

YO UNI



Prepared By

Xue Luo (xl3213)
Namira Suniaprita (ns3646)
Zhiyi Tao (zt2308)
Eric Wang (zw2913)
Erwin Xaverius (ex2132)



Background

Helping students in New York City to find a good neighborhood that meets budget constraint and convenience level

Our main target audience are Sophomore and transfer students

Solution

Collaborate with Universities in New York City area by providing a website for students in New York City to help students rent a place that is match with their budget and convenience enough for them to live



Rating System

Affordability Rating

- Measures rental affordability for 2022 - 2023 across regional and property types
- The rating uses A to D scale, with A being the most affordable

YouCompare

Convenience Rating

- Evaluates rental property convenience based on 2022-2023 complains received
- A being the most convenient

Data Sources - 311 Data



Description

- The dataset contains information requests made by New York City residents through the 311 hotline, including incident address, complaint type, complaint date, and resolution date.
- The types of service requests submitted involve issues such as noise complaints, street conditions, and more

About

- Velocity : Updated daily to ensure the most recent information is available.
- Data Creation : The dataset was first created on October 10, 2011.
- Data Provider : The data is provided by the 311 service and the Department of Information Technology and Telecommunications.
- Size : The dataset comprises 32.4 million rows and 41 columns.
- Data Structure: The data is structured and represented as CSV and JSON.

Extraction

- Data is extracted through CSV from the website from 2022 - 2023

Transformation

- Add and divide area into 10 parts, in-house rating and divide complaints by zip code

Load

- Divide complaints by zip code

Data Sources - Zillow



Description

- The dataset contains information on rental prices in New York City, collected from Zillow's extensive database of property listings.
- It includes details such as the neighborhood, median rent price, price per square foot, and historical price trends.

About

- Velocity : Zillow updates this dataset on a monthly basis to reflect the most current market conditions.
- Data Creation: Zillow has been collecting rental price data since its inception in 2006.
- Data Provider : The data is provided by Zillow, an online real estate marketplace that offers a wealth of information on rental properties, sales listings, and housing market trends.
- Size: The dataset size may vary depending on the specific data requested (e.g., time frame, neighborhoods, or property types).
- Data Structure: The data is available in a structured format, and users can download it as CSV (Comma-Separated Values) or Excel files for further analysis and visualization.

Extraction

- Data is extracted through CSV from the website from 2022 - 2023

Transformation

- Average 2022 price, recent price using February 2023, add longitude and latitude using Geolocation API, and add score

Load

- Load the database as it is since the data itself already straightforward.



Design Implementations

Back-End Scheme



Extract

- Fast-processing: distributed processing capability
- Horizontal scalability: handle large dataset
- Fault tolerance: handle software failures



Transform & Load

- Normalize and create tables using SQL
- Use PySpark to transform data and write the transformed data to SQL databases

Front-End API



Visualize

- Connect the SQL database: PySpark, Pandas
- Create visualizations with Mapbox
- Deploy the web application with Flask

A Glimpse of the Data



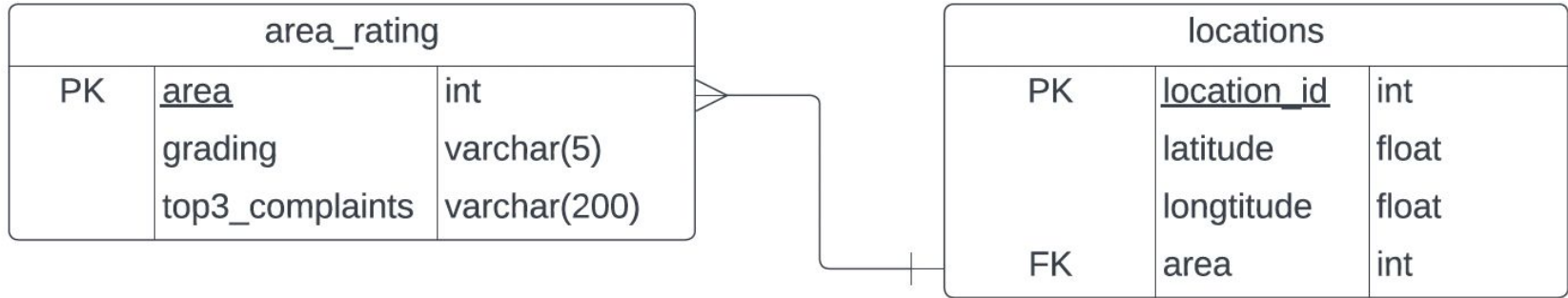
Zillow Data

City	RegionName	Borough	Neighborhood	AvgRentalPrice_2022	RecentRentalPrice	normalized	Latitude	Longitude	A_rating
0 New York	11385	Queens	West Central Queens	2754.70904	2798.324183	0.88238935	40.7039074	-73.892442	B
1 New York	10467	Bronx	Bronx Park and Fordham	1916.843626	1913.763889	0.61400401	40.8744717	-73.867987	A
2 New York	11226	Brooklyn	Flatbush	2534.483763	2567.450203	0.81184671	40.6460751	-73.955107	B
3 New York	11220	Brooklyn	Sunset Park	2141.206077	2277.866667	0.68587187	40.6411306	-74.013113	A
4 New York	11207	Brooklyn	East New York and New Lots	2769.975243	2823.458261	0.88727943	40.672341	-73.894845	B
5 New York	11214	Brooklyn	Southwest Brooklyn	1964.232803	2101.277778	0.62918373	40.6020542	-73.996624	A
6 New York	10025	Manhattan	Upper West Side	3703.685947	3715.296008	1.18636596	40.7976842	-73.967636	C
7 New York	10314	Staten Island	Mid-Island	1949.527109	2047.777778	0.6244732	40.6040228	-74.147161	A
8 New York	11221	Brooklyn	Bushwick and Williamsburg	2950.637198	3005.915873	0.94514913	40.6909045	-73.928436	B
9 New York	10458	Bronx	Bronx Park and Fordham	1999.576592	2092.555556	0.64050506	40.8614014	-73.888784	A
10 New York	11235	Brooklyn	Southern Brooklyn	2383.738069	2520.095238	0.7635598	40.5862936	-73.951726	A
11 New York	11206	Brooklyn	Bushwick and Williamsburg	3359.766338	3441.584111	1.07620151	40.7014247	-73.94292	B
12 New York	11229	Brooklyn	Southern Brooklyn	2157.987911	2410.4	0.69124743	40.6002584	-73.944368	A
13 New York	11355	Queens	North Queens	1982.295416	2147.773016	0.63496955	40.750443	-73.819955	A
14 New York	11230	Brooklyn	Borough Park	2349.86317	2400.997436	0.75270898	40.6215203	-73.965621	A
15 New York	11233	Brooklyn	Central Brooklyn	2590.653137	2728.696416	0.82983891	40.6789935	-73.920361	B
16 New York	10457	Bronx	Central Bronx	1745.76099	1969.016667	0.55920276	40.8479645	-73.897766	A
17 New York	10462	Bronx	Southeast Bronx	1980.84445	2090.555556	0.63450478	40.8428421	-73.857973	A
18 New York	10029	Manhattan	East Harlem	2629.047899	2683.664087	0.84213753	40.7920076	-73.94452	B
19 New York	11218	Brooklyn	Borough Park	2463.204171	2513.094444	0.78901441	40.6436449	-73.977288	A

311 Data

Area	Complaint Type	normalized	s rating	grading	Top Three Complaints
1	73043	0.18644472	5.36352012	B	['Electronics Waste Appointment', 'Illegal Parking', 'Missed Collection']
2	72970	0.18625838	5.36888584	B	['Illegal Parking', 'Noise - Residential', 'Blocked Driveway']
3	31929	0.08149985	12.2699615	B+	['Illegal Parking', 'Noise - Residential', 'Blocked Driveway']
4	16213	0.04138423	24.1637945	A-	['Noise - Residential', 'HEAT/HOT WATER', 'UNSANITARY CONDITION']
5	44379	0.11327889	8.82776989	B	['Illegal Parking', 'Noise - Residential', 'Electronics Waste Appointment']
6	557636	1.42338468	0.70255077	C	['Illegal Parking', 'HEAT/HOT WATER', 'Noise - Residential']
7	555482	1.41788652	0.70527506	C	['Illegal Parking', 'Noise - Residential', 'Blocked Driveway']
8	87756	0.22400015	4.46428278	B-	['Illegal Parking', 'Noise - Residential', 'Blocked Driveway']
9	75	0.00019144	5223.568	A	['New Tree Request', 'Electronics Waste Appointment', 'Street Light Condition']
10	2478193	6.32567114	0.15808599	C	['Noise - Residential', 'Illegal Parking', 'HEAT/HOT WATER']

ER Diagram



Evaluation Criteria



Applicable Data Governance Policies

Dashboard Leading Time

Tested how long does it take for the link to works in the user's search engine

Data Availability

Assessing the completeness of the data used in the website.

Data Consistency

Evaluating the consistency of the data template

Data Quality

- 311 Data : We extracted up-to-date data with CSV file directly from NYC Open Data website
- Zillow Data : the data is extracted directly from the website

Data Privacy

- Our data only include data that publicly available
- 311 data: complaints' data is published under their willingness
- Zillow data: there is no sensitive information about customers in the data

Recommendations



Integrate APIs

- Utilize the 311 API and Zillow API to fetch data
- Implement dynamic function in the cloud for transformation process

Automate Data Fetching

Set up cron jobs to automate data collection, with daily updates for 311 data and monthly updates for Zillow data.

Integrate 911's Data

Incorporate 911 data to analyze the safety of a neighborhood based on the frequency and types of incidents reported.

Cost Implications



Price: \$12/month/user

Storage Capacity: 2TB

Encryption: AES-256



Price: \$20/month/user

Storage Capacity:
Unlimited

Encryption: AES-256



Price: \$12/month/user

Storage Capacity: 2TB

Encryption: AES-256