This raw data was scraped from the Zameen.com website. The data presents website listings for Lahore (1273), Islamabad (1250) and Karachi (1106) into a structured CSV file. The data scraping organized the listings information into the following column headings:

1) Title
2) Label
3) Type
4) Price
5) Location
6) Baths
7) Area
8) Purpose
9) Bedrooms
10) Added
11) Description
12) Main Features
13) Rooms
14) Business and Communication
15) Healthcare Recreational
16) Nearby Locations and Other Facilities
17) Other Facilities
18) Popular
19) Source
20) Time Stamp

Figure 1 and **Figure 2** show a screenshot of a Zameen listing numbered with where most of the information for each of these column headings can be found for ease of understanding. The scope of this data pertains to the attributes of each house and its impact on price
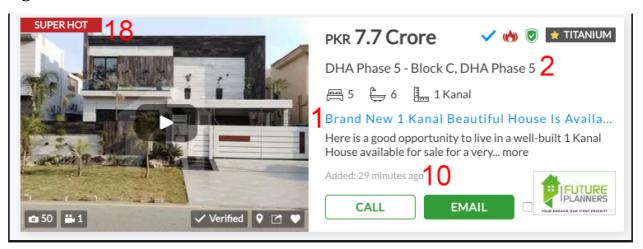
Figure 1

SUPER HOT 18

50 1 ✓ Verified

PKR **7.7 Crore** ✓ 🔥 🛡 ⭐ TITANIUM

DHA Phase 5 - Block C, DHA Phase 5 2

🛏 5   🛁 6   📏 1 Kanal

Brand New 1 Kanal Beautiful House Is Availa... 1

Here is a good opportunity to live in a well-built 1 Kanal House available for sale for a very... more

Added: 29 minutes ago 10

FUTURE PLANNERS

CALL   EMAIL

Figure 2

| Overview | Location & Nearby | Home Finance | Price Index | Trends |

Overview

Details

3	Type	House	7	Area	1 Kanal
4	Price	PKR 7.7 Crore	8	Purpose	For Sale
5	Location	DHA Defence, Lahore, Punjab	9	Bedroom(s)	5
6	Bath(s)	6	10	Added	23 minutes ago

Description 11

Here is a good opportunity to live in a well-built 1 Kanal House available for sale for a very affordable Rs 74,000,000.

The House is located in DHA Defence, Lahore. It features 5 bed, designed with comfort and space in mind.

Future planners offers Mazhar Munir Designer

<u>For the project you need to do the following:</u>

Processing of Data
- Data cleaning
- Data preparation as required

Understanding data
- Identify the problem and set the questions for data analysis (what broader questions can be answered for this dataset?)
- Exploration of data

Selection of technique
- How the problem will be solved using the selected technique
- Justify the use of technique to address the identified problem

Results
- Interpretation of results
- Discussion and analysis of the results
- Conclusive arguments to answer the identified problem

Recommendations & Limitations

Presentation
- Record a 15-minutes presentation covering all of the above topics. You need be selective in your choice of selecting tables, visualizations and results.

<u>Submission requirement:</u>

Video and script file