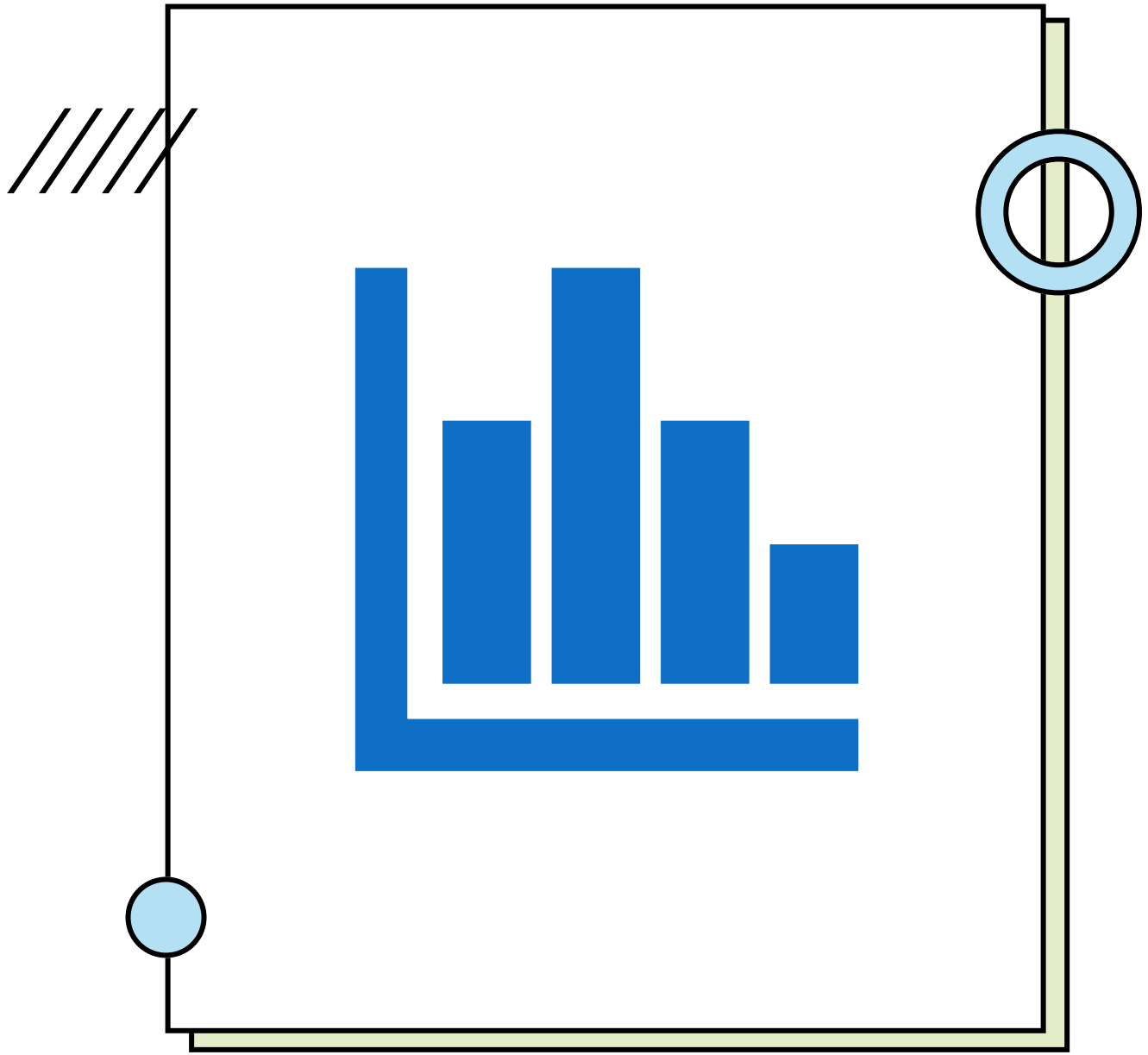
Creating an Algorithm to predict house pricing



Analysis and Interpretation by Leticia Fernandes

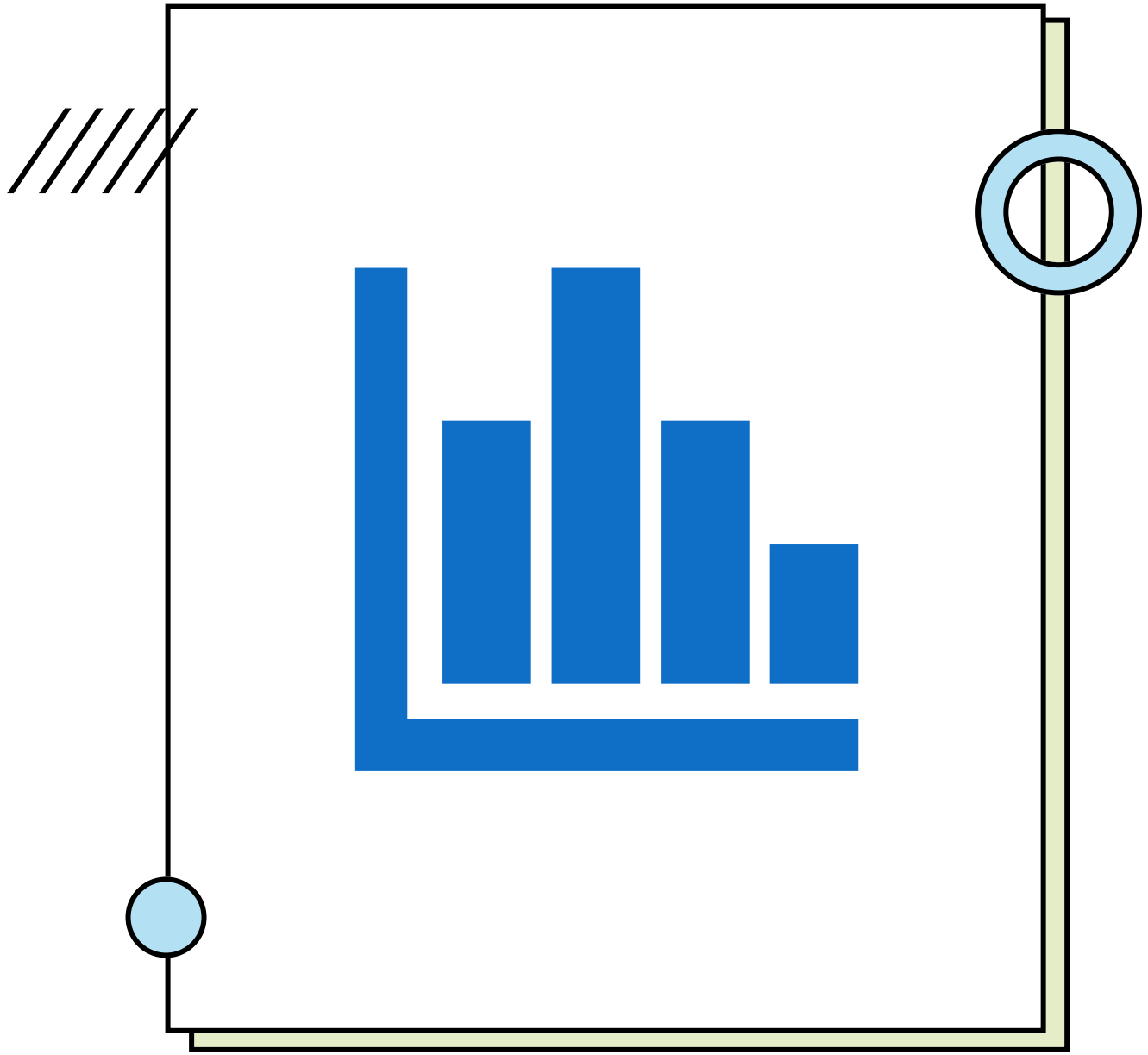
Business Problem

- Help a startup company to predict house prices in the King county region to compete with Zillow in the consumer information marketing business.

Summary

Create an Algorithm to predict prices on properties in King county. Noted important factors included:

- Zip Code,
- Sqft
- Grade (Condition)
- Renovated



Data

Data provide by Flatiron School corresponding zip codes taken from King County Open Data Source.

Our variables include:

- View;
- Waterfront;
- Number of Bedrooms;
- Number of Bathrooms,
- SQFT living;
- Grade (condition);
- Year Built;
- Renovated, etc.

Modeling and Methods

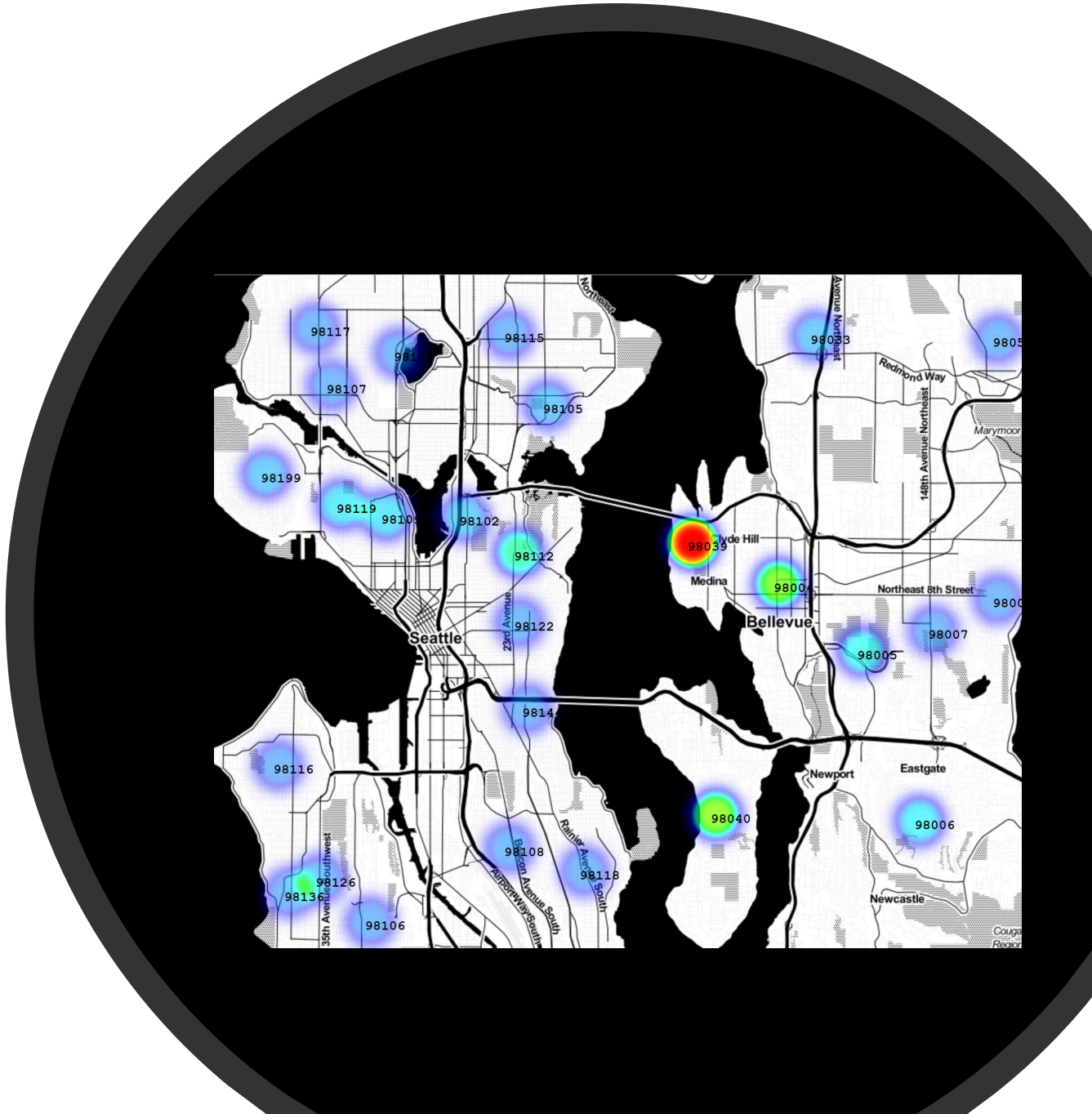
- Linear Regression - OLS model
- Baseline Model
- Log Transformation
- R^2
- Dummy variables

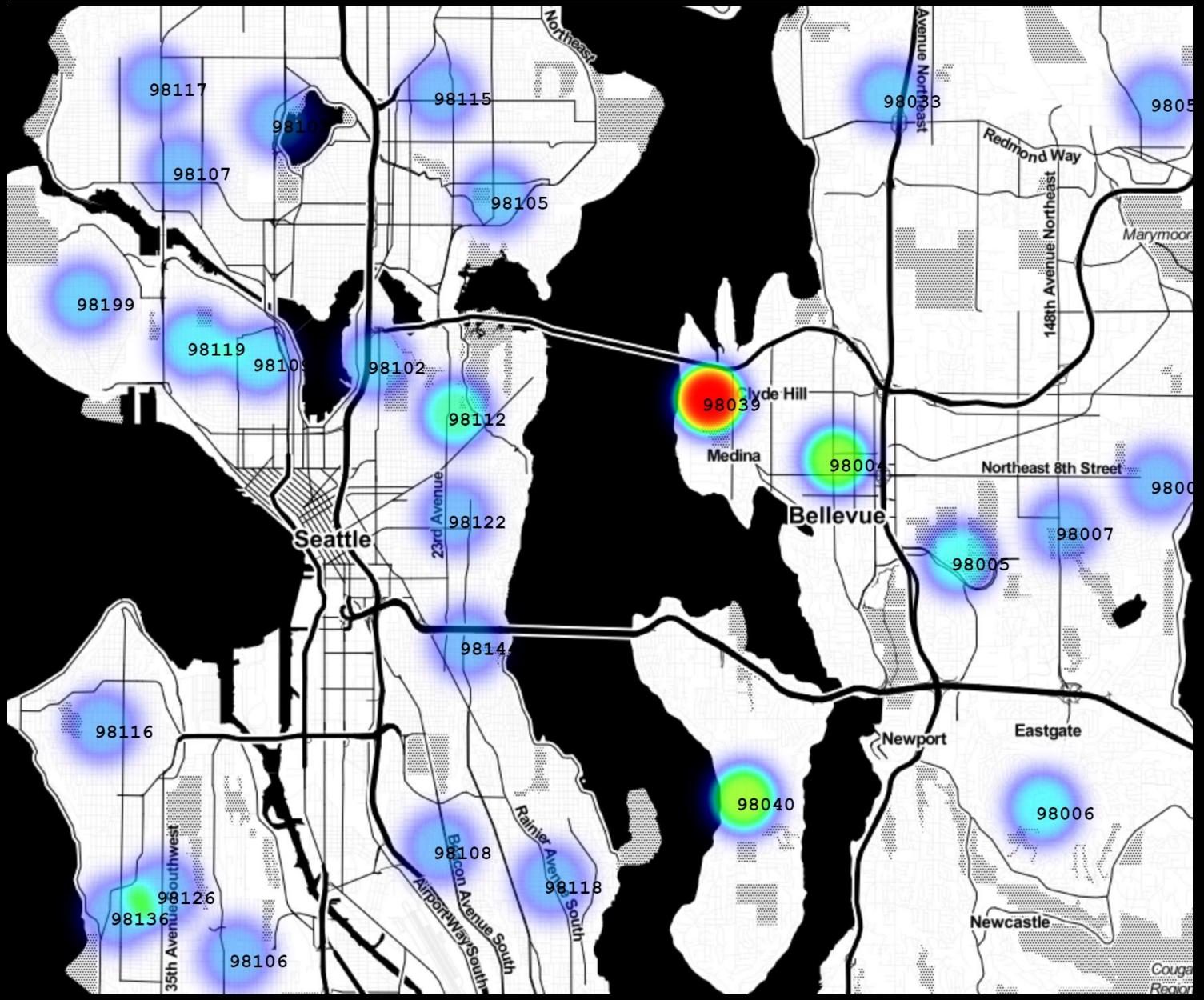
Results

Important parameters in the model, as:

- Waterfront properties;
- Centrally located;
- Zip code

proved increase prices significantly.




98117
98115
98033
98107
98105
98199
98119
98102
98039
98112
Medina
98004
Northeast 8th Street
Bellevue
98122
Seattle
98007
98005
98116
Newport
Eastgate
98040
98006
98108
98118
98126
98136
98106
Newcastle
Redmond Way
148th Avenue Northeast
23rd Avenue
Rainier Avenue South
Beacon Avenue South
35th Avenue Southwest

Technical Results

- R^2 final model

  0.826

- Train RMSE
  - $ 117,768.73

- Test RMSE
  - $ 117,182.99

Future Work

To better predict prices:

- Larger, recent, accurate data.
- More parameters.

Thank you!

Leticia Fernandes