**Use Case Description:**

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But a marketing statistique study proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

*GreatHouse Holdings* is a real estate company located in the west coast of the US. As a part of digital transformation *GreatHouse Holdings* wants to take advantage of internal and external data to improve their sales by matching the offer and the demand on reals estate market.

**Goals:**

To achieve their objectives, *GreatHouse Holdings* will use the following data :

* Stocks of houses to sell (nb of rooms, bedrooms and of course the prices)
* District referential containing information about the district : address of house (longitude, latitude)
* Real estate ad from competitors

## Agence ask you to design an end to end application to display all the properties on sale? The goal is to help the real estate agency know and negotiate better the final price of each home, and to help the home buyer to find efficiently their dream house.

For Avanade Data Engineer, we will accomplish the following steps to achieve the client needs.

**Step 1:**

Build the following Infra

* + Create an account storage,
  + Databricks workspace,
  + AKV,
  + Git,
  + Back AKV and mount Account storage
  + Azure Data factory

All useful links :

**Mount azure blob storage**: *https://docs.databricks.com/\_static/notebooks/data-sources/mount-azure-blob-storage.html*

**Security- secret scopes:** [*https://docs.microsoft.com/en-us/azure/databricks/security/secrets/secret-scopes*](https://docs.microsoft.com/en-us/azure/databricks/security/secrets/secret-scopes)

**Delta lake** : *https://delta.io/*

**Step 2:**

After years of operations, the business of Agence works wells, then it decided to extend their services.

*GreatHouse Holdings* bought another real estate company *Roger&Brothers*, the company wants to integrate data into one place this data should be conformed and cleansed.

For this objectif, Avanade should propose a stategie for data integration and data clansing as follow:

*Data integration Scenario:*

The data should be **integrated from both companies**.

The data should be traited to drop **duplication** and **wrong** information.

The datasets are:

* Stocks of houses to sell
* District referential containing information about the district

**Step 3:**

*GreatHouse Holdings* would like to have a better understanding about the market, it helps them to design next year’s strategy about houses on sale.

For this use case, Avanade has some exploration scenarios to treat and KPI’s to put in place:

* Analyse the average price per nomber of rooms
* Analyse the different indicators the lowest, highest, average of price from each district
* Which kind of house is most in the market (such as how many rooms, how many bedrooms)?
* For different price range, analyse the difference of houses (nb of rooms, bedrooms)?
* (optionnal) Could you build a price map by the address of house (geolocation information)?
* What other interesting analysis could you think of?

**Step 4:**

*GreatHouse Holdings* company would like to share these analysis results with all the members of the company.

Could you choose an user-friendly way to visualize all your results?

**(optional) Step 5:**

The data engineering team created a data flow to collect in real time new houses proposed by the competition. *GreatHouse Holdings* wants to analyse :

**About the study referentiel of district file**

1. longitude: A measure of how far west a house is; a higher value is farther west  
  
2. latitude: A measure of how far north a house is; a higher value is farther north  
  
3. housingMedianAge: Median age of a house within a block; a lower number is a newer building  
  
4. totalRooms: Total number of rooms within a block  
  
5. totalBedrooms: Total number of bedrooms within a block  
  
6. population: Total number of people residing within a block  
  
7. households: Total number of households, a group of people residing within a home unit, for a block  
  
8. medianIncome: Median income for households within a block of houses (measured in tens of thousands of US Dollars)  
  
9. medianHouseValue: Median house value for households within a block (measured in US Dollars)  
  
10. oceanProximity: Location of the house w.r.t ocean/sea

**Use Case Goals :**

* Exercise 1: Setup infra
  + Create an account storage,
  + Databricks workspace,
  + AKV,
  + Git,
  + Back AKV and mount Account storage
* Exercise 2: Load Source data
  + Create bronze area and drop source data
  + Create a silver area to store integrated clean data
  + Load data from bronze to silver
* Exercise 3: Operations
  + Create calculation process and store result in gold area
* Exercise 4: Visualization
  + Query the gold area to create visualizations
* Exercise 5:
  + Stream analytics