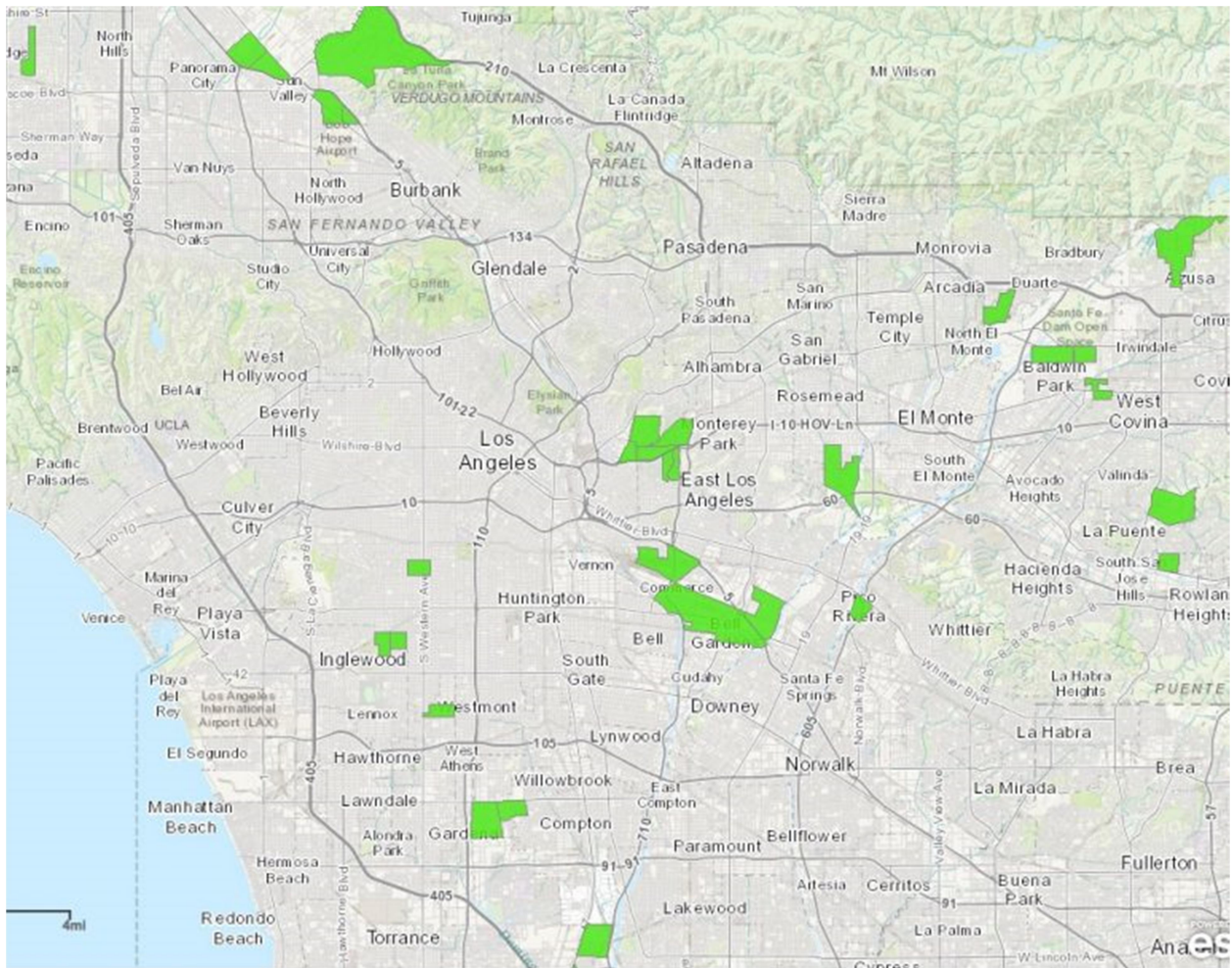Q: What areas of Los Angeles have the greatest nutritional needs?

USDA defines a Food Desert as an area more than a mile away from a supermarket and they estimate 23.5million Americans fall under this category.

Map from Foodforward.org

11.5 million of them live at or below federal poverty thresholds.

Low income and low access populations are usually at the greatest risk for health problems.

Data: Data.city.org & USDA.gov & census.gov

Hypothesis: Using USDA and census data can we identify the areas with the greatest need?

🏆 Featured Prediction Competition

Zillow Prize: Zillow's Home Value Prediction (Zestimate)

Can you improve the algorithm that changed the world of real estate?

$1,200,000
Prize Money

Zillow · 361 teams · 8 months to go (4 months to go until merger deadline)

Overview   Data   Kernels   Discussion   Leaderboard   More   **Submit Predictions**

Overview

- **Description**
- Evaluation
- Prizes
- Competition Overview
- Timeline

Zillow's Zestimate home valuation has shaken up the U.S. real estate industry since first released 11 years ago.

A home is often the largest and most expensive purchase a person makes in his or her lifetime. Ensuring homeowners have a trusted way to monitor this asset is incredibly important. The Zestimate was created to give consumers as much information as possible about homes and the housing market, marking the first time consumers had access to this type of home value information at no cost.

"Zestimates" are estimated home values based on 7.5 million statistical and machine learning models that analyze hundreds of data points on each property. And, by continually improving the median margin of error (from 14% at the onset to 5% today), Zillow has since become established as one of the largest, most trusted marketplaces for real estate information in the U.S. and a leading example of impactful machine learning.

111 Archer Ave,
New York, NY 10031
4 beds · 3 baths · 3,410 sqft

● FOR SALE
$1,175,000
Zestimate®: $1,275,448

EST. MORTGAGE
$4,461/mo 🏠
Get pre-qualified

Built in 2009, perfectly blending elegance with functional living space. Excellent floor plan with 3 beds up and 1 on main. Open living, kitchen & dining w/ huge fireplace & Sound views. Spacious kitchen w/ slab granite surfaces and center island. Huge master suite with Jacuzzi tub & separate shower. Features: hdwd floors, all

In this competition, Zillow is asking you to predict the log-error between their Zestimate and the actual sale price, given all the features of a home. The log error is defined as

$$logerror = log(Zestimate) - log(SalePrice)$$

and it is recorded in the transactions file **train.csv**. In this competition, you are going to predict the logerror for the months in Fall 2017. Since all the real estate transactions in the U.S. are publicly available, we will close the competition (no longer accepting submissions) before the evaluation period begins.

Train/Test split

- You are provided with a full list of real estate properties in three counties (Los Angeles, Orange and Ventura, California) data in 2016.
- The train data has all the transactions before October 15, 2016, plus some of the transactions after October 15, 2016.
- The test data in the public leaderboard has the rest of the transactions between October 15 and December 31, 2016.
- The rest of the test data, which is used for calculating the private leaderboard, is **all** the properties in October 15, 2017, to December 15, 2017. This period is called the "sales tracking period", during which we will not be taking any submissions.
- You are asked to predict 6 time points for **all** properties: **October 2016 (201610), November 2016 (201611), December 2016 (201612), October 2017 (201710), November 2017 (201711), and December 2017 (201712).**
- Not all the properties are sold in each time period. If a property was not sold in a certain time period, that particular row will be ignored when calculating your score.
- If a property is sold multiple times within 31 days, we take the first reasonable value as the ground truth. By "reasonable", we mean if the data seems wrong, we will take the transaction that has a value that makes more sense.

File descriptions

- **properties_2016.csv** - all the properties with their home features for 2016. Note: Some 2017 new properties don't have any data yet except for their parcelid's. Those data points should be populated when **properties_2017.csv** is available.
- **properties_2017.csv** - all the properties with their home features for 2017 (will be available on 10/2/2017)
- **train_2016.csv** - the training set with transactions from 1/1/2016 to 12/31/2016
- **train_2017.csv** - the training set with transactions from 1/1/2017 to 9/15/2017 (will be available on 10/2/2017)
- **sample_submission.csv** - a sample submission file in the correct format

Loan Data

This dataset includes customers who have paid off their loans or not

by **Zhijin** · last updated 2 months ago

Overview   Kernels   Discussion   Activity

Download (6 KB)   **New Kernel**

Kernels	>		Discussion	>		Top Contributors	>
Loan Data EDA (+ simple mo... run a month ago	3 votes		Loan Analysis 13 days ago	2 replies		satadru5	1st
first notebook run a day ago	2 votes		Loan Status a month ago	0 replies		Net	2nd
Loan Data - Worksheet run a month ago	1 vote		first notebook a month ago	1 reply		Zhijin	3rd

Context

This data set includes customers who have paid off their loans, who have been past due and put into collection without paying back their loan and interests, and who have paid off only after they were put in collection. The financial product is a bullet loan that customers should pay off all of their loan debt in just one time by the end of the term, instead of an installment schedule. Of course, they could pay off earlier than their pay schedule.

Content

Loan_id A unique loan number assigned to each loan customers

Loan_status Whether a loan is paid off, in collection, new customer yet to payoff, or paid off after the collection efforts

Principal Basic principal loan amount at the origination

terms Can be weekly (7 days), biweekly, and monthly payoff schedule

Effective_date When the loan got originated and took effects

Due_date Since it's one-time payoff schedule, each loan has one single due date

Paidoff_time The actual time a customer pays off the loan

Pastdue_days How many days a loan has been past due

Age, education, gender A customer's basic demographic information