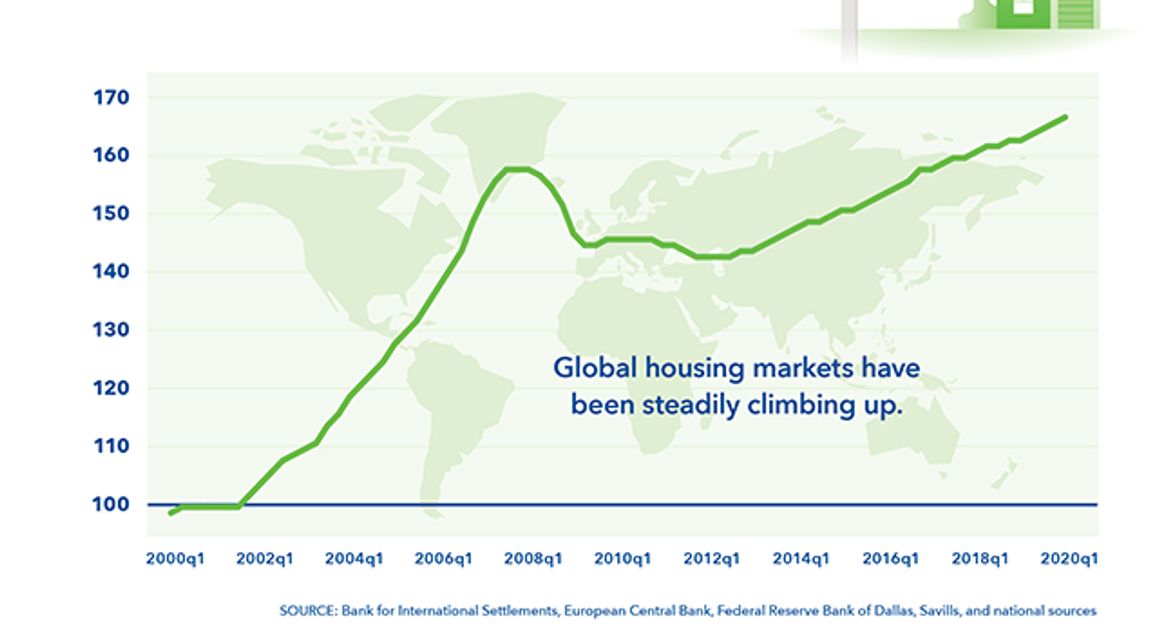
Predicting House Prices Using Machine Learning.

**Introduction:**

One of the world's most significant investments was the housing market. Everyone is in this world thinks to take ownership of a home. According to IMF(International Monetary Fund), the Real House Price index was Jumped 100 billion to 170 billion from 2000 - 2020 and rapidly increasing. This steady housing growth is directly related to the population it is estimated significant growth on the housing market.



**Figure 1**: International Monetary Fund.

**Source:**

Imf.org. 2021. *IMF Global Housing Watch*. [online] Available at: <https://www.imf.org/external/research/housing/> [Accessed 26 April 2021].

**Safe Investment:**

Real estate was one of the oldest investments. So, people always believe it is a safe investment. The government has been encouraging to buy it and providing tax benefits for purchasing it.

**Problems in Housing Market:**

If the buyer wants to buy the home, he must find an agent to show the perfect home. If the buyer does not know what the price is currently going on. Soon, he will fall into the trap because the agent will increase the price too much as the buyer was unaware of this. The buyer will agree to pay for what the agent said.

**How This project Will Help Buyer:**

This project will help the buyer suggest the estimated price of the home based on the previous data and machine learning model.

Complex machine learning models will analyse the previous selling data, transform, and learn to predict the price.

**Aim:**

The overall aim of the project is to find the best price for the home.

**Research Questions:**

* + What are algorithms should be used?
  + How much model will be accurate?

- Does missing data will be any impact for the project?

- Any data transformation is needed and why?

**Data Source:**

Whole project based on previous selling data and applying the complex mathematical model to it.

For the data source, this project was dependent on Kaggle housing prices data. Kaggle is the world's largest data science community where every data science user tries to find a solution/improvement for the problem based on the data. The current housing data source was copyrighted © by GPL v2 (General Public License). According to the license, this data free to use, manipulate and distribute for free without user consent.

**Sources:**

kumar, A., 2021. *House Price Prediction Challenge*. [online] Kaggle.com. Available at: <https://www.kaggle.com/anmolkumar/house-price-prediction-challenge> [Accessed 22 April 2021].

Gnu.org. 2021. *GNU General Public License v2.0 - GNU Project - Free Software Foundation*. [online] Available at: <http://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html> [Accessed 22 April 2021].

**Methods and practices using data science:**

This project was to give an estimation of the house. For this, we would require using Data science practices to make the project successful.

1. Data Cleaning / Wrangling
2. Brief Insights and analysis
3. Applying Machine Learning models
4. Predicting and visualisation

**Data Cleaning / Wrangling:**

Data have to be prepared in this process, and any missing values have to be checked and managed. Data should be transformed if needed.

**Brief Insights and analysis:**

After cleaning and transforming, data will be visualised to understand how the data relationships were. Moreover, what needs to change.

Most of the techniques will be used as explained in the reference 2

**Applying Machine Learning principles:**

Machine Learning is the science of getting computers to learn and act like humans do and improve their learning over time in an autonomous fashion by feeding them data and information in the form of observations and real-world interactions.

Most of the techniques will be used as explained in the reference 1

**Regression:**

Regression is a statistical method used in finance, investing, and other disciplines that attempt to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

There will be several methods that will be applied to predict the correctly.

**Predicting and Visualisation:**

Finally, once it is predicted, the likelihood will show using the graphs and charts using visualisations.

**GANTT Chart:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Implementation in Weeks** | **2** | **4** | **6** | **8** | **10** | **12** | **14** |
| Data Cleaning |  |  |  |  |  |  |  |
| Visualisation |  |  |  |  |  |  |  |
| Implementation |  |  |  |  |  |  |  |
| Results |  |  |  |  |  |  |  |
| Final Documentation |  |  |  |  |  |  |  |

**Legal:**

Project data was using total under licensed. It is fair to use and distribute according to license.

**Social:**

No people were used in this project for experimentation, or any others practices as well.

**Ethics:**

Strictly this project does not intend to hurt any feelings or beliefs of any person—this project aimed to help the person's financial decisions stability itself.

**Summary**

In conclusion, Using previously sold houses and analysing will tell a lot about the pricing estimations and more about the homes as well.

Besides by applying more machine learning concepts will make a solid understanding of the price of the home.

**References:**

1.-Géron, A., 2019. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. 2nd ed. O'Reilly Media, pp.322-415.

2. - Ozdemir, S., 2016. *Principles of data science*. Packt Publishing, pp.157-180.