



AWS FOR DE MOCK PROJECT

HOME PRICES IN CALIFORNIA

NGUYỄN XUÂN THÁI HÒA





HOME PRICES IN CARLIFORNIA



INTRODUCTION

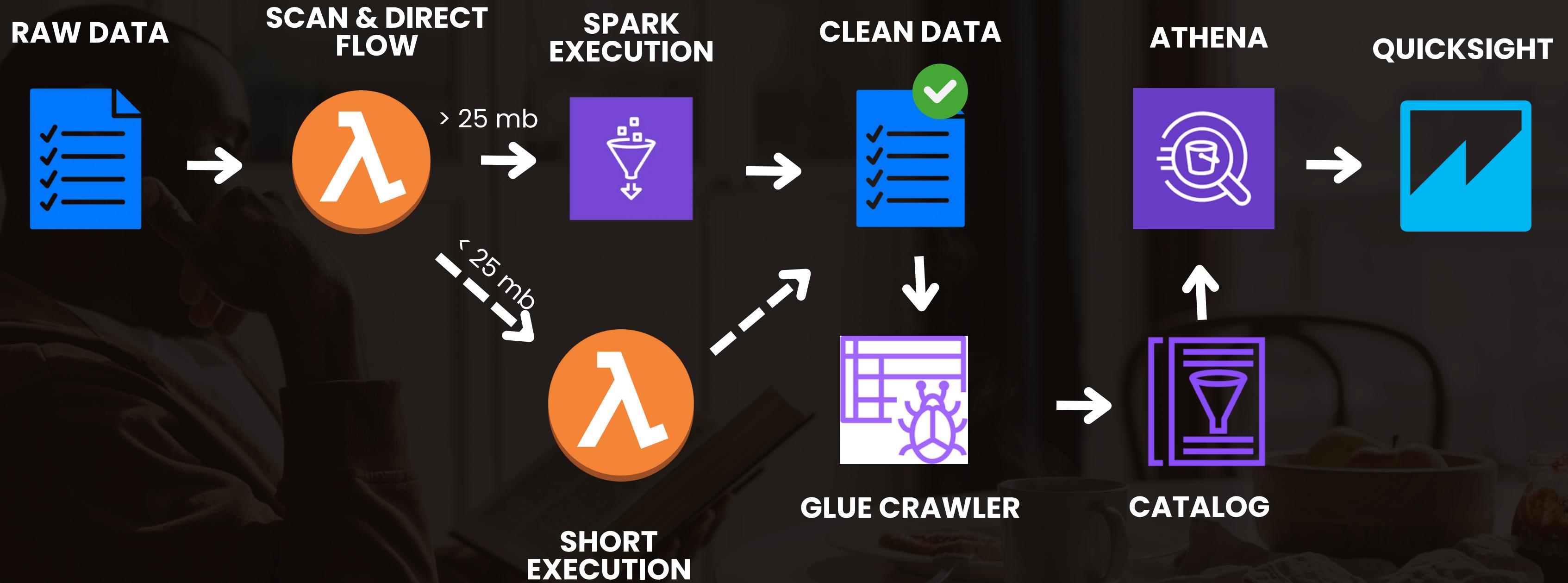
- There are different types of houses that are being sold in California.
- The dataset contains different information of houses on the online market.
- **Purpose:** Conducting an analysis and comparison of recent home prices in California.





HOME PRICES IN CARLIFORNIA

PIPELINE





HOME PRICES IN CARLIFORNIA

DATASET

The home prices dataset contains home informations:

- Id
- Summary
- Sold price
- Type
- Year built
-
- Heating
- Cooling
- Garage
- Area

Id	Address	Sold Price	Summary	Type	Year built
0	540 Pine Ln	3825000.0	540 Pine Ln, Los Altos, CA 94022 is a single f...	SingleFamily	1969.0
1	1727 W 67th St	505000.0	HURRY, HURRY.....Great house 3 bed and 2 bat...	SingleFamily	1926.0
2	28093 Pine Ave	140000.0	'THE PERFECT CABIN TO FLIP! Strawberry deligh...	SingleFamily	1958.0

DATA FLAWS

There are some problems with the dataset:

- Too many categorical values in a column
- Some unreasonable values
- Contain null values

NUMBER OF ROWS: 47539

NUMBER OF COLUMNS: 41

DATA PREPROCESSING



HOME PRICES IN CARLIFORNIA

SELECTING RELEVANT FEATURES

Only relevant columns are kept for more convenient data cleaning and saving storage



- Id
- Sold price
- Type
- Year built
- Area
- Listed price
- City
- Bedrooms
- Bathrooms
- Garage
- Heating system
- Cooling system

DATA PREPROCESSING



HOME PRICES IN CARLIFORNIA

Some values are shown to be unreasonable:

- AREA < 10 sqft
- Year built > 2024
- Sold price < 100 \$

}

Discard!



MOVING OUTLIER

DATA PREPROCESSING

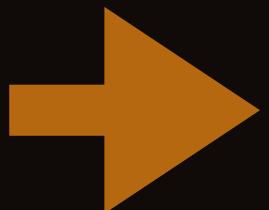


HOME PRICES IN CARLIFORNIA

HANDLING COLUMNS WITH TOO MANY CATEGORICAL VALUES

In house type, there are too many complicated values (52 types)

- Detached, 2 Story
- Detached, 4+ Story
- SingleFamily
- Detached, 3 Story
- 4+ Story
- 3 Level, Live/Work
- Townhouse
- 3 Story, Duplex, House & Apartments, Townhouses
- Luxury, Townhouse
- Manufactured Home
- Mobile Home
- MobileManufactured
- Studio
- Top Floor, Studio
- ...



Categorize them into 6 main types



Single-family



Appartement



Mobile house



Condominium



Townhouse



Tiny house

DATA PREPROCESSING



HOME PRICES IN CARLIFORNIA

HANDLING NULL VALUES

SingleFamily	1926.0	Combination	Wall/Window Unit(s), Evaporative Cooling, See ...
SingleFamily	1958.0	Forced air	Nan
SingleFamily	1947.0	Central	Central Air

Since the number of nulls/NaN is not too significant



Drop the null/NaN values



VISUALIZATION

I will show the visualization on QuickSight

CONCLUSION



HOME PRICES IN CARLIFORNIA



- The cost of a house is significantly influenced by its square footage, home type and geographic location.
- Home prices are higher in big cities.
- The listed price of the house is usually 20% higher than the final sold price.
- The average house price changes through years, reaches highest in 2019 and it is prone to decrease in the next years.