# Serverless Data Engineering

**Boston Rentals Craigslist**

## DATASET

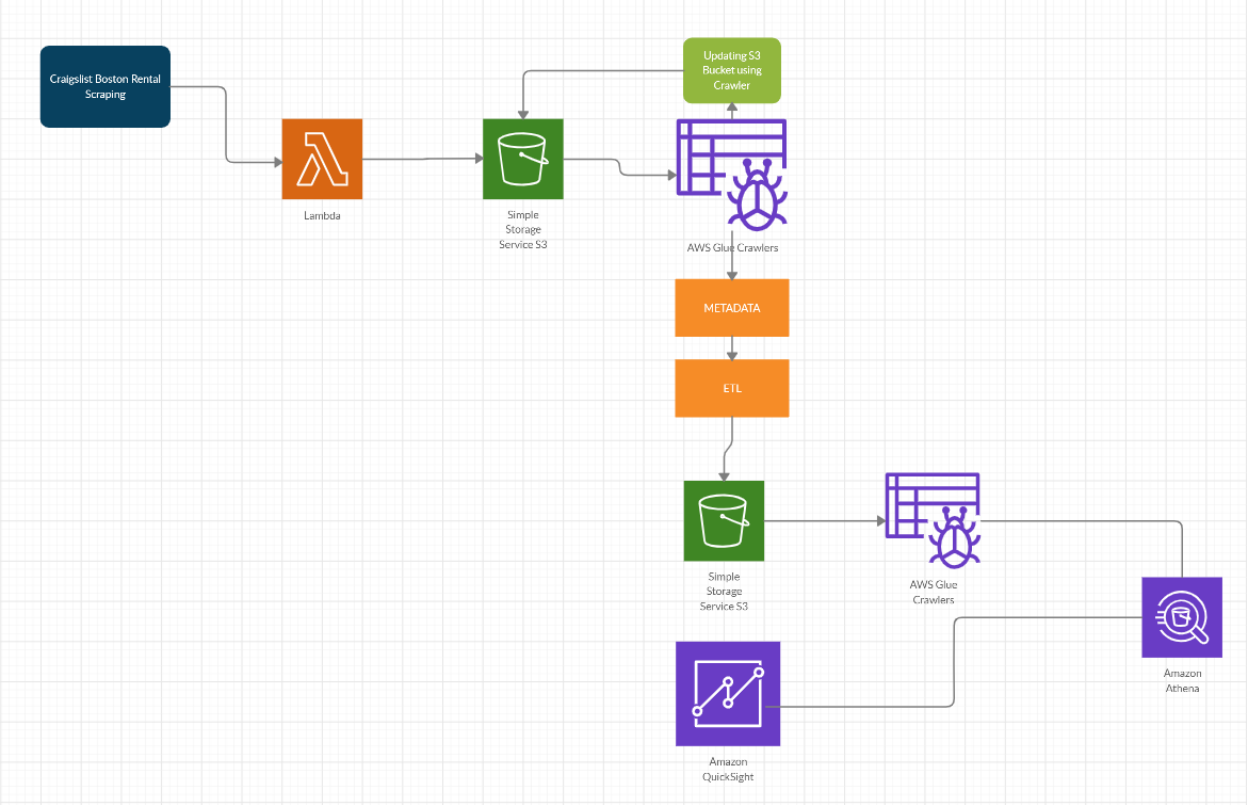
Dataset link: https://boston.craigslist.org/search/apa

**Craigslist** is an American classified advertisements website with sections devoted to jobs, housing, for sale, items wanted, services, community service, gigs, résumés, and discussion forums.
The section of data that is being used is housing. The chunk of data that we are concentrating on is to retrieve the listings based on number of bedrooms, price and square feet.

# PROJECT GOALS

- Scrape data from Craigslist for housing based on Neighborhood.
- Create a pipeline for Data Extraction, Ingestion and Inference
- Visualize the insights from the Data using QuickSight

# DATA PIPELINE




## Technologies used:

1. Python
2. AWS Lambda
3. AWS GLUE
4. AWS ATHENA
5. AWS Quick Sight

## STEP - 1 : Data Scraping

- The data set was programmatically downloaded from Craigslist
- Beautiful Soup is used to scrape data into S3 with AWS LAMBDA as a compute service

- The zipped files are then extracted and pushed to a **Landing S3 Bucket** which serves as the source for the processing aspect of the pipeline

## STEP - 2 : Data Pre-Processing

- Used AWS Glue for Cleaning and Transformation of the data
- The raw data has 8 Columns
- Transformed and cleansed data is written to S3 bucket using ETL job into Parquet format
- AWS Crawler is being used to read data from S3 to Athena

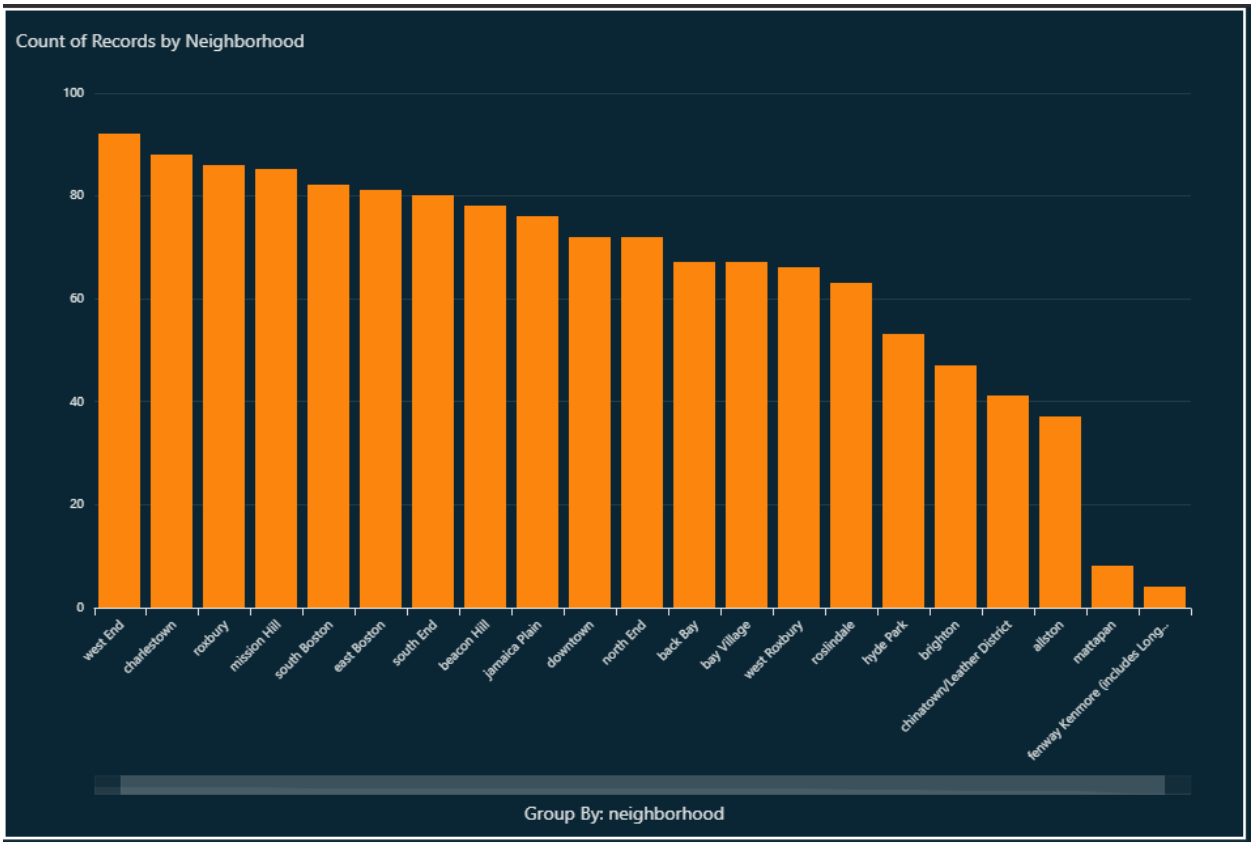
## STEP - 3 : Visualizations using AWS QuickSight

- We have used QuickSight to visualize some important features for the end users
- Users get detailed insights from the Dashboards

## STEP - 4 : Analysis on Visualization

1. Count of Records by Neighborhood

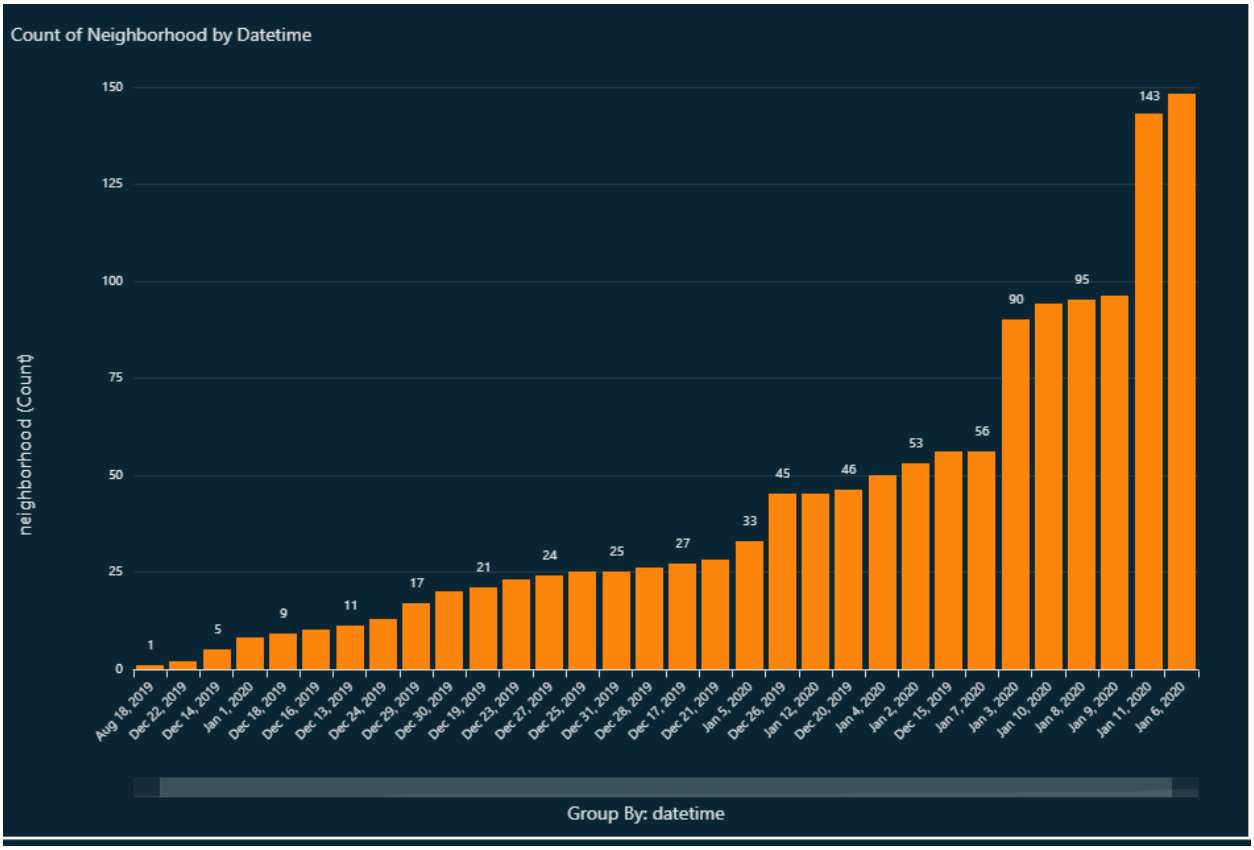
   The visualization shows number of listings for each neighborhood.




2. Count of Records by Datetime

   The visualization shows number of listings based on Date.

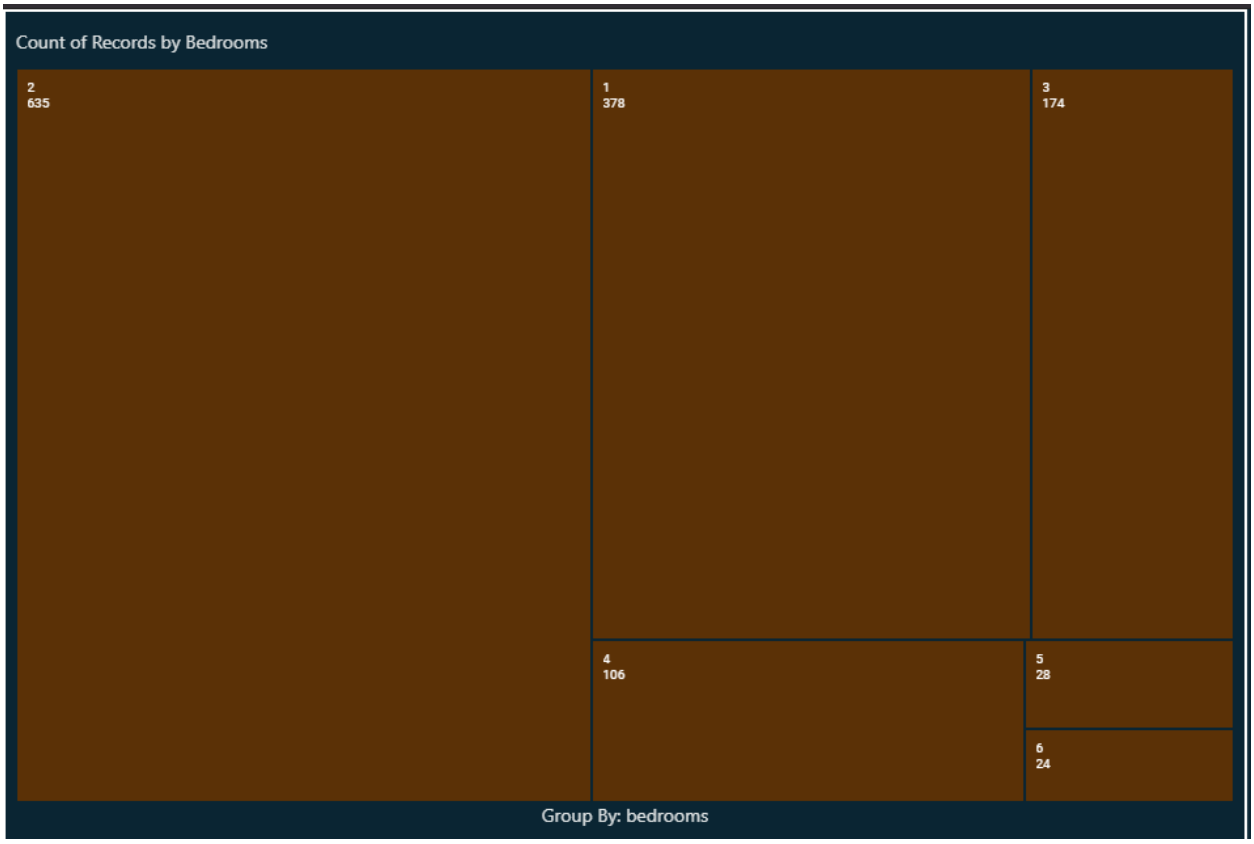
   The number of listings are higher in the month of January.




3. HeatMap of the listings based on the number of bedrooms

   The visualization shows count of records based on the number of bedrooms.

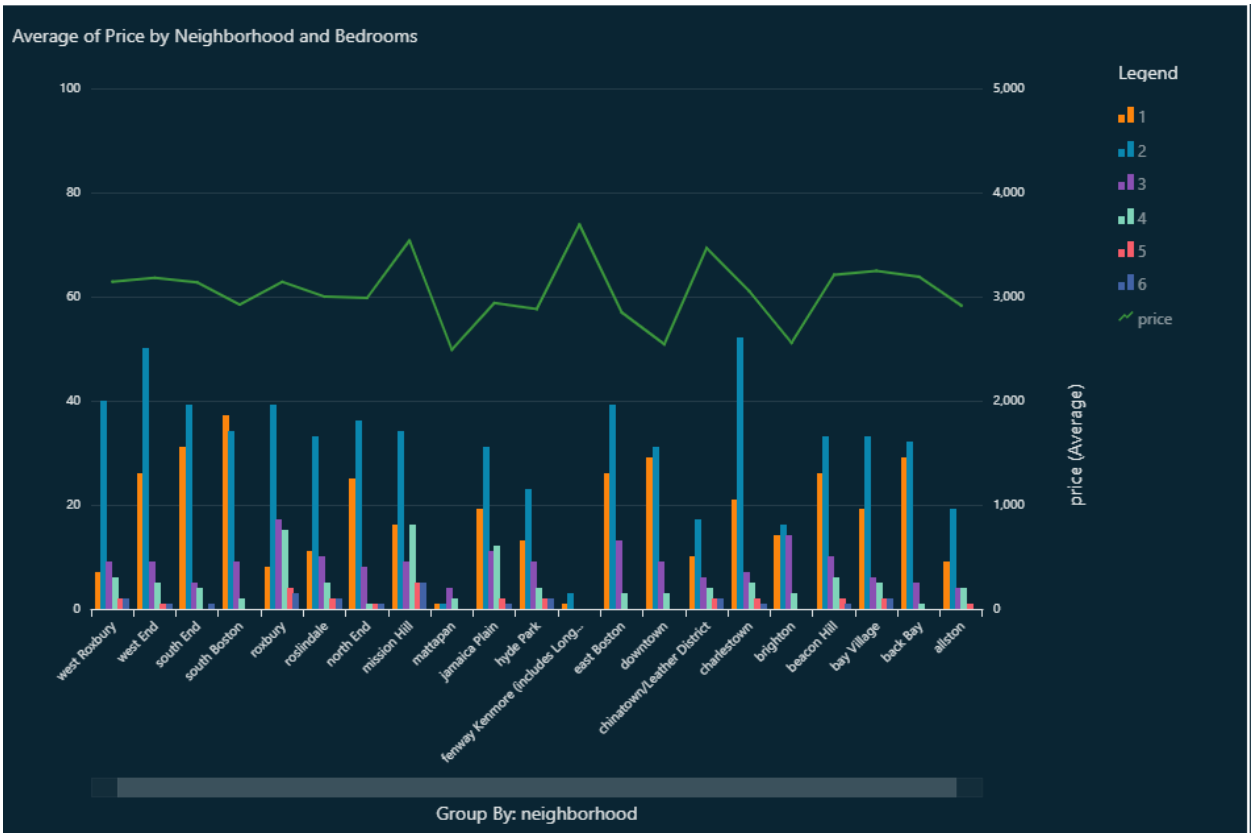
   The number of listings for the 2 Bedrooms are highest and 6 bedrooms are least.




4. Average of Price by Neighborhood and Bedroom

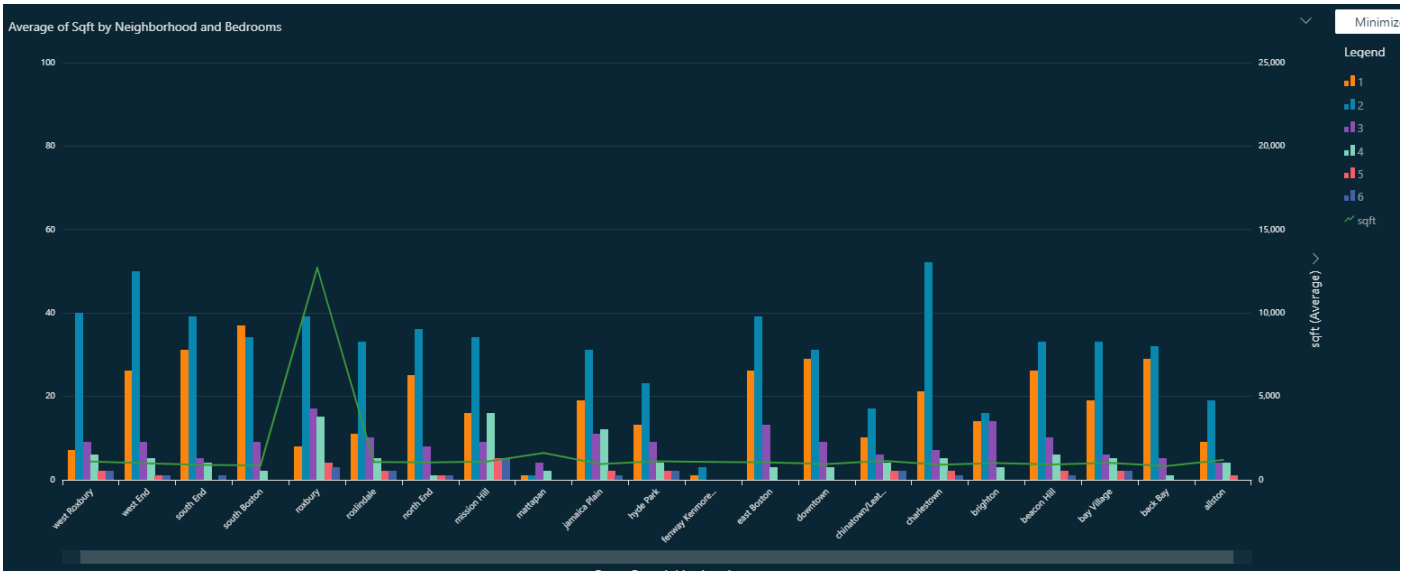
   The visualization shows average price by Neighborhood and Bedrooms.

   From the graph, Fenway has the highest average price for the rent.




1. Average of Sqft by Neighborhood and Bedrooms

   The visualization shows average sqft by Neighborhood and Bedrooms. From the graph, there is a listing that has been false information for average square feet for the listing in Roxbury.


Average of Sqft by Neighborhood and Bedrooms
Minimiz
Legend
1
2
3
4
5
6
sqft
100
80
60
40
20
0
25,000
20,000
15,000
10,000
5,000
0
sqft (Average)
west Roxbury
west End
south End
south Boston
roxbury
roslindale
north End
mission Hill
mattapan
jamaica Plain
hyde Park
fenway kenmore...
east Boston
downtown
chinatown/Lwst...
charlestown
brighton
beacon Hill
bay Village
back Bay
allston