**Project Title**: *HAPPY INDEX*

**Project Summary**

The project aims to develop an interactive web application that estimates happiness levels across various regions in the United States, down to the zip code level. By utilizing publicly available datasets on factors like crime rates, income, air quality, and mental health services, the application will provide users with a detailed analysis of happiness in specific areas along with updates on recent news near the selected location. Users can interact with a geo-coded heatmap, search for happiness scores by location, and update their own happiness ratings over time. Moreover, city officials can utilize the application to get deeper insight into the well-being of their residents and make more precise decisions about the resources allocation.

**Description of Application**

Happiness is both a subjective and relative measure. Yet, we often see convincing news articles on why certain countries or states are happier places to live than others. The USA encompasses a vast amount of land, of which individuals often can only visit a fraction of during their lifetime. Most often people tend to live near their childhood neighborhoods for their whole lives. Unfortunately, this means that people are likely passing out opportunities to live in places that would make them happier. Furthermore, when people are choosing a house to live in from different neighborhoods, assessing these areas for future happiness can be difficult. Conversely, city officials might not have the tools needed to properly understand the happiness of their residents in different regions and, thus, are unable to properly allocate resources in improving the lives of their residents.

Our application seeks to provide a unique and interactive approach to estimating the happiness of individuals living in certain regions. By using our application, users can explore different areas to live in, and possibly optimize their future happiness. We plan to have more granular analysis on locations than existing sources, which are on a country, state, or city level. We will provide users an interactive experience in which they can query an analysis of any location based on the zip code, city, or state. We will also use our user base as another data source to answer surveys upon sign up. City officials can then use this user base to perform studies on their residents' happiness levels and allocate resources in improving conditions where needed.

The main data sources for the application will be publicly available datasets based on living conditions, such as climate, cost of living, safety, education, etc. Users of the application will be able to interact and query location based happiness analytics via a

browser. The website will feature a heatmap of happiness overlaid on a map of the USA, with the ability to search for specific locations and see a short summary of specific regions, and their living conditions. To provide an interactive and up-to-date feel, we will also implement a “Recent News” feature, with web scraping functionality for recent news near the chosen location.

**Creative Component**

The applications will have three creative components. First, a summary of the overall happiness analytics will be displayed via a geo-coded **heatmap** that is overlaid on a map of the USA. This provides users a visually pleasant and easy way to quickly gauge the happiness of different regions. Our second creative feature will try to provide a short list of newly added news articles that our application has **web scraped** regarding the location that the user has selected for focus. Finally, if plausible, we consider utilizing a real-time **API connection** to some of the data sources for up-to-date data. For example for fetching transportation proximity or entertainment options within a close proximity of the chosen location. Additionally, our vast amount of data sources will require thorough data engineering to provide aggregated data at a zip code level in a scalable manner.

**Usefulness**

Our application is useful because it allows users to have a quantitative measure of happiness within different locations. For example, when considering future cities to live in, or when trying to decide between two neighborhoods to rent from. Additionally, city officials and planners can utilize the data gathered by our application to allocate development resources where needed the most.

Our application provides a more granular level of observation for happiness than existing applications. The application can provide a happiness score for any location in the USA. Other existing websites are limited to a selection of pre-researched locations, like cities or states (see Niche.com or World Happiness Report), or provide an analysis on large generalized areas (see Sharecare.com). Our application will be able to provide details up to a zip code level. After all, it is unlikely that different neighborhoods within one city would experience the same level of happiness. We plan to enable this granular analysis via zip code based datasets and complex features, such as heat maps, web scraping, and API usage, all of which are explained in more detail under the previous and following sections. However, the application’s usefulness is impacted by our poor ability to objectively provide a happiness analysis. Our analysis is influenced by our subjective thinking, and our application will likely incorporate a generalized model for assessing happiness.

Users are also encouraged (or required) to sign up before use of the applications. With a sign up feature, we will be able to perform a short survey on all users and provide more value for city officials if they want to engage their local residents. Whilst city officials are expected to use the same view as regular users, they have the ability to submit requests to local residents to fill in additional questionnaires via our application management team communication channels. Thus, our application also serves as a convenient way for interaction between residents and city officials.

**Realness**

We will use up to 7 different datasets to calculate happiness in a region. By looking at factors such as crimes, income & poverty, home value, air quality, and people served in mental health facilities we can evaluate how happy a specific location is.

Crime Data: The crime data will be sourced from the "Crimes - 2001 to Present" dataset provided by the City of Chicago. This dataset is available in CSV format and contains millions of rows, as it tracks criminal activities in Chicago from 2001 onwards. The dataset captures essential details such as the type of crime, location (with latitude and longitude), and the time of the incident. By using this data, we can evaluate the public safety of different areas, which plays a crucial role in assessing the happiness of residents. Although this dataset is limited to the Chicago area, similar datasets from other regions could be incorporated into our application in future versions.

Income and Poverty Data: The income and poverty data will come from the American Community Survey (ACS), available through the U.S. Census Bureau. This data is offered in both CSV and XLS formats and contains detailed demographic information about household income, poverty rates, and economic conditions across various regions. With attributes like income distribution and poverty levels by zip code, this dataset provides insight into the financial well-being of residents, a key factor in determining regional happiness.

Housing Data: Housing data will be sourced from Zillow Research, which offers a wealth of information about home values, price trends, and rental costs. Available in CSV or XLS format, this dataset has moderate to high cardinality depending on the region and time frame selected. Housing affordability and market trends directly affect quality of life, making this data a vital component for evaluating happiness in specific areas, particularly in relation to cost of living.

Air Quality Data: Air quality data will be accessed through the AirNow API, which provides real-time air quality index (AQI) measurements across different geographical locations. This data is available in JSON format via the API and captures important metrics like particulate matter, ozone levels, and other pollutants. Air quality has a direct impact on both physical health and overall well-being, making it a key indicator when assessing the environmental quality and happiness of an area.

Mental Health Services Data: Mental health services data will be obtained from the National Mental Health Services Survey (N-MHSS) 2020, which is available in CSV format. This dataset includes information on the number of mental health facilities, the population served, and the types of services available in different regions. Access to mental health services is crucial for maintaining mental well-being, and this data will provide valuable insights into how mental health support contributes to happiness levels in various locations.

Nearby Search (New) | Places API | Google for Developers: The Google API for places will provide an additional data source for evaluating access to key resources such as grocery stores, transportation, and public parks. This data is accessible in JSON format through requests from Google's APIs, and includes data on place types, URLs and coordinates. However, since we already have multiple other data sources, we consider this as a lower priority and reserve the option to implement this based on available resources: time and money.

Beautiful Soup: This is a Python library that allows users to scrape and parse HTML and XML documents easily. We will use this to find recent articles/news, which will act as a data source, about the specific location or zip code we are looking at. Due to the time constraint as well as the various amount of datasets we already have, this feature might not be implemented.

**Description of Website Functionality**

*Users on the app can interact with the application in many ways. Users will create a profile when signing up, providing information like their residential location (zip code), gender, and their happiness level based on where they live (see Figure 1). Users will be identified based on their emails and are allowed to create only one account per email. A possible CAPTCHA check will be implemented to prevent bots from creating multiple accounts and consequently manipulating the user-sourced data. Users could have the ability to update their happiness rating over time, reflecting any changes in their perception of their environment.*

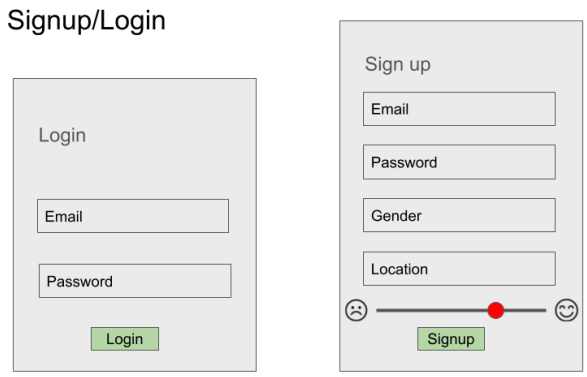


*Figure 1. Signup view for users to access our application.*

*After signup, users can search for specific zip codes to check happiness scores or browse a map that displays happiness scores for different areas (see Figure 2). These happiness scores are computed using the publicly available data outlined before. The total happiness score of an area will be a sum of multiple factors, which we will optimize to reflect happiness as accurately as possible. In case the source data is updated and uploaded into our applications back-end, these happiness scores will be updated as well.*

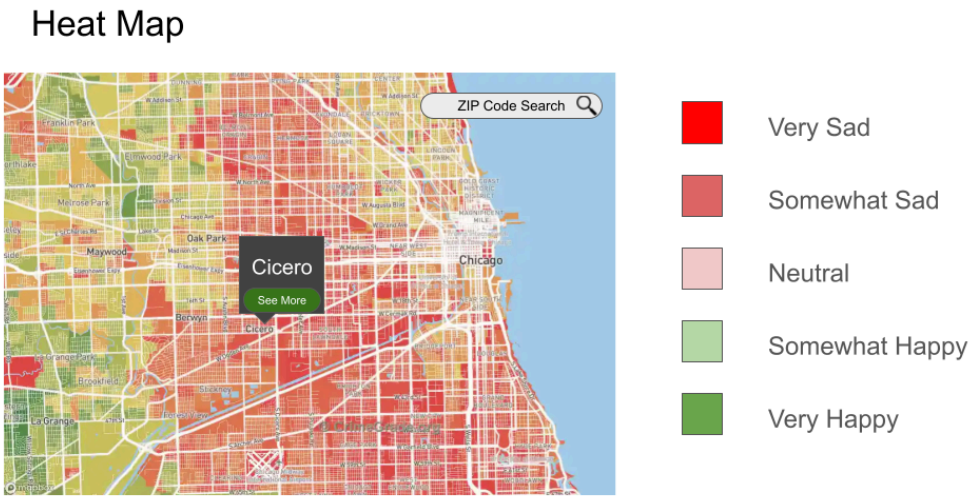


*Figure 2. Heat map displaying happiness levels across residents in different zip codes. (adapted from:* https://crimegrade.org/safest-places-in-chicago-il/)

*Additionally, users can zoom in and click "See More" from within the map to view detailed statistics about each region (e.g., income, housing costs, mental health services; see Figure 3). This view is also complimented by news articles scraped from the web. Finally,*

*the users could have the option to delete their profile if they choose to stop using the app, upon which sensitive information related to them will be deleted.*

Heat Map pt.2 (See More Page)

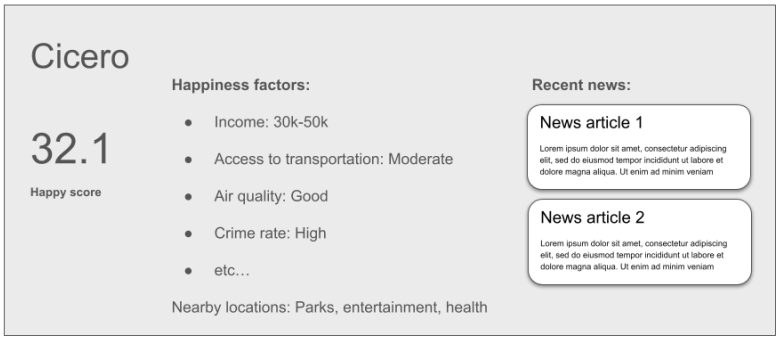


Figure 3. Additional details from user selected location (after pressing “See More” button)

**Project work distribution**

As part of this course, the entire team will work on the back-end, especially the database. In addition, we have certain functionalities for which we have named responsible teammates. However, we will utilize our teammates' strengths to the best of our abilities and being named as responsible in this first stage does not mean that person will complete the full functionality of the said feature. Each teammate is responsible for helping other team members where needed.

Responsibilities:

***Subash**: Database development and web scraping for recently published news articles near selected locations.*

***Jiahe**: Database development and main front-end functionalities.*

***Jaakko**: Database development and data engineering as well as possible API connections to external services.*

***Yun-Yang**: Database development and data engineering as well as web scraping for recently published news articles near selected locations.*