Identifying Real Estate Opportunities

In San Francisco

Through Big Data Analysis

Methodology

Establish Business Case	EDA and modelling	Shortlist Opportunities
- Are we at the peak of a RE cycle? - Even if you get a "bargain" property, will the market crash as a whole? - What are the rental yields like?	Setting the context to identify "bargains" - What does our data set look like? - What should I further study? - What variables should I use to estimate value?	Fine tuning and putting the model to use - Assessing coefficients (ridge, lasso) - Testing assumptions - Translate model into practical insights

Establish Business Case

Real Estate Prices & Returns:

Peer Comparison against other cities

Most Expensive Homes in The World

City	90 sqm apartment price	price to household income	Gross Rental Yield	Price To Rent Ratio City Centre
Hong Kong, Hong Kong	$2,765,065	47.5	1.8%	56.5
Singapore, Singapore	$1,788,108	22.3	2.3%	43.0
London, United Kingdom	$1,544,135	21.2	2.9%	34.7
Seoul, South Korea	$1,409,990	24.0	1.4%	72.8
Beijing, China	$1,352,845	44.2	1.7%	60.4
New York, NY, United States	$1,325,715	10.8	4.9%	20.4
Shenzhen, China	$1,239,971	44.9	1.3%	77.6
Shanghai, China	$1,217,843	41.5	2.0%	50.8
Geneva, Switzerland	$1,191,403	10.5	3.4%	29.4
Taipei, Taiwan	$1,183,366	33.1	1.0%	96.7
San Francisco, CA, United States	$1,170,781	7.8	5.9%	17.1
Zurich, Switzerland	$1,134,909	8.2	3.3%	30.6

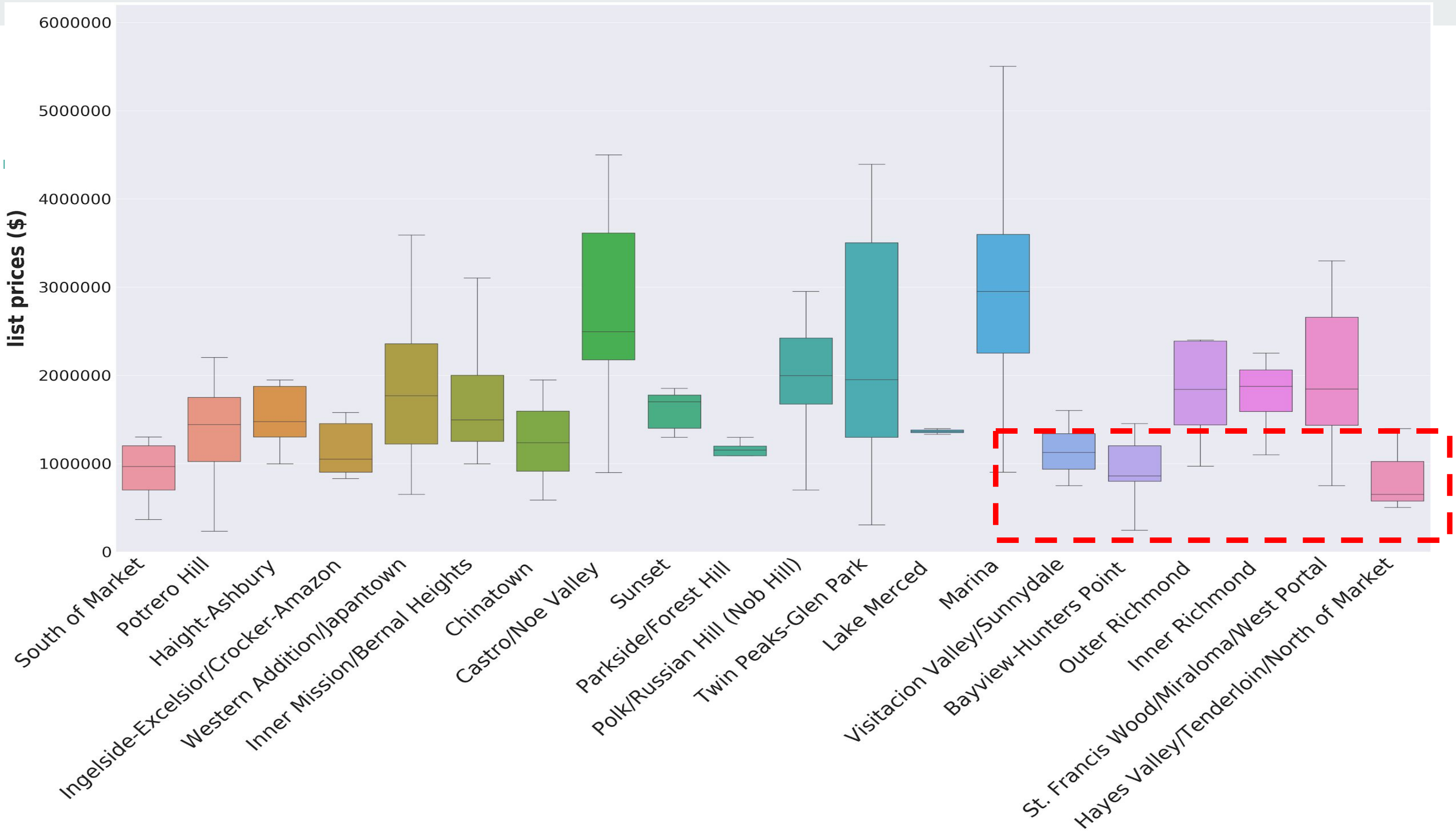
Most Expensive Homes in The World

City	90 sqm apartment price	price to household income	Gross Rental Yield	Price To Rent Ratio City Centre
Hong Kong, Hong Kong	$2,765,065	47.5	1.8%	56.5
Singapore, Singapore	$1,788,108	22.3	2.3%	43.0
London, United Kingdom	$1,544,135	21.2	2.9%	34.7
Seoul, South Korea	$1,409,990	24.0	1.4%	72.8
Beijing, China	$1,352,845	44.2	1.7%	60.4
New York, NY, United States	$1,325,715	10.8	4.9%	20.4
Shenzhen, China	$1,239,971	44.9	1.3%	77.6
Shanghai, China	$1,217,843	41.5	2.0%	50.8
Geneva, Switzerland	$1,191,403	10.5	3.4%	29.4
Taipei, Taiwan	$1,183,366	33.1	1.0%	96.7
San Francisco, CA, United States	$1,170,781	7.8	5.9%	17.1
Zurich, Switzerland	$1,134,909	8.2	3.3%	30.6

Comparably valued in terms of price, potential upside in terms of price to income

EDA and Modelling

Understanding the data and its context to model effectively

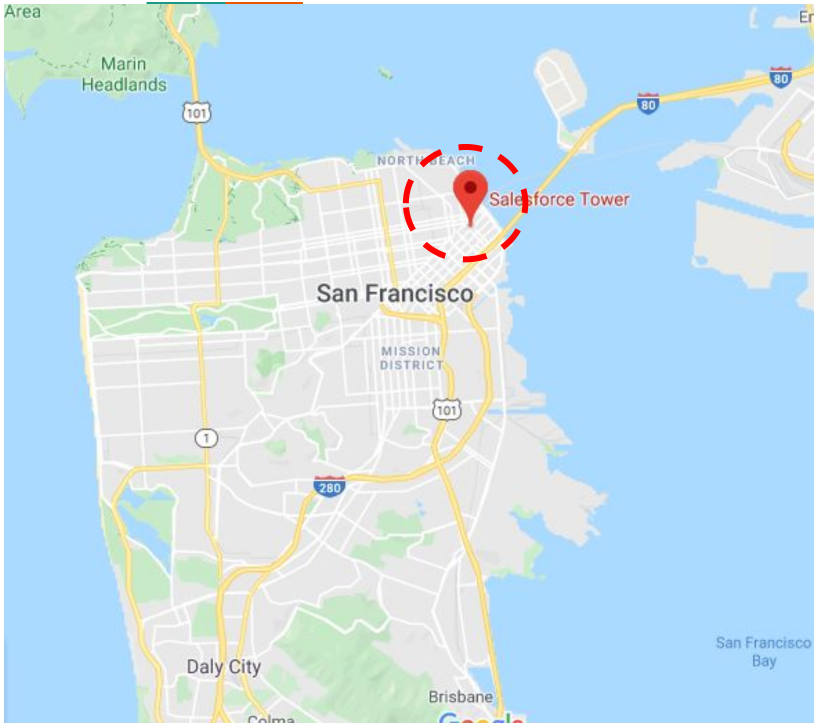


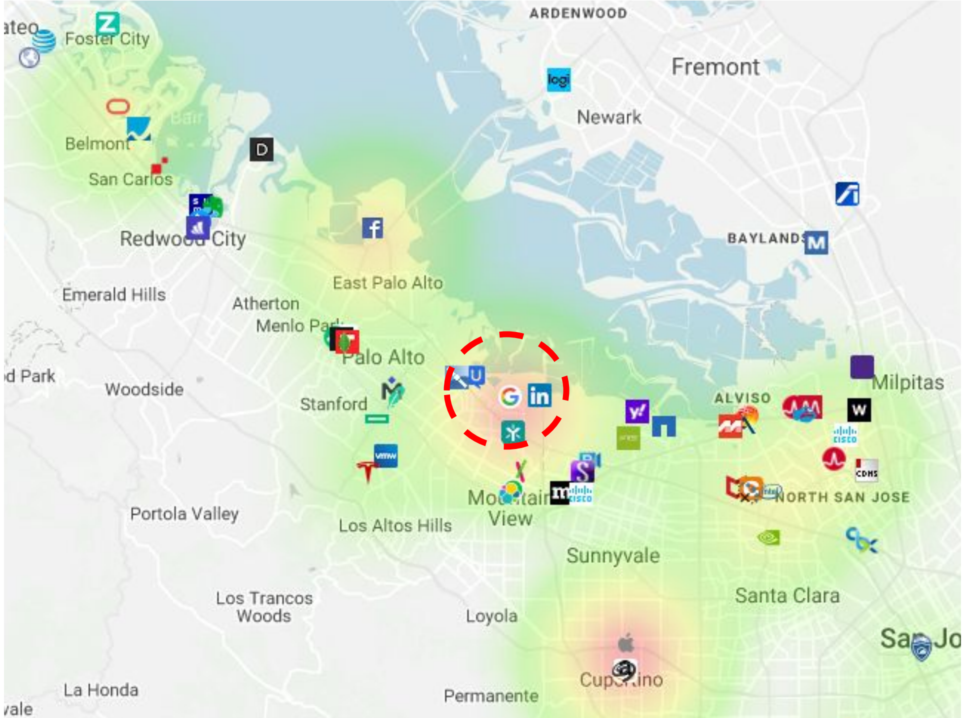
More Data Needed to Determine Best Value for Money

Factors to Home Prices

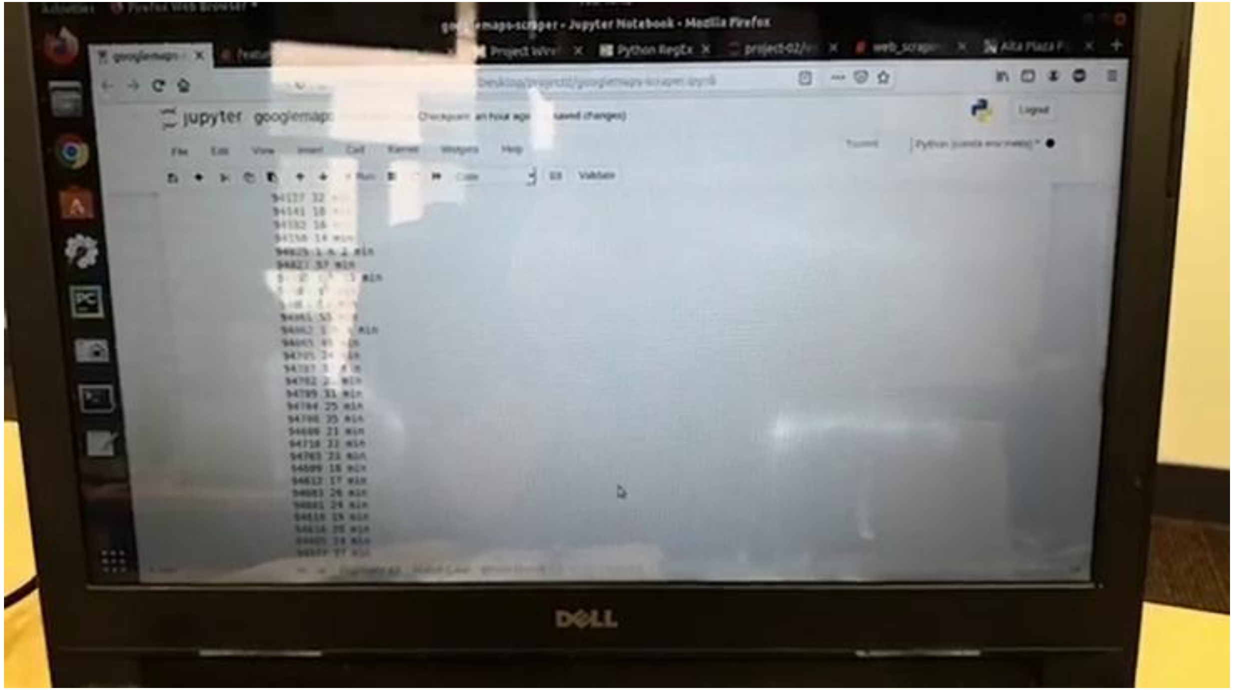
External Factors		Internal or Direct Factors	
Factor	**Measurable**	**Factor**	**Measurable**
Distance to Centers of Commerce	**Travel times**	**Home size**	**Square Feet listed**
Crime	Crime Statistics	**Age**	**Year Built /Refurbished**
Closeness to amenities	Walk Scores	**Layout**	**1BR, 2BR, # bathrooms**
Neighborhood Quality	Zip code household income	Internal Finishings	Descriptors (NLP), pictures (IP)

Computing Distance to Centers of Commerce





Computing Distance to Centers of Commerce



~250 Iterations, then
Lowest of Travel TImes Between SalesForce Tower and Googleplex Used

Assessing Model Fit

Data set size: ~1600 rows
Variables: Travel times, year built, home size, layout (1BR, 2BR, #bathrooms)
Methodology: Ordinary least square regression applied, standardize data, ridge and lasso used to drop variables

Dep. Variable:	PRICE	**R-squared (uncentered):**	0.724
Model:	OLS	**Adj. R-squared (uncentered):**	0.723
Method:	Least Squares	**F-statistic:**	632.5
Date:	Fri, 24 Jan 2020	**Prob (F-statistic):**	0.00
Time:	07:59:13	**Log-Likelihood:**	-18568.
No. Observations:	1211	**AIC:**	3.715e+04
Df Residuals:	1206	**BIC:**	3.717e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
SQUARE FEET	893.4146	44.149	20.236	0.000	806.798	980.032
YEAR BUILT	183.2160	53.232	3.442	0.001	78.778	287.654
shortesttime	-1.815e+04	2609.236	-6.955	0.000	-2.33e+04	-1.3e+04
BEDS	-1.617e+05	3.35e+04	-4.828	0.000	-2.27e+05	-9.6e+04
BATHS	1.51e+05	5.41e+04	2.791	0.005	4.49e+04	2.57e+05

Omnibus:	1458.122	**Durbin-Watson:**	1.950
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	422354.801
Skew:	5.743	**Prob(JB):**	0.00
Kurtosis:	93.766	**Cond. No.**	5.14e+03

Test Set

R^2 = ~0.488

Shortlist Opportunities

Refining the data to provide sharper insights

Shortlisted Opportunities

ADDRESS	CITY	PRICE **(less than $2M)**	SQUARE FEET	$/psf
2641 Yuba St	El Cerrito	$899,000	8360	349.54
2637 E 16th St	Oakland	$850,000	4192	202.77
1725 Estudillo Ave	San Leandro	$1,449,000	4750	305.05
1112 CHUAUCER #2	Berkeley	$1,499,000	4800	312.29
915 Grosvenor Pl	Oakland	$1,250,000	4163	300.26
1225 VIENNA Dr #976	SUNNYVALE	$525,000	2600	201.92
1985 Tunnel Rd	Berkeley	$1,495,000	4083	366.15

Mobile Home

On Redfin for 4 days,
viewed 1,766 times
Redfin Estimate: $1,016,985

Redfin Estimate: 1,593,751

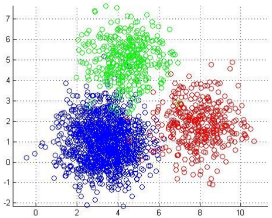
Further Studies

Time series Analysis

- How do prices evolve over time? Do we observe any neighborhoods with high price increases?
- How has demand shifted over time? More need for studios?

Expand Scope

- Include additional parameters like household income, crime rates, school quality, walkability etc.
- Compare trends across cities (LA, NY etc.) - How do we see variables shifting?



Greater Statistical Analysis

- E.g K means clustering - do positive attributes have a tendency to cluster? Do negative attributes compound?
- Points to the effect of “market making”

Acknowledgements

Redfin
target_companies_list.remove('Zillow')
Target_companies_list += ['Redfin']

Michael Boles (ex Metis Student!)
https://towardsdatascience.com/@michaeladamboles

John Joo
https://blog.dominodatalab.com/exploring-us-real-estate-values-with-python/

Stack Overflow and Google

Disclaimer

Slides expresses my own opinions. I have no business relationship with any company which is mentioned in this article.

The information contained herein is my personal opinion and does not constitute an investment advice, an offer or solicitation to subscribe for, purchase or sell the homes/investment product(s)/companies mentioned herein.