Housing Market Sales Dashboard

Salma Katri

April 20, 2022

Background

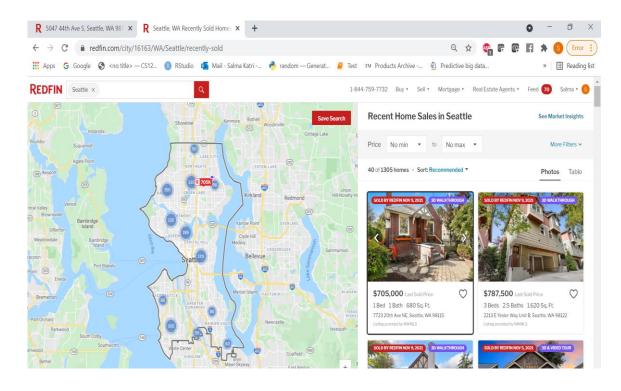| Motivation |
|---|
| • Housing prices have seen unprecedented y-o-y increases

• Homebuyers need real-time data to understand market tends

• Redfin is one of the most widely used website by homebuyers

• While data is available, there is no aggregate view |

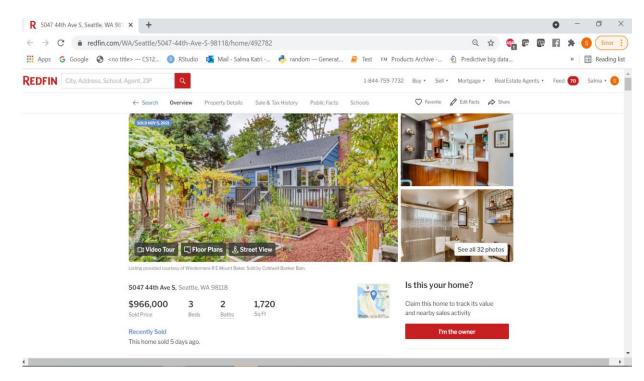| Objective |
|---|
| • Build interactive dashboards to display up-to-date housing market trends, key statistics and graphs |

Data Ingestion

Recently Sold Main Page   **Individual House Listing**

Scraping pipeline designed to download sold homes data from Redfin, then scrape additional home specific information from house listing link

Pipeline Design

Data Ingestion
Build scraping pipeline in Python to collect Redfin sold houses data

Testing/Robustness
- Use schedule to scrape new Redfin data month
- Add tests to check newly ingested data for errors

Data Storage
Organize scraped data into tables format and load into SQL

Processing
Clean data in Python, aggregate and calculate statistics using NumPy and Pandas and download processed data as csv

Deployment
Visualize the data using Plotly in Streamlit to create an interactive dashboard showing useful trends and summary statistics

Deployment