

PREDICTING DAMAGE DONE BY NEPAL EARTHQUAKE--



RESEARCH PROSPECTS -

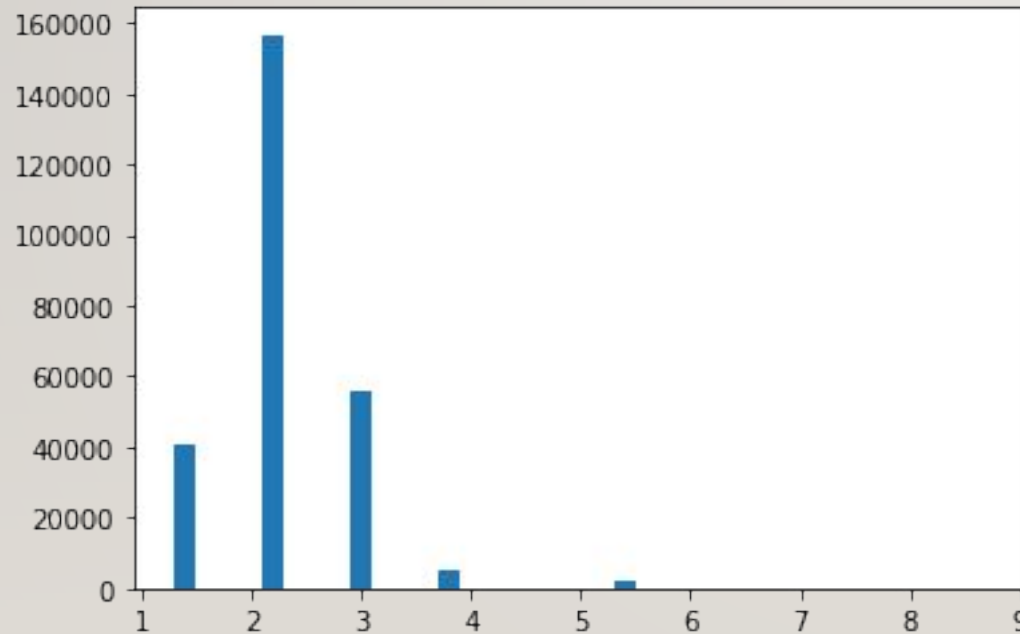
- Goal of this project is to predict the damage to the buildings caused due to earthquakes in Nepal .
- The research will help predict damage to all the buildings and hence we will be able to know which building is more susceptible to damage in the future or have faulty construction .
- This will help predict buildings which have high chances of getting damaged and plan their re-construction accordingly and prevent damage in case of future earthquakes .

METHODS

- The sample is taken from www.drivendata.com and I only used the initial few columns for making my model to prevent overfitting the model
- In this research first exploratory analysis was done to find various correlation between various variables
- And finally Random Forest model was generated to make predictions .

RESULTS--

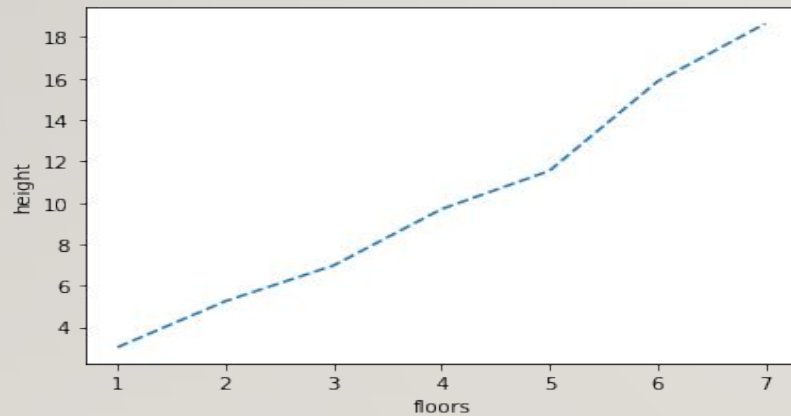
- Univariate analysis showed most of the building in Nepal had only 2 floors and most of the buildings were below 140 years of age .



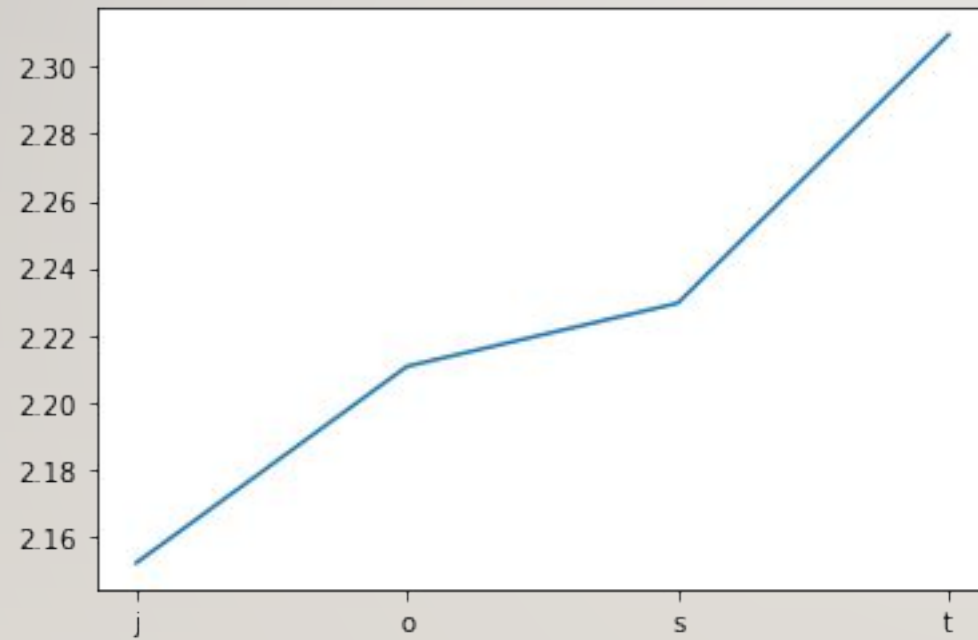
Floors histogram

FLOORS VS HEIGHT

- Floors and height came to be almost perfectly correlated .
- Even an ANOVA test was done to provide evidence in favour of this . ANOVA test had a low p value high f-static and floors even explained 65% of the variability



DAMAGE(Y) VS POSITION(X)-



MODEL – RANDOM FORESTS

- Since a perfect correlation for damage was not clear in the data analysis so random forests was chosen as model for prediction with $n=40$ which gave approximately 70% accuracy on our dataset .
- Using this prediction on test set we were able to make suitable predictions and hence ended our research here .

CONCLUSION / LIMITATIONS -

- In this many of the columns were not used during research which could have also been used to make a better model .
- Although outliers were removed some might still have been left which can cause errors .
- Since no exact dependence was found on any of the independent variables an high accuracy model could not have been made without risking overfitting .
- Since random forests takes a lot of time to fit increasing number of trees proved to be an limitation as well