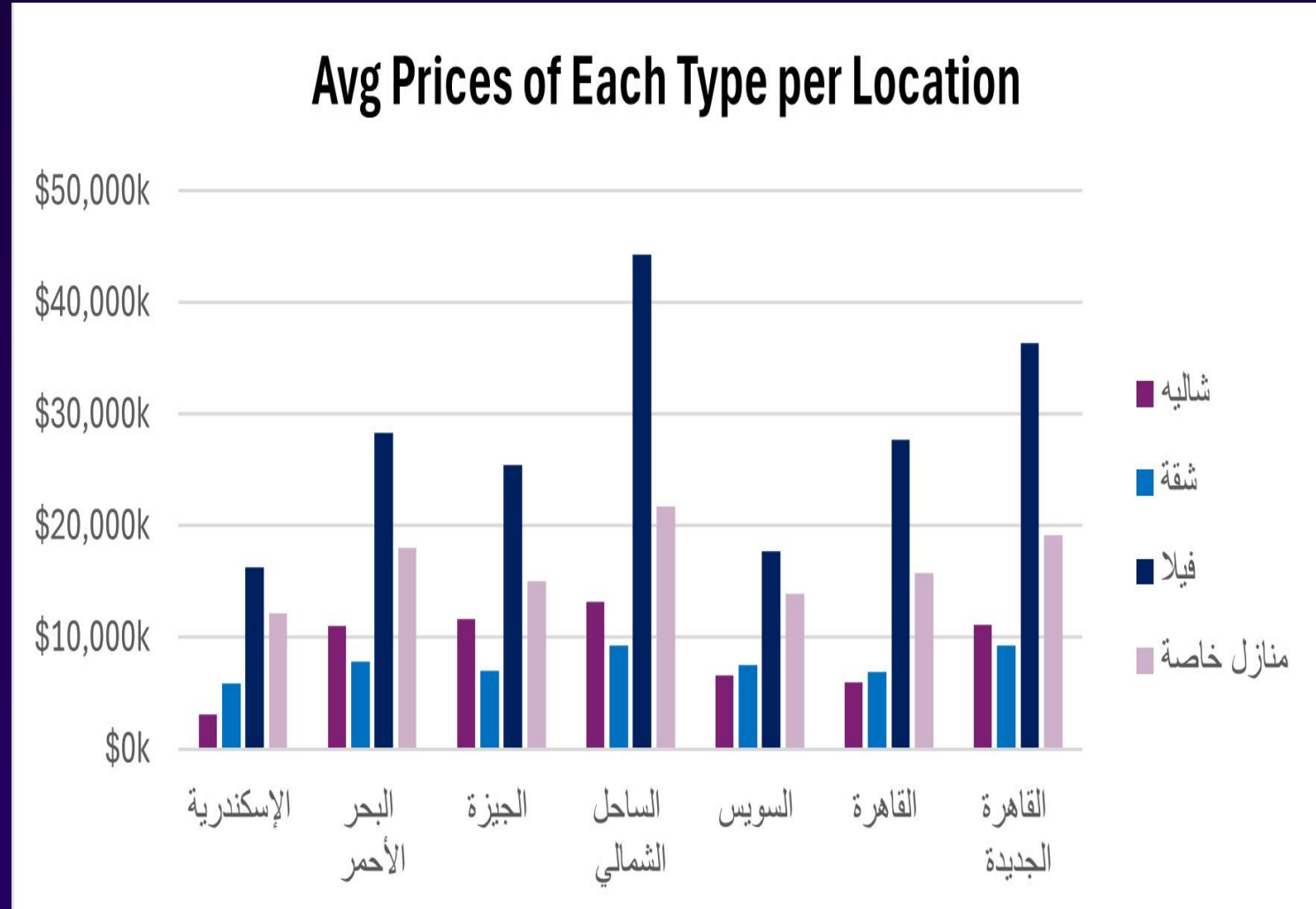
Average Prices per Sub-Location

Key Findings

– Saʿid Abdelrahman area records the highest average property prices compared to all other locations

– Some sub-locations show extremely high prices, especially for villas (e.g., New Cairo, North Coast).

– Villas display the widest price range across sub-locations, while apartments show more stable prices.

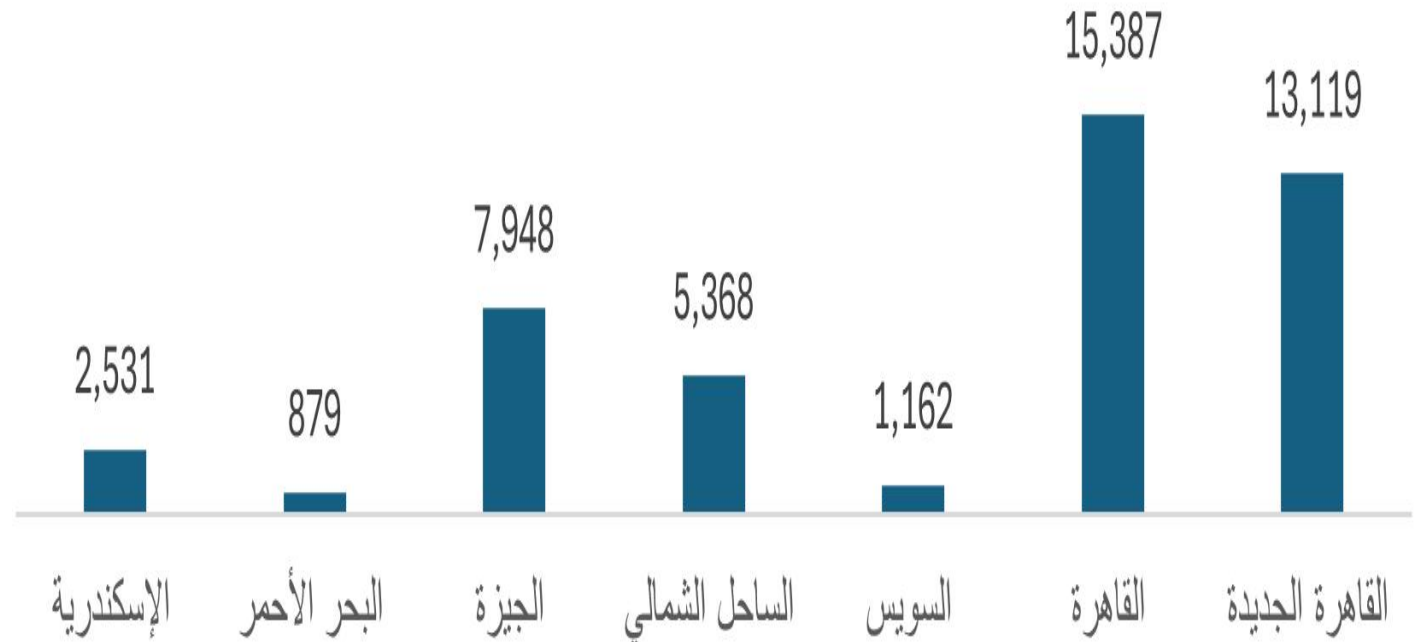The analysis highlights which areas are high-end markets versus mid/low-end markets for each property type.

Avg Prices per Sub Location
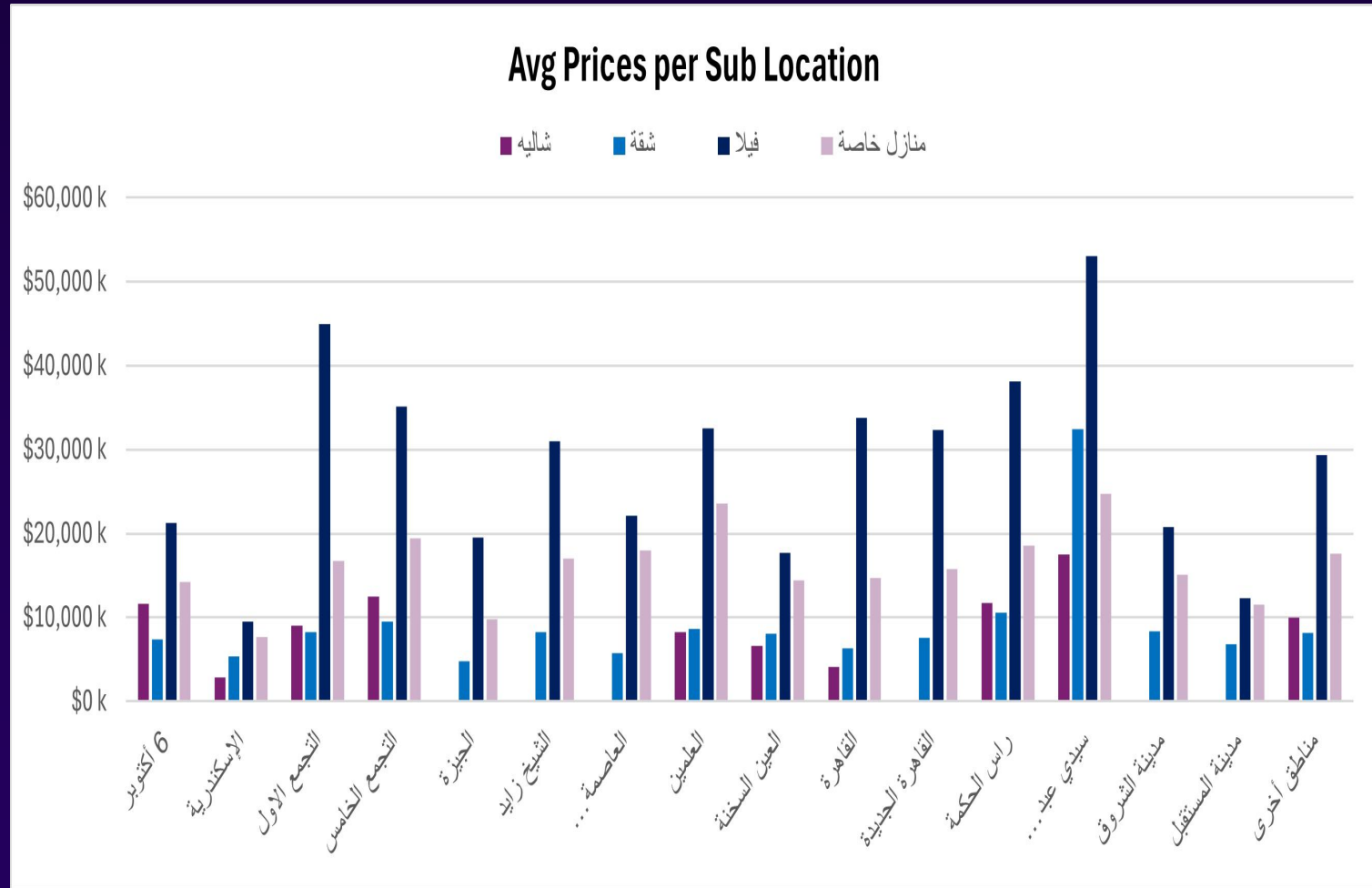
Number of Houses per Location

Key Findings

- Cairo and New Cairo have the largest number of listings, showing the highest market activity.

- GIZA ,North Coast (Sahel) and Alexandria follow, but with significantly fewer listings than Cairo.

- Suez and Red Sea governorate appear as smaller markets with limited housing data.

A few locations (like Damietta or Dakahlia) have very small sample sizes and can be ignored for modeling.

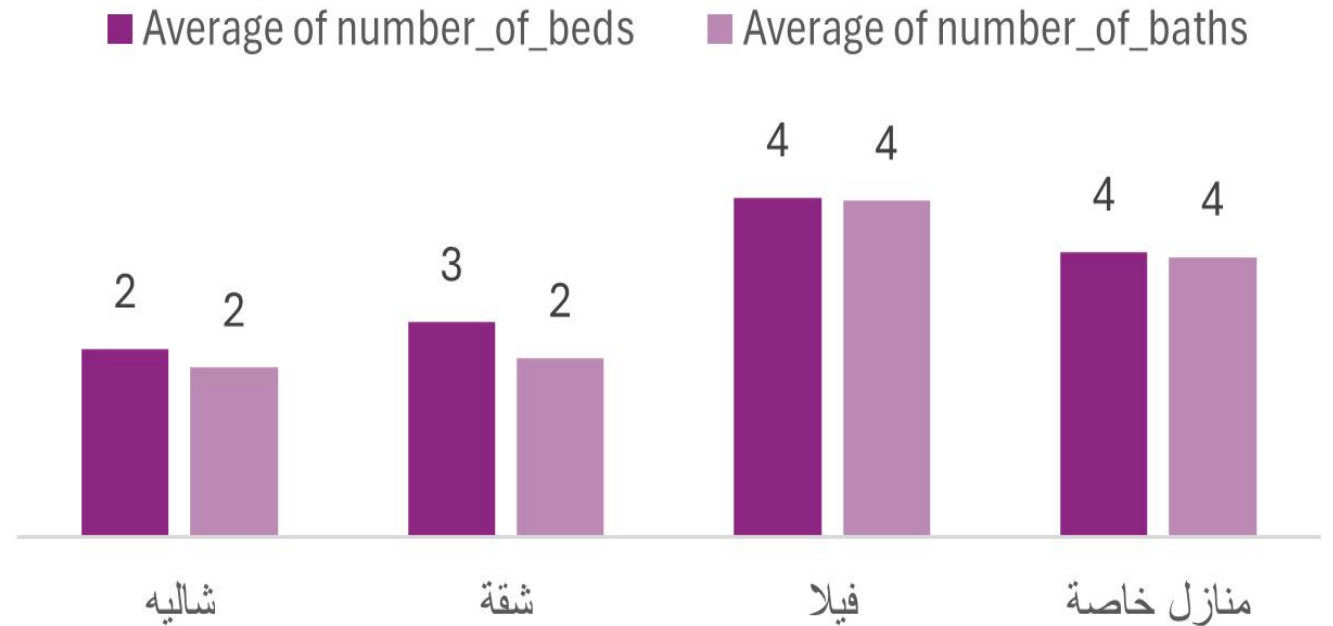

Number of Hoses per Location

2,531	879	7,948	5,368	1,162	15,387	13,119
الإسكندرية	البحر الأحمر	الجيزة	الساحل الشمالي	السويس	القاهرة	القاهرة الجديدة

Number of Houses per Location

Key Findings

- Most properties in the dataset have 2–3 bedrooms and 2–3 bathrooms..

- Properties with 4–5 rooms exist but are less common; 6+ rooms are rare outliers.

- More rooms = higher price trend is clearly visible in the data.

Outlier values (6+ bedrooms or bathrooms) have been grouped to improve model stability

Avg Num of Bed&Bath Rooms per Type

■ Average of number_of_beds   ■ Average of number_of_baths

	شاليه	شقة	فيلا	منازل خاصة
beds	2	3	4	4
baths	2	2	4	4

05

Current Dataset Status

Current Dataset Status

1 - 46,000 cleaned rows of ready-to-sell properties.

2 - Main features prepared:

- Property Type (grouped)

- Location (governorate & area, rare values removed)

- Bedrooms & Bathrooms (outliers grouped)

- Price (to be log-transformed for modeling)

3 - No duplicates or major missing values remain.

Dataset now suitable for feature encoding and model training.

06

Next Steps

Next Steps

- Encode categorical features (property type, location).

- Normalize / log-transform skewed numerical features (price).

- Train/test split for model evaluation.

- Build baseline models (Linear Regression, Random Forest, etc.).

- Evaluate and tune to improve accuracy.

THANK YOU

Mohamed Waleed EL-masry   e-mail :

mwezzat16@gmail.com