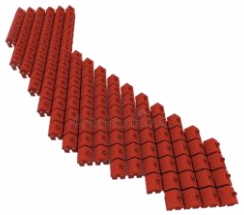


California Housing Market Application

Group: Data Works

- Nupur Pathak
- Revathi Boopathi
- Sree Divya Cheerla
- Vani Bhat

Agenda

I. **Introduction**
   a. Project Goals and Objective
   b. Problem and Motivation

II. **System Design**
   a. Architecture Design
   b. Date Collection
   c. Data Model

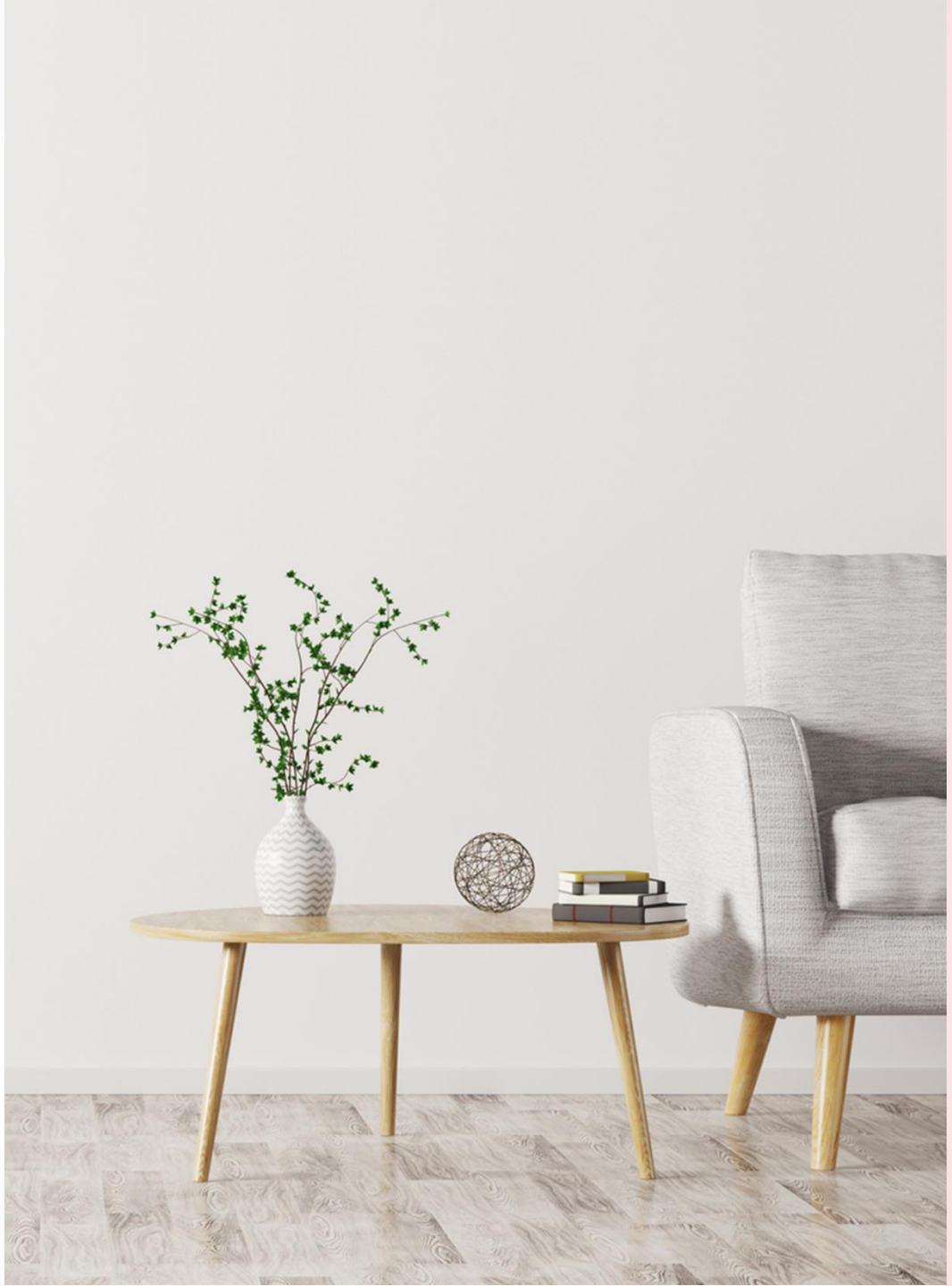
III. **System Implementation**
   a. Setup AWS resources
   b. Data Storage
   c. Extract, Transform, and Load (ETL)
   d. Modeling
   e. Analytics and Visualization

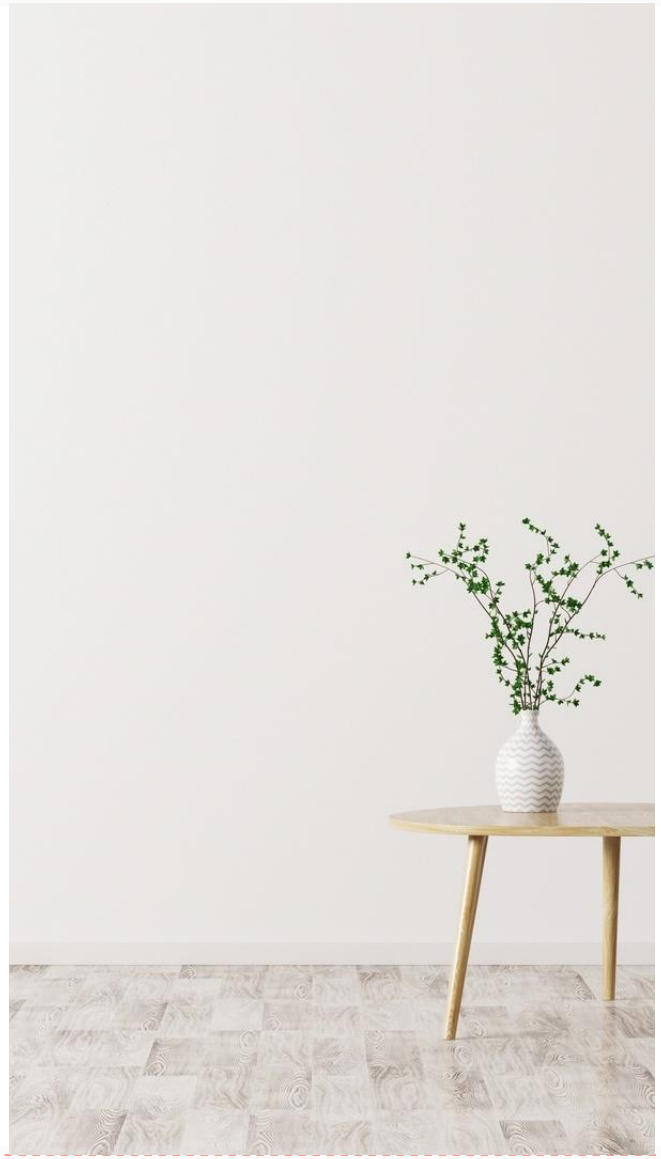
IV. **Conclusion and Future Work**

I. Introduction

a. Project Goals and Objective

b. Problem and Motivation





Project Goals and Objectives

- Build California housing market application to ease the process of housing property search for the customers based on their preferred buying criteria
- Influence on house/property values by the presence of public schools with higher ranking in the region

Problem and Motivation

The surge in demand for the US housing market and shortage of inventory has made it difficult for people to find their dream home with all the features they desire with ease. Hybrid work culture resulting from the pandemic is fueling the demand in the housing market.

There are various parameters that are impacting property values such as:

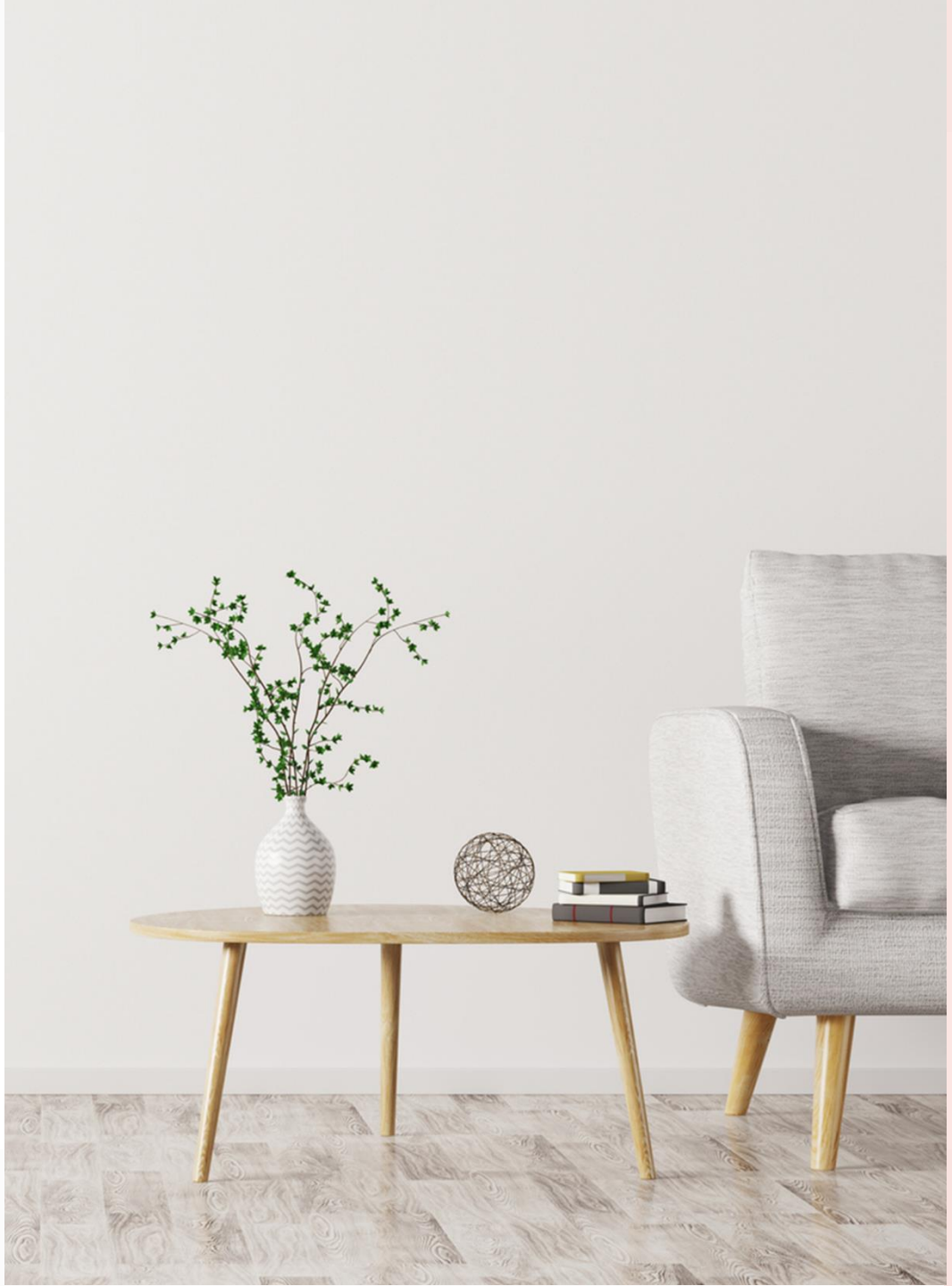
- School proximity
- Migration of people impacting the population density
- Crime rate
- Job opportunities
- New construction
- Recreational centers
- Medical care facilities
- Access to public transportation, and others

II. System Design

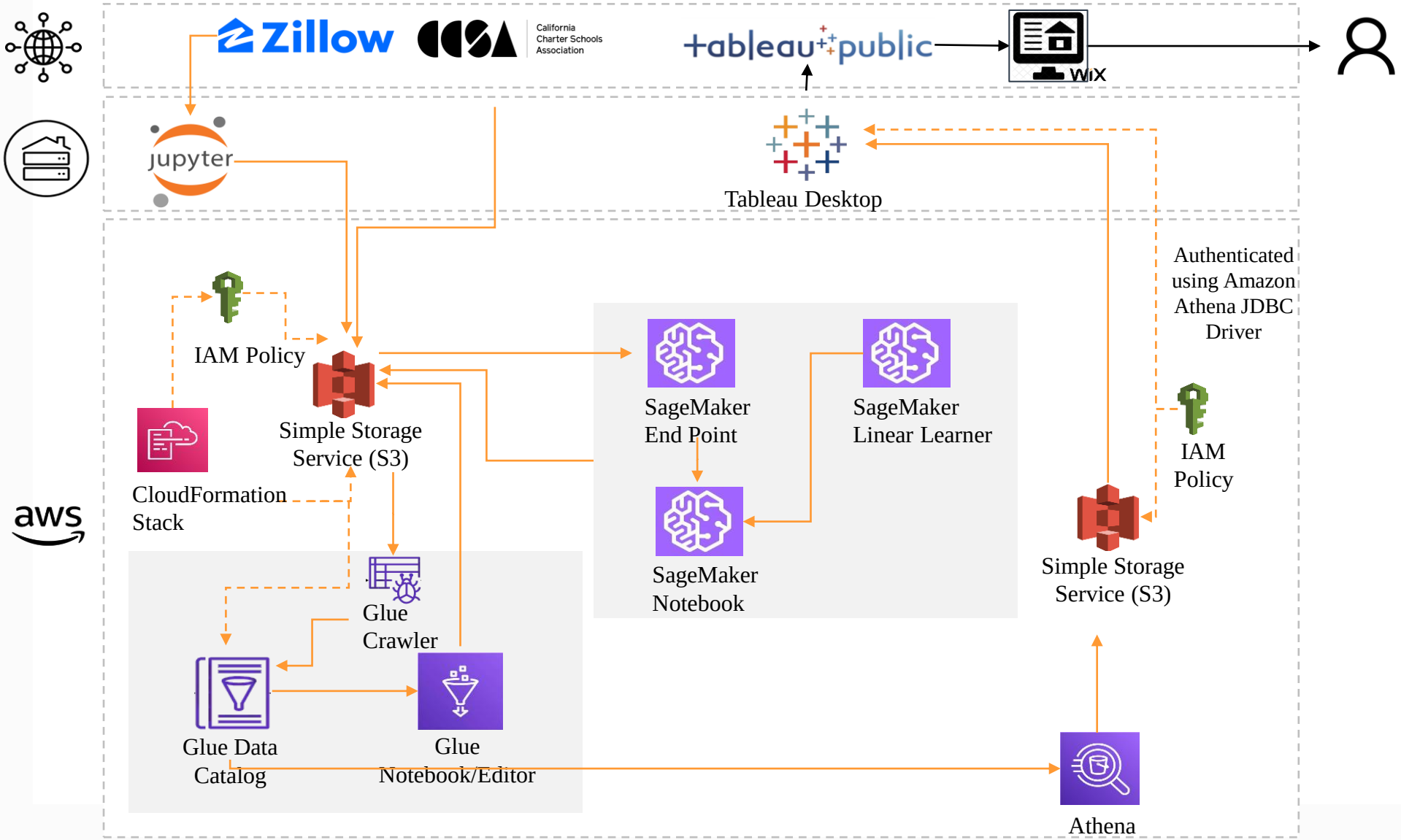
a. Architecture Design

b. Data Collection

c. Data Model



Architecture Design



Data Collection: Housing Data + School Ranking

Zillow

jupyter

Housing Data:

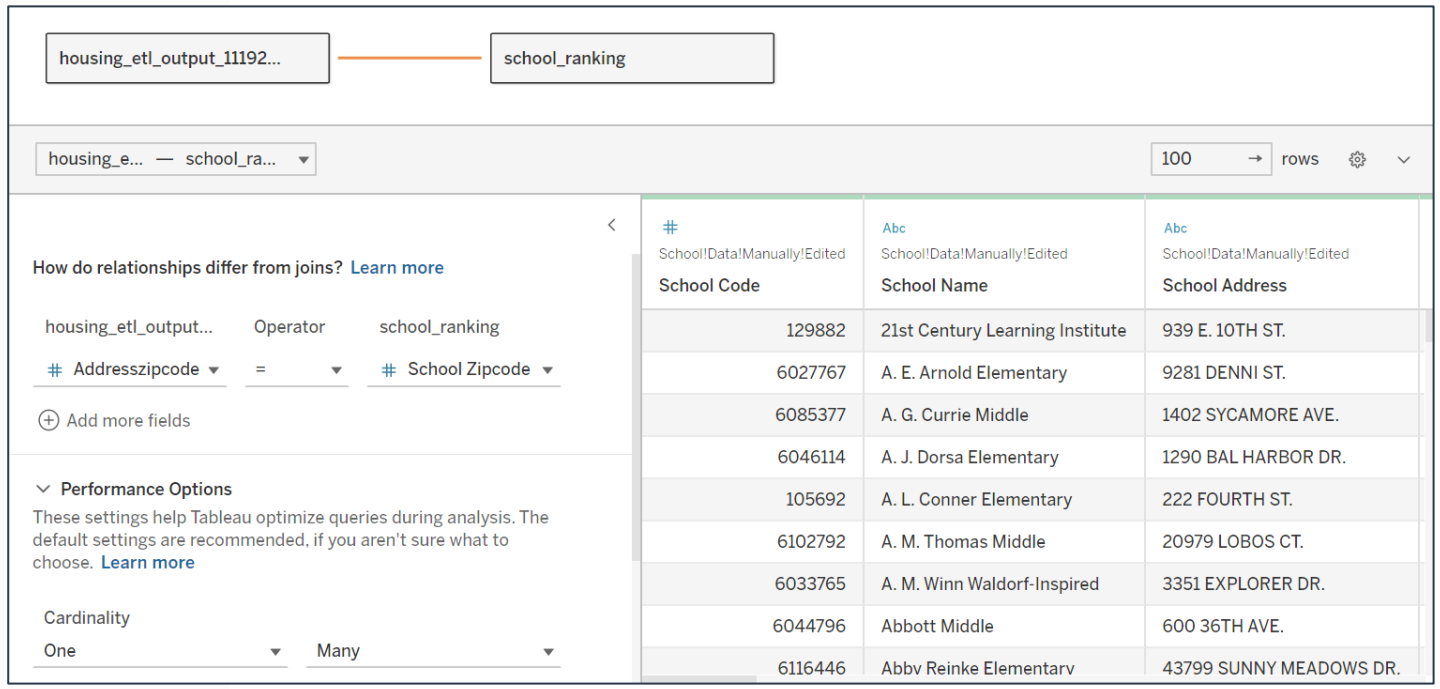
- Data on the house listings is collected with the help of web scraper
- The data is scraped for cities in California iterating listings on each page for the search results
- Listings sourced from the site are collated in Jupyter notebook

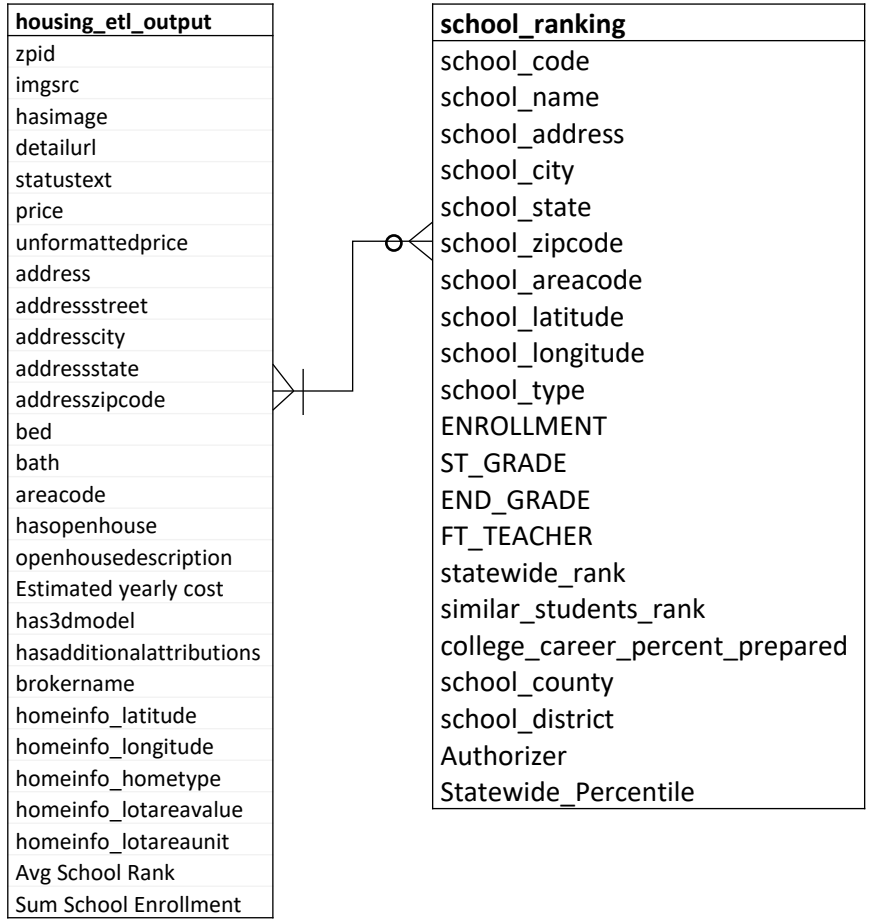
CCSA California Charter Schools Association

School Ranking Data:

- California Charter Schools Association provides the rating for public and private schools in California
- The school state-wide rankings are sourced from CCSA for 2019. Each school is ranked on a scale of 10.

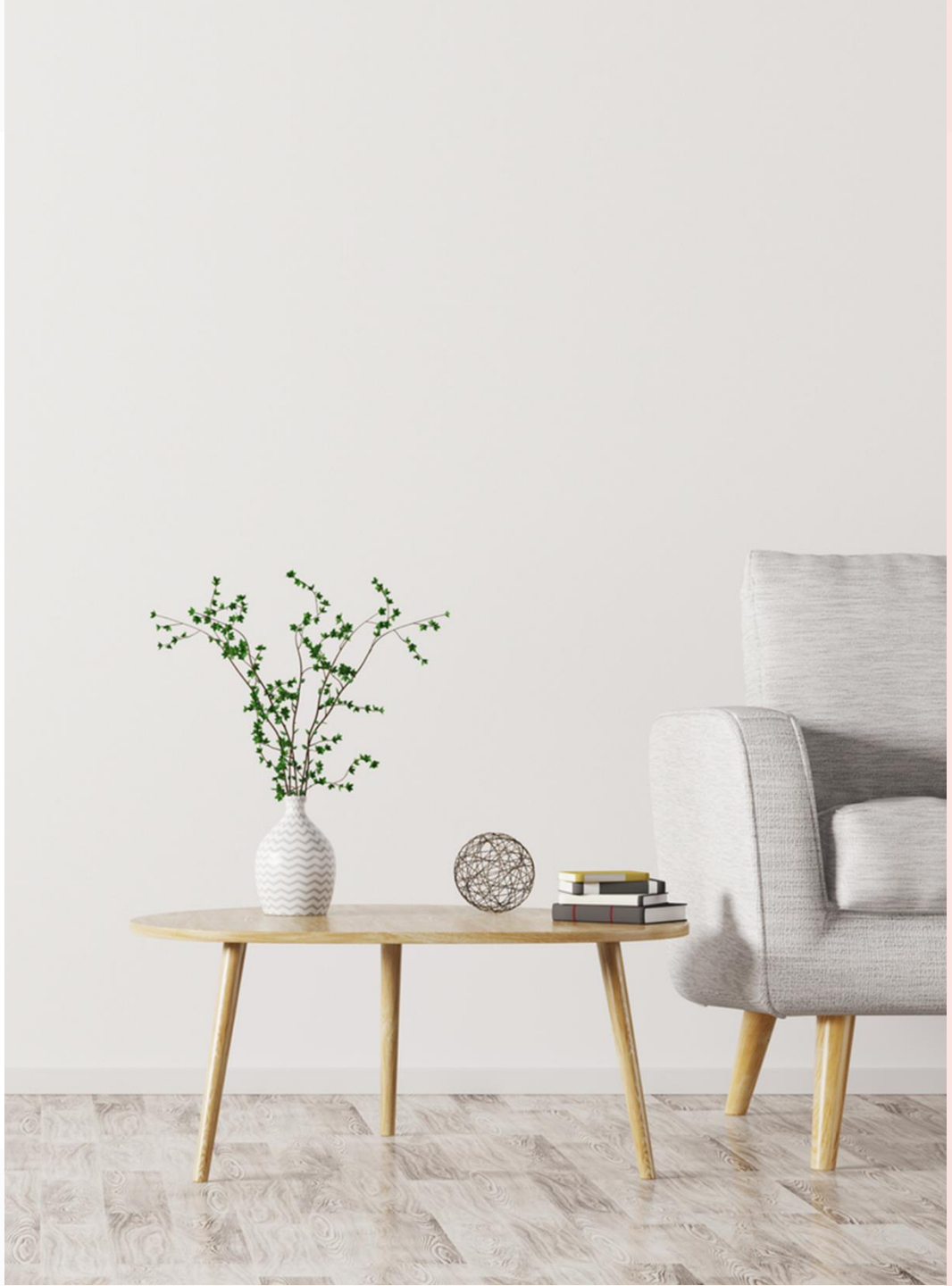
Data Model


housing_etl_output_11192...
school_ranking
housing_e... — school_ra...
100 rows
How do relationships differ from joins? Learn more
housing_etl_output...
Operator
school_ranking
Addresszipcode
=
School Zipcode
Add more fields
Performance Options
These settings help Tableau optimize queries during analysis. The default settings are recommended, if you aren't sure what to choose. Learn more
Cardinality
One
Many
School!Data!Manually!Edited
School Code
School Name
School Address
129882 21st Century Learning Institute 939 E. 10TH ST.
6027767 A. E. Arnold Elementary 9281 DENNI ST.
6085377 A. G. Currie Middle 1402 SYCAMORE AVE.
6046114 A. J. Dorsa Elementary 1290 BAL HARBOR DR.
105692 A. L. Conner Elementary 222 FOURTH ST.
6102792 A. M. Thomas Middle 20979 LOBOS CT.
6033765 A. M. Winn Waldorf-Inspired 3351 EXPLORER DR.
6044796 Abbott Middle 600 36TH AVE.
6116446 Abby Reinke Elementary 43799 SUNNY MEADOWS DR.


housing_etl_output
zpid
imgsrc
hasimage
detailurl
statustext
price
unformattedprice
address
addressstreet
addresscity
addressstate
addresszipcode
bed
bath
areacode
hasopenhouse
openhousedescription
Estimated yearly cost
has3dmodel
hasadditionalattributions
brokername
homeinfo_latitude
homeinfo_longitude
homeinfo_hometype
homeinfo_lotareavalue
homeinfo_lotareaunit
Avg School Rank
Sum School Enrollment
school_ranking
school_code
school_name
school_address
school_city
school_state
school_zipcode
school_areacode
school_latitude
school_longitude
school_type
ENROLLMENT
ST_GRADE
END_GRADE
FT_TEACHER
statewide_rank
similar_students_rank
college_career_percent_prepared
school_county
school_district
Authorizer
Statewide_Percentile

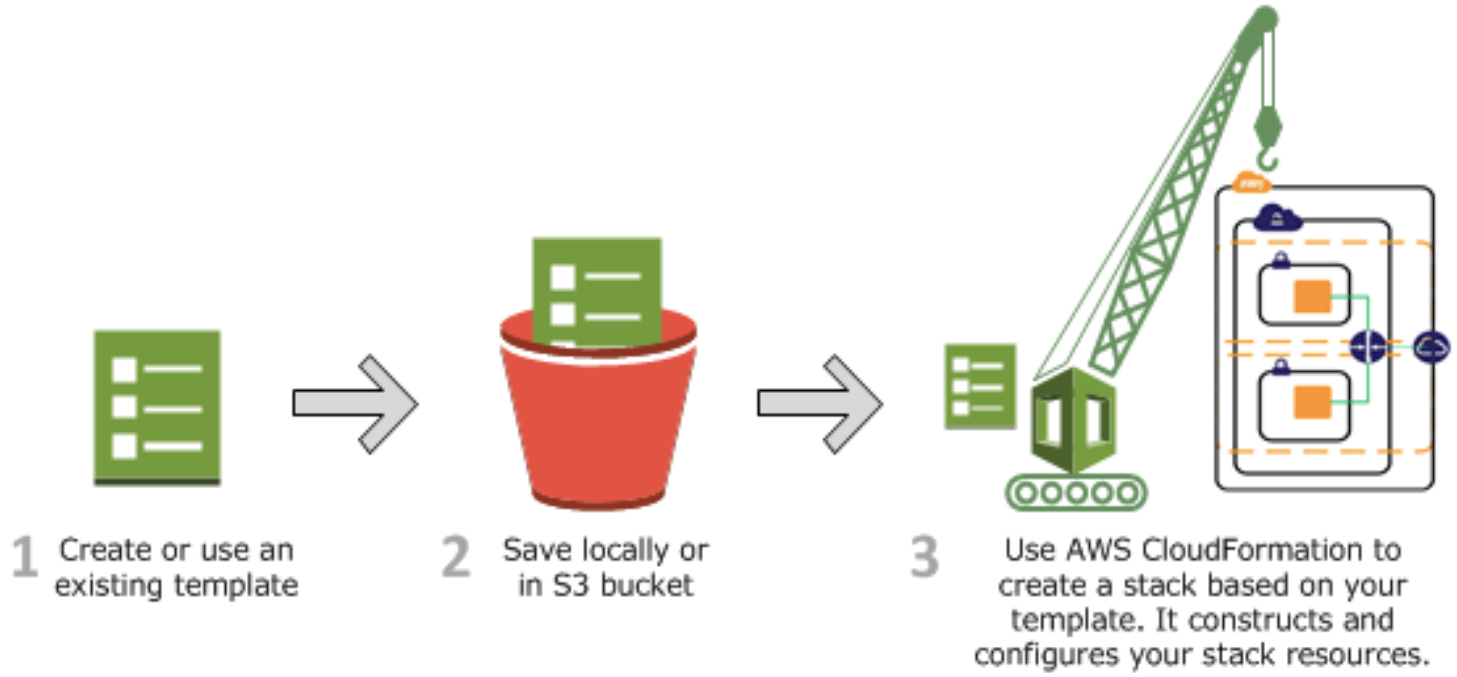
III. System Implementation

a. Setup AWS resources
b. Data Storage
c. Data Transformation (ETL)
d. Modeling
e. Analytics and Visualization



Setup AWS resources: AWS Cloud Formation

Automate resource creation for AWS GLUE



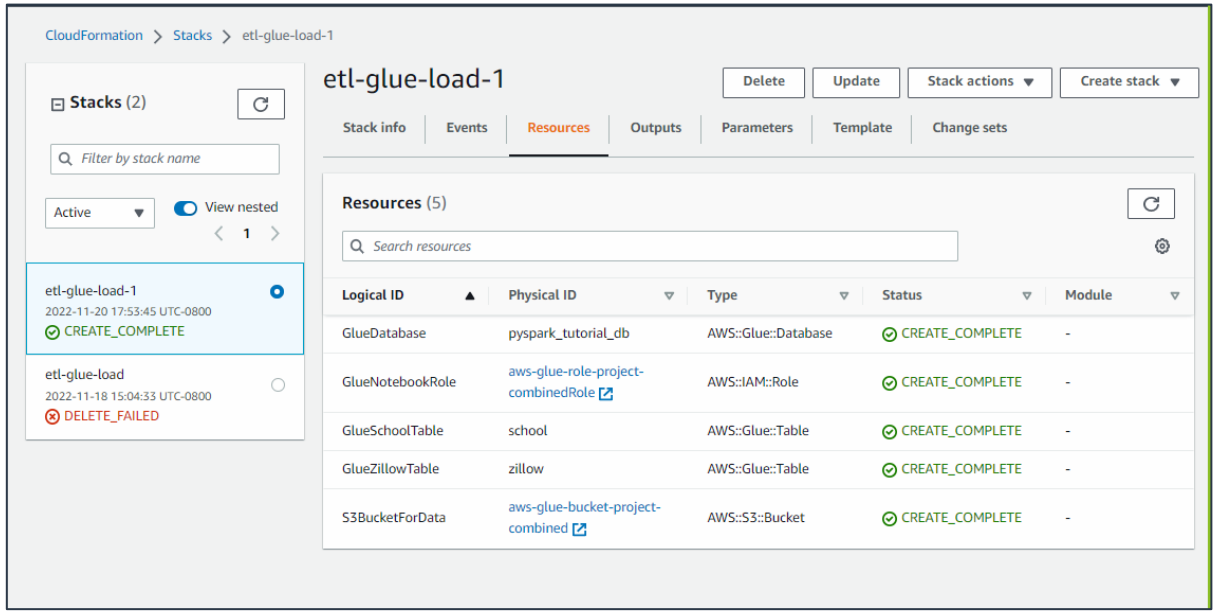
Setup AWS resources: AWS Cloud Formation

.yaml file to configure the resources

school_zillow_combined.yaml

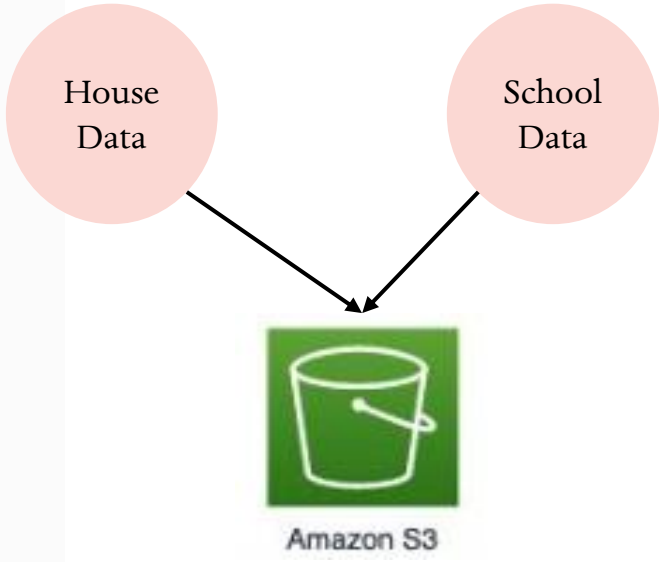
```
- Effect: Allow
  Action:
    - "glue:*"
    - "s3:*"
    - "ec2:DescribeVpcEndpoints"
    - "ec2:DescribeRouteTables"
    - "ec2:CreateNetworkInterface"
    - "ec2:DeleteNetworkInterface"
    - "ec2:DescribeNetworkInterfaces"
    - "ec2:DescribeSecurityGroups"
    - "ec2:DescribeSubnets"
    - "ec2:DescribeVpcAttribute"
    - "iam:ListRolePolicies"
    - "iam:GetRole"
    - "iam:GetRolePolicy"
    - "cloudwatch:PutMetricData"
    - "ec2:CreateTags"
    - "ec2:DeleteTags"
    - "logs:CreateLogGroup"
    - "logs:CreateLogStream"
    - "logs:PutLogEvents"
  Resource: "*"
- Effect: Allow
  Action:
    - "s3:*"
  Resource:
    - "arn:aws:s3:::{S3PySparkBucketName}/*"
    - "arn:aws:s3:::{S3PySparkBucketName}/"
- Effect: Allow
  Action:
  - "iam:GetRole"
  - "iam:PassRole"
```

Loads database, IAM roles, data tables with schema, S3 bucket

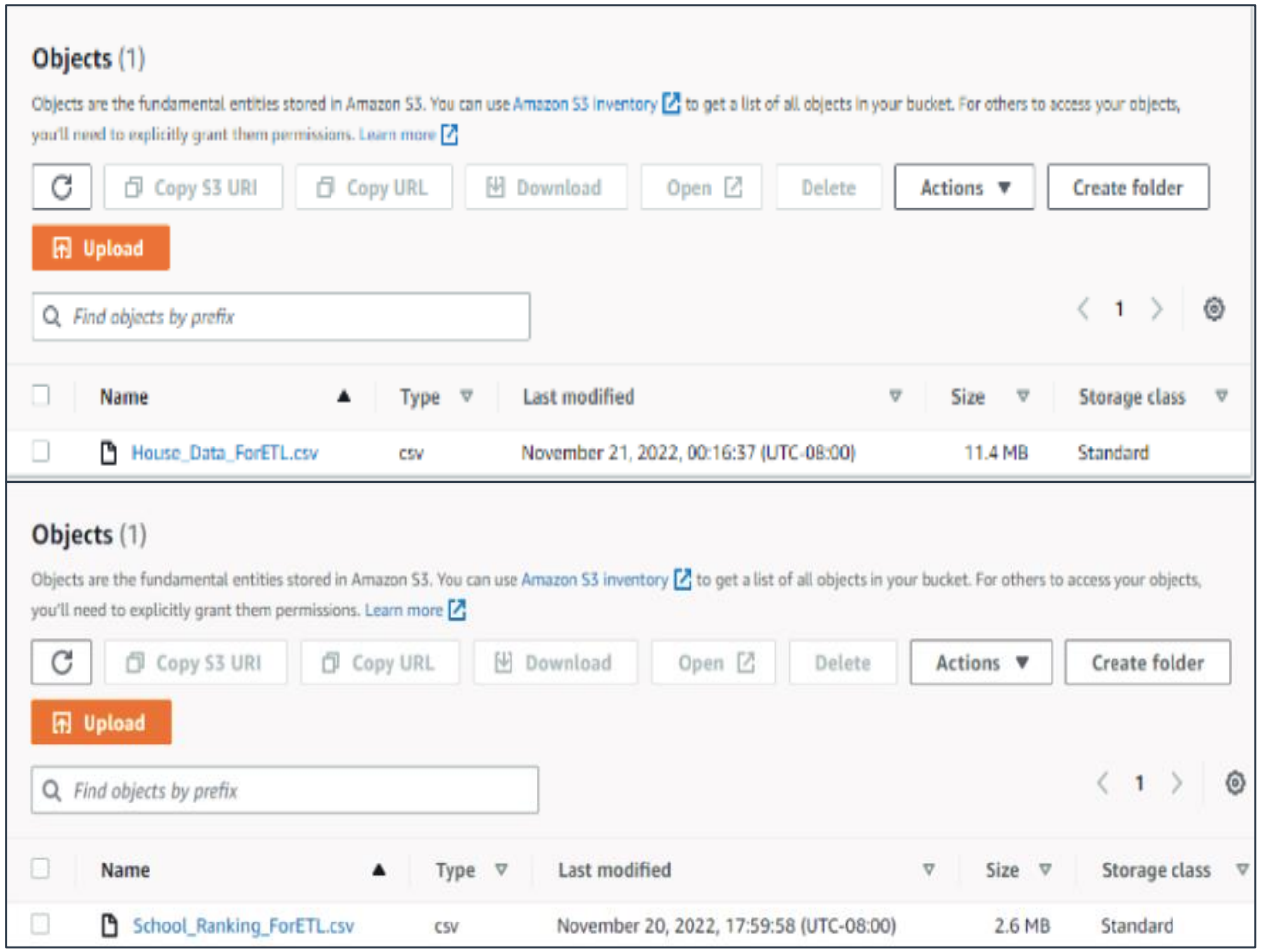


Data Storage: AWS S3

Add school and house data into S3 bucket

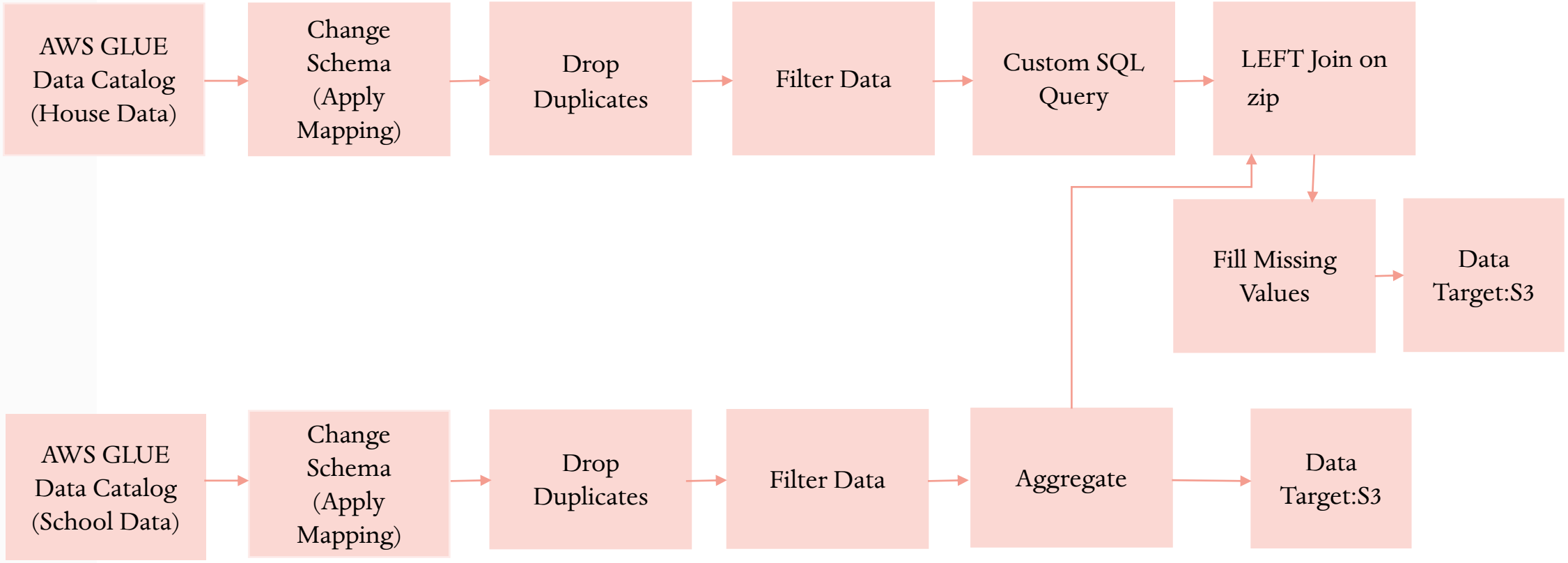


aws-glue-bucket-project-combined



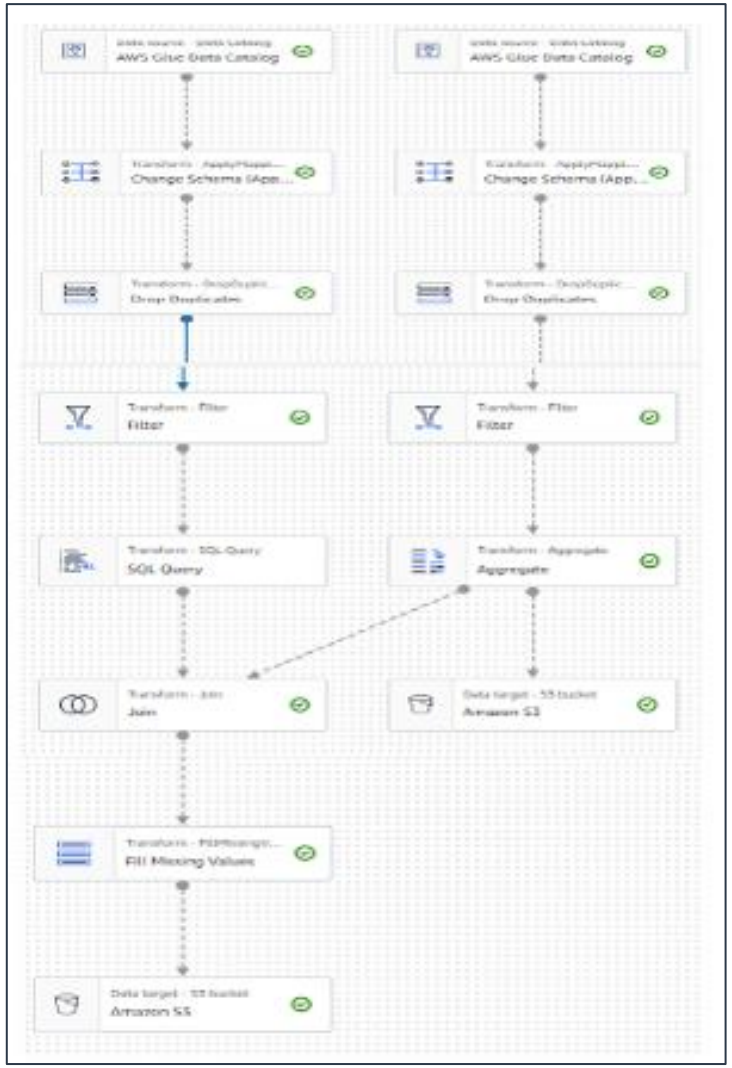
Data Transformation (ETL): AWS GLUE

AWS GLUE visual editor flow to transform the datasets and load to data target S3

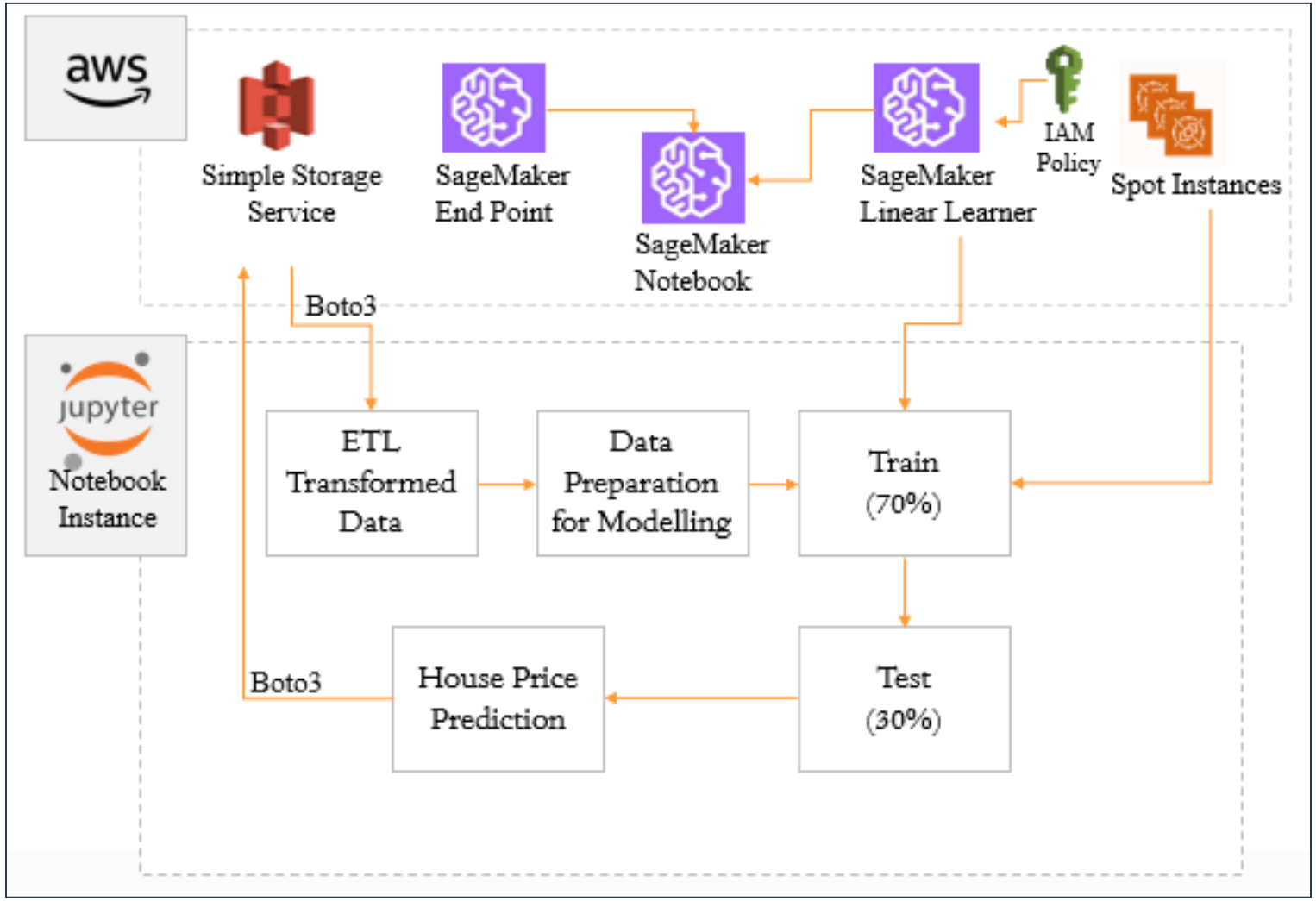


Data Transformation (ETL): AWS GLUE Visual Editor Flow Diagram

Transforming data from AWS GLUE Data Catalogue using AWS Glue's ETL Job



Modeling:
SageMaker Architecture and Process Flow



Modeling:
Model Features and Evaluation metrics

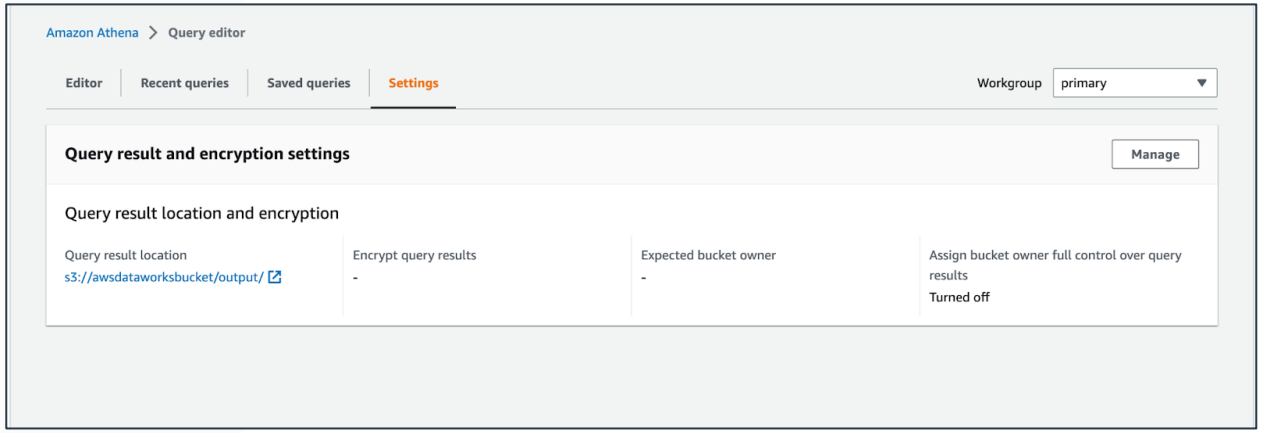
Features

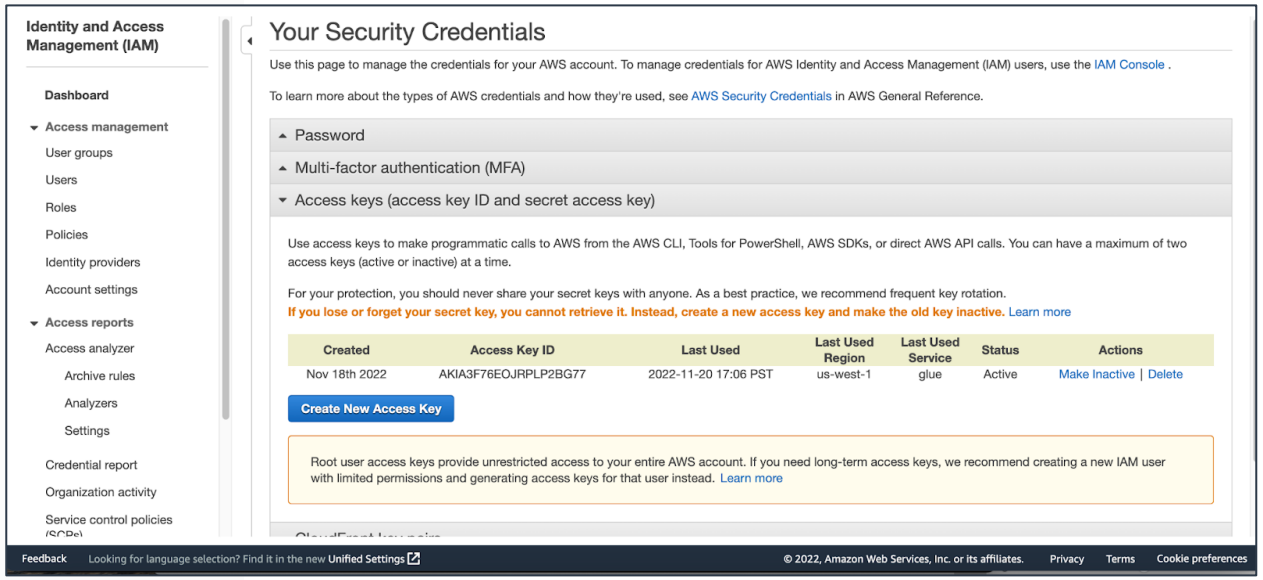
Bed	Bath	Additional Attributes	Home type	Price per square feet	Avg School Rank	School Enrollment	Estimated Yearly cost	Home price
2	1	False	Townhouse	$ 687	8.9	600	750 k	980 k
3	1	True	Single Family	$ 300	6	300	380 k	1.1 M
1	1	True	Condo	$ 200	2.3	250	250 k	200 k
…	…	…	…	…	…	…	…	…

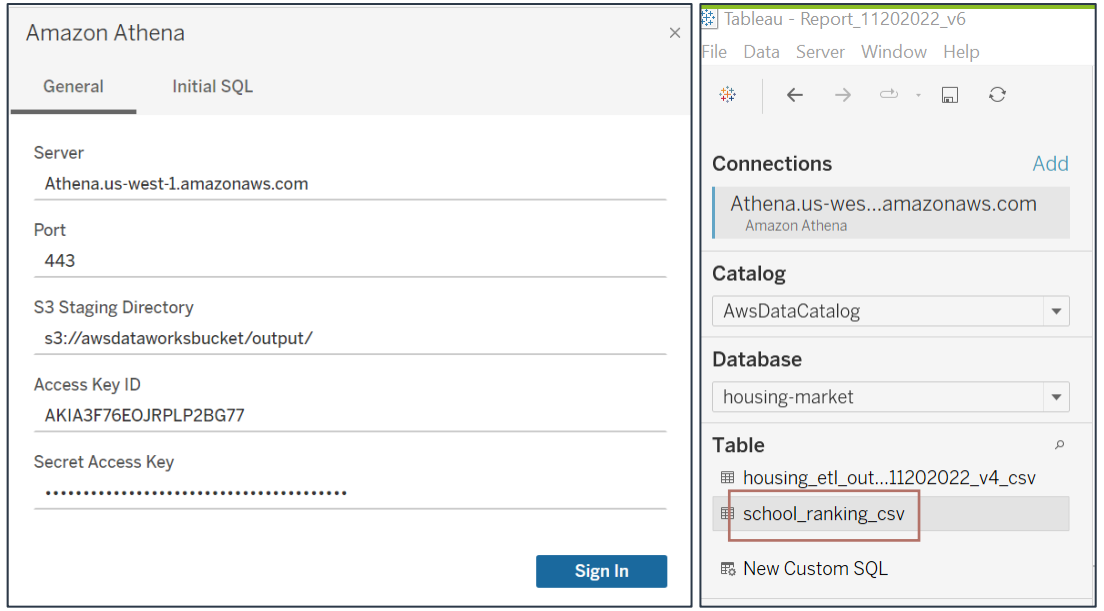
Evaluation

- R2 – 0.6
- Adjusted R2 – 0.59
- Explained Variance Score – 0.6

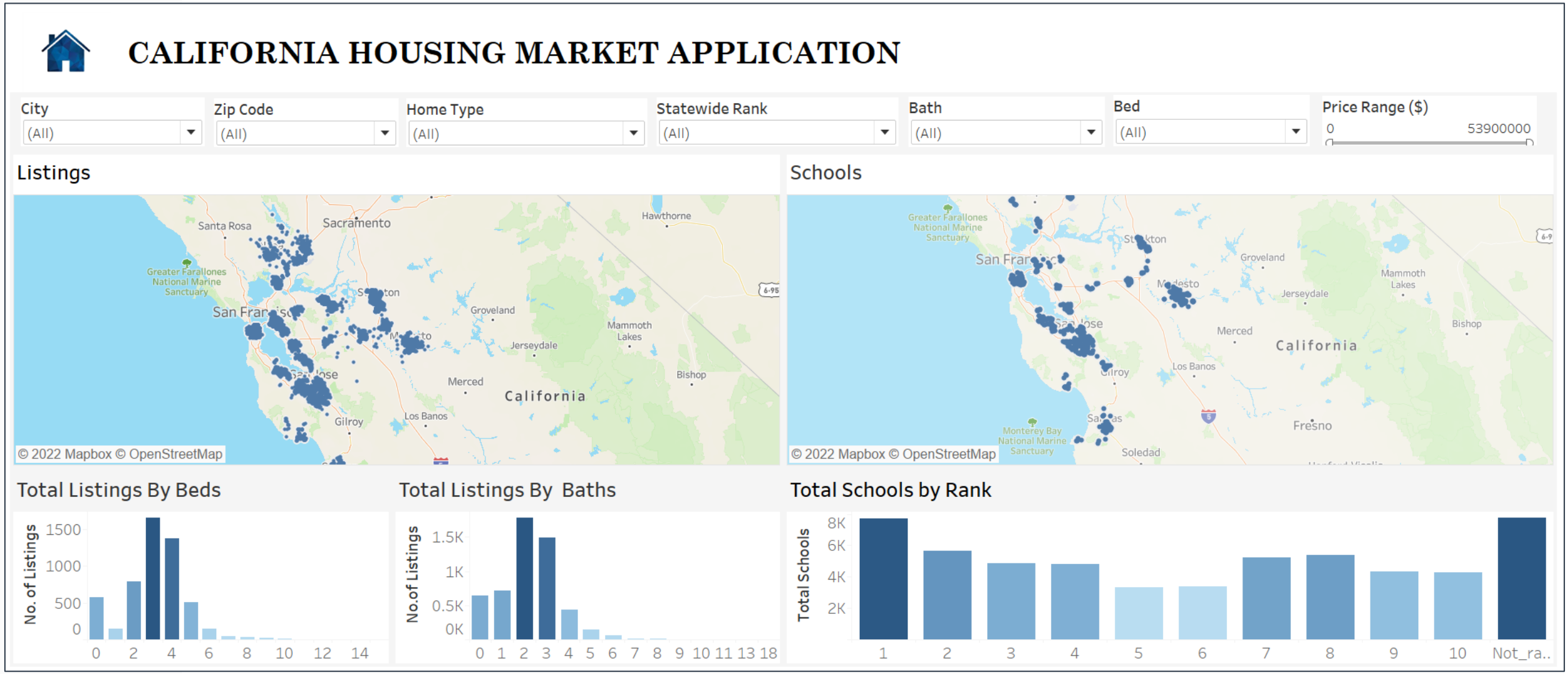
Analytics and Visualization: Amazon Athena to Tableau Dashboard







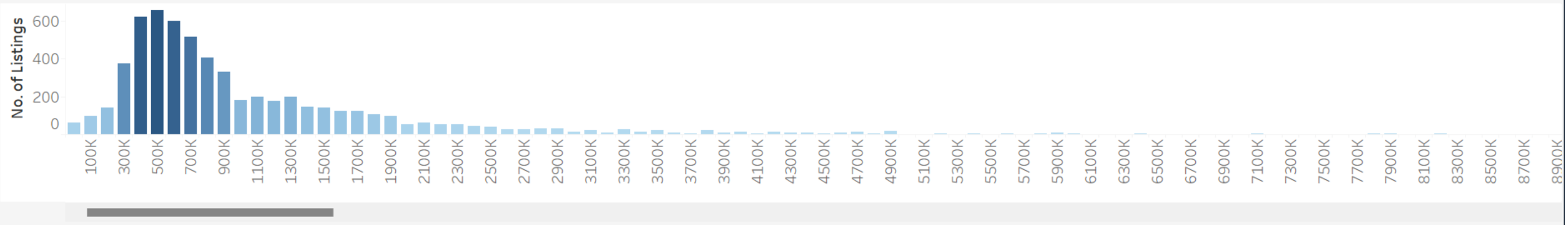
Dashboard: Home Page



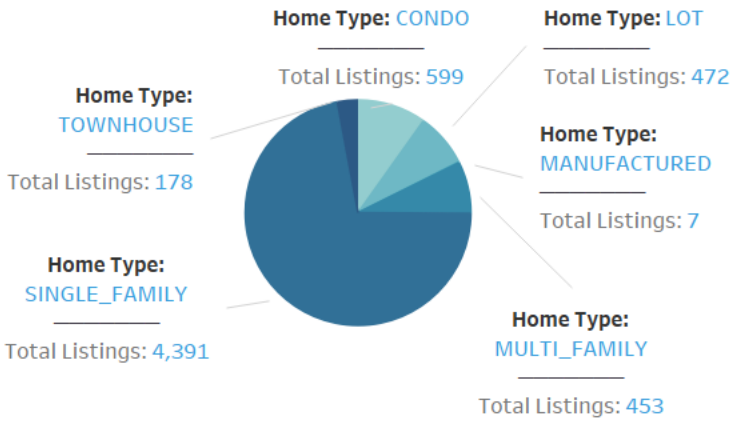
Dashboard: Listings Details

LISTINGS DETAILS

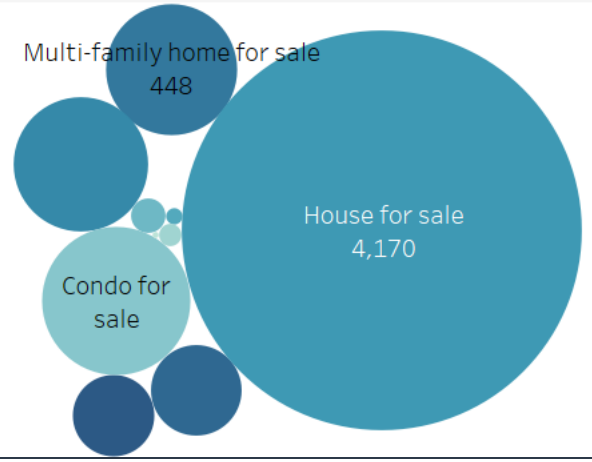
Total Listings by Price Range



Total Listings by Home Type



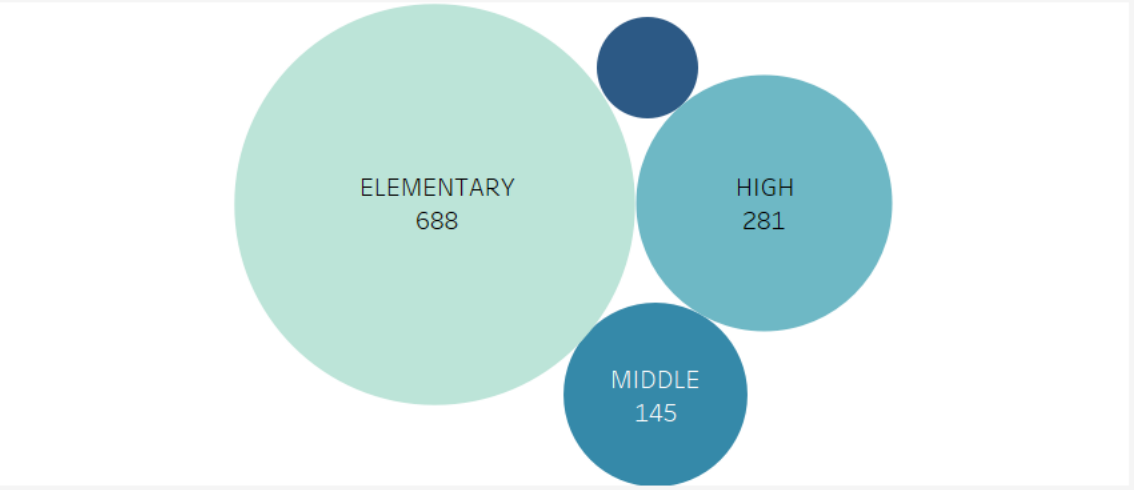
Status of Listings



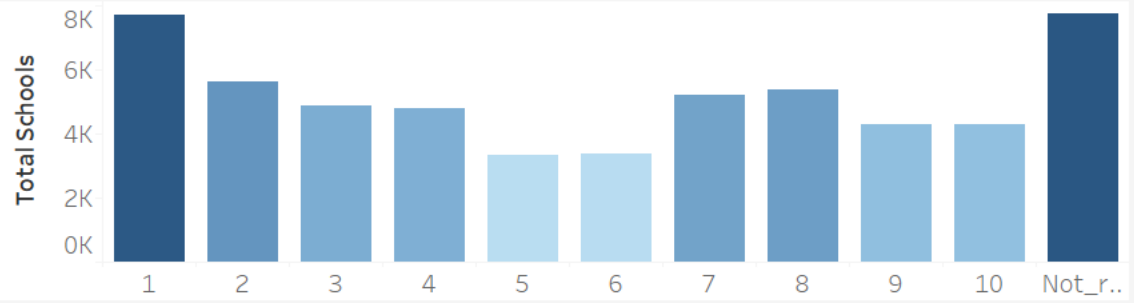
Dashboard: School Details

SCHOOL DETAILS

Total Schools by School Type



Total Schools by Rank

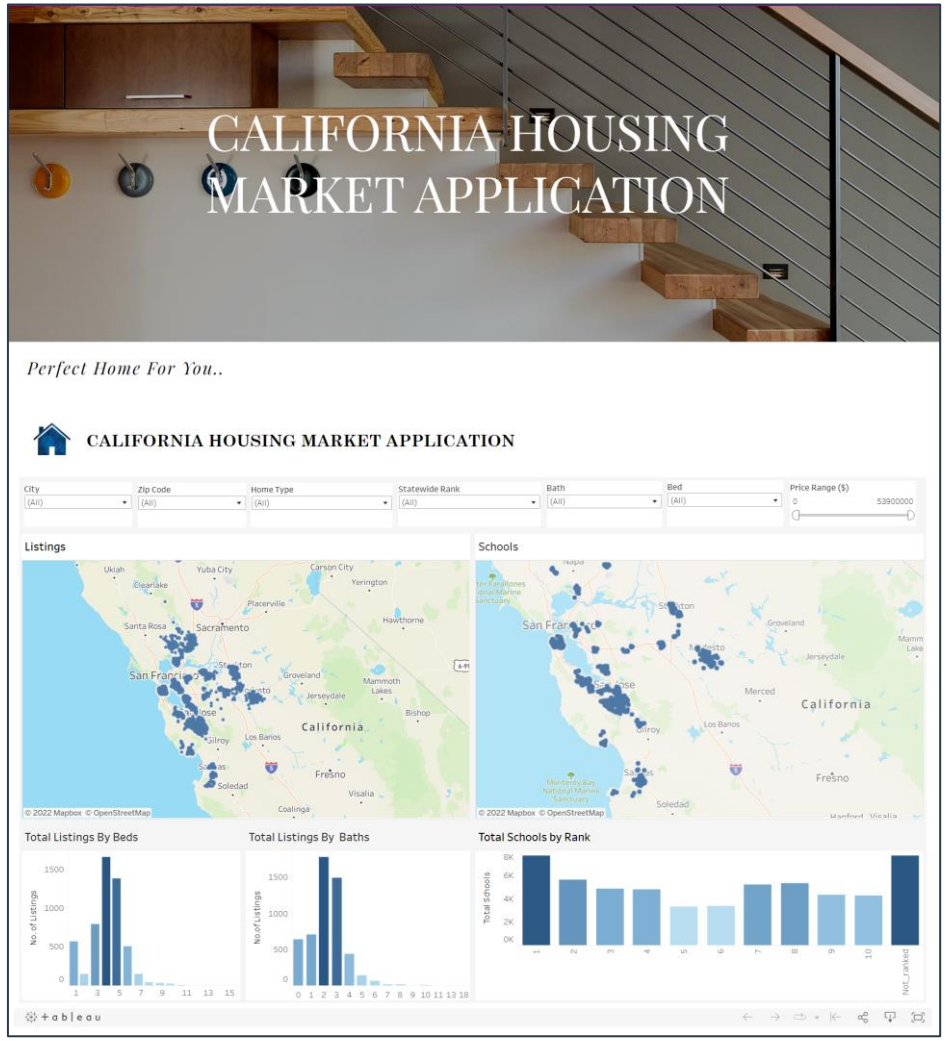


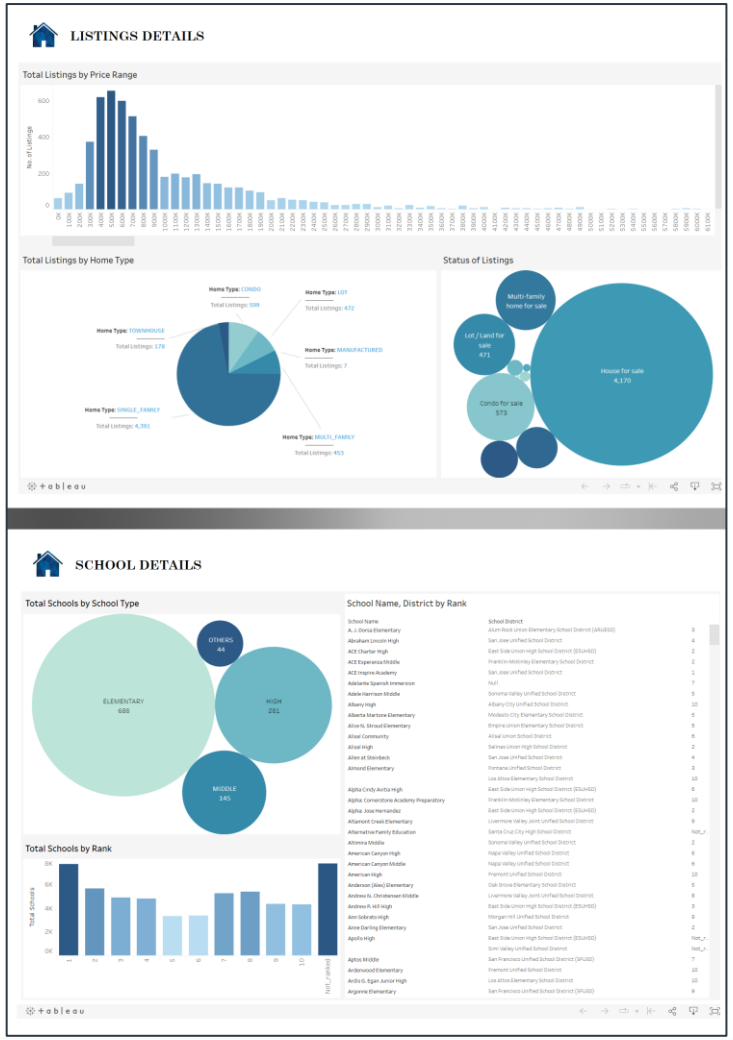
School Name, District by Rank

School Name	School District	
Aptos Middle	San Francisco Unified School District (SFUSD)	7
Ardenwood Elementary	Fremont Unified School District	10
Ardis G. Egan Junior High	Los Altos Elementary School District	10
Argonne Elementary	San Francisco Unified School District (SFUSD)	9
Argus High (Continuation)	Ceres Unified School District	Not_ranked
Ark Independent Studies	Santa Cruz City High School District	Not_ranked
Art Freiler	Tracy Joint Unified School District	6
Aspire East Palo Alto Charter	Ravenswood City Elementary School District	4
Aspire Langston Hughes Academy	Stockton Unified School District	3
Aspire Port City Academy	Stockton Unified School District	6
Aspire Summit Charter Academy	Ceres Unified School District	4
Aspire Vanguard College Preparatory ..	Empire Union Elementary School District	4
August Boeger Middle	Mt. Pleasant Elementary School District	3
August Elementary	Stockton Unified School District	4
Bagby Elementary	Cambrian Elementary School District	9
Balboa High	San Francisco Unified School District (SFUSD)	7
Baldwin (Julia) Elementary	Oak Grove Elementary School District	6
Barbara Spratling Middle	Keyes Union Elementary School District	5
Bardin Elementary	Alisal Union School District	2
Barrett Elementary	Morgan Hill Unified School District	3
Barron Park Elementary	Palo Alto Unified School District	9
Bay View Academy	Monterey Peninsula Unified School District	7
Bear Creek High	Lodi Unified School District	5
Belle Haven Elementary	Ravenswood City Elementary School District	1
Ben Painter Elementary	Alum Rock Union Elementary School District (ARUESD)	6
Berkeley Arts Magnet at Whittier	Berkeley Unified School District	9
Berkeley Technology Academy	Berkeley Unified School District	Not_ranked
Bernal Intermediate	Oak Grove Elementary School District	7
Bernard L. Hughes Elementary	Empire Union Elementary School District	4
Berryessa Union Elementary	Berryessa Union Elementary School District	Not_ranked
Big Sur Charter	Monterey Peninsula Unified School District	9

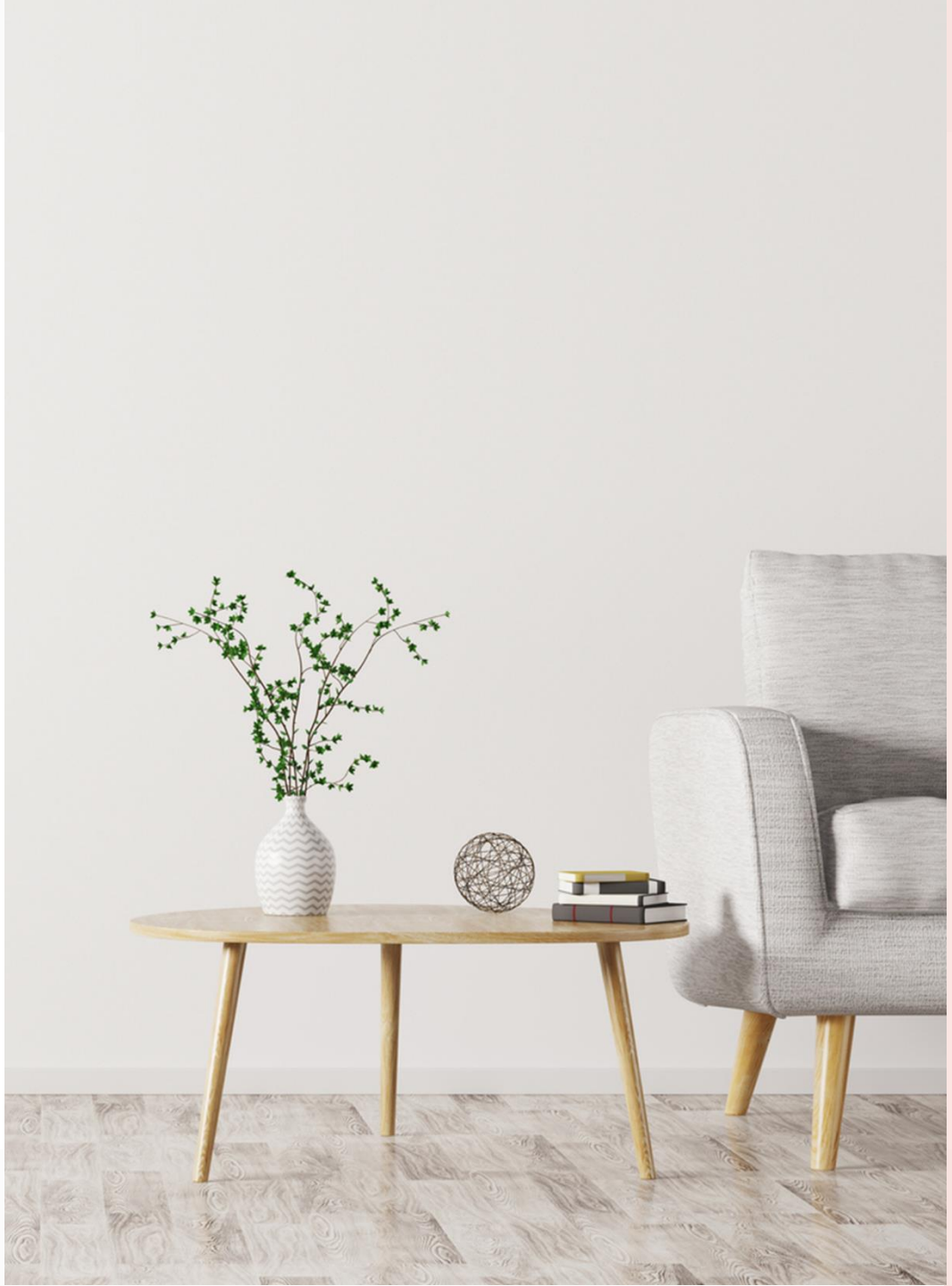
California Housing Market Application Site

Website Link





IV. Conclusion and Future Work



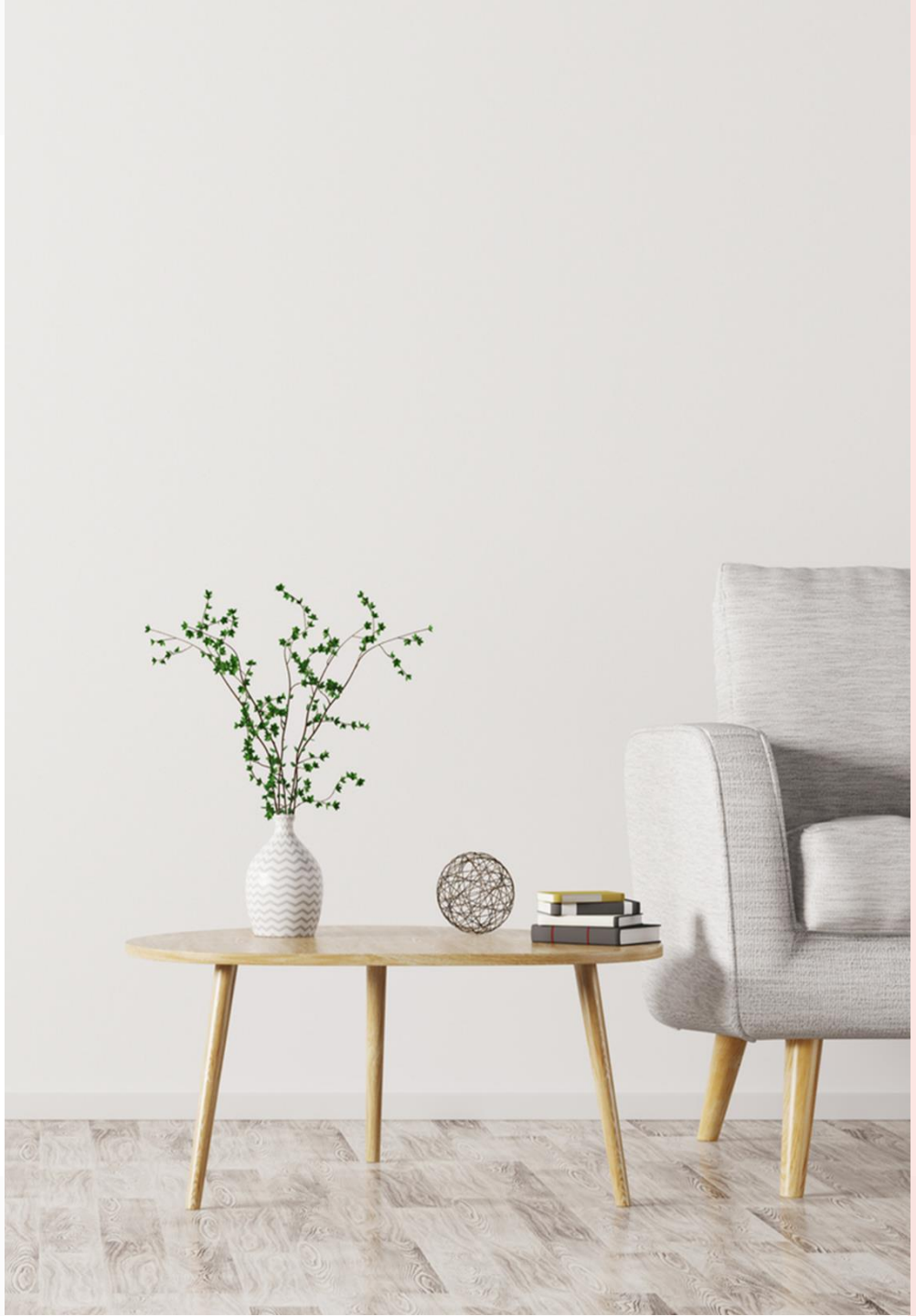
Conclusion and Future Work

Overall,

- Built a scalable ‘California Housing Market’ big data application leveraging various services in AWS cloud platform.
- It serves as a search platform for the home buyers to view the properties, get the best price quote and insights about house price variation due to the presence of schools

As a future work,

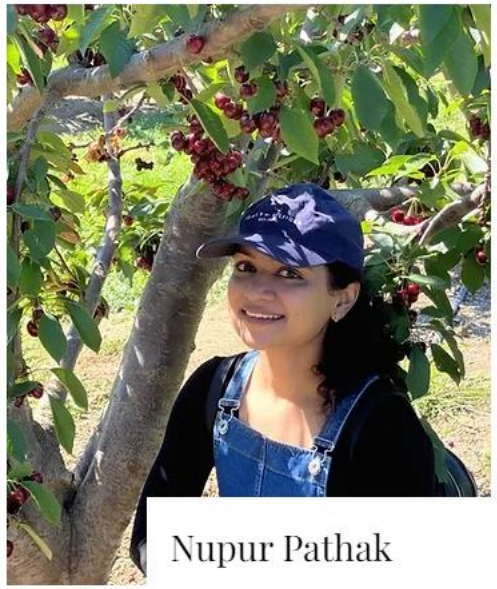
- The application can be improvised to use AWS Kinesis to live capture the data and analytics
- Event triggering through AWS Lambda to connect various AWS componet’s and data pipeline for continuous deployment and execution in a single click



References

- https://www.dreamstime.com/illustration/california-houses.html
- https://templates.office.com/en-us/presentations
- https://homeshots.us/
- https://www.ccsa.org/what-we-do/student-success
- https://www.zillow.com/homes/for_sale/
- https://www.tableau.com/
- https://public.tableau.com/app/discover
- https://www.zillow.com/https://jupyter.org/
- https://www.wix.com/
- https://www.dreamstime.com/illustration/california-houses.html
- https://templates.office.com/en-us/presentations
- https://homeshots.us/
- https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/cfn-whatis-howdoesitwork.html

Team



Nupur Pathak

Revathi Boopathi

Sree Divya Cheerla

Vani Bhat

Thank You