

Team Project Proposal

Part 1. General Information

Name: Nicholas Souris(ns3205), Yanhong Yang (yy1553), Zeleng Zhuang(zz1135)

Project Title: Recommendation System for NYC Residential Community

Project Description: Our project aims to use historical data regarding housing prices along with several other parameters that might affect the standard of living of residents in an effort to create a recommendation system for NYC residential communities based on zip code. We will collect data including housing price, income, safety, infrastructure and other commonly used factors in evaluating a residential community. We will learn about kinds of techniques and algorithms to build a recommendation system and apply them to practice. If time allows, we will build a platform for our recommendation system to collect data from anonymous users that are willing to provide information to enrich our database.

Who will benefit from the analytic: all and potential New Yorkers

Part 2. General Data Source Information

Data Sources	Data Source Description	Data Size
www.city-data.com	Statistical data on population, house income, house value and etc.	MB
http://maps.nyc.gov/crime	Crime data.	MB
Yelp	Data regarding transportation, restaurants, shopping malls, supermarkets and so on.	GB
Trulia, Zillow, Realtor	These are websites listing rent and mortgage prices provided by sellers.	MB ~ GB

Part 3. Detailed Data Source Information

Data Sources	Data Characteristics	Data Frequency
www.city-data.com	Need to use API.	Statistical and static data
http://maps.nyc.gov/crime	Need to use API.	Monthly
Yelp	Need to use API.	Real-time
Trulia, Zillow, Realtor	Need to use API	This data source is currently in our backup plan. Daily data should be good enough for the project.