

# City-scalable Destination Recommender System for On-demand Senior Mobility

---

## Data description

Describe the data that will be gathered in the course of the research project, including whether the data should be preserved for long-term access.

One of the objectives of this project is to better understand the mobility needs of the elderly across different cities. This is accomplished via a joint survey conducted with collaborators from the University of Texas, El Paso (UTEP), on elderly living in El Paso, TX, and in New York, NY. Aggregated data and the codes used in this project are preserved for long-term access.

Additionally, researchers used Yelp's dataset to extract collaborative and content-based features to identify customer and restaurant profiles. With the help of the Yelp data, it is possible to run the destination recommender system with user observations and recommendations over time.

## Data format and metadata standards

Describe the standards and machine-readable formats that will be used in the course of the research project.

Although the survey was created in Qualtrics, the same questions were copied in Microsoft Word. Then, hard copies were brought to the senior centers to have the seniors fill out. Each completed survey form was manually uploaded to Qualtrics at the end of each survey day.

Two students from NYU, with proper IRB certification to survey human subjects, conducted a survey at senior centers. The questionnaire (see final report from UTEP Year I project) was conducted at several senior centers in New York City (NYC). Based on NYCHA Facilities and Service Centers data from December 2012 (NYCHA, 2012), there are 116 occupied senior centers in NYC. The surveys took place at five different senior centers (Figure 3) from February 5 to February 16, 2018.

The Yelp dataset contains five different tables: User, Business, Review, Check-In, and Tips. This information contains actual business, user, and users' review data from >700 cities and JSON files. In total there are 156,000 businesses, 1,100,000 users, and 4,700,000 reviews. The ratings users give to businesses range from 1-5 as discrete values as a number of review stars. This dataset spans more than 10 years of Yelp reviews.

## Policies for access and sharing

Discuss the access policies that will apply to the data, so as to protect against the disclosure of identities, confidential business information, national security information, etc. and whether public use files may be generated from the data.

The individual survey responses specified as sensitive by the respondents is not shared with the public.

Final datasets, aggregated in a way to remove private data, are shared with the public via Zenodo, or the National Transportation Library repository.

The codes for a prototype recommender system and computational experiments are publicly available on Zenodo.

The Yelp dataset used in this study is already publicly available.

## Policies for re-use, redistribution, derivatives

Discuss the policies for re-use, re-distribution and derivative projects.

The NYU PI Joseph Chow and the UTEP PI Kelvin Cheu are the only team members with the rights to manage the data.

The aggregated dataset will also be preserved in the NTL repository, where it will remain under public domain. The data are under the public domain and free for reuse, redistribution, and derivative products.

## Plans for archiving and preservation

Outline the plans for archiving and preservation, specifying where research data will be deposited, and specify that data will be deposited at the time of initial publication of any related peer-reviewed journal article.

The dataset will be archived on Zenodo. Prior to archiving, the data is stored on C2SMART Data Warehouse which is backed up regularly.

The research team will mint persistent Digital Object Identifiers (DOIs) for each dataset stored in Zenodo. These DOIs will be associated with

dataset documentation as soon as they become available for use.

The dataset will be retained in perpetuity. Zenodo meets all of the criteria outlined on the "Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan" page: <https://ntl.bts.gov/publicaccess/evaluatingrepositories.html>