



# STROOM

REIMAGINING COMMERCIAL REAL ESTATE



Using ML to estimate upside and value - add opportunity

# Product Overview



## What We Do

Stroom is a software platform that provides commercial real estate acquisitions teams the ability to screen and underwrite deals faster



## Data Science

Stroom leverages advanced data science and economics models to turn Real Estate data into decisions.

By aggregating data from multiple sources including alternate data, the platform creates a unique view of any property and accurately estimates its value-add potential.



## Investment Integrations

Stroom takes a very focused approach on automating acquisition workflows.

The platform streamlines the acquisition process and offers proprietary insights to screen and forecast both on and off market properties that meet an investors' buy box. Put simply, the platform addresses the bottleneck by reducing time spent on administrative tasks related to sourcing and analyzing deals.



# Target Customer



## 01 Acquisition Teams

- Average Acquisition Deal Size: \$15M - \$30M
- Often acts as GP/Sponsor in acquisitions in capital stack
- Small/Midsized firm of fewer than 20 professionals
- Strong operator within local geography + invest in technology

## 02 Developers

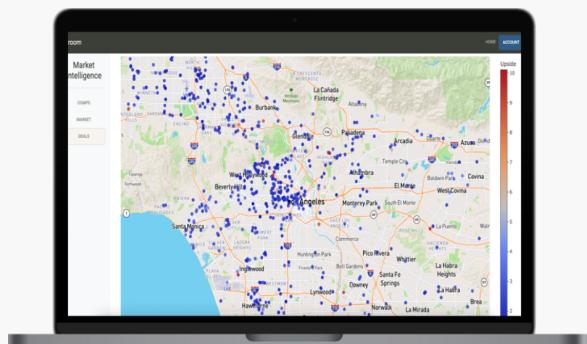
- Typically for the initial phases of development due diligence process
- Regional players on the precipice of launching projects in new markets
- Small/midsized firms building 50-units or fewer per property

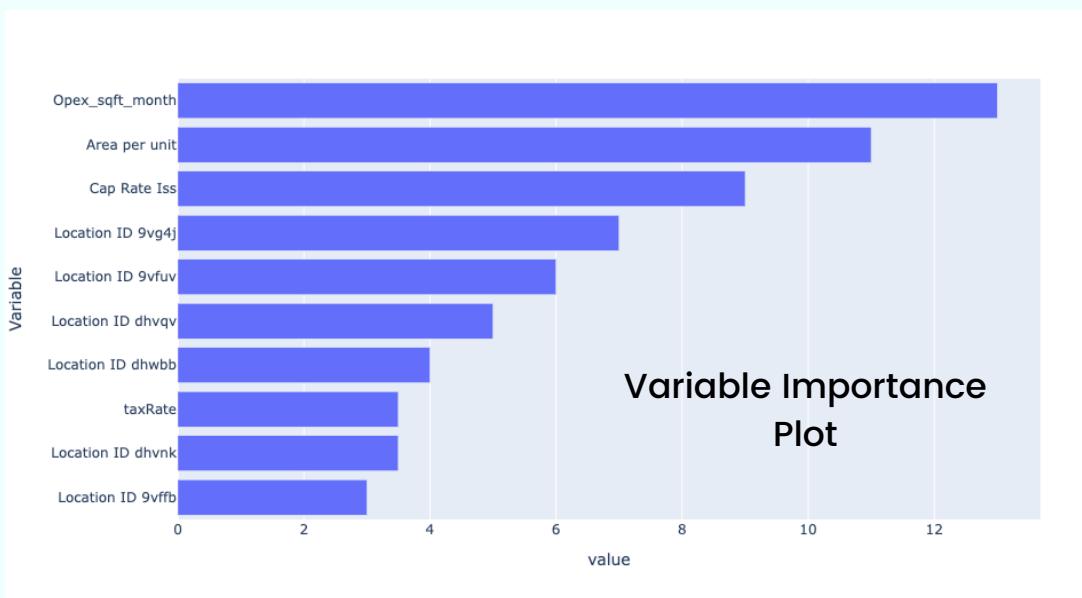
# Data Science

## Model Interpretability and Transparency

Our training data includes thousands of properties from different markets. We obtain data on physical characteristics of property, financial, operational data such as Revenue, Expenses, and Occupancy. We also gather traditional and alternate data on local economy, demand, demographics and location specific features. Through a set of API and data pipelines, we've created a rich dataset for training our model.

Our models are powered by raw data from nationwide 3rd party Real Estate data providers, location data and several other sources of data. We place special emphasis on the quality of data and ensure that we maintain high standards. Each data source is thoroughly examined for outliers and missing-ness before we ingest data into our platform and use it in training our ML models.





## Model Specification

We perform a series of steps to train the model.

- Split data into training and test. We use 80-20 split.
- Check for bad, missing data. We use statistical distributions to identify outliers and use business logic to drop certain rows with bad data. In some cases, we use geographic proximity to impute missing data or replace bad data.
- We use geohashes (location identifiers) in our models to control for some of the unobservables and location specific characteristics.
- Categorical variables are encoded using OneHotEncoder.
- Hyper parameters are tuned using GridSearch CV to obtain the best combination for variables, and subset of training data.
- Metrics like Mean squared error and Mean Absolute Percentage Error are used to evaluate the model.

**25K**

# of Properties

**100+**

Features Analyzed

**18%**

Mean Abs. Error