# Fintech Assignment 2

Due Date: March 6th

Project Proposal: House Price Predictor

This assignment will use historical Australian house price sale data from 2018 - 2020 sourced from Kaggle.

Predict what the median house prices are by suburb for 2019 and compare the predicted values to the actual house prices.

Output visuals to a webpage using either a dashboard / interactive dashboard or **Flask**

Predictive models:

* Predictive Scatter plot
* Simple Random forest
* Linear Regression
* BalancedRandomForestClassifier
* EasyEnsembleClassifier
* LogisticRegression

Questions to ask the buyer

* Which suburb
* No. of bedrooms
* No. of bath
* Property type e.g. house, townhouse or unit

Data Sources:

Kaggle: <https://www.kaggle.com/htagholdings/aus-real-estate-sales-march-2019-to-april-2020>

Github (Nipune)

* Share the repository with every1
* Barebones import into jupyter file

Code:

1. Import libraries
2. Load data
3. Drop columns
4. Clean Data:
   1. na
   2. unique
   3. Anomalies / outlier removal
5. Split into training & data
6. Build machine learning model
7. Data pre-processing (standard scaler)
8. Test the model
9. Export the model
10. Save PNGs / Exports etc (pickle joblib h5)
11. Dashboard
12. Website/Flask

<https://hackersandslackers.com/plotly-dash-with-flask/>

