|  |
| --- |
| TECHNICAL REPORT |

|  |
| --- |
| **Distributed and Scalable Data Engineering**  **(DSCI-6007)** |

|  |  |
| --- | --- |
| SPRING 23 |  |



Contents

[Project Name 2](#_Toc96341549)

[Executive Summary 2](#_Toc96341550)

[Technical Report 3](#_Toc96341551)

[Highlights of Project 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341552)

[Submitted on: 3](file:///C:\Users\ardia\Desktop\TEACHING\SPRING%2022\DSCI6002%20Data%20Science\Projects\Technical%20report%20Template.docx#_Toc96341553)

[Abstract 4](#_Toc96341554)

[Methodology 5](#_Toc96341555)

[Results Section 9](#_Toc96341556)

[Discussion](#_Toc96341557) 11

[Conclusion 11](#_Toc96341558)

[Contributions/References 12](#_Toc96341559)

|  |
| --- |
| **Residential Property Sales Analysis** |

|  |  |  |
| --- | --- | --- |
| Executive Summary Before COVID, the housing market showed stability and consistent price growth. The pandemic disrupted the market, causing economic uncertainties and altering buyer behavior. The prediction model adapted during COVID, considering changing feature importance and market dynamics. Post-COVID, the market exhibited signs of recovery, influenced by shifted buyer preferences. The adaptive model remains crucial for ongoing monitoring and informed decision-making in the evolving real estate landscape | | |
| person at a table writing in a notebook with people around | | |
| **Team Members:**  **Roshini Bandi (Team leader) - Data Engineer**  **Saikrishna Chigicherla – Data Scientist**  **Vamshi Reddy Thatikonda - App Development**  **Geethika Myneni - Data Modeling** | **Questions?**  Contact: schig4@unh.newhaven.edu |  |

|  |
| --- |
|  |

|  |  |
| --- | --- |
| **Project Title:**  **Residential Property Sales Analysis** |  |
| Highlights of Project  * The project illuminates residential property sales trends, leveraging predictive modeling for strategic decision support in real estate investments, considering demographic impacts and regulatory compliance. * We have configured the pipeline and model to directly read the data from snowflake database and execute the prediction result in a web application.  Submitted on: 05/12/2023 |

## Abstract

This project explores residential property sales, employing descriptive statistics and visualizations to reveal insights into market trends and property values. Findings empower real estate stakeholders with data-driven decision-making, underscored by transparent code documentation for reproducibility.

**A scheduling approach to prevent security vulnerabilities**

Introductory Section

* House Price prediction models are crucial for data-driven decision-making, guiding strategic choices and resource allocation.
* They play a pivotal role in optimizing operations across various industries, ensuring efficiency and cost-effectiveness.
* These models provide a competitive edge by enabling businesses to anticipate and respond to market dynamics, influencing success in areas like marketing, sales, and real estate.
* GitHub link: <https://github.com/saikrishna9899/DSE_project>

Data source Link

Through the following link, web scraping is being used to collect data.

<https://opendata.dc.gov/datasets/DCGIS::computer-assisted-mass-appraisal-residential/explore>

## 

## Methodology

The curriculum employs the CRISP DM methodology, which covers subjects including business understanding, data understanding, data preparation, modeling, evaluation, and model deployment.

Business Understanding

* As a data scientist, it is crucial to understand the type of business your company is in, the industry it operates in, how it functions, and basically everything else related to the business. Only by having a thorough understanding of the business will you be able to pinpoint its current challenges, assess them, and find a solution or solutions, as well as identify the tactics to achieve the company's objectives.
* Focused on enhancing user/buyer experience by providing valuable insights into retail price values.

Data comes from **Open Data source** and describes the sale history for active properties listed among the District of Columbia’s real property tax assessment roll. The dataset contains 108,552 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.

Data Understanding:

Data Preparation

Knowing what data, you already have, where to get more data, how much data is available, and what tools to employ to get it are all necessary during this phase. Your data science project will make more sense if you comprehend your data from the beginning

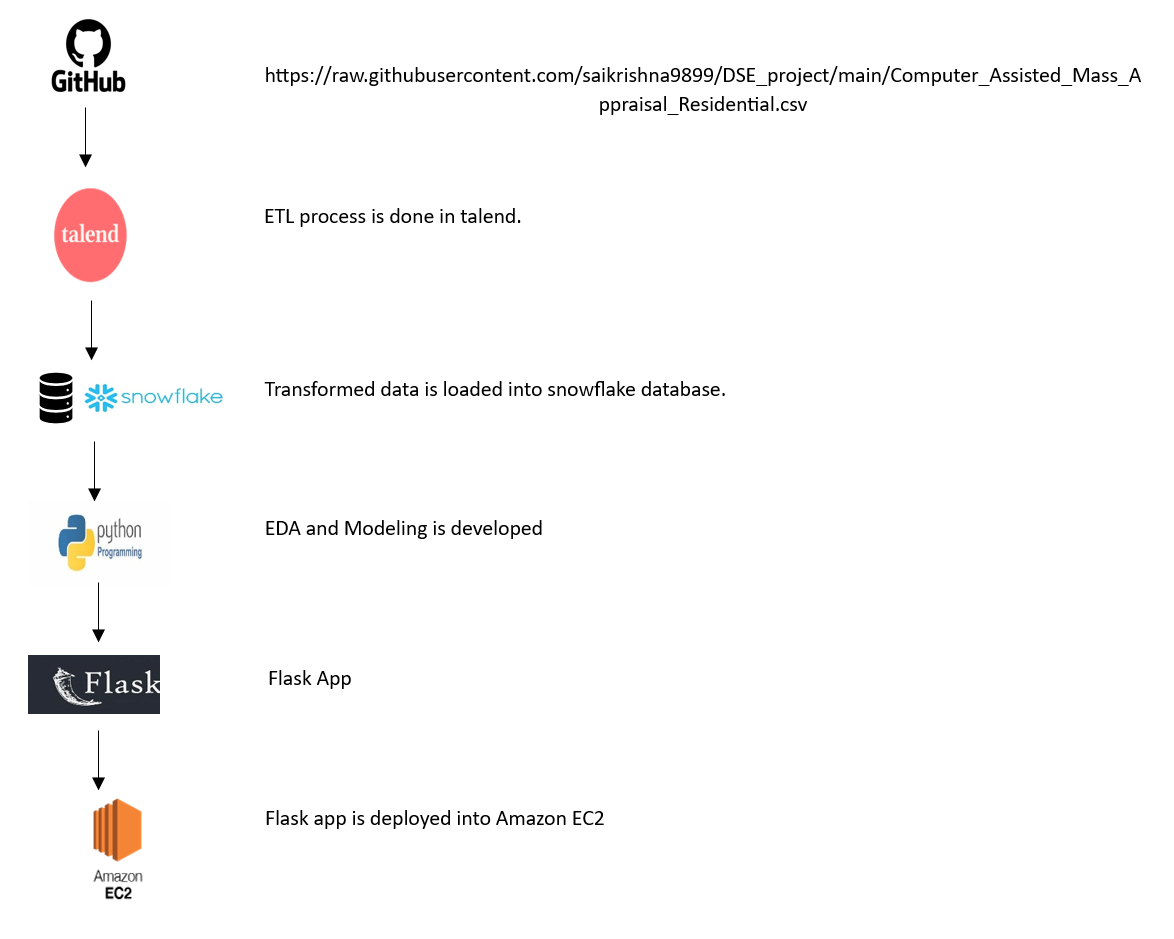
In our Project,

* We have collected the data from open source and uploaded into git.
* We have used Talend ETL to read the data from git and transform the data and then upload modified data into snowflake.
* We also performed EDA (Exploratory Data Analysis) on the input data.

**Modeling and Evaluation**

* Read the data from snowflake and did EDA (Exploratory Data Analysis) on the data.
* Build a Linear regression model, by taking different variable to predict the accuracy of sales price.
* Used a Flask application, so that user can enter the input values (House Variables) and get the output (Predicted house price)

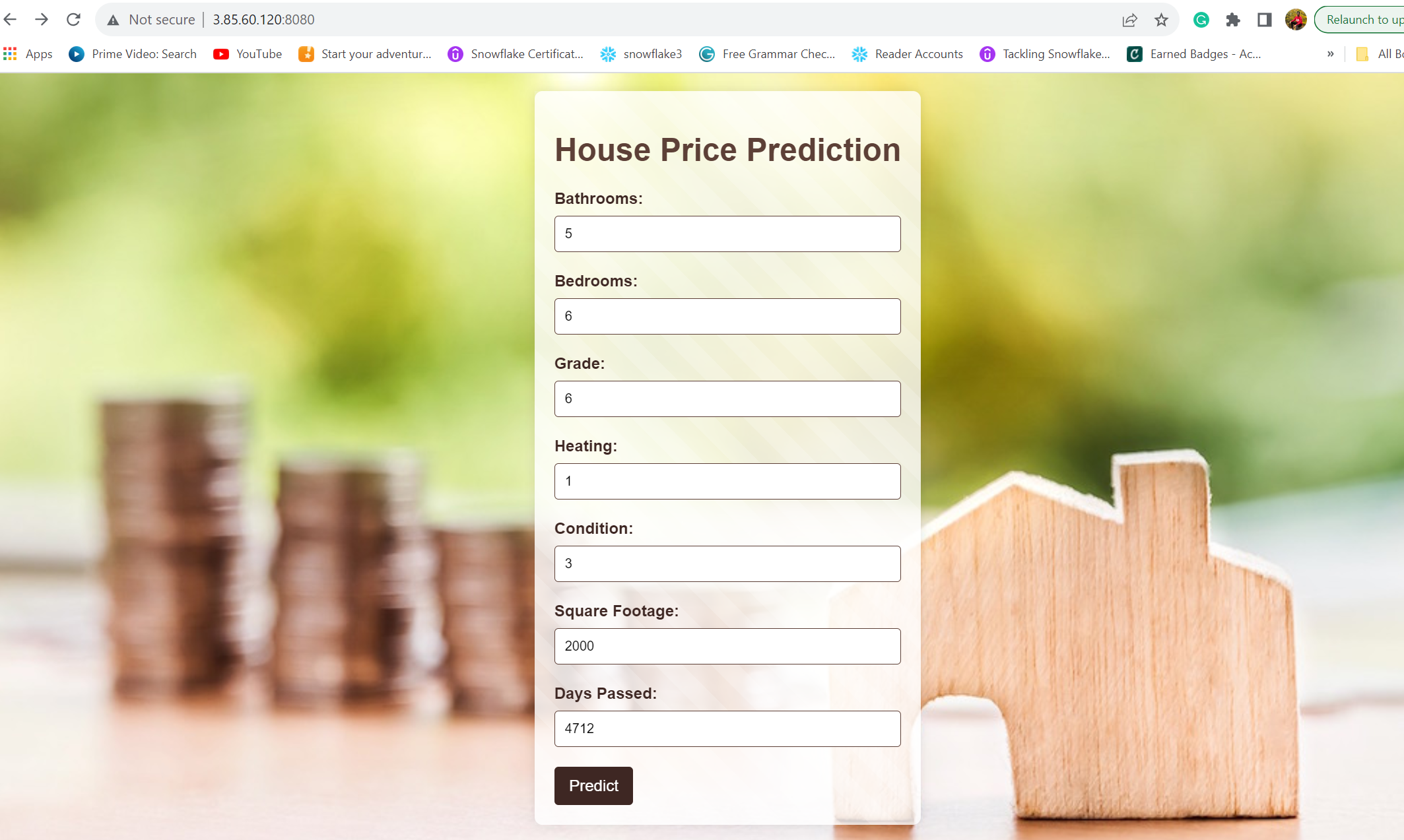
## Data Pipeline Architecture

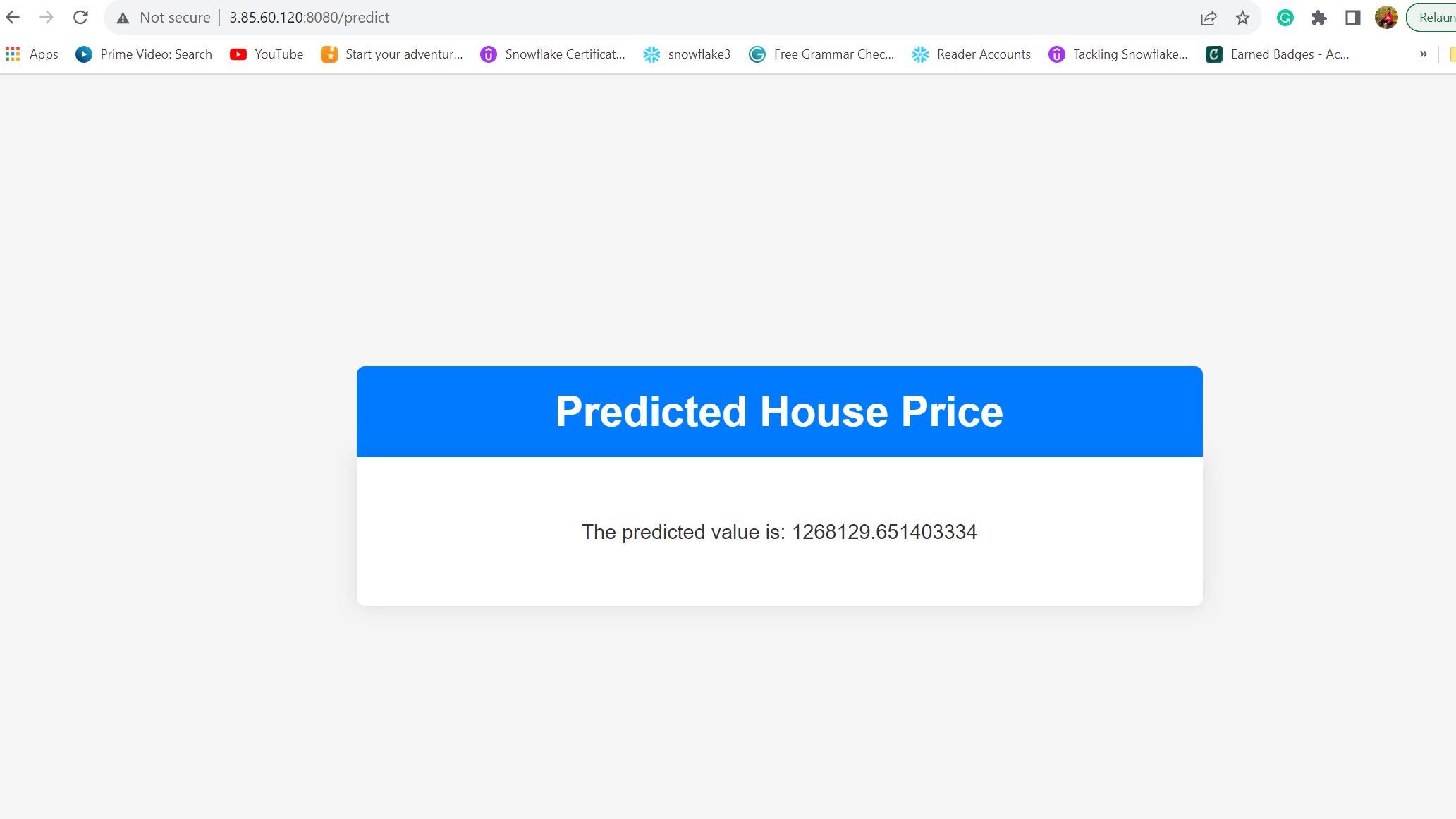


## Results Section

The output is deployed

<http://ec2-3-85-60-120.compute-1.amazonaws.com> or <http://3.85.60.120:8080/>





## Discussion

We had an idea to build a linear regression model into a Flask web app to address residential property sales complexities, faced few hurdles. Subsequent reassessment revealed the refined approach as the most effective solution, utilizing linear regression for insightful market analysis and presenting findings through a user-friendly Flask web interface.

## Conclusion

In conclusion, our residential property sales analysis, developing a linear regression model and integrating to flask web app and deployed to Amazon EC2, provides valuable insights for decision-makers, despite initial technical challenges. This refined approach stands as an effective solution, contributing to informed strategies in the real estate market.

Git hub Link

<https://github.com/saikrishna9899/DSE_project>

## 

## 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>