

StratusHere Location Intelligence Engine

Idris Kuti, Gautam Matta, Akash Nikam,
Junfei Xia, Chris Walker

Team 105

Introduction

StratusHere is an interactive web application which allows small business owners to determine the viability of locations for their business type and annual sales volume. This helps small businesses and non-technical stakeholders make data driven decisions about their location.

Using only simple inputs (zip code, business category, and sales estimate) we generate recommendations for properties using stored data. The application serves all Atlanta zip codes.

We categorize properties into Okay, Good, and Great based on how they compare to nearby properties using a random forest model.

How to Use our App?

Step 1 : Enter the zip code of interest

Step 2 : Provide a category and sales estimate

Step 3: View recommendations from the interactive map for your business!

You can also switch between Location and Heat Map view. The Heat Map view highlights Good and Great concentrations.

Our Approach

We narrowed the available businesses in the Atlanta region to 150k. This allows us to focus on small retail businesses with under 100 employees.

Existing business data is used to predict the business category. By successfully predicting the category, we apply the model to new properties.

If a property has an above median number of trees in the random forest which indicate predict a business category it is considered to be Great. If it is at the median it is Good and everything else is Okay.

Under the Hood

Data comes from three primary sources; (a) Mergent Intellect business database, (b) the US census, and (c) data generated from sub-models for traffic estimation. Data is cleaned and merged for model fit. Data is 100 Mb in size before compression on-disk and data preprocessing for modeling. Our app hosts a slice of this data.

Modeling Experiments

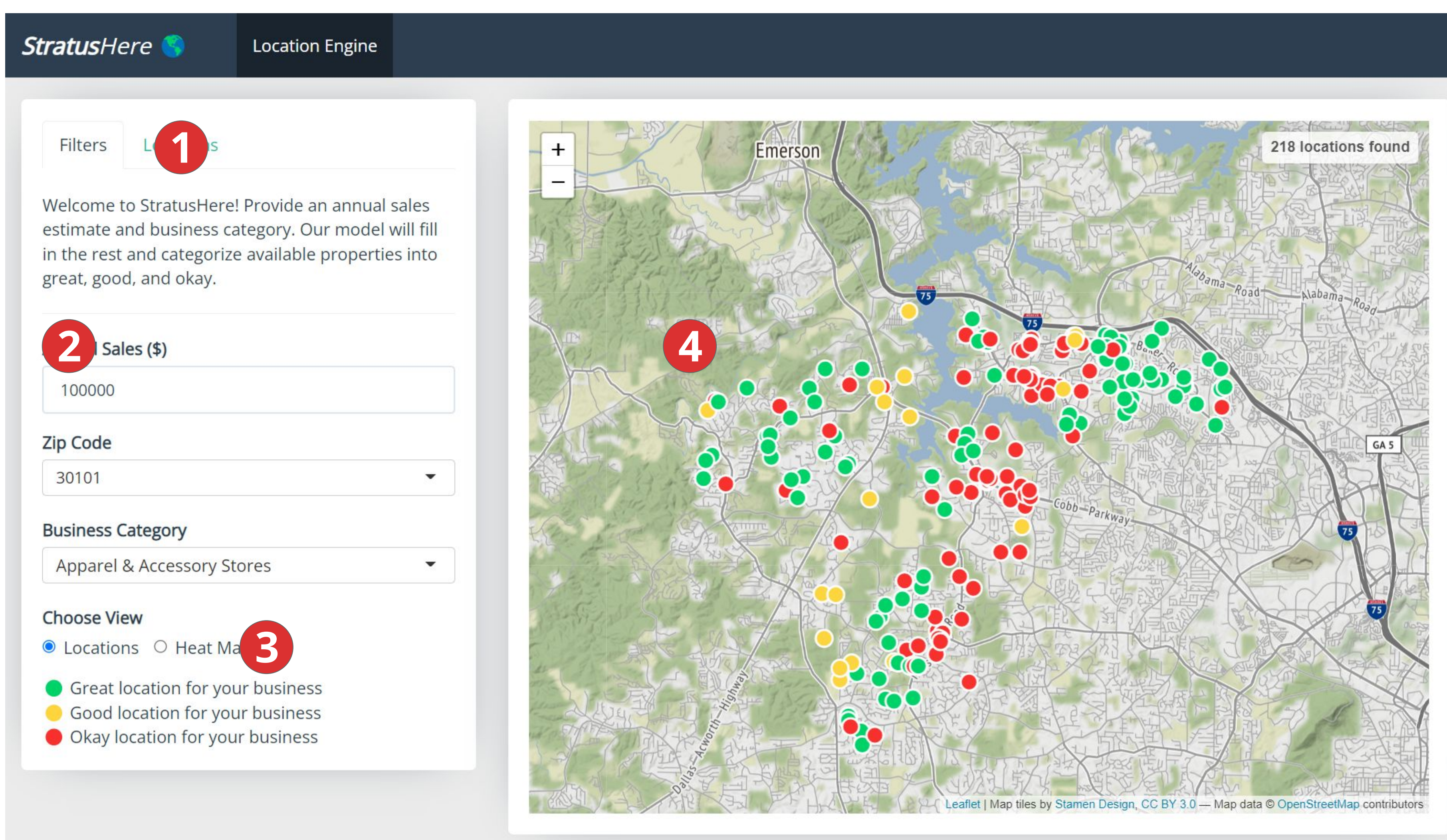
From a selection of classification models like KNN, XGboost, etc., RandomForestClassifier is chosen as the Machine Learning Algorithm to drive the app. Accuracy of prediction, speed and size of the model are considered when making this selection.

Web App

This R Shiny web application is hosted on shinyapps.io. The model serves a random forest which is carefully researched and estimated in Python before embedding it into the production app.

Users input basic details about their business. The web application merges inputs with stored census data. At model prediction C++ is used to compute 10 million recursive calls in under 0.2 seconds.

We fashioned the web application to resemble real estate search engines like Redfin, and Zillow.



1. Property listings table tab
2. Control panel + user inputs

3. Location/heat map selector
4. Interactive map + hover tooltips