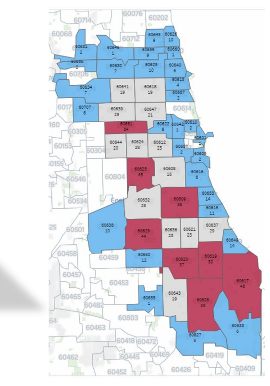
Crime Data Analysis for the City of Chicago

Team 008 – Krishna Kumar, Shilpi Jain, Christopher Manley, Kalvin Edwards, Sumit Machwe, Aditya Tadakaluru



Introduction - Motivation

- Neighbourhood safety is crucial for any new homeowner or renter [1]. Many times, an area that is considered safe may or may not be safe for an individual or a group of individuals [2].
- Current apps such as Trulia, Zillow, Crime grade and City Data provide overall general safety rating of a zip code without any personalization i.e., they do not take the factors associated with given user's profile into account. These websites use factors such as school ratings, dwelling attributes and comparable etc. that are limited in scope when it comes to personalization.
- Unlike general safety ratings, we want to personalize the safety ratings based on the user's profile i.e., find the areas that are safe or unsafe for an individual associated with specific user profile.

Novelty and Community Impact

- The key novelty of our approach is user centric personalized safety rating supported by a powerful and intuitive map-centric dashboard.
- Theories such as "clear victimology theory" is used in creating user profiles and identifying associated factors to evaluate the safety rating of a zip code.
- According to this theory the individual demographics are key to the probability of being a victim where factors such as age, sex, ethnicity, race, lifestyle [3][4] affluency can impact the type and probability of vulnerability [5].
- Our solution is beneficial to anyone considering moving to Chicago, current residents of Chicago, city officials and public policy makers, law enforcement agencies, insurance agencies [6], education boards and prospective businesses planning on investments in the area [7] can benefit from this application.
- Having the ability to give more insight on any type of crime for a particular location is invaluable information for prospective or current residents. For e.g., to be successful, street robbers must attack at the right time at the right place [8].

Solution Approach

GUI and Visualization

- Tableau dashboards are chosen as the main platform for implementing GUI and visualization while a combination of Tableau, MySQL, R and Python are used for the modelling and analysis component.
- **Personalized safety rating app** is a map-centric **tableau dashboard** and consists of three main components – dropdown controls, map control, floating panels.

Modelling and Analysis

- **Zip code rating using Crime Type** : Chicago dataset is spatially joined with Chicago zip code layer to extract a zip code for each crime record in the dataset. The zip codes are then mapped to a matrix of simulated user profiles and crime profile by the recommender engine discussed below using the probability distribution for being victimized in a zip code derived from the summary statistics calculated from historical crime data.
- **Mapping Profile susceptibility to a Crime Type**: Probabilities are calculated for each demographic attribute (e.g., age, income) against each crime type using the year 2019 from the Bureau of Justice website and stored in a reference table. We found that there were seven crime types that we could use that had a probability distribution for different classes in each attribute of demographic profile available in survey. We made certain assumptions that each attribute is independent(naïve bias) and the crime susceptibility of a user profile is a multiple of probability of each attribute.
- **Optimization of safety rating between profiles**: We used the probabilities calculated in the previous step to choose the crimes that are of high priority to a given user. We then built a user profile recommender system to aid reuse of safety ratings generated for one profile for other similar profiles using cosine similarity resulting in efficient use of current user profiles and less manual curation of profiles.

Data

- Because of its high crime rates and availability of large crime dataset, the city of Chicago is chosen as the region of interest. The original dataset downloaded from Chicago data portal contains 1.67 million crime records for the past five years
- Performed Feature engineering, leveraged Database SQL query, python/pandas, and geopy to create new features.
- The susceptibility of a profile is mapped to a crime type using survey responses from the National Crime Victimization Survey, a publication of the Inter-University Consortium for Political and Social Research.
- This data has 171 fields that we have re-engineered for features curated to regress the models for specific type of crimes against the demographic data available.

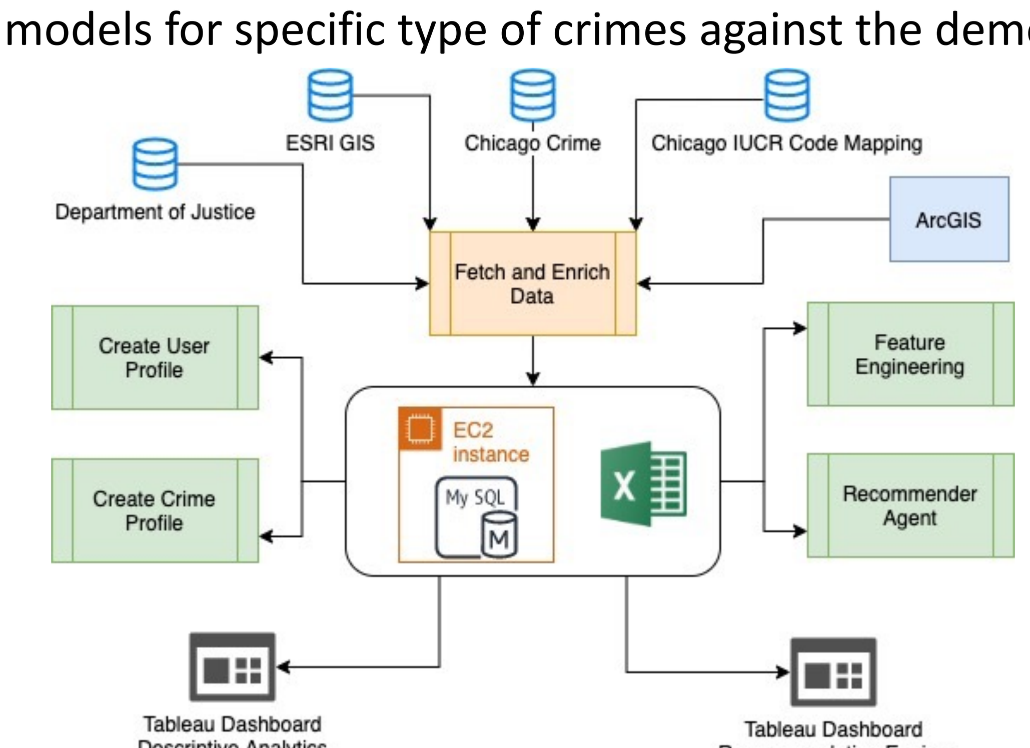


Fig 1: Solution Architecture

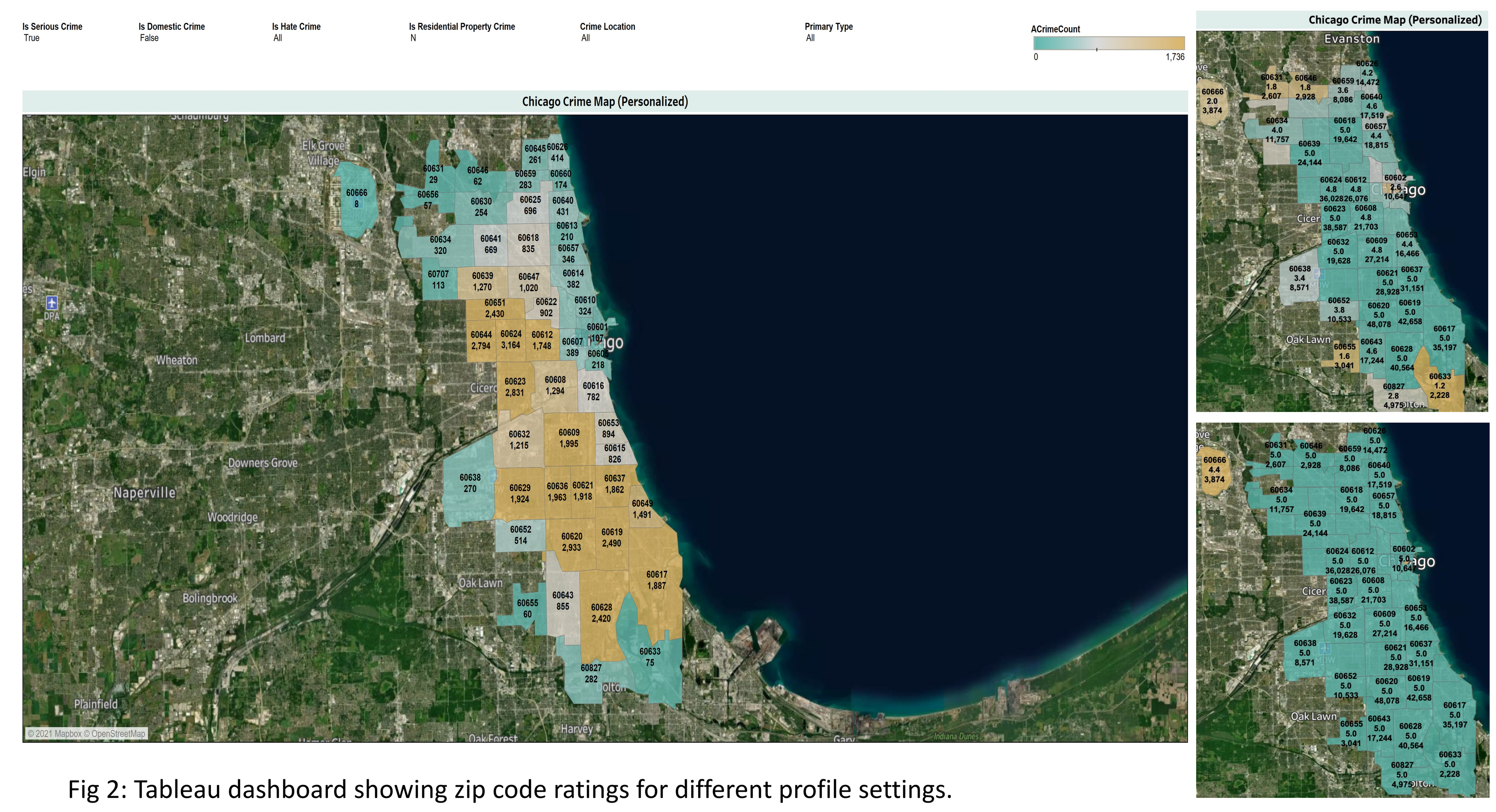


Fig 2: Tableau dashboard showing zip code ratings for different profile settings.

Experiments and Results

Crime as a factor of user demographic: Bureau of Justice statistics publishes a publicly available, standardized survey database across the US population. Our initial approach was to use this database for feature engineering by regressing each crime type as factor of demographics. During this process we found that the dataset required substantial amount of data engineering. We shifted our strategy to using summary statistics from the database as explained in our solution approach section. The main assumption (supported by literature review) since the inception is that the safety rating is a factor of demographic profile. We were able to objectively validate this assumption and the same can be observed on the map through noticeable differences in zip code symbology for both general rating of a zip code (Fig 2. center) and the personalized rating for two different user profiles (Fig 2. Smaller pictures on the right). Our key results are as follows:

1. Our recommendation agent provides zip code ratings in the dataset on a scale of 1 - 5 based on average rating given by other users who are deemed similar
2. For a user without assigned profile, Tableau dashboard will identify crime hot-spots across zip codes. Based on descriptive analytics, we conclude that:
   - 80% of zip codes have serious crime. Less than 50% of crimes occur around Residential Properties.
   - Hate crimes and Financial Crimes (near ATM and Banks) are sparse and exhibit no recognizable pattern
   - Assault and Robbery are the most predominant crime types
3. Not all zip codes of Chicago are equal in terms of crime distribution, underscoring the fact that user susceptibility depend on personal attributes like demographics, nature of work and lifestyle preferences
4. It is possible to assign a susceptibility score to a user profile at crime type and zip code level. Our implementation calculates probabilities across available demographic features for each individual crime type and derives a combined probability (product) for each user profile and each crime type.
5. Our tool can generate sophisticated user profiles of prospective renters, real estate home buyers via feature engineering large datasets. This is unique to our application and is not available with any existing market tools (Zillow etc.)

Future work

Integration with real estate solutions

By extending our standalone application into a cloud-based web application, it will allow for deep linking from a real estate website like Zillow etc. and serve as a good supplementary tool for users who have already shortlisted the houses and wants to compare the respective zip codes based on personalized safety.

Further personalisation based on lifestyle/demographic info

Create an interactive webpage (chatbot etc.) that can ask the users demographic and lifestyle questions, then dynamically map the user inputs to a profile with the help of recommendation engine.

Further personalisation on tableau

The map functionality can be further extended to give user the ability to unselect some hotspots from a zip code based on their lifestyle choices to recalculate the crime susceptibility based on their modified preferences.

References

1. Samarin, Mikhail and Sharma, Madhuri (2020) "A Neighborhood-level Perspective of Real Estate Determinants in Three U.S. Cities," International Journal of Geospatial and Environmental Research: Vol. 7 : No. 3 , Article 2.
2. Madero-Hernandez, A. (2021). Lifestyle Exposure Theory of Victimization. In The Encyclopedia of Women and Crime (eds F.P. Bernat and K. Frailing). https://doi.org/10.1002/9781118929803.ewac0334
3. Bunch, Jackson, Jody Clay-Warner, and Man-Kit Lei. "Demographic Characteristics and Victimization Risk: Testing the Mediating Effects of Routine Activities." Crime & Delinquency 61, no. 9 (November 2015): 1181–1205. https://doi.org/10.1177/0011128712466932.
4. Deng L. Comparing the Effects of Housing Vouchers and Low-Income Housing Tax Credits on Neighborhood Integration and School Quality. Journal of Planning Education and Research. 2007;27(1):20-35. doi:10.1177/0739456X07301467
5. Wolfgang, Marvin E., and Simon I. Singer. "Victim categories of crime." J. Crim. L. & Criminology 69 (1978): 379. https://scholarlycommons.law.northwestern.edu/jclc/vol69/iss3/8
6. Hefei Wen, Jason M. Hockenberry, Janet R. Cummings: The effect of Medicaid expansion on crime reduction: Evidence from HIFA-waiver expansions. Journal of Public Economics, ISSN 0047-2727, https://doi.org/10.1016/j.jpubeco.2017
7. Greenbaum RT, Tita GE. The Impact of Violence Surges on Neighbourhood Business Activity. Urban Studies. 2004;41(13):2495-2514. doi:10.1080/0042098042000294538
8. Bernasco, Wim, et al. "Do Street Robbery Location Choices Vary Over Time of Day or Day of Week? A Test in Chicago." The Journal of Research in Crime and Delinquency, vol. 54, no. 2, SAGE Publications, 2017, pp. 244–75, https://doi.org/10.1177/0022427816680681