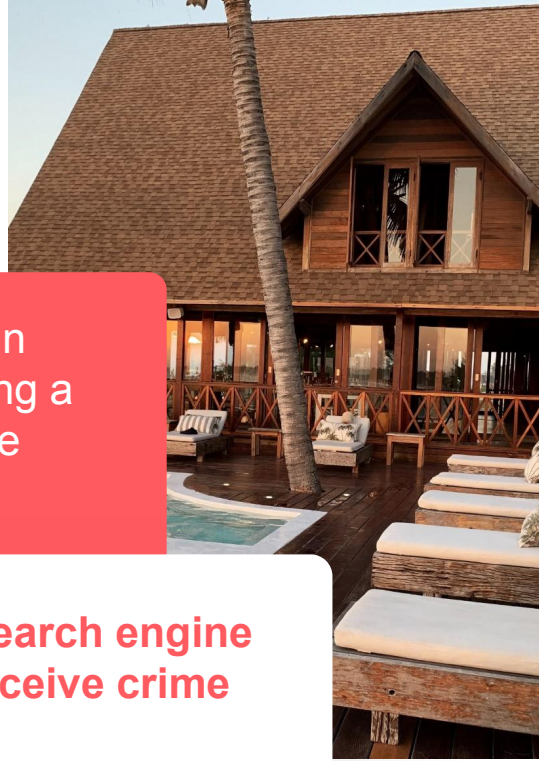
Crime Information Search Engine for Airbnb Users

Group 3
Jianxiao Chen, Kuan-Chung Chen, Jes Shang, Elina Xu, Jingshu Zhang

- Background
- Data Source
- Design & Technologies
- Policies & Costs
- Conclusion

1. Background & Definition



Airbnb is a popular platform for travelers to book lodging in various locations across the world. However, when booking a listing in an unfamiliar area, it can be difficult to assess the safety of the neighborhood.

This project aims to address this issue by developing a **search engine that allows users to input the name of a listing and receive crime information for the area**.
The search engine will serve for use in five major U.S. cities: **New York, Chicago, Los Angeles, Austin, and San Francisco**.

Search by listing

2. Data Source

Airbnb

http://insideairbnb.com/get-the-data/

- ❑ Name of Listing
- ❑ Latitude & Longitude

- ❑ Crime Occur Date
- ❑ County/Zip Code
- ❑ Crime Type: Robbery/Rape/Burglary/Murder/Shooting
- ❑ Population by Zip Code

Crime Rate

(2022-present)

- **Austin:** https://data.austintexas.gov/Public-Safety/Crime-Reports/fdj4-gpfu
- **LA:** https://data.lacity.org/
- **NYC:** https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc
- **Chicago:** https://data.cityofchicago.org/Public-Safety
- **San Francisco:** https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783
- **Census Data**: https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/

3. Design & Technologies

Extract

APIs are interfaces provided by websites that allow us to access data in a structured way. It's useful when we need to access real-time data and when we want to interact with a website programmatically. APIs provide a controlled way to access crime data from external systems.

Web scraping involves using a script to extract data from a website. This can be useful when the Airbnb website doesn't provide an API.

Transform

Apache Spark's distributed computing architecture allows it to process large datasets much faster than traditional ETL tools, which rely on serial processing.

Apache Spark provides a wide range of built-in data processing functions, as well as support for custom user-defined functions, which helps join crime and Airbnb datasets, clean and filter the useful information for users.

- Delete columns in Airbnb and crime data, such as room type, price, etc.
- Convert the latitude and longitude in the data to zip code

Transform

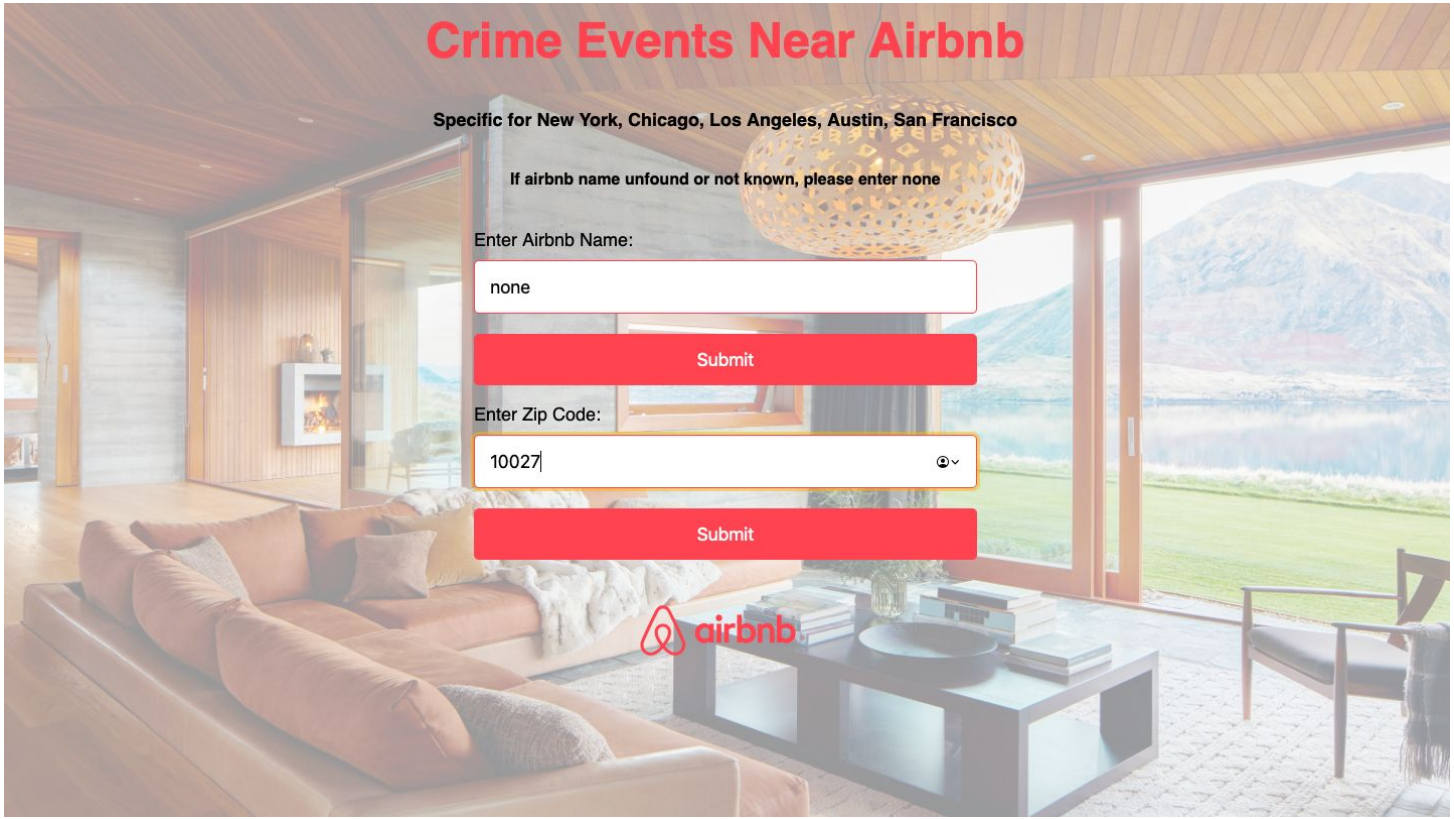
MongoDB is a document-oriented NoSQL database and has the characteristics of consistency and partitioning.

MongoDB allows us to store, manage Crime and Airbnb datasets from Spark and transform it into *Crime data summary, Day/Night time crime info, and Top5 crime event of the city* by using filter, sort, and aggregate function

Interactive web

Flask allows us to build an interactive web dashboard to let user input an Airbnb name and output the related crime information in the given area.

Search Engine Demo



4. Data policies & Cost Projection

Data Policies:

- Data security and privacy: both Inside Airbnb and crime datasets are from open data sources

Cost Projections:

Item	Annual Cost	Description
Storage costs	~1,380	Data size: 5TB including backup storage Storage type: AWS S3 Intelligent Tiering - **Storage costs**: 5,000 GB x $0.023 per GB x 12 months = $1,380 - **Monitoring and automation fees**: $0.0025 (< 1,000 objects in total)
Licensing fees	0	We will use an Open APIs

Salaries expense	18,880	Data Engineer salaries: for regular maintenance and future scaling projects
		- 1 part-time engineer for regular maintenance
		  - 8 hrs/month x 12 months x $30/hr = 2,880
		- 1 full-time junior engineer for project scaling (2 months)
		  - 40 hrs/week x 8 weeks x $50/hr = 16,000

Total (USD)	20,260	

4. Evaluation Metrics

Quantitative metrics:

Accuracy of Data

The accuracy of crime data obtained from the search engine

Speed of Search Engine

The response time of the search engine to process and display the results

Qualitative metrics:

User Satisfaction

User's feedback on the search engine, including ease of use, accuracy of information, and overall satisfaction

5. Conclusion & Recommendation

Conclusion

Successfully developed a Crime Information Search Engine to assist Airbnb users in assessing the safety of neighborhoods.
Created a scalable and cost-effective solution that improves the user experience when booking Airbnb listings.

Recommendation

1. Expanding the search engine to additional cities or countries
2. Integrating more data sources such as real-time crime updates or user-generated content
3. Enhancing the search engine with advanced filtering or mapping capabilities.
4. Exploring partnerships with Airbnb or other platforms to seamlessly integrate our search engine into their services