**Introduction**

Real estate investors are those who purchase real estate to hold it as a long-term buy-and-hold investment. For this reason, they are always interested in knowing the new houses that are for sale and what their historical price has been. The digital era has brought forth online marketplaces that offer a wide range of homes for sale. For example, Zillow.com. The company was founded in 2006 by Rascoff and Humphries with the goal of estimating the value of every home in the United States. Zillow seeks to make information available about “all the homes” in the United States, saying that “data wants to be free”. Zillow has constructed a captivating operational context that supports the use of data to buy, rent, or sell housing (Loukissas,2018).

In New York City, Zillow has enormous properties listed available for sale that investors could purchase. According to Zillow. Com, It has 23155 homes for sale in New York NY. This is a big problem for investors who have many properties to choose from on Zillow but don't have time to review each one and their price history. New York real estate investors need to find new houses to invest in as quickly as possible because they want to find the best option.

As a result, developing a relational database helps efficiently manage properties in New York and aids investors in selecting houses faster. Companies rely on databases and must maintain them consistently (Gregurec, 2011). The New York real estate database allows the investors to select what can be purchased, taking into account the historical price increase, knowing the neighborhood, the agent, and the agency where they can buy the house and its entire price history.

A real estate investors process involve the following steps:

* Finding all available homes in New York City.
* Finding the agents and agencies that were listing those homes.
* Finding the neighborhoods.
* Analyzing the house's historical price.
* Selecting the houses that have the greatest increase in price.
* Deciding the houses that investors should buy.
* The real estate investors can buy the properties.

In conclusion, this project aims to develop a real estate database for investors. The use cases are: Analyzing the increase in housing prices. Selecting the houses that have the greatest increase in price and that could be purchased. Identifying where the selected houses are located.Finding the agents and agencies that have the selected houses.

**Objectives**

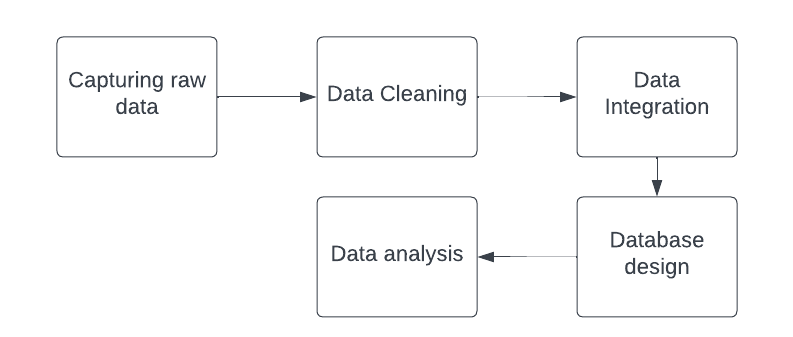
General objective: To design and build a real estate database for investors using data from Zillow.com.

**Specific objectives:**

* To design a conceptual real estate database.
* To design a logical real estate database.
* To design a physical real estate database.
* To analyze and evaluate the real estate investors database highlighting strengths and limitations.
* To analyze the legal and compliance requirements applicable to the build a database using web scraping.
* To fulfill real estate investors' use cases.

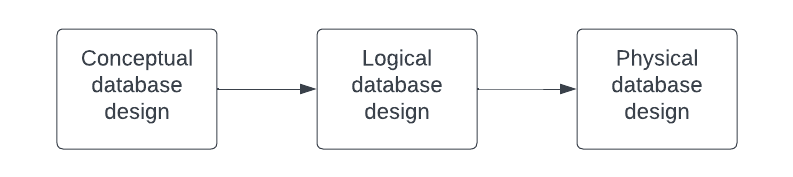
**Methodology**

The methodology is presented step by step in the main phases of the data management process, as shown in Figure 1 (EMC, 2015).



1. Capturing raw data.
   1. Data scraping.
2. Data cleaning.
3. Data integration.
4. Database Design

The database design is based on that developed by (Connolly, Thomas, et al,2014). The main steps as shown in Figure 2.



* 1. Conceptual database design
     1. Build a conceptual data model.
        1. Identify entity types.
        2. Identify relationship types.
        3. Identify and associate attributes with entity or relationship types.
        4. Determine attribute domains.
        5. Determine candidate, primary, and alternate key attributes.
        6. Check the model for redundancy.
        7. Validate conceptual data model against user transactions.
        8. Review conceptual data model with user.
  2. Logical database design
     1. Build a logical data model.
        1. Derive relations for logical data model.
        2. Validate relations using normalization.
        3. Validate relations against user transactions.
        4. Check integrity constraints.
        5. Review logical data models with users.
        6. Check for future growth.
  3. Physical database design
     1. Translate logical data model for target DBMS.
        1. Design base relations.
        2. Design representation of derived data.
        3. Design general constraints.
     2. Design file organizations and indexes.
        1. Analyze transactions.
        2. Choose file organizations.
        3. Choose indexes.
        4. Estimate disk space requirements
     3. Design user views.
     4. Design security mechanisms.
     5. Consider the introduction of controlled redundancy.
     6. Monitor and tune the operational system.

1. Data analysis

Connolly, Thomas, et al,2014,Database Systems: a Practical Approach to Design, Implementation, and Management, Global Edition.

EMC E. (2015) Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and

Presenting Data.

(Loukissas,2018) All the Homes: Zillow and the Operational Context of Data.

(Gregurec,2011) <https://www.circleinternational.co.uk/wp-content/uploads/2021/01/IJMC_13-4-new-1.pdf#page=167>

Zillow. Com https://www.zillow.com/new-york-ny/