Beachfront Bungalows

(Final Project)

Raj Ray

Introduction

- Beachfront Bungalows International – A new realty-services start-up
- Target Market – Tech-Savvy high net-worth individuals
- Services Offered (using Machine Learning and AI) -
  - Property Appraisal for Buyers and Sellers
  - House Price Valuation through Analysis of Existing Listings (Datasets)

Project Goal

o Live Demo - Home appraisal by means of automated object recognition on Smartphones using TensorFlow

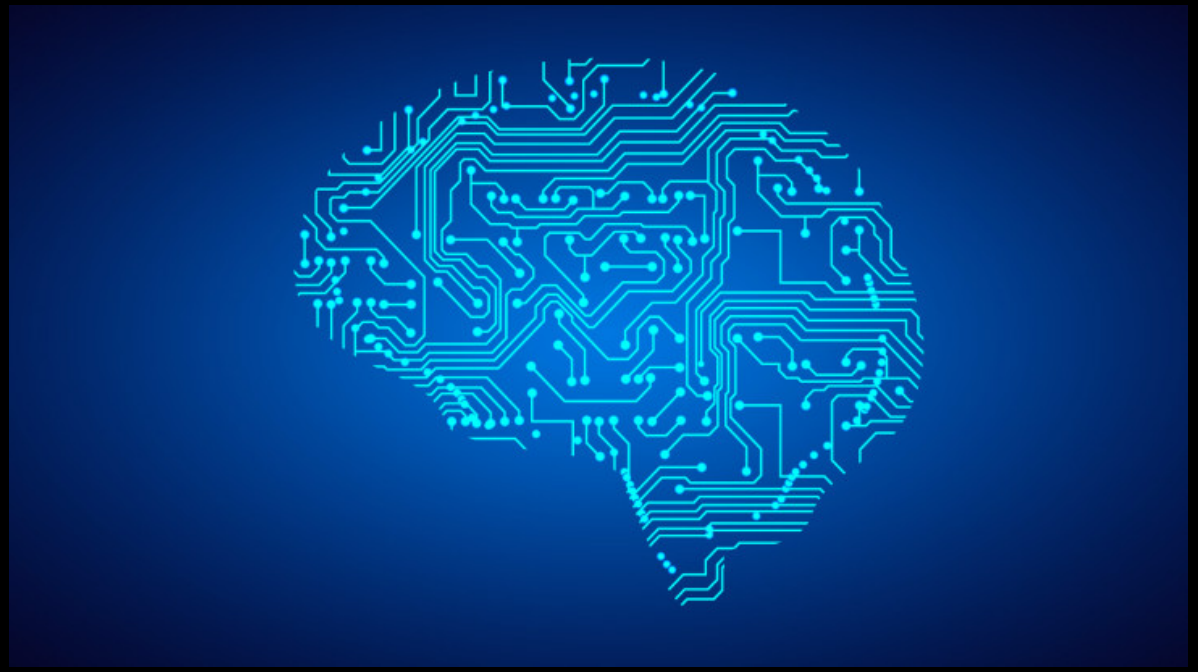
o Computational Analysis – Predictive Analyses of Home Prices using Machine Learning models

- Create, train and test multiple “Supervised ML Models” for regression analysis of past and existing listings
- Derive insights into the performance of different ML models by analyzing prediction accuracies for home prices
- Determine which ML model is best suited for house price valuation based on multiple parameters.

Demo

GitHub Link: https://github.com/tensorflow/tensorflow/tree/master/tensorflow/examples/android

Link to APK File: https://ci.tensorflow.org/view/Nightly/job/nightly-android/lastSuccessfulBuild/artifact/*zip*/archive.zip



Machine Learning Models

Models:

1. Linear Regression
2. Multivariate Regression
3. Decision trees
4. Random Forests
5. Support Vector Machines (SVM)
6. Lasso
7. Ridge
8. K- Nearest Neighbor (K-nn)
9. Gradient Descent Boosting
10. K-fold Grid-Search with Cross Validation

(Note: One-Hot Encoding to transform categorical parameters)

Performance Metric:

1. Correlations
2. Mean-squared error
3. Accuracy

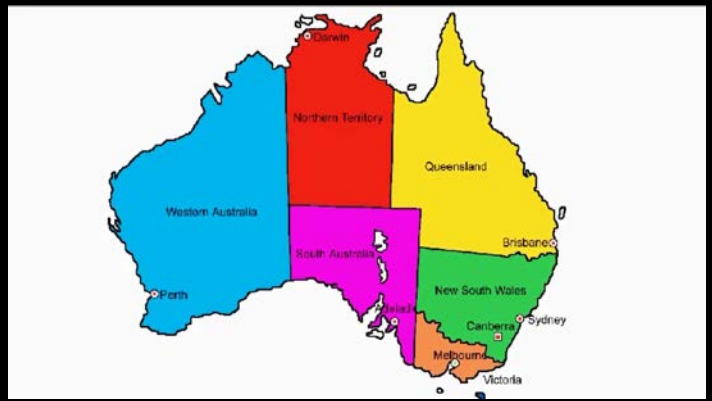
Core Phases:

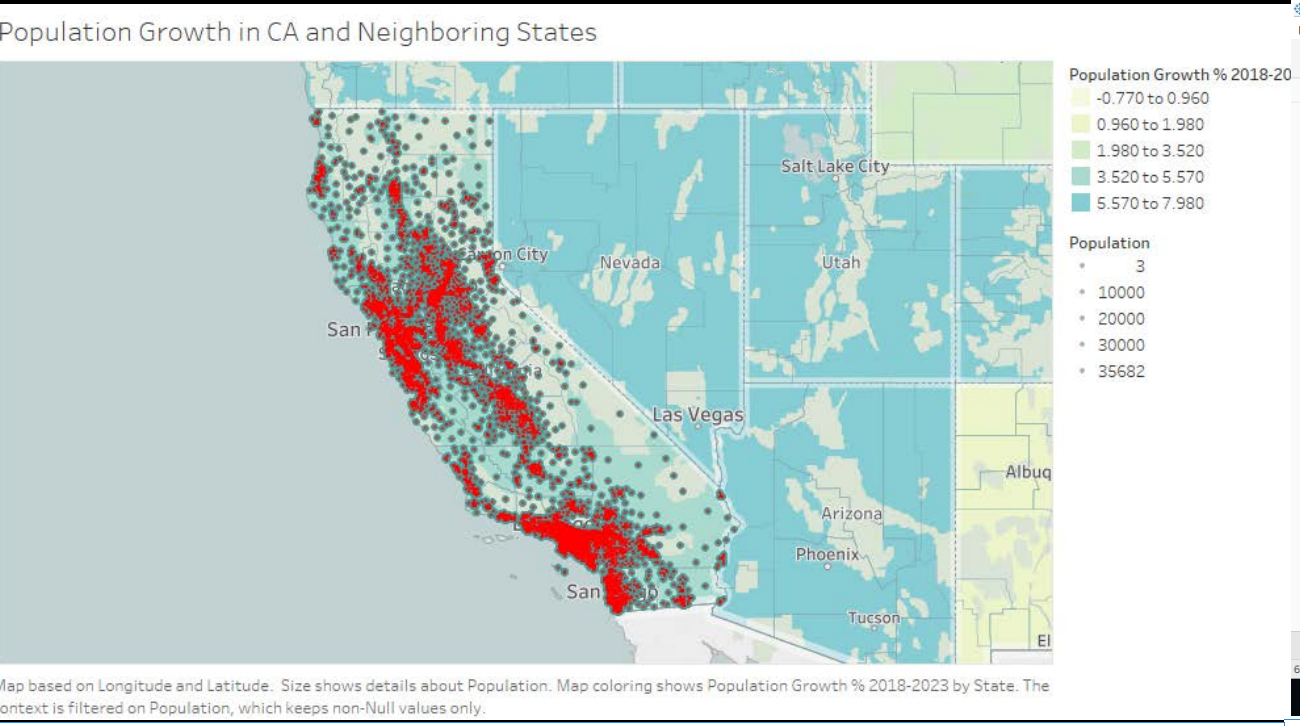
1. Data Input
2. Exploratory Analysis
3. Features and Labels
4. Data Preparation / Cleaning
5. Model Creation
6. Training and Testing
7. Model Fine-Tuning
8. Accuracy Evaluation

Datasets

- US:
  - Coastal Areas Housing Prices in California
  - Data from 1990 California census
  - 20,640 property listings
  - 10 variables each including longitude, latitude, total rooms, proximity to ocean, price etc.
  - Source: Paper published in Statistics & Probability Letters 33.3 (1997): 291-297)

- AUSTRALIA:
  - Coastal Region Housing Prices in Melbourne
  - Data scraped from publicly available weekly listings on www.domain.com.au
  - 14,242 property listings
  - 21 variables each including address, longitude, latitude, suburb, total rooms, bedrooms, bathrooms, lot-size, price, postal code etc.
  - Source – Kaggle.com




Population Growth in CA and Neighboring States
Population Growth % 2018-20
-0.770 to 0.960
0.960 to 1.980
1.980 to 3.520
3.520 to 5.570
5.570 to 7.980
Population
3
10000
20000
30000
35682
Salt Lake City
Nevada
Utah
Las Vegas
Arizona
Phoenix
Tucson
Albuq
Map based on Longitude and Latitude. Size shows details about Population. Map coloring shows Population Growth % 2018-2023 by State. The
ontext is filtered on Population, which keeps non-Null values only.

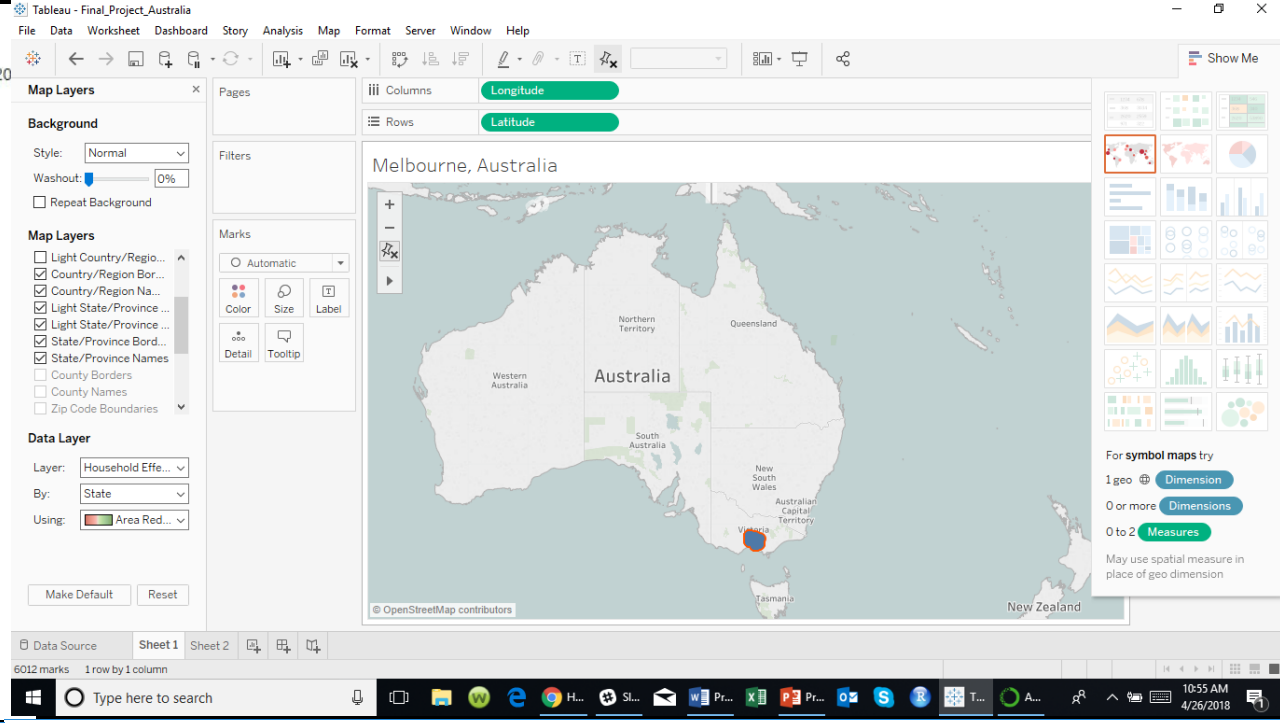


Tableau - Final_Project_Australia
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Map Layers
Background
Style: Normal
Washout: 0%
Repeat Background
Map Layers
Light Country/Regio...
Country/Region Bor...
Country/Region Na...
Light State/Province ...
Light State/Province ...
State/Province Bord...
State/Province Names
County Borders
County Names
Zip Code Boundaries
Data Layer
Layer: Household Effe...
By: State
Using: Area Red...
Make Default
Reset
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Columns
Longitude
Rows
Latitude
Melbourne, Australia
Northern Territory
Queensland
Western Australia
Australia
South Australia
New South Wales
Australian Capital Territory
Victoria
Tasmania
New Zealand
© OpenStreetMap contributors
For symbol maps try
1 geo Dimension
0 or more Dimensions
0 to 2 Measures
May use spatial measure in place of geo dimension
Data Source
Sheet 1
Sheet 2
6012 marks
1 row by 1 column
Type here to search
10:55 AM
4/26/2018

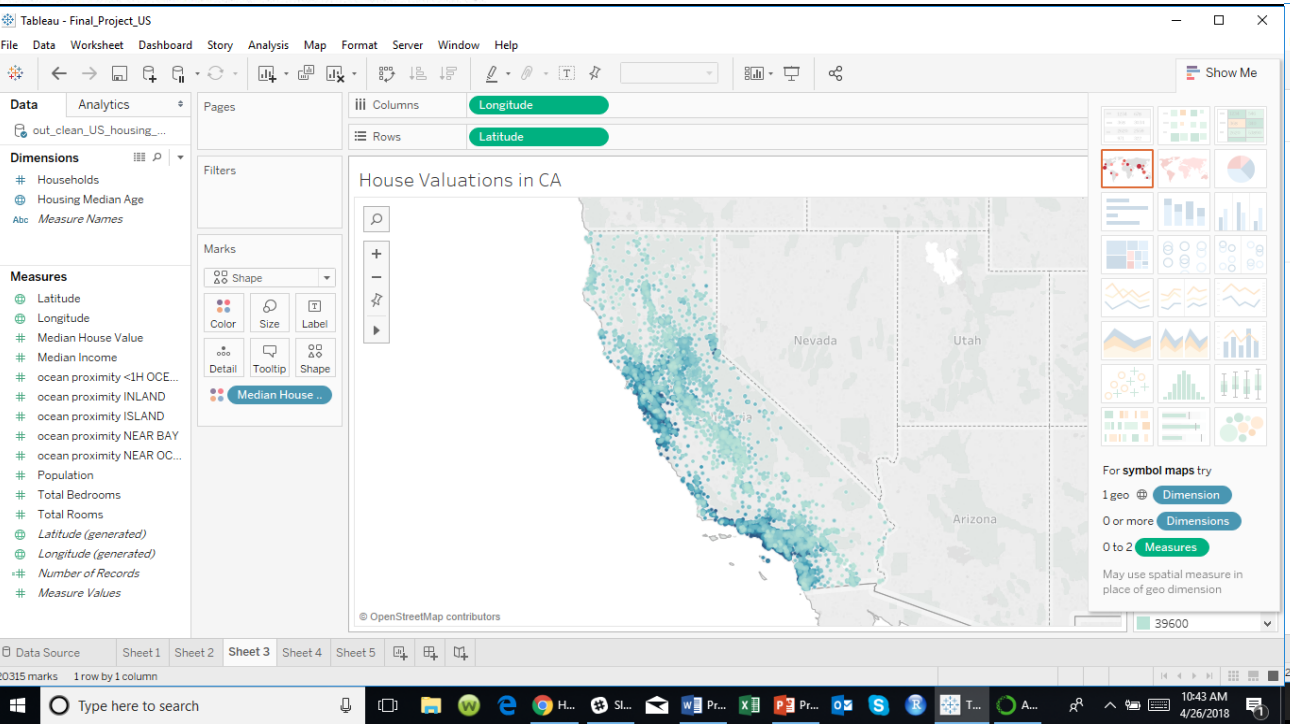


Tableau - Final_Project_US
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
out_clean_US_housing_...
Dimensions
Households
Housing Median Age
Measure Names
Measures
Latitude
Longitude
Median House Value
Median Income
ocean proximity <1H OCE...
ocean proximity INLAND
ocean proximity ISLAND
ocean proximity NEAR BAY
ocean proximity NEAR OC...
Population
Total Bedrooms
Total Rooms
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Shape
Color
Size
Label
Detail
Tooltip
Shape
Median House ..
Columns
Longitude
Rows
Latitude
House Valuations in CA
Nevada
Utah
Arizona
© OpenStreetMap contributors
39600
For symbol maps try
1 geo Dimension
0 or more Dimensions
0 to 2 Measures
May use spatial measure in place of geo dimension
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 4
Sheet 5
20315 marks
1 row by 1 column
Type here to search
10:43 AM
4/26/2018

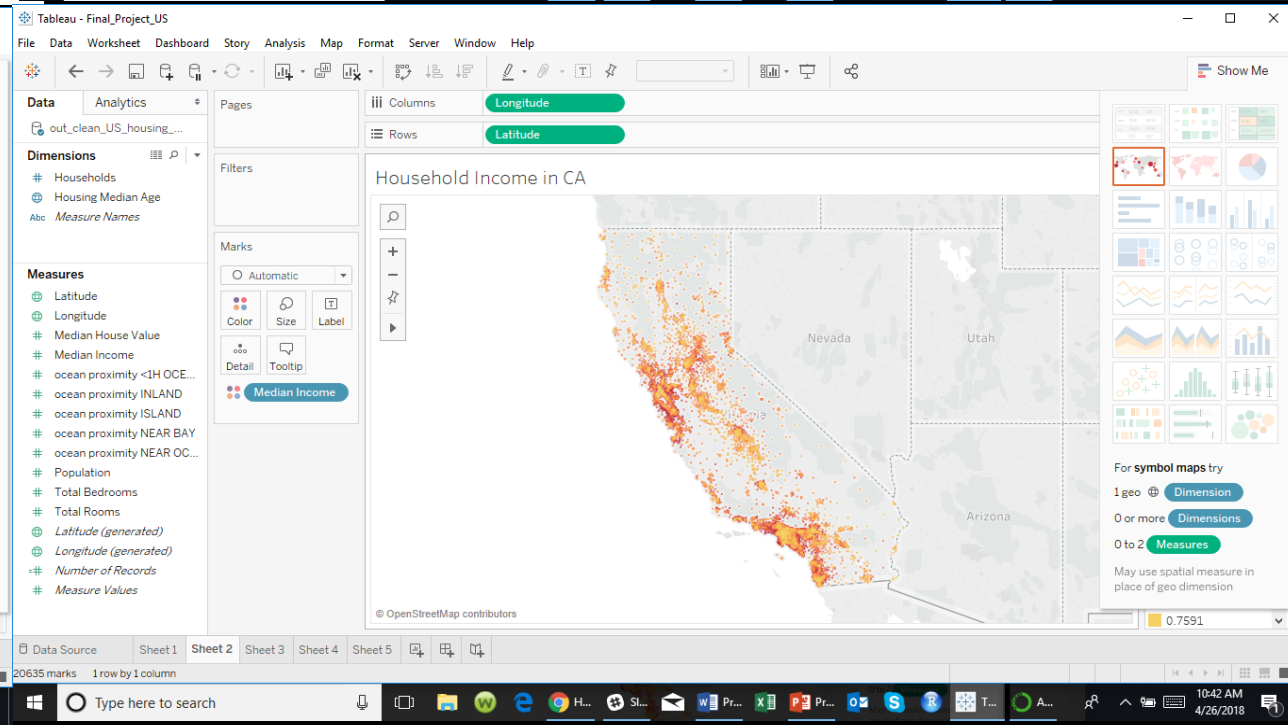


Tableau - Final_Project_US
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
out_clean_US_housing_...
Dimensions
Households
Housing Median Age
Measure Names
Measures
Latitude
Longitude
Median House Value
Median Income
ocean proximity <1H OCE...
ocean proximity INLAND
ocean proximity ISLAND
ocean proximity NEAR BAY
ocean proximity NEAR OC...
Population
Total Bedrooms
Total Rooms
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Median Income
Columns
Longitude
Rows
Latitude
Household Income in CA
Nevada
Utah
Arizona
© OpenStreetMap contributors
0.7591
For symbol maps try
1 geo Dimension
0 or more Dimensions
0 to 2 Measures
May use spatial measure in place of geo dimension
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 4
Sheet 5
20635 marks
1 row by 1 column
Type here to search
10:42 AM
4/26/2018

Tableau - Final_Project_Australia

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Show Me

Data | Analytics

out_clean_Aussie_housi...

Dimensions

- Abc Council Area
- Postcode
- Abc Suburb
- Abc Type
- # Year Built
- Abc *Measure Names*

Measures

- # Bathroom
- # Bedrooms
- # Building Area
- # Car
- # Distance
- Latitude
- Longitude
- # Lotsize
- # Price
- # Rooms
- *Latitude (generated)*
- *Longitude (generated)*
- *Number of Records*
- # *Measure Values*

Pages

Filters

Marks

Shape

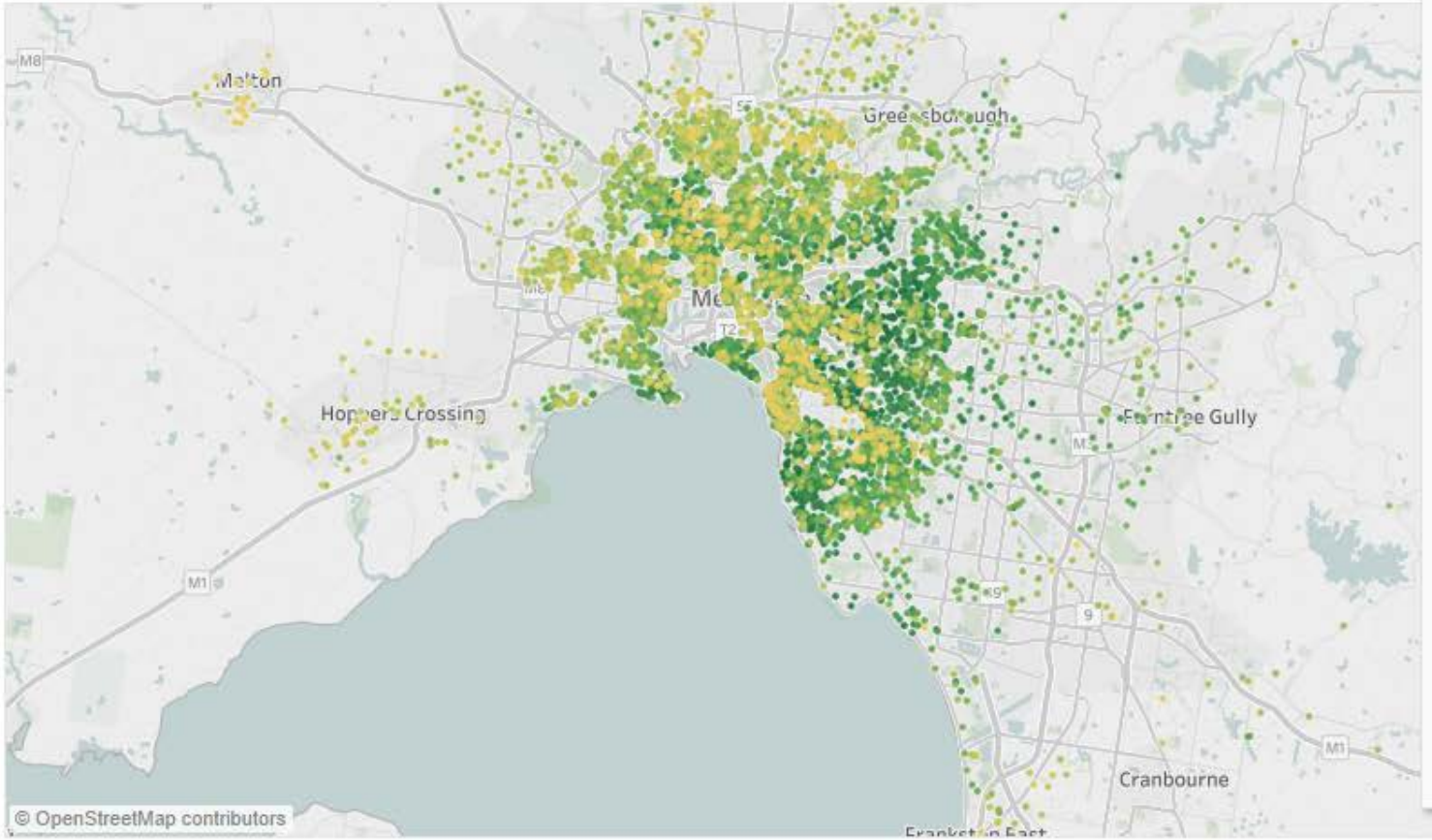
Color | Size | Label | Detail | Tooltip | Shape

Price

Columns: Longitude

Rows: Latitude

High-Priced Homes in Melbourne



For **symbol maps** try

1 geo Dimension

0 or more Dimensions

0 to 2 Measures

May use spatial measure in place of geo dimension

257500

Data Source | Sheet 1 | Sheet 2 | Sheet 3

6179 marks 1 row by 1 column

Type here to search

11:18 AM
4/26/2018

Tableau - Final_Project_Australia

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Show Me

Map Layers

Background

Style: Dark

Washout: 0%

Repeat Background

Map Layers

- [x] Base
- [x] Land Cover
- [ ] Coastline
- [ ] Streets and Highways
- [ ] Light Country/Regio...
- [ ] Light Country/Regio...
- [x] Country/Region Bor...
- [x] Country/Region Na...
- [ ] Light State/Province ...

Data Layer

Layer: Households

By: State

Using: Blue-Gre...

Make Default Reset

Pages

Filters

Marks

Automatic

Color Size Label

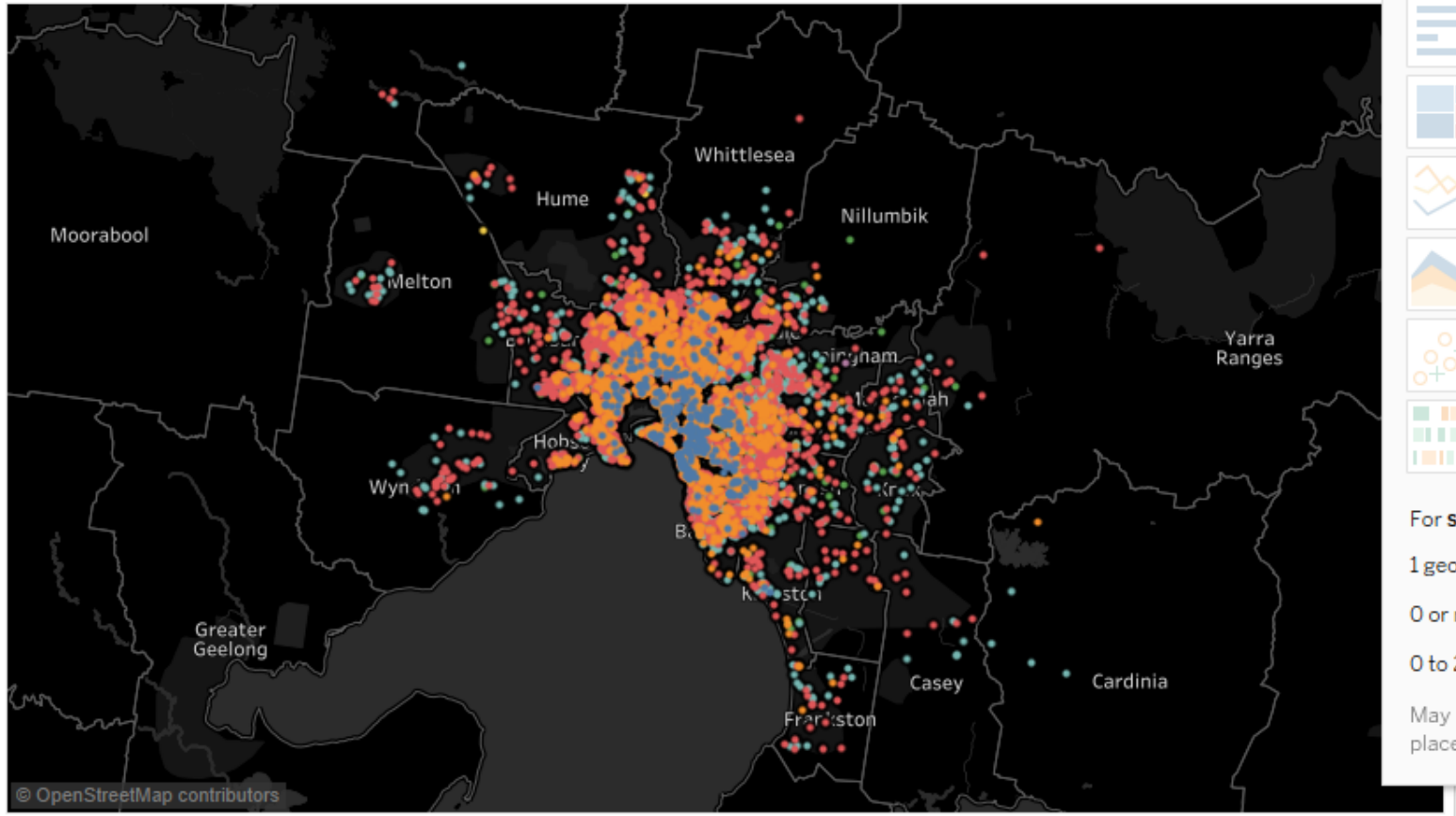
Detail Tooltip

Rooms

Columns: Longitude

Rows: Latitude

# of Rooms per Household



For **symbol maps** try

1 geo Dimension

0 or more Dimensions

0 to 2 Measures

May use spatial measure in place of geo dimension

Data Source | Sheet 1 | Sheet 2 | Sheet 3

052 marks 1 row by 1 column

Type here to search

11:16 AM
4/26/2018

Results

ML Model	Mean Squared Error	Accuracy (R2)
Linear Regression	0.4503	0.549
Multivariate Regression	2.17e-13	1.0
	0.0156	0.998
Random Forests	0.0020	0.999
Support Vector Machines	0.8227	0.323
Lasso	0.0008	0.999
Ridge	0.0019	0.998
K-nn	0.81	1.0
Gradient Descent with Boosting	0.1668	.78
K-Fold Grid-Search with Cross-Validation	0.0216	1.0

Conclusions

- Key parameters affecting house price valuation?
  - Location (Zipcode / postal code, proximity to beach)
- Main challenges of machine learning?
  - Large unclean datasets and computational time - Needs GPU or Cloud
- How to improve accuracy and reduce overfitting?
  - Cross-Validation, Lasso / Ridge Regression (Regularization), Scaling, Boosting
- Does overall performance depend on feature selection?
  - Correlated Features and One-Hot Encoding. Drop unrelevant / uncorrelated features
- Effects of k-fold Cross-Validation?
  - Improves over-fitting, Effectively uses small data-sets
- Did certain algorithms perform poorly on both datasets? Why?
  - Yes, Linear Regression because most natural processes tend to be non-linear
- Which supervised models are better for predictive analysis?
  - Yes, Multivariate Regression, Lasso, Ridge, Random Forests performed well.


Thank You