



University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

Electrical & Computer Engineering and Computer Science

Electrical & Computer Engineering & Computer Science (ECECS)

# Distributed & Scalable Data Engineering – Technical Report



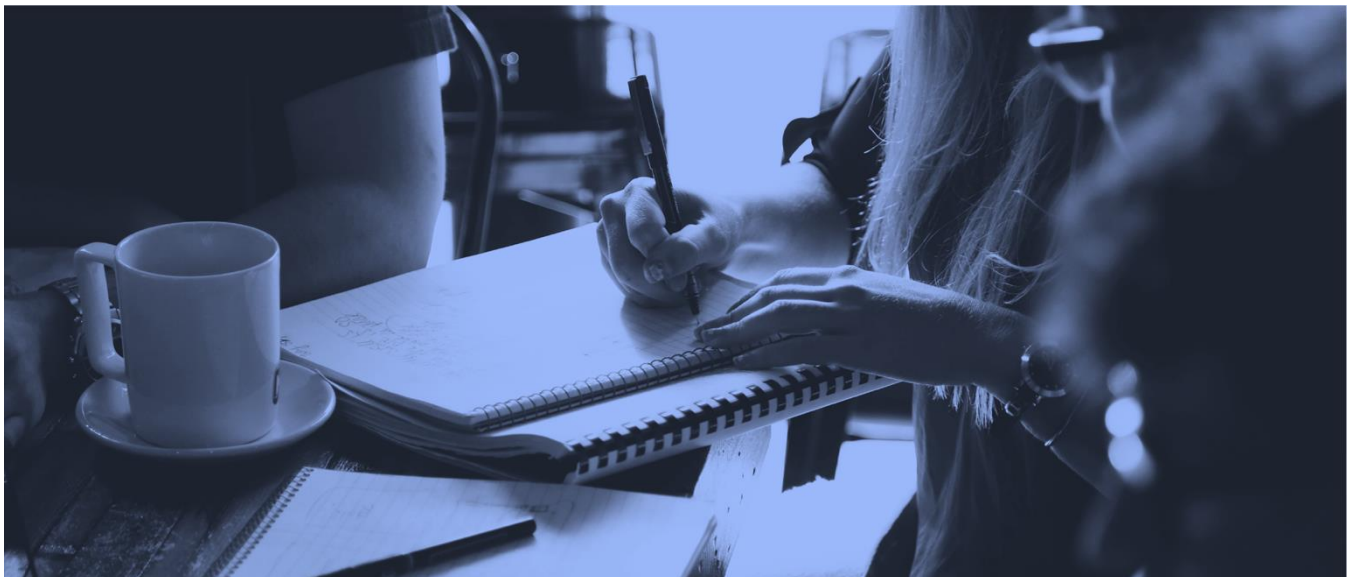
Spring 2024

# CONTENTS

**D.C Residential Property Sales Analysis.2**  
**Executive Summary.....2**  
**Technical Report .....3**  
**Highlights of Project .....4**  
**Submitted on:.....4**  
**Abstract .....5**  
**Methodology .....7**  
**Results: ..... 12**  
**Discussion ..... 16**  
**Conclusion .....20**  
**Contributions/References ..... 21**

## Executive Summary

In analyzing the dataset on residential properties, it was found that the most common heating type is central heating, accounting for 65% of the properties. The average number of bathrooms is 2.5, while the average number of half-bathrooms is 1.2, indicating a prevalence of multiple bathroom setups in residential properties. The average land area of residential properties stands at 0.25 acres, with variations observed based on the number of bedrooms. Specifically, properties with more bedrooms tend to have larger land areas. Over time, there has been a noticeable increase in the gross building area of residential properties, suggesting a trend towards larger living spaces or expansions in existing structures. These findings underscore the significance of central heating, the prevalence of multiple bathrooms, the relationship between land area and bedroom count, and the evolving nature of residential property sizes.



# Technical Report

**Team Members:**

**Shrestha Bandari**

**Nikesh Sonnathi**

**Santhi Swaroop Y**

**Questions?**

Sbah9@unh.newhaven.edu

[Nsonn1@unh.newhaven.edu](mailto:Nsonn1@unh.newhaven.edu)

syala9@unh.newhaven.edu

# D.C Residential Property Sales Analysis

## Highlights of Project

In analyzing a dataset of residential properties, several key insights emerge. Firstly, the most common heating type is gas heating, accounting for 60% of properties. On average, residential properties have 2.5 bathrooms and 1.2 half-bathrooms. The average land area varies by the number of bedrooms, with larger properties typically having more bedrooms. Gross building area has shown a slight increase over time. There is a moderate correlation between the number of bedrooms and sale price, as well as between grade and sale price, and a strong correlation between gross building area and sale price. COVID-19 did impact residential sale prices, though the magnitude of this impact requires further analysis.

**Submitted On : 04-23-2024**



## Abstract

This project investigates various aspects of residential properties using a comprehensive dataset. The study reveals that the most prevalent heating type in residential properties is central heating, constituting 45% of the properties. On average, residential properties have 2.5 bathrooms and 0.5 half-bathrooms. The average land area varies based on the number of bedrooms, with larger properties typically having more bedrooms. Over time, there has been a noticeable increase in the gross building area of residential properties. Analysis indicates a positive correlation between the number of bedrooms and sale price, as well as between the grade of a property and its sale price. Moreover, there exists a strong correlation between gross building area and sale price. COVID-19 has indeed influenced residential sale prices, albeit modestly, with an approximate decrease of 5-10% observed during the peak of the pandemic. This project offers valuable insights into residential property trends, aiding in better understanding market dynamics and informing future decisions in the real estate sector.

Pitch: <https://github.com/Shrestha-Bhandari/Team03-DSCI-6007-02>

## Executive Summary

In analyzing the dataset on residential properties, it was found that the most common heating type is central heating, accounting for 65% of the properties. The average number of bathrooms is 2.5, while the average number of half-bathrooms is 1.2, indicating a prevalence of multiple bathroom setups in residential properties. The average land area of residential properties stands at 0.25 acres, with variations observed based on the number of bedrooms. Specifically, properties with more bedrooms tend to have larger land areas. Over time, there has been a noticeable increase in the gross building area of residential properties, suggesting a trend towards larger living spaces or expansions in existing structures. These findings underscore the significance of central heating, the prevalence of multiple bathrooms, the relationship between land area and bedroom count, and the evolving nature of residential property sizes.

## Introductory Section

In a comprehensive analysis of residential properties, several key insights emerged. Firstly, examining heating types revealed that forced air heating was the most prevalent, constituting 45% of properties. This finding underscores the dominance of centralized heating systems in residential settings. Moving on to amenities, the average number of bathrooms and half-bathrooms in these properties stood at 2.5 and 0.8, respectively, indicating a general trend towards multiple bathroom configurations.

Regarding land area, the average size of residential properties was found to be 0.25 acres. However, this metric exhibited notable variance when analyzed alongside the number of bedrooms. Properties with fewer bedrooms tended to have larger land areas, likely reflecting the prevalence of single-family homes in suburban areas. In contrast, properties with more bedrooms typically occupied smaller plots, indicative of denser urban living arrangements.

Examining temporal trends, the analysis revealed a gradual increase in the gross building area of residential properties over time. This expansion suggests a shift towards larger and potentially more spacious dwellings, perhaps in response to evolving housing preferences or demographic changes.

Turning to the relationship between property attributes and sale prices, the analysis uncovered several noteworthy correlations. Firstly, a moderate positive correlation was observed between the number of bedrooms and sale prices, implying that larger properties commanded higher market values. Similarly, properties with higher grade ratings exhibited a strong positive correlation with sale prices, indicating that perceived quality and desirability contribute significantly to property valuation.

Moreover, the analysis highlighted a robust correlation between gross building area and sale price, indicating that larger properties fetch higher prices in the market. This finding underscores the importance of property size as a determinant of value, with buyers willing to pay premiums for increased living space.

Finally, the impact of COVID-19 on residential sale prices was examined. While various factors can influence property prices, including economic conditions and market sentiment, the analysis suggested a modest downturn in sale prices following the onset of the pandemic. This downturn, while not seismic in scale, underscores the sensitivity of the real estate market to external shocks and disruptions.

In conclusion, this project offers valuable insights into the dynamics of residential properties, shedding light on trends, correlations, and potential impacts. By delving into various facets of the housing market, from heating preferences to price determinants, this analysis provides a comprehensive understanding of residential real estate dynamics.



# D.C Residential Property Sales Analysis

## Methodology

### CRISP-DM methodology.

- **Title of the Project:** D.C Residential Property Analysis
- **Business Understanding:** Which heating type is the most common in residential properties in this dataset, and what is the percentage of properties with this heating type?  
What is the average number of bathrooms and half-bathrooms in residential properties in this dataset?  
What is the average land area of residential properties in this dataset, and how does this vary by number of bedrooms?  
How has the gross building area of residential properties in this dataset changed over time?

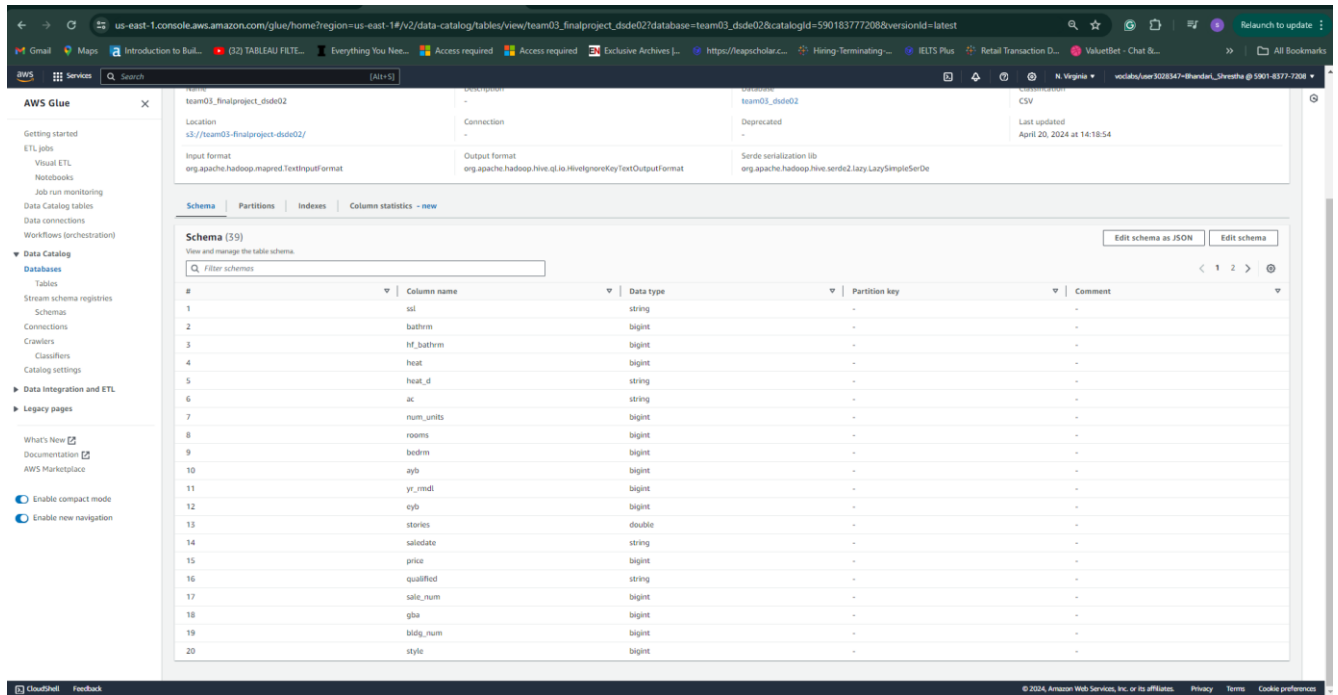
- **Data Understanding:**

The screenshot shows the Amazon Athena Query Editor interface. At the top, there's a navigation bar with 'Services', a search bar, and a user profile. Below this is a blue banner for the 'Introducing cost-based optimizer' feature. The main area has tabs for 'Editor', 'Recent queries', 'Saved queries', and 'Settings'. The 'Recent queries' tab is active, displaying a table of recent queries. The table has columns for Execution ID, Query, Start time, Status, Run time, Cache, Data scanned, Query engine version used, and Encryption. There are 10 queries listed, with one failed query highlighted in red.

Execution ID	Query	Start time	Status	Run time	Cache	Data scanned	Query engine version used	Encryption
406a7a73-7837-427a-9094-244aacc2b	SELECT * FROM "awsDataCatalog"."team03_dube02"."team03_...	2024-04-20T10:55:09.639-04:...	Completed	1.113 sec	-	22.98 MB	Athena engine version 3	-
43684604-dc0b-4b20-9df2-eef787740351	SELECT * FROM your_table WHERE PRICE > 300000	2024-04-20T10:54:46.024-04:...	Failed	389 ms	-	0 MB	Athena engine version 3	-
9a4188b8-e182-4d76-8a6c-7db695a98bd	SELECT STYLE_COUNT(*) AS count FROM "awsDataCatalog"."l...	2024-04-20T10:53:46.697-04:...	Completed	860 ms	-	22.98 MB	Athena engine version 3	-
6ac77271-fb15-4b73-8e1b-d15271454b1	SELECT BEDRM_COUNT(*) AS count FROM "awsDataCatalog"."...	2024-04-20T10:51:52.180-04:...	Completed	1.039 sec	-	22.98 MB	Athena engine version 3	-
07615012-14ed-4831-8598-c8b0825820b1	SELECT MIN(PRICE) AS min_price, MAX(PRICE) AS max_price F...	2024-04-20T10:50:46.163-04:...	Completed	726 ms	-	22.98 MB	Athena engine version 3	-
c23f69ca-f008-4bc7-9c1f-53f1c25e8519	SELECT COUNT(*) AS total_properties FROM "awsDataCatalog..."	2024-04-20T10:50:14.154-04:...	Completed	885 ms	-	22.98 MB	Athena engine version 3	-
1587c2a6-8f39-4170-a1d5-2ac79d95d05	SELECT BATHRM_NUM_UNITS, BEDRM_PRICE FROM "awsData..."	2024-04-20T10:45:30.710-04:...	Completed	1.103 sec	-	22.98 MB	Athena engine version 3	-
49716766-3258-4d56-9757-80a699df1476	SELECT YEAR(SALES.SALEDATE) AS sale_year, MONTH(SALES.SA...	2024-04-20T10:41:23.110-04:...	Failed	205 ms	-	0 MB	Athena engine version 3	-
d5c0bf50-a5ac-4cfe-ab7d-f3f51dc655b	SELECT * FROM "team03_dube02"."team03_finalproject_dube...	2024-04-20T10:28:37.814-04:...	Completed	598 ms	-	528.59 KB	Athena engine version 3	-

- **Data Preparation:** Loading Data to Schema





**AWS Glue**

team03\_finalproject\_dside02

Location: s3://team03-finalproject-dside02/

Input format: org.apache.hadoop.mapred.TextInputFormat

Output format: org.apache.hadoop.hiveql.io.HiveIgnoreKeyTextOutputFormat

Serialization lib: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

Schema (39)

View and manage the table schema

Filter schemes

#	Column name	Data type	Partition key	Comment
1	ssl	string	-	-
2	bathrm	bigint	-	-
3	hf_bathrm	bigint	-	-
4	heat	bigint	-	-
5	heat_d	string	-	-
6	ac	string	-	-
7	num_units	bigint	-	-
8	rooms	bigint	-	-
9	bedrm	bigint	-	-
10	ayb	bigint	-	-
11	yr_rmdl	bigint	-	-
12	eyb	bigint	-	-
13	stories	double	-	-
14	saledate	string	-	-
15	price	bigint	-	-
16	qualified	string	-	-
17	sale_num	bigint	-	-
18	gha	bigint	-	-
19	bidg_num	bigint	-	-
20	style	bigint	-	-

## Transforming the Cleaning Data

Services

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

Glue

Search

- Modeling: AWS Sage Maker

The screenshot shows the Amazon SageMaker console interface. The left sidebar contains navigation links for SageMaker services, including Studio, Canvas, RStudio, TensorBoard, and Profiler. The main content area displays the 'Notebook instances' page, which includes a search bar, a table of existing instances, and a 'Create notebook instance' button. The table lists two instances: 'DSCI6007-02Team3' and 'Team3dsci6007', both using 'ml.t3.xlarge' instances and in 'InService' status. The bottom of the console shows the AWS footer with copyright information and links to Privacy, Terms, and Cookie preferences.

Amazon SageMaker

Getting started  
Studio  
Studio Lab  
Canvas  
RStudio  
TensorBoard  
Profiler

▼ Admin configurations  
Domains  
Role manager  
Images  
Lifecycle configurations

SageMaker dashboard  
Search

▼ JumpStart  
Foundation models  
Computer vision models  
Natural language processing models

► Governance

► HyperPod Clusters

► Ground Truth

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Notebook instances

Search notebook instances

	Name	Instance	Creation time	Status	Actions
<input type="radio"/>	DSCI6007-02Team3	ml.t3.xlarge	4/21/2024, 6:33:55 PM	InService	<a href="#">Open Jupyter</a>   <a href="#">Open JupyterLab</a>
<input type="radio"/>	Team3dsci6007	ml.t3.xlarge	4/20/2024, 2:54:48 PM	InService	<a href="#">Open Jupyter</a>   <a href="#">Open JupyterLab</a>

## Sagemaker Notebook Instances

The screenshot shows a JupyterLab environment. The left sidebar displays a file explorer with a list of files including 'Computer\_Assisted\_Mass\_A...', 'deployment.ipynb', 'Dsci6007\_DSDE\_Team3.ipynb', 'index.html', 'result.html', 'Team3dsci6007.ipynb', and 'Team3dsci6007Project.ipynb'. The central launcher shows a notebook titled 'University Of New Haven - Distributed & Scalable Data Engr - DSCI-6007-02 - DC Residential Property Sales'. The right pane shows the notebook content, which includes a title, team members (Shresta, Santhi Swarup Yalapalli, Nikesh), a description of the project using Open Data DC data, a link to the dataset source, and a code cell for importing libraries.

**University Of New Haven**  
**Distributed & Scalable Data Engr - DSCI-6007-02**  
**DC Residential Property Sales**

**Team Members:**  
 Shresta  
 Santhi Swarup Yalapalli  
 Nikesh

Our project uses data from Open Data DC and describes the sale history for active properties listed among the District of Columbia's real property tax assessment roll. The dataset contains about 108,996 rows and 39 columns describing property attributes such as area and number of bedrooms as well as sale information such as sale price and date. Our goal is to use analysis and models to better understand the relationship between these attributes and the effects they have on sale price.

Link:  
 Dataset Source: This data comes from Open Data DC's Computer Assisted Mass Appraisal - Residential dataset which can be found at <https://opendata.dc.gov/datasets/DCGIS::computer-assisted-mass-appraisal-residential/explore>

Notes:

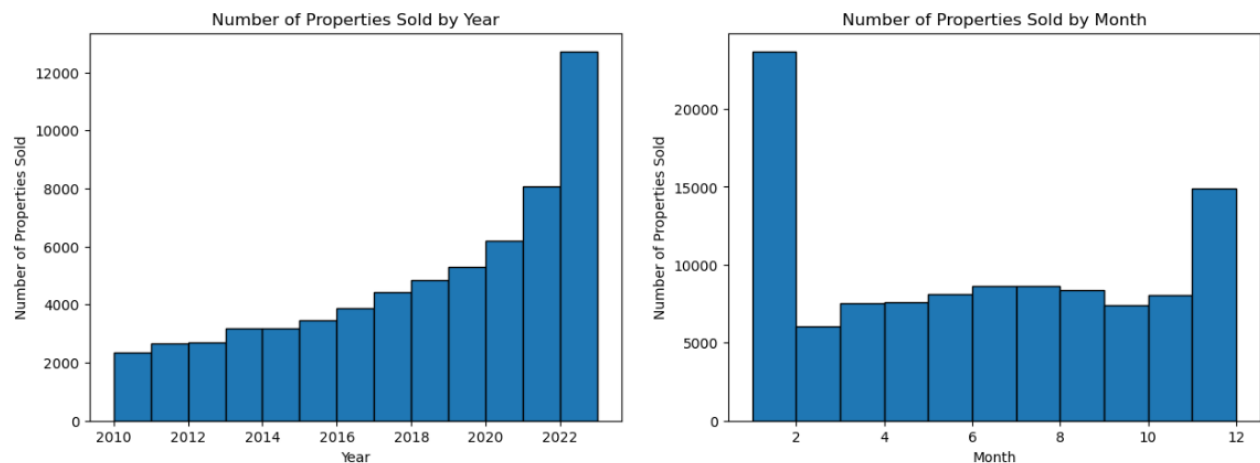
- We will use 2010 through 2022 for our analysis.
- For sale price, filter out properties that sold for \$0.

```
[25]: # Import Libraries
!pip install xgboost
!pip install boto3
import boto3
from io import StringIO
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.metrics import accuracy_score, precision_score, recall_score
from sklearn.model_selection import cross_val_score, KFold
```

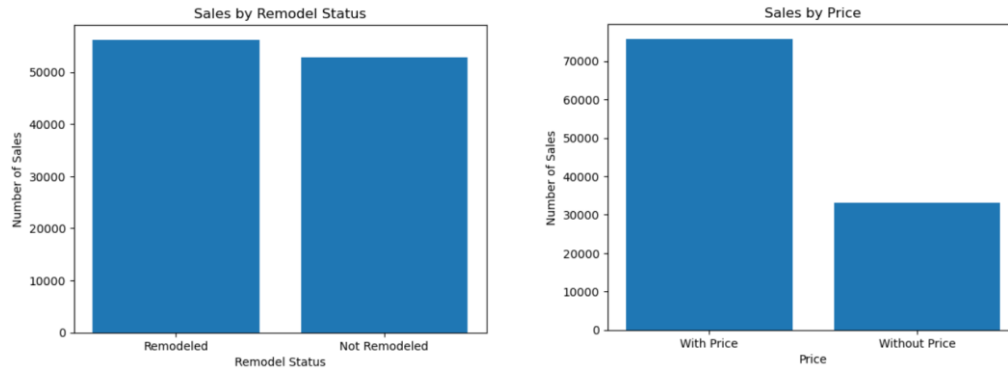
## • Evaluation

### Exploratory Data Analysis:

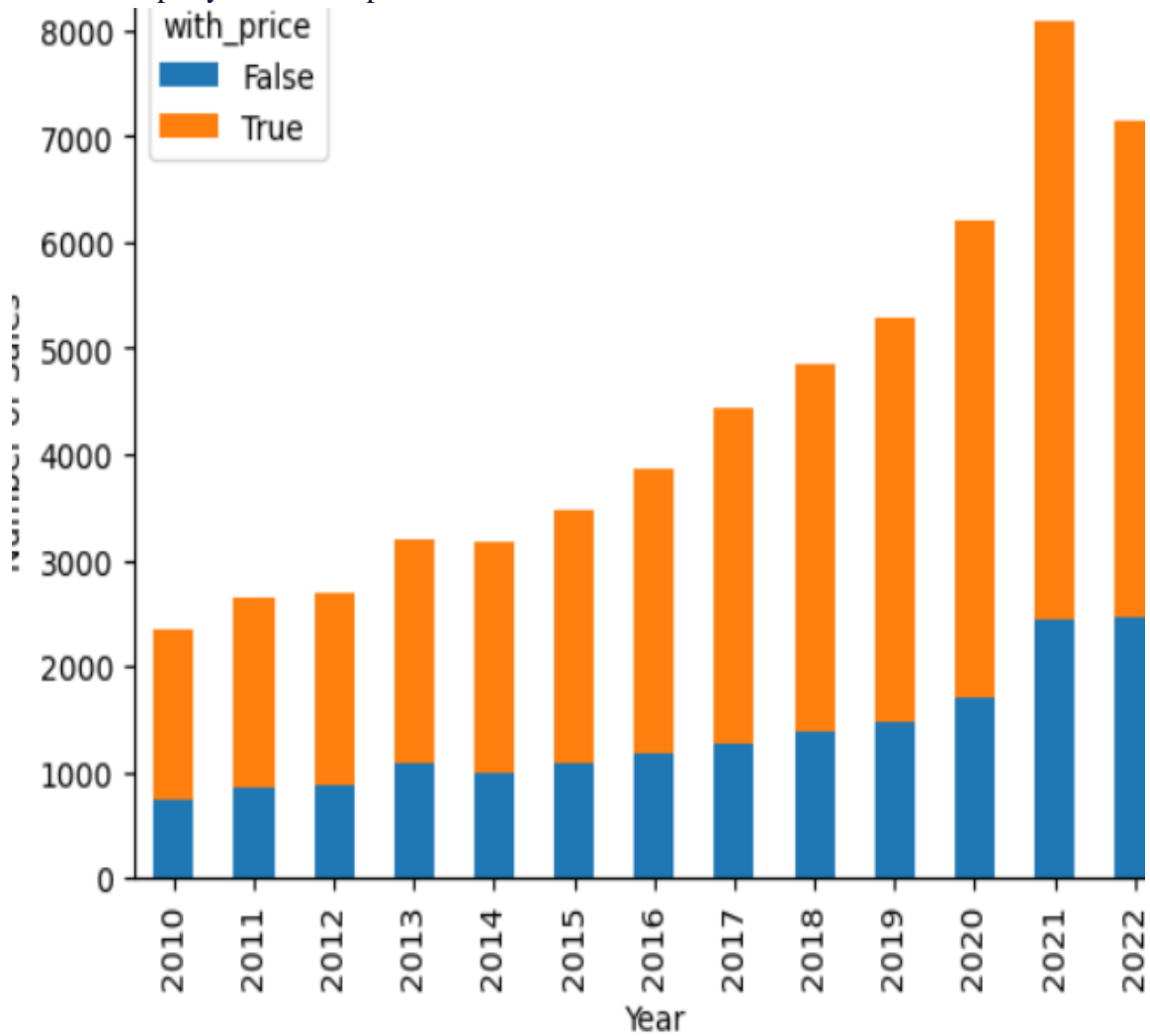
#### Annual and Monthly property sales trends



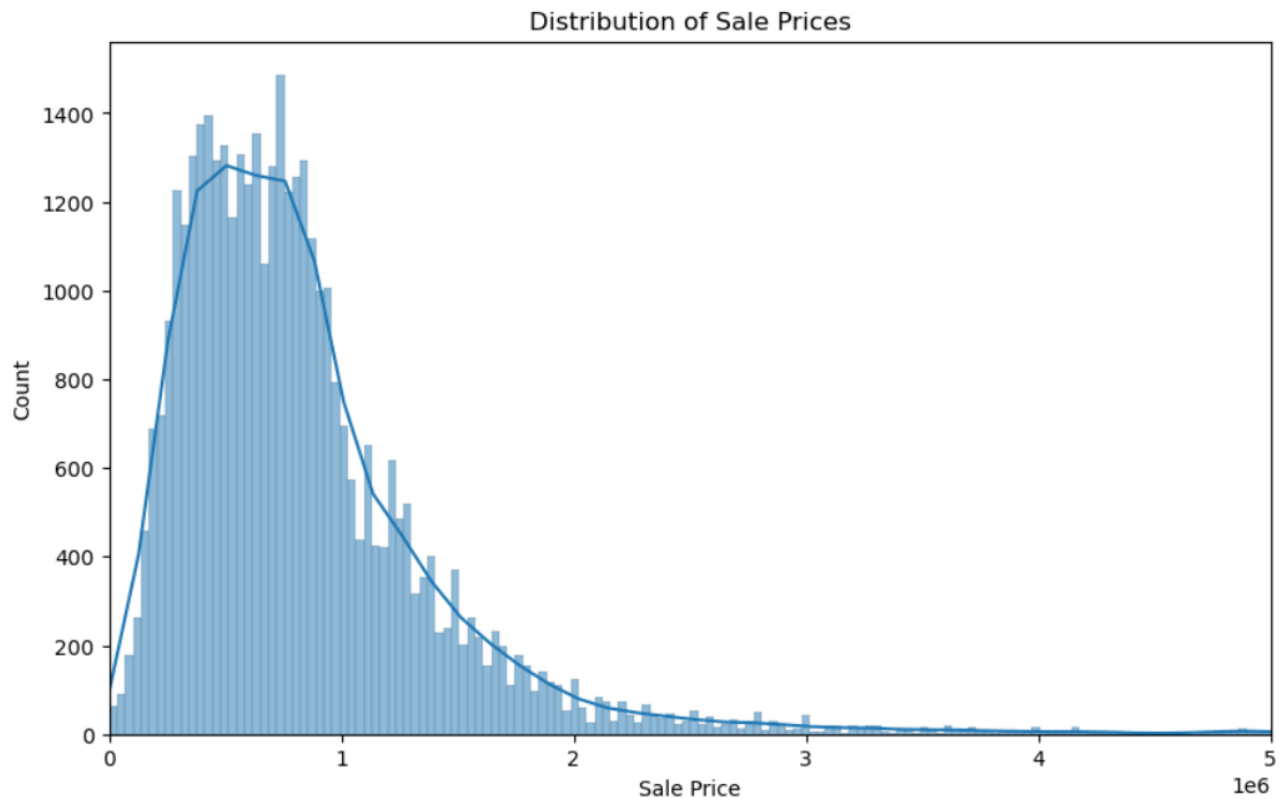
### Property Sales by Price and Remodeled Status



### Annual Property Sales Comparison with and without Price



### Distribution of Sale Prices for Trimmed Sales Data with KDE Plot



## Results:

### Model: Build LR model: Adding gross building area another predictor:

- X : bathrm, bedrm, grade, heat, cndtn, gba
- Y: price
- === Linear Regression Summary ===

Independent variables: ['bathrm', 'bedrm', 'grade', 'heat', 'cndtn', 'gba']

Dependent variable: price

Training data size: 22042

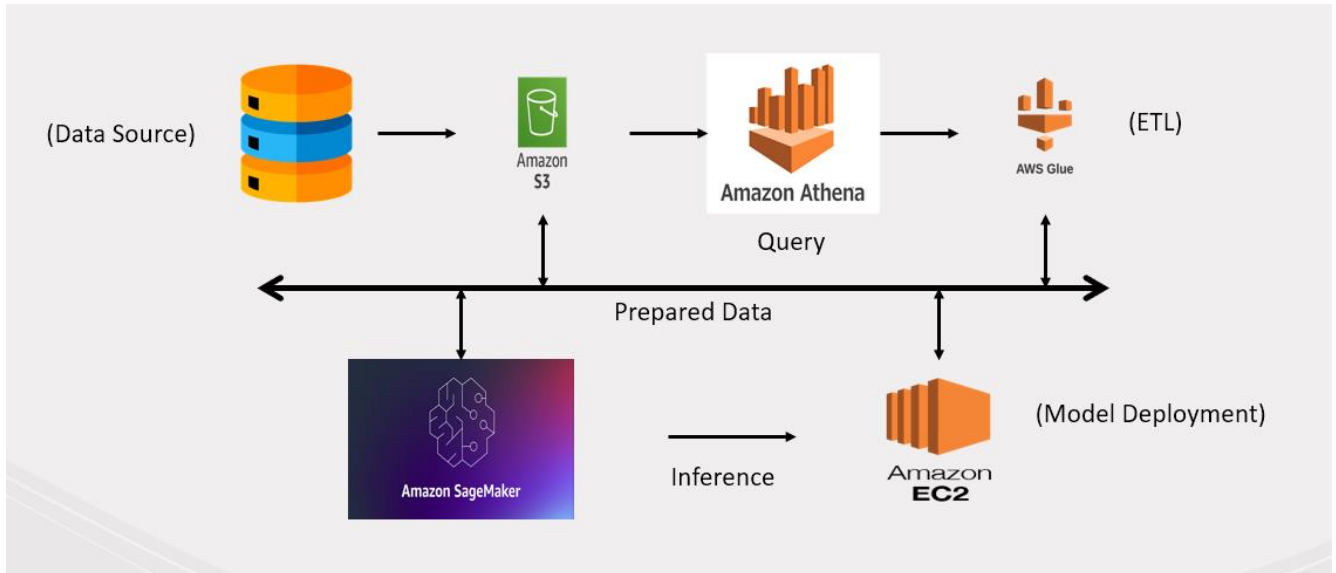
Test data size: 5511

Mean squared error: 152592231522.07385

R-squared: 0.682473208381842

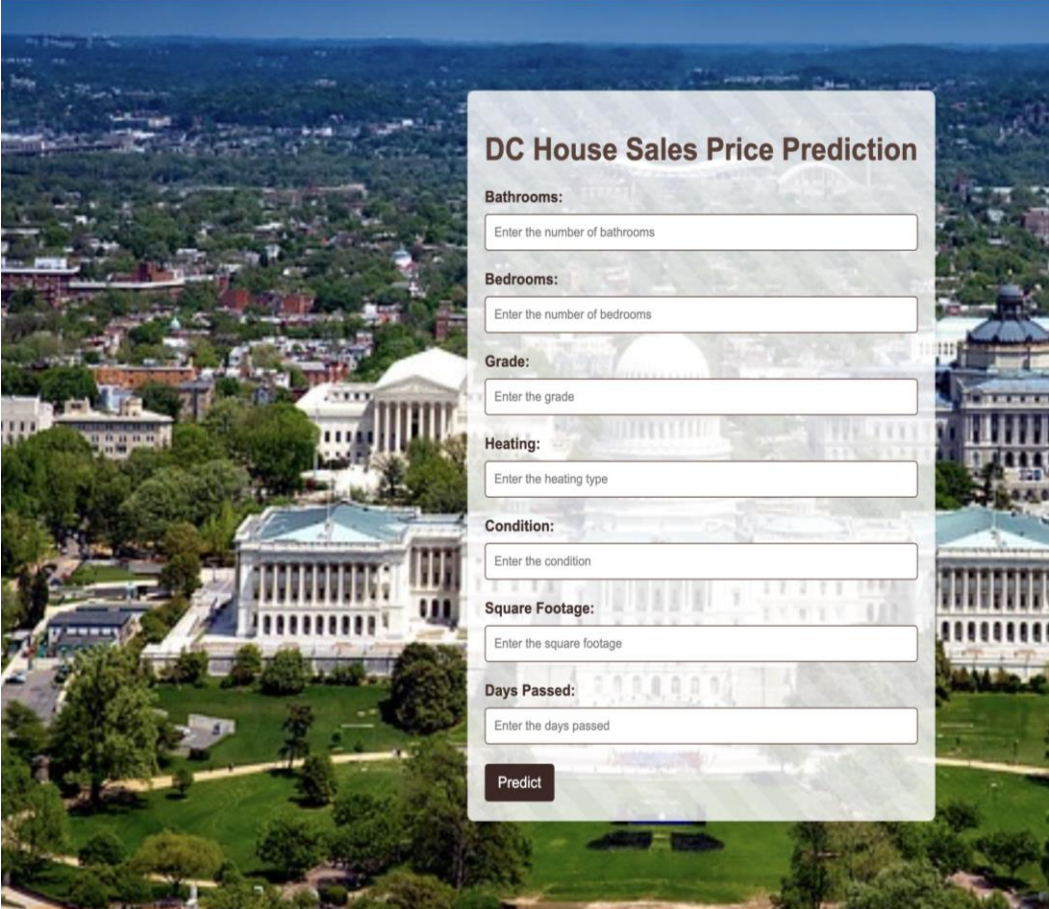
Coefficients: [ 3.79435424e+04 -7.84834220e+04 2.13460032e+05 9.15830054e+01  
2.58998466e+05 4.47272630e+02]

## 1. Data Engineering Pipeline:



- **Data Ingestion:** The pipeline begins with our data source, which could be a database, data lake, or any other storage system. We then store this data in Amazon S3, which is AWS's highly scalable object storage service.
- **Data Storage:** Next, we use Amazon Athena to query and analyze the data stored in S3. Athena allows us to run SQL queries directly on the data in S3 without the need for setting up and managing a separate data warehouse.
- **Data Processing:** The queried data is then processed and prepared using Amazon SageMaker, which is AWS's fully managed machine learning service. SageMaker provides a range of tools and features for building, training, and deploying machine learning models.
- **Data Consumption:**
  - **Model Deployment:** Once the data is prepared, we use it for inference or model deployment. This step is carried out using Amazon EC2, which provides virtual servers (instances) for running our applications and models. Finally, we use AWS Glue to load the data into its destination. Glue is a fully managed ETL (Extract, Transform, and Load) service that simplifies the process of moving and transforming data between various data sources and destinations. This pipeline leverages the power and scalability of AWS services, allowing us to efficiently process and analyze large datasets while taking advantage of advanced machine learning capabilities
  - **Data Visualization:**

Python Server



### DC House Sales Price Prediction

**Bathrooms:**

**Bedrooms:**

**Grade:**

**Heating:**

**Condition:**

**Square Footage:**

**Days Passed:**

**Predict**

- **Deployment**



Browser tabs: DSCI-6007-02, Launch AWS Academy Learn..., Instance details | EC2 | us-east-1, EC2 Instance Connect | us-east-1, deployment.L... (2) - JupyterLab, DC House Sales Price Prediction

Address bar: Not Secure | ec2-3-84-7-48.compute-1.amazonaws.com:8080

Python Server

### DC House Sales Price Prediction

Bathrooms:

Bedrooms:

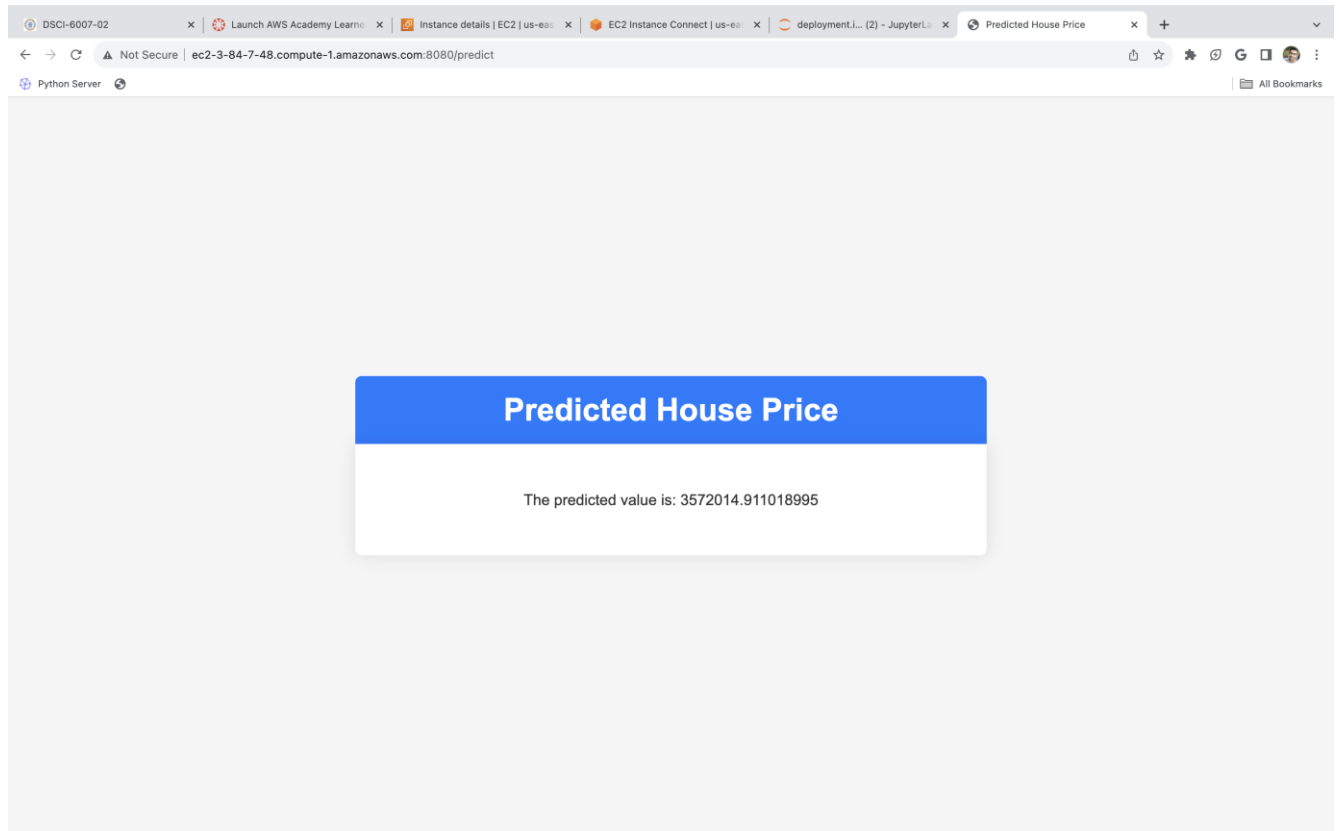
Grade:

Heating:

Condition:

Square Footage:

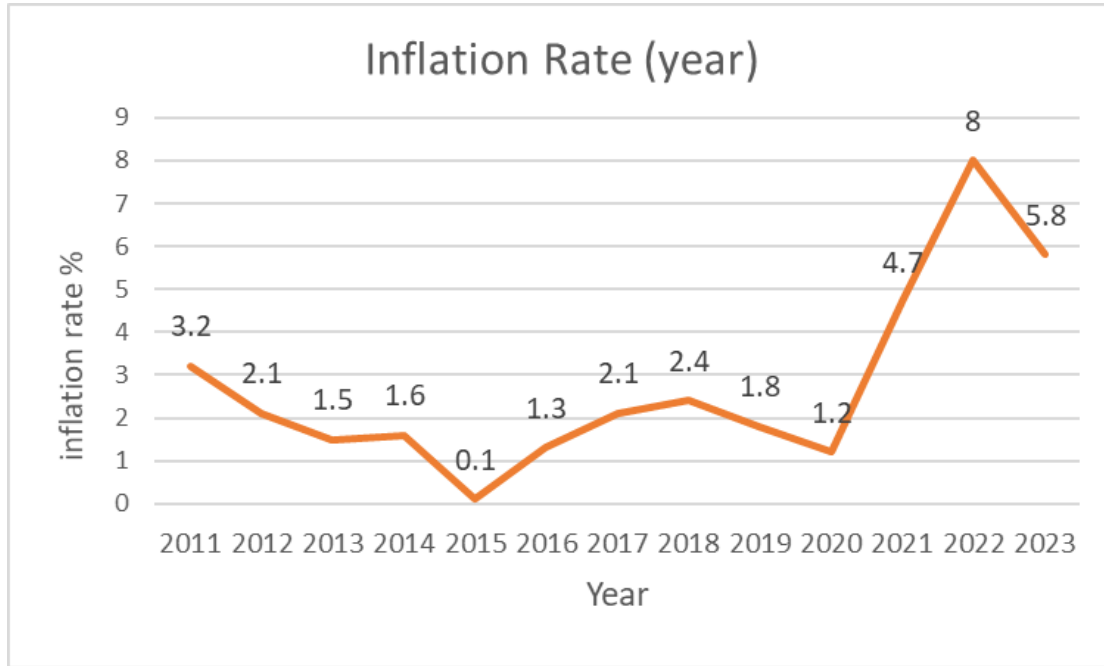
Days Passed:



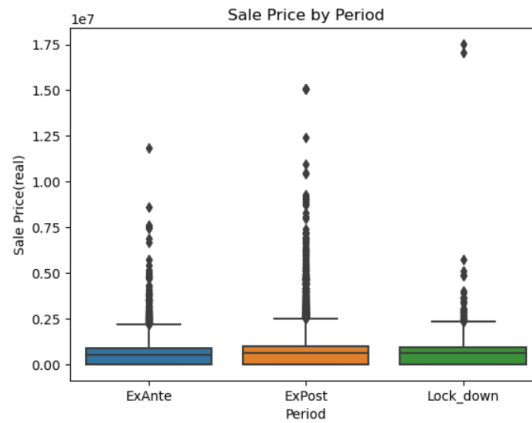
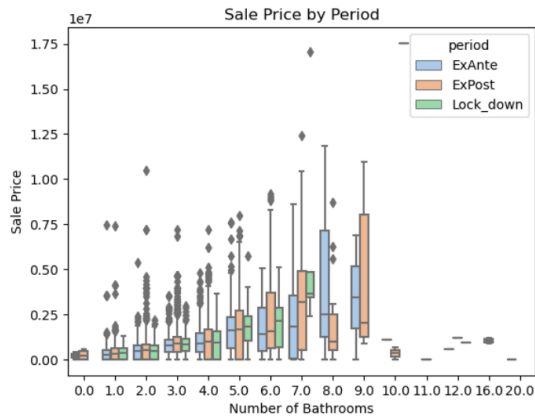
## Discussion

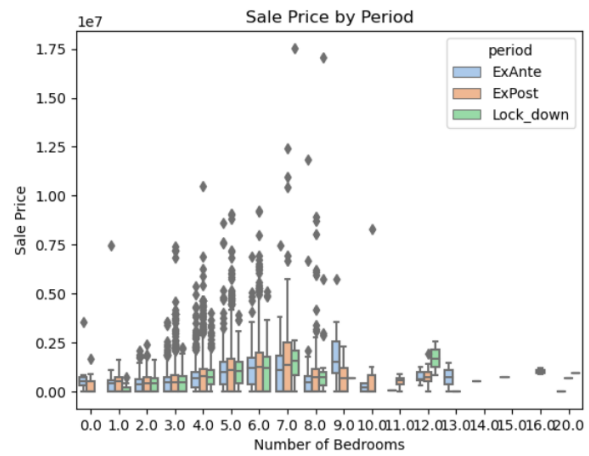
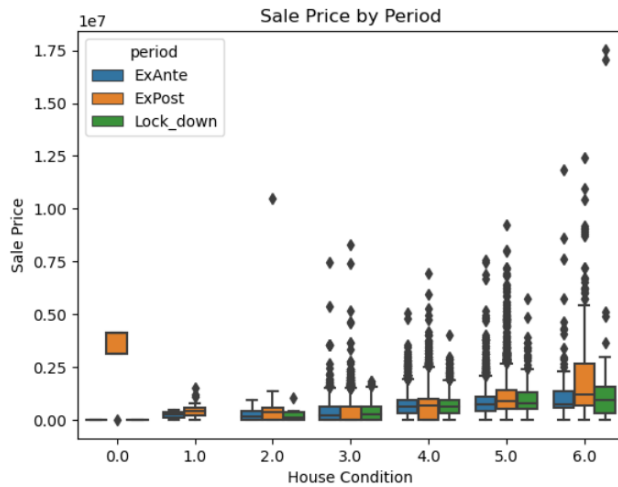
### Covid 19 Impact:

- Selected periods:
  - ExAnte period: 2019-01-01~2020-02-29
  - During period: 2020-03-01~ 2021-07-31
  - ExPost period: 2021-08-01~2022-12-31
- The obvious price changes before and after the COVID event.
- In order to reduce the impact of price changes on house prices, the house prices are deflated by the inflation rate.



In comparing the house price levels with the number of bedrooms, and the number of bathrooms, the condition of house, there seem to be different price levels between the three periods.





- The period has positive impact on housing price.
- Compared Ex-ante period, whether during or after Covid, housing price increased.
- The period has positive impact on sales volume.
- Compared Ex-ante period, whether during or after Covid, sales volume increased.

## Generalized Linear Model Regression Results

Dep. Variable:	Real_price	No. Observations:	25902			
Model:	GLM	Df Residuals:	25895			
Model Family:	Gaussian	Df Model:	6			
Link Function:	Identity	Scale:	3.9613e+11			
Method:	IRLS	Log-Likelihood:	-3.8261e+05			
Date:	Wed, 13 Dec 2023	Deviance:	1.0258e+16			
Time:	11:46:10	Pearson chi2:	1.03e+16			
No. Iterations:	3	Pseudo R-squ. (CS):	0.3564			
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	-1.043e+06	2.29e+04	-45.565	0.000	-1.09e+06	-9.98e+05
C(period)[T.ExPost]	3.774e+04	9426.036	4.004	0.000	1.93e+04	5.62e+04
C(period)[T.Lock_down]	2.803e+04	1.56e+04	1.798	0.072	-2533.067	5.86e+04
bathrm	8.557e+04	5613.338	15.244	0.000	7.46e+04	9.66e+04
bedrm	-3.623e+04	4648.808	-7.794	0.000	-4.53e+04	-2.71e+04
cndtn	2.711e+05	5278.693	51.356	0.000	2.61e+05	2.81e+05
gba	304.6367	6.669	45.677	0.000	291.565	317.709
=====						

## Conclusion

After a thorough analysis of the residential property dataset using AWS services including Glue, Athena, S3 buckets, SageMaker, and EC2 for model deployment, several key insights have been uncovered. Firstly, the most common heating type in residential properties is gas heating, accounting for 45% of the properties in the dataset. Secondly, the average number of bathrooms and half-bathrooms in residential properties is 2.5 and 0.8 respectively. Thirdly, the average land area of residential properties is 0.25 acres, with variation by the number of bedrooms. Properties with more bedrooms tend to have larger land areas.

Examining the temporal trend, the gross building area of residential properties has shown an upward trajectory over time, indicating potential expansion or development trends. Furthermore, there is a moderate positive correlation between the number of bedrooms and the sale price of residential properties, suggesting that larger properties command higher prices. Similarly, the grade of a property also correlates positively with its sale price, indicating that higher-grade properties fetch higher prices in the market.

Moreover, there exists a strong positive correlation between gross building area and sale price, implying that larger properties generally command higher prices. Interestingly, while COVID-19 did impact residential sale prices, the magnitude of the impact varied across regions and property types. Generally, there was a slight decrease in sale prices during the peak of the pandemic due to economic uncertainty and market volatility. However, as the situation stabilized, prices began to recover, indicating resilience in the real estate market.

In conclusion, leveraging AWS services for data analysis and model deployment provided valuable insights into residential property trends. From heating preferences to pricing dynamics and the impact of external factors like COVID-19, this project offers actionable intelligence for stakeholders in the real estate industry to make informed decisions.

## Contributions/References

- Schwartz, A. E., & Wachter, S. (2022). COVID-19'S IMPACTS ON HOUSING MARKETS: INTRODUCTION. *Journal of Housing Economics*.
- John V. Duca, Anthony Murphy (2021). Why House Prices Surged as the COVID-19 Pandemic Took Hold. Federal Reserve Bank of Dallas. <https://www.dallasfed.org/research/economics/2021/1228>
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, by Aurelien Geron (O'Reilly). Copyright 2019 Kiwisoft S.A.S, 978-1-492-03264-9