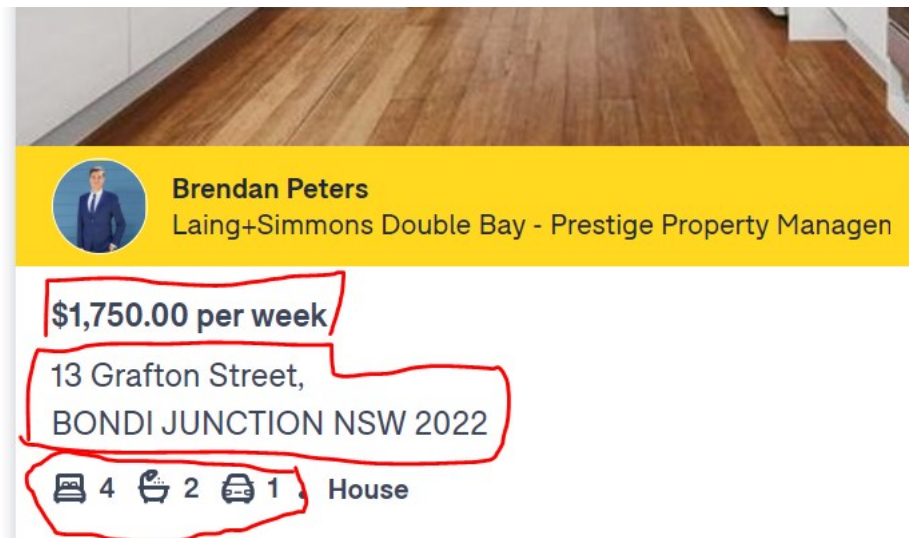


## Taking Geodata to the Next Level: Machine Learning & distance to the beach for predicting Sydney House Renting Prices

Hi , Last few days I being browsing many real estate webpages because I am looking for a new place in Sydney. I wanted to automate this process and see how each of the apartment features impact in its price, because I am a beach lover I also wanted to see which is the distance of the apartment to the sea and how that feature impacts on the renting price

For generating my Dataset I have web scrapped 5000 apartment listings from [www.domain.com.au](http://www.domain.com.au) and extracted their price and their features



After doing some Data Cleaning we end up with this Data Frame:

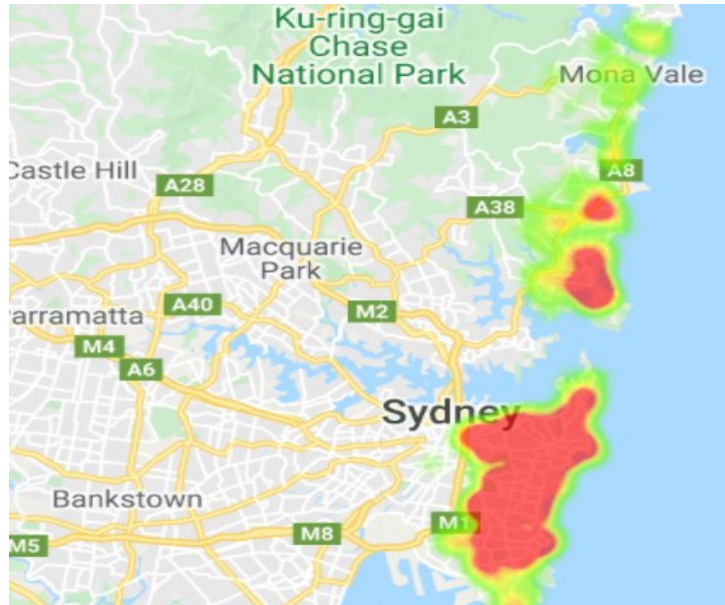
price	weekly_rent	classs	bedrooms	bathrooms	parkings	adress	neigh
\$1900 Per week	1900.0	House	3.0	3.0	1.0	1 Carlton street, MANLY NSW 2095	MANLY NSW 2095
\$1290 per week	1290.0	Apartment / Unit / Flat	2.0	2.0	1.0	49 North Steyne, MANLY NSW 2095	MANLY NSW 2095
\$1,250 pw	1250.0	Apartment / Unit / Flat	2.0	2.0	1.0	2 Denison Street, MANLY NSW 2095	MANLY NSW 2095
		Apartment / Unit /				27 The Crescent, MANLY NSW	

But in order to get the distance of each of the 5000 apartments to the sea we would need to translate the address into Geocoordinates and then somehow calculate which is the distance from those address to the sea.

With Google Maps API we can get the Geocoordinates of each address:

weekly_rent	classs	bedrooms	bathrooms	parkings	adress	neigh	lat	lon	point
1900.0	House	3.0	3.0	1.0	1 Carlton street, MANLY NSW 2095	MANLY NSW 2095	-33.792945	151.286346	POINT (151.2863456 -33.7929453)
1290.0	Apartment / Unit / Flat	2.0	2.0	1.0	49 North Steyne, MANLY NSW 2095	MANLY NSW 2095	-33.796189	151.287348	POINT (151.2873478 -33.7961886)
1250.0	Apartment / Unit / Flat	2.0	2.0	1.0	2 Denison Street, MANLY NSW 2095	MANLY NSW 2095	-33.794231	151.286388	POINT (151.2863879 -33.7942311)

Now that we have the Geocoordinates we can plot the apartments using google maps heatmap

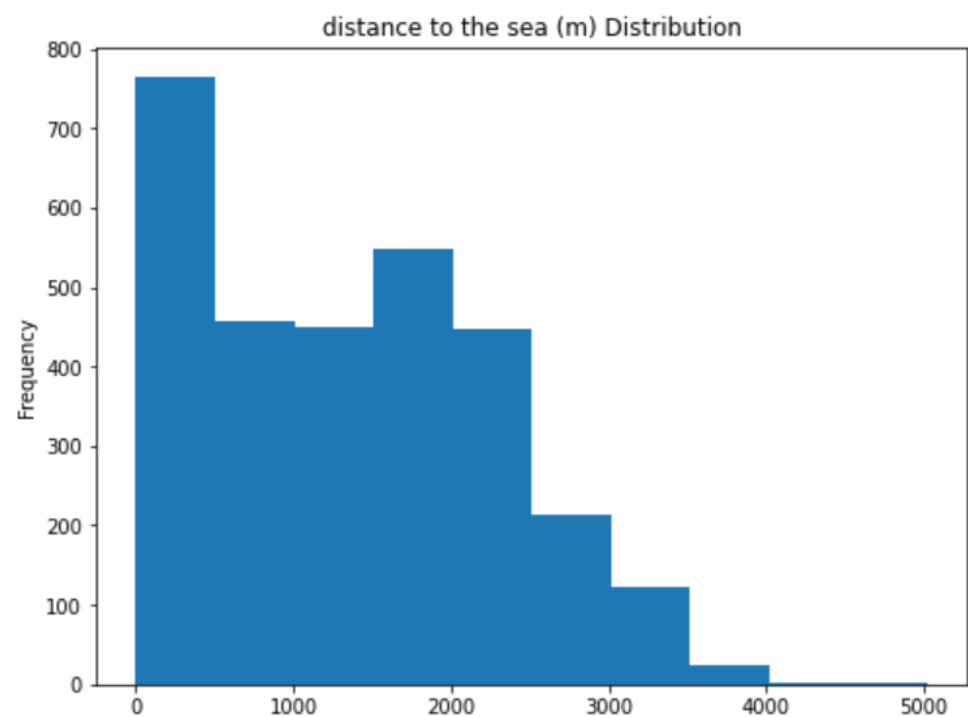


Now comes the harder part as we need to compute the distance of each apartment to the sea in meters, for doing so we would need to download NSW coastline geocoordinates and compute the closest distance of each point to the coast.

In NSW Data Spatial Catalogue webpage we can download NSW coastline as a shapefile and use it to calculate the distance to the sea. After some complex calculation (which are all detailed in my Jupyter notebook) we end up with this nice DataFrame:

	class	adress	postcode	bedrooms	bathrooms	parkings	distance_m	weekly_rent
	House	1 Carlton street, MANLY NSW	2095	3.0	3.0	1.0	63.947924	1900.0
	Apartment / Unit / Flat	49 North Steyne, MANLY NSW	2095	2.0	2.0	1.0	35.458528	1290.0
	Apartment / Unit / Flat	2 Denison Street, MANLY NSW	2095	2.0	2.0	1.0	79.877208	1250.0
	Apartment / Unit / Flat	37 The Crescent, MANLY NSW	2095	3.0	1.0	1.0	91.949958	1150.0

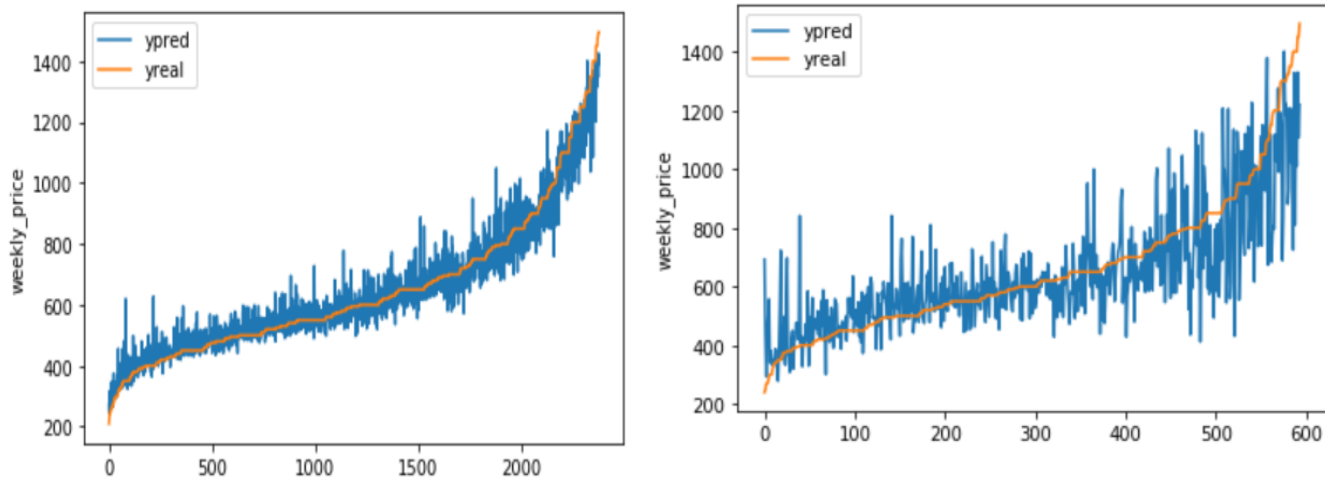
And this is the Distribution of the Distance to the Sea



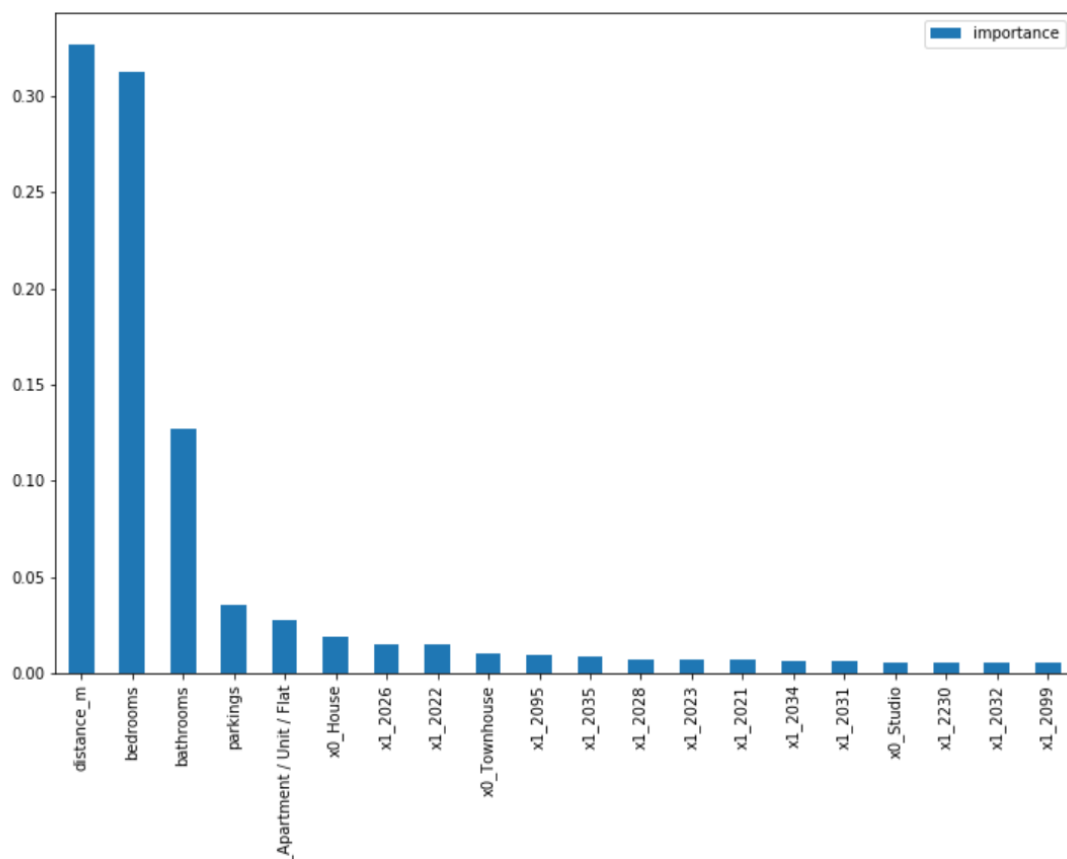
Now that we have all of our data neat and tidy we can use Machine Learning to build the model. The model I have chosen is **Random Regression Forest**. We would first have to do some feature engineering to handle categorical values and to normalize our data. After splitting our data into train and test set we can evaluate our model:

Train Score : 0.9417669706898879

Test Score : 0.8507742838060278



Although results in the Test set are a little bit lower , both sets yield really good results. As a Final step we can see how each individual feature impacts in the apartment renting Price and which are the most relevant in determining an apartment price.



Wow! As we can see distance to the sea is the most relevant feature in determining the renting price of a Sydney Apartment, followed by the N of rooms and bathrooms (The other variables are the POSTCODE and the apartment class)

The Jupyter notebooks are available on my Github webpage : <https://github.com/santiagorush/JupyterScripts> . If you have any questions do not hesitate to reach out.