1. **Introduction**

Every day many people are looking for a house or apartment to rent. They try to find a good deal by searching for each one at home or in specialized companies.

It takes a lot of time to do this search.

To support these people I tried to create a predictive model to identify the **total rent amount** based on some characteristics of the property.

In addition to trying to predict the total rent, we can answer some questions:

- Is the total rent amount in line with other similar properties?

- What characteristics should I disconsider in order to lower the total rent amount?

- If I add, for example, an additional bathroom what happens to the total rent amount?

2. **Data**

I will use a dataset from Kaggle (https://www.kaggle.com/rubenssjr/brasilian-houses-to-rent).

This dataset have 6079 houses to rent with 13 features:

- city: anonimyzed city where the property is located;
- area: property area (m^2);
- rooms: quantity of rooms;
- bathroom: quantity of bathroom;
- parking spaces: quantity of parking spaces;
- floor: floor;
- animal: accept animals?;
- furniture: furniture?;
- hoa (A): homeowners association tax;
- rent amount (B): rent amount;
- property tax (C): property tax;
- fire insurance (D): fire insurance;
- total: total rent (A + B + C + D)

3. **Methodology**

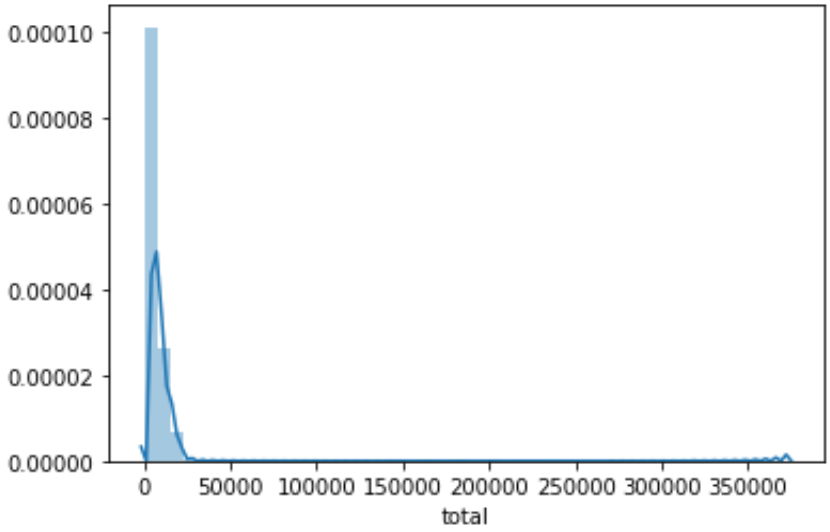
The first step is the data exploration to understand the data, checking missing data, data balance and adjusting some features for better analysis.

Next I will verify if exists any relationship among features and total.

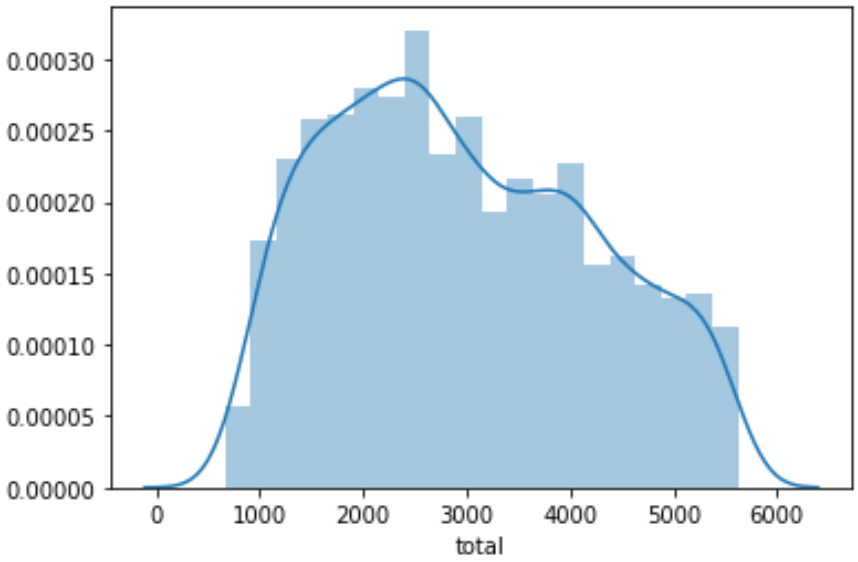
Then I will try predictive models to identify total rent amount and identify features that more influences it and the correlations.

4. **Results and discussion**

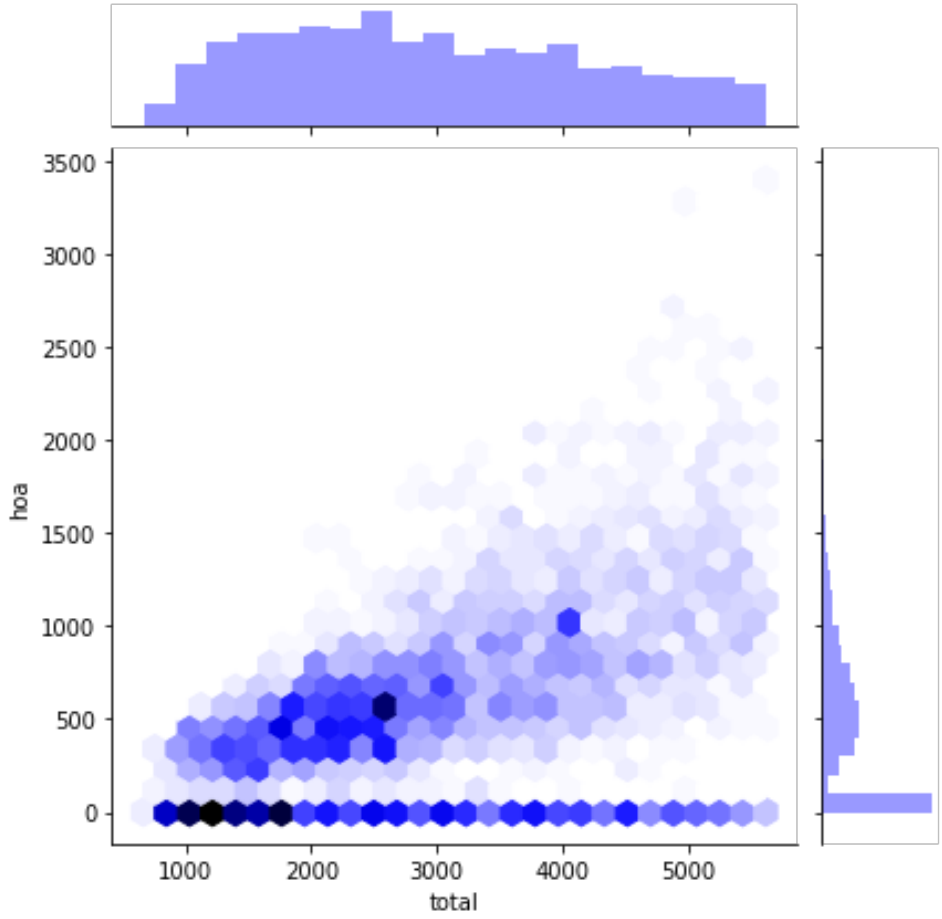
After data analysis I detect the data has no missing data and the balance is not OK



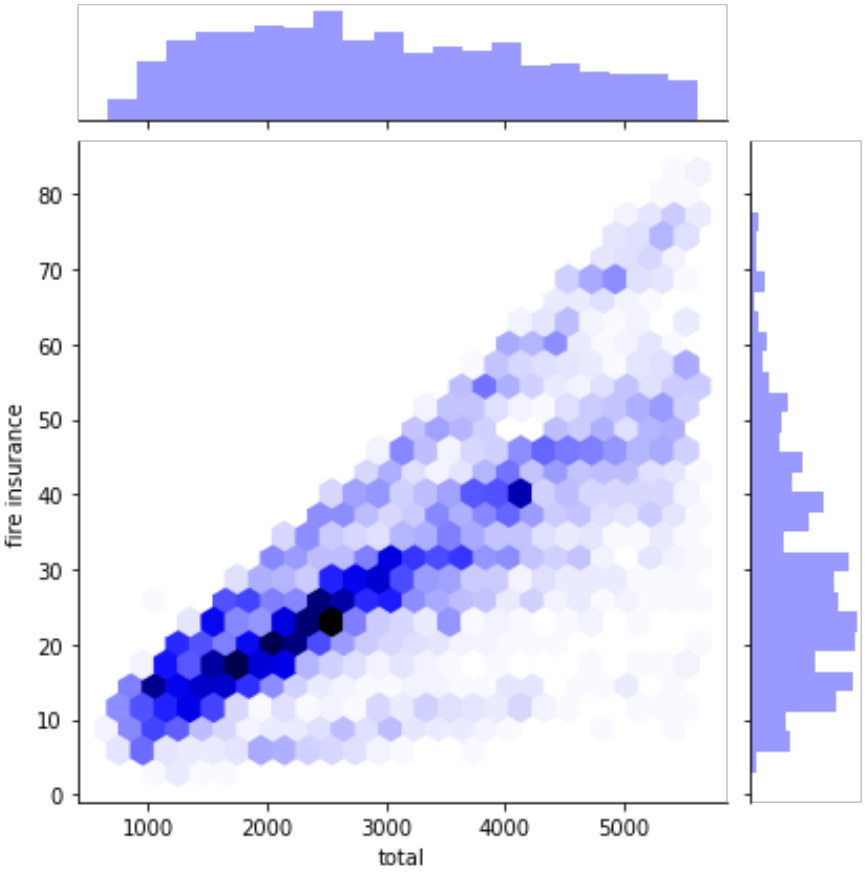
So, I remove outliers and normalize data:



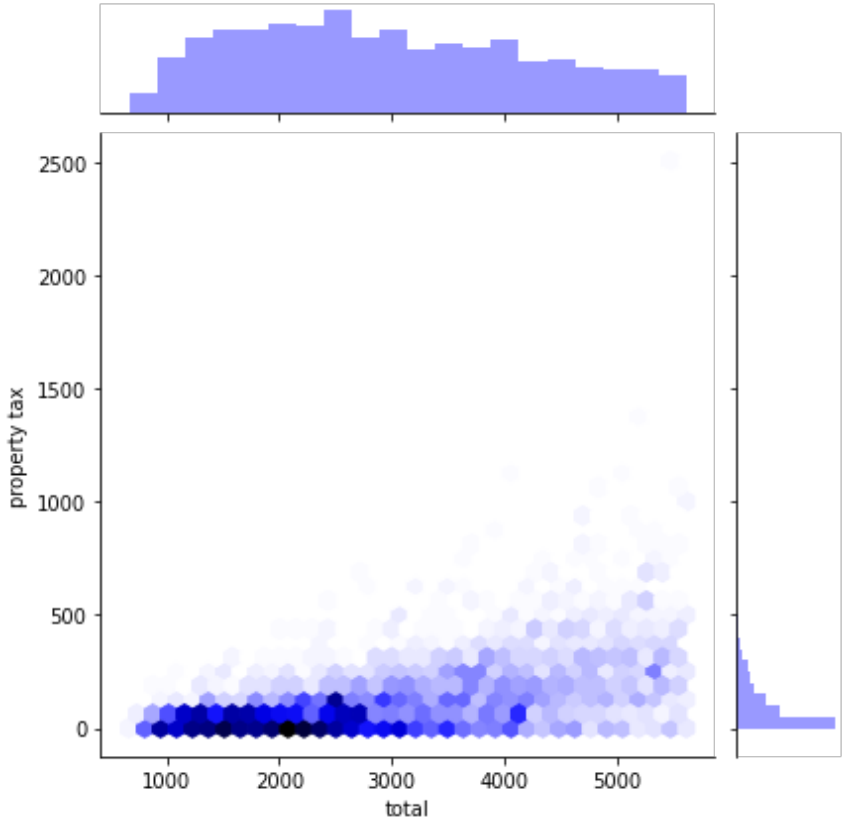
Then I verify the relationship among some features that compose the total feature.



total x hoa



total x fire insurance



total x property tax

Then I try to predict the **total rent amount** with some regressors like KNN, Decision Tree, Extra Tree and Randon Forest.

5. Conclusion

The regression's accuracy table is:

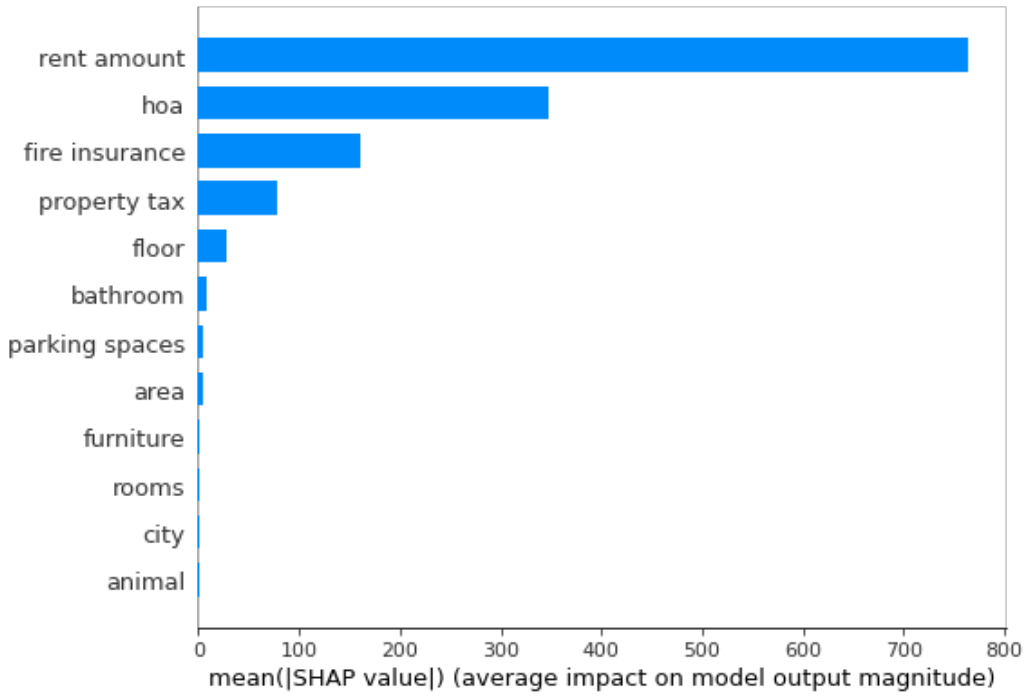
Model	MAE
KNN Regressor (k=5)	211.2968831168831
Decicion Tree Regressor (n=15)	73.74767470624613
Extra Tree Regression	105.2414730375486
Random Forest Regression	42.4924805194805

With Random Forest Regression I have better results and using SHAP I can demonstrate that some features influences the total rent amount.

I understand some features that compose the total rent amount have more influence on it.

We can see the “floor”, “bathroom”, “parking spaces” and “area” influence. So these house’s characteristcs have a good influence into total rent amound and must be considered by tenant.

Below we can see features importance in Random Forest Regression:



I create the correlation features heatmap, too.