ORIGINAL RESEARCH

Classification of Effect on Real Estates in Mumbai During Covid-19

S. Ananda Kumar[1] · Hemkar Goswami[1] · Aman Ashesh Kumar[1] · R. Sangeetha[2] · Pooja Jain[3]

Abstract
At the point when the global pandemic hit, we saw one of the greatest monetary accidents in late history. With crashes in the securities exchange and irregularities in all business sectors and ventures, one thing that never appears to change is the increment in worth of homes. This pandemic had made the land area take a significant misfortune. The sentiment scores of the real estate partners had arrived at an unequalled low score of 31. The crash of the real estate put many companies and brokers into bankruptcy. People stopped investing in real estate while others withdrew any sort of the ongoing investments. The conditions worsened with the second wave. The second wave, unlike the first, had a significantly higher death rate and spread very quickly instilling fear in the minds of the people. This became a factor in the huge economic crash that this pandemic has provided. The government has been desperately looking for measures to come back stronger but have failed. This called for research into this domain and start predictions with various models to help prepare for any foreseeing losses/crashes. Within the past 10 years, the housing market has gotten more and more modest with houses that used to be affordable being valued at absurdly high amounts. This increase in value has seen a reduction in home proprietorship percentage in proportion to populace of the country. Our project aims to project values for real estate in the absence of covid. The difference in values are analyzed and generalized. Furthermore, solutions to improve the current condition will also be given.

Keywords COVID-19 pandemic · Real estate · *K*-means clustering · Correlation · Linear regression

Introduction

The effect of COIVD-19 pandemic on India has intensely flustered the whole Indian economy as well as a massive loss of human lives. Needless to say, all the major sectors were badly shaking as importations and exportations sharply plummeted with some noteworthy exceptions where above average growth was discerned. Major foreign or domestic business ventures straightaway suspended or fundamentally reduced undertakings in multiple sectors all over the nation. These resulted in massive inflation in.

Needless to say, real estate world also suffered the slumps of COVID-19 during the cross-country lockdown. As constructions faced an unexpected halt, migration of workers fashioned the recommencement of function even bothersome [1]. Real estate developers challenged severe liquidity necessities and house buyers lost essential desire of securing an estate as the market got extremely chomped. New deals of real estate's underwent an implausible layout as the country prepared to fight the pandemic. Half a year into the instance, real estate need is by all means getting loomed by reliable house buyers eager to manipulate the weakened

✉ S. Ananda Kumar
   s.anandakumar@vit.ac.in

   Hemkar Goswami
   hemkargoswami@gmail.com

   Aman Ashesh Kumar
   amankumar7506@gmail.com

   R. Sangeetha
   sangeemathi2012@gmail.com

   Pooja Jain
   poojajain.som@itmuniversity.ac.in

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

[2] SITE, Vellore Institute of Technology, Vellore, India

[3] School of Management, ITM University, Gwalior, Madhya Pradesh, India

awareness and thinned out exchange potential [2]. Convincingly, with fast spread of the coronavirus in our country, the real estate business did not prepare enough for an enormous impact. Inferable since the peril of a disease, the real estate has essentially seen a waning in estate visits in addition to the buyer interest.

Indian real estate market has reached a stagnate state over the last year or so, suffered by a second wave of this coronavirus pandemic further crumpled the real estate developers' stimuli. In spite of all the analysis done for every sector, the precise influence of COVID-19 in the real estate business has not been document properly [3]. To accomplish the best solutions, an in-depth investigation is required in which the effect on the prices of the houses in absence and presence of COVID-19 will be related at lengths [4]. This will shed light on the exact effects of COVID-19. Thus, improved solutions can be discovered built on the outcome of the research. All in all, this project will provide explanations to help the country's economy by focusing on one of its vital sectors and in future. By the help of policy measures of the Government, the improvement expanse is on the way to recuperation, conversely at a lethargic speed [5]

Problem Statement

The effect of the novel COVID-19 on local Indian real estate market has been unprecedented. In conjunction with the national GDP discerning into negative numbers, the real estate sector tapped to the rock bottom of lows during the just about three-month-long nationwide lockdown. Development work across construction sites, paced at a slower pace. Then the second wave of COVID-19 spread delayed the economic recovery further. A major swing in the real-estate sector occurred, technology has now become a major player renting and buying properties. In the meantime, physical spot visits has declined though it continue be very vital process of buying a real-estate.

Real estate pricing took a major hit due to the COVID-19 pandemic as well as the emotions of people, as they have finally begun to understand the real situation or coming to terms with it. Thus, we try to crack the trend set by the pandemic for real estate. Analysis and prediction through machine learning models in order to visualize the outcome to help the industry grow more.

Literature Survey

Examination utilizes genuine exchange information which contains traits and exchange costs of estate which is utilized by the AI models. This study makes use of four machine learning models. The study conclusively proves that machine learning based model is a viable way to estimate real estate prices and also provides a reasonable analysis among these models. These models are easy to program and implement. All the models generate the desired output. The only change is in the efficiency. The models are flexible and can give desired outputs accordingly. The models are sensitive to noisy data. The time required to train these models is considerably high. The performance of the models depends a lot on the input values [6].

The model uses a diverse enhanced model based on IPSO-BPNN. The real estate market of Changsha has been taken as the dataset for this model and IPSO is used to enhance the original and the threshold values. It has both theoretical and practical significance in the research on real estate price [7]. Improved accuracy over previous such indexes by avoiding the drawbacks. This model also provides an improved IPSO algorithm for get better outputs. Sensitive to noisy input data. The IPSO-BPNN model is highly complex and difficult to code. The BPNN is a time-consuming model [8].

The study uses data that are collected from 1.2 million chartered property listing noted over the course of 2 months. The ANN-GWO algorithm is then trained using this data and the predictions are engendered. The method is further evaluated by using MSE and Precision Indexes. This model showed an outstanding accuracy of 98.79%. It has a huge future prospect since increasing the neurons will provide even better results. Can give desired results even with huge datasets. This model is highly hardware dependent. This model is time consuming when input dataset is huge. Debugging the problems in the network with huge lot of numerical data [9].

The data implemented are from the Melbourne Housing Market dataset. Machine learning techniques are applied to explore historical property transactions in Australia to discover useful models for brokers. Further experimentation reveals that the amalgamation of Stepwise and Support Vector Machine is a viable approach. Various algorithms comparatively combined to get the best results. Improving the performance and accuracy of the models. Proposed solution can be applied on all types of dataset irrespective of their size and factors influencing them. Sometimes trade-offs are to be done to achieve the desired outcome. Method is time consuming in order to find right algorithms. Comparing results from various models makes the process more complex [10].

Gradient Boosting Model is employed to predict estate prices. The public dataset containing 38,961 records of Karachi city is obtained and used in this paper. Their proposed house price prediction model is able to envisage with 98% accuracy. This model showed a very high rate of accuracy. It considers a whole of factors in determining the result. Uses XGBBoost algorithm to get speed, performance and flexibility. Accuracy is prioritized which may

result in dropping of few factors which might be essential to some people. Performance increase results in large time consumption. Metrics measurement can also result in a slow process [11].

Data taken into account include various characteristics including security perception to predict the real estate pricing. It involves a dataset containing 70,000 houses in Italy. This paper basically concludes that the surroundings/neighborhoods provide an irreplaceable reserve to evaluate the monetary value of the real estate. Very essential factors of daily life have been incorporated. Outcome is generic process applicable to any neighborhood. Can give optimal results even with huge datasets. Not all the factors considered only essentials has been chosen like the neighborhoods. Multi step dynamic process makes it complex. A trade-off between urban well-being, the economic success of cities and affordable housing happens [12].

The real estate prediction in this paper is done for Moscow Residential Market. This paper basically shows the dependence of the value of real estate on the dollar-rubble exchange with the help of a scheduled model. New factors like international influence are also considered. A deep monetary insight is reckoned. It has a huge future prospect since an unorthodox approach is used. This model is highly specific to countries with a sound economy. Use of various real estate models increases the complexity. Due to less flexibility relatively simple prediction models are utilized [13].

This model uses two famous metrics for its evaluation which are RMS (root mean squared) and MAE (mean absolute error). The use of these metrics helps in the prediction as well as in the determination of the performance index. Strictly calculation based. Accurate price calculation. It has a huge future prospect. Might not satisfy price increases. May vary with the pecuniary conditions. Prices may bounce steeply [14].

Partial datasets are used for prediction and utilizes "Hybrid Lasso and Gradient boosting regression model" for forecasting each and every house price giving a highly detailed result. Improved accuracy over previous such indexes by avoiding the drawbacks. This model showed a very high rate of accuracy. Output is applicable everywhere. Might not work with large dataset. Metrics measurement can also result in a slow process. Processing time might increase with increase the performance of the algorithm [15].

A geostatistical model for the forecast of the real estate in Italy predicts two things at the same time: Spatial and Temporal development of Italian NNT. Furthermore, the spatio-temporal association is researched upon and the temporal model is evaluated. The convenience and usefulness of using spatio-temporal geostatistics. Tools to describe socio-economic phenomena are good and have a high accuracy rate. But some metrics measurement can also result in a slow process. Performance increase results in large time consumption. Sometimes trade-offs are to be done to achieve the desired outcome [16].

Proposed Solution

Proposed Algorithm

Our project will be implemented via the following steps:

Step-1: Correlation between the area and price in the dataset is determined and Linear Regression Model is applied on the dataset. A hypothesis is generated and proved.

Step-2: The statistics acquired from correlation and values from the original dataset are fed to the *K*-means clustering algorithm directly.

Step-2: In *K*-means clustering, value of *K* is determined via hit and trial.

Step-3: Data points are allocated to the neighboring centroids which are responsible for the formation of *K* clusters obtained in Step-2.

Step-4: Variance is premeditated and new centroid is positioned for every cluster obtained.

Step-5: The steps are repeated till the best clusters are found.

Step-6: The data are given a graphical representation to encompass the information obtained by the *K* Clustering Algorithm, correlation and linear regression model.

Step-7: Solutions are given on how to improve the current situation of real estate based on the information obtained.

Algorithm Architecture

See Fig. 1.

Models and Technologies to be Used

1. **Correlation**

   Correlation is just the measure of how two variables are related to each other. For our project, this measure is important for data cleaning as well as analyzing. The correlation is done between the price and apartment's area covered.

2. *K-Means Clustering*

   K-means clustering is an algorithm responsible for grouping the dataset into *k* clusters. The *k* clusters are obtained manually in our project. The clusters formed contain similar properties and hence we obtain groups/categories without the requirement of any training.

   In our project, we will be using this *K*-means clustering to group apartment on the basis of their price and area, thus each cluster contains similar properties.
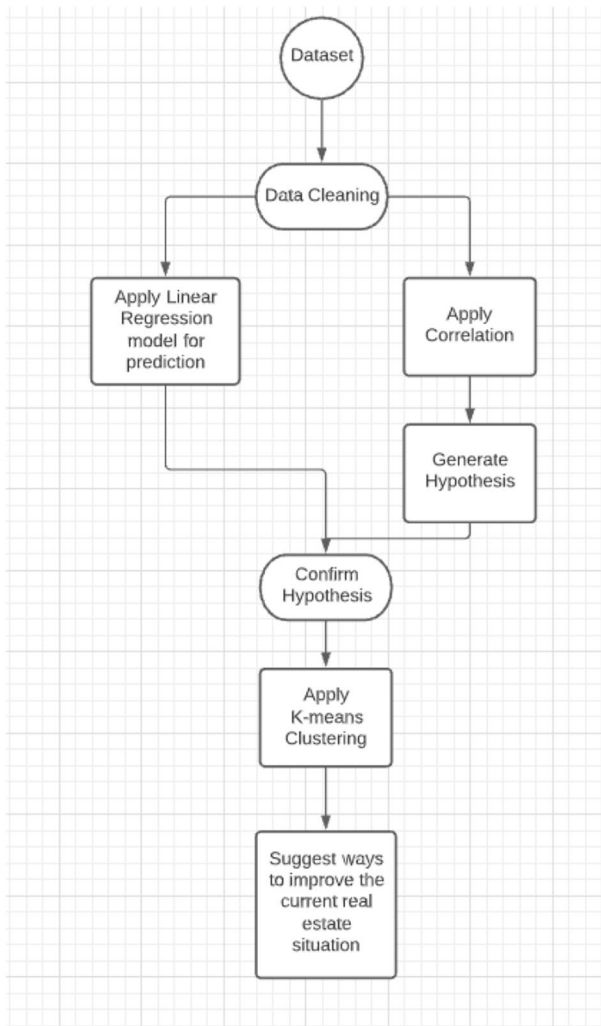
3. **Linear Regression Model**

Fig. 1 Architecture of the algorithm

This model is used predict linear data trend based on the past data available. It accounts for both, dependent as well as independent data which might influence the predictions.

Tools

1. Pandas: Analysis of data along with manipulation
2. NumPy: For working with arrays, linear algebra and matrices.
3. *K*-means clustering—a clustering algorithm for clusters with similar properties.
4. matplotlib: Plotting graphs to obtain information visually.
5. Seaborn: Library which incorporates its traits from matplotlib
6. Cufflinks are a library used for visualization straight from Pandas.

7. ArcGIS API for Python, a Python library for working with maps and geospatial data, powered by web GIS.

Applications

1. Predict the trend based on the current and past real estate pricing.
2. Get a grasp of the aftermath of the catastrophes on real estate pricing.
3. Keep the brokers updated with the current situation and spread awareness.
4. Provide valuable and verified information so that people can make informative decisions.

Dataset

https://www.kaggle.com/sameep98/housing-prices-in-mumbai

This dataset includes the housing prices in Mumbai before COVID. We will run our model on this dataset and generate the report along with the solutions for the same.

Implementation

(i) Data cleaning and splitting

The input dataset has a lot of values with no input ("NA") along with unnecessary symbols as well as signs such as ":" and ",". Hence it is important that the data are cleaned before we start analyzing the trend and impact of the pandemic on the real estate. Furthermore, there are outliers which will hinder our analysis and lessen our accuracy so we eliminate them in our next step. We have removed 71 values from our dataset which would have given us a less accurate analysis. The final dataset contains 388 rows and 4 columns as shown above. To get a better understanding of the impact of Covid, we have split the dataset into two sets. Set A contains values of the real estate market after the pandemic hit, while Set B contains values of the real estate before the pandemic hit.

(ii) Visualization of the price and area trend pre and post pandemic

Before we go into detail with correlation and *K*-means clustering, visualizing the data available to us already is a must. This gives us a rough idea of the impact of the pandemic on the real estate (Fig. 2).

The plot shown above gives us an idea of the changes in the price in the same location. As seen, there are hardly any abrupt changes in the value in the same location as seen. For example, Malad has roughly the same price even on different dates (Fig. 3).
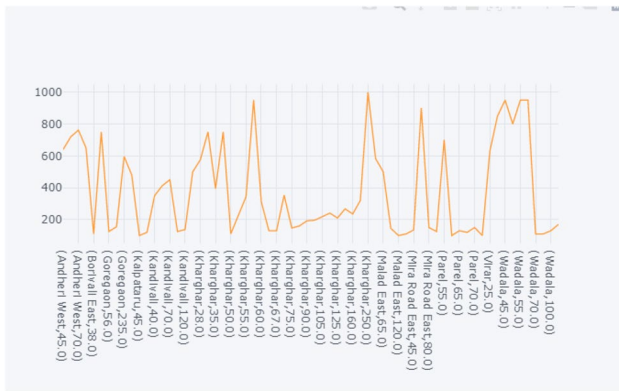
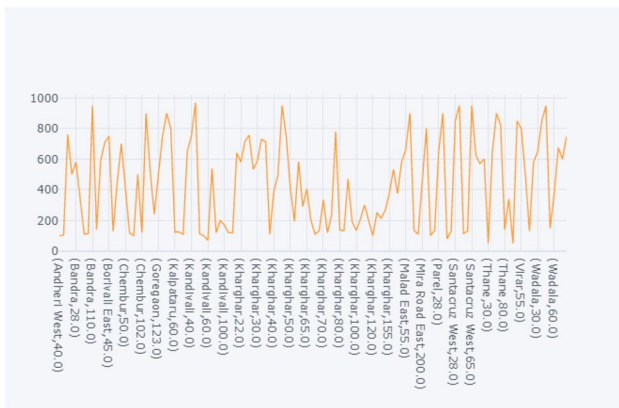**Fig. 2** Price and area trend before the pandemic

Fig. 3 Price and area trend after the pandemic

The plot above shows a highly volatile real estate market wherein places even in the same location show varying values on different dates. It is clear from the graph that the reason for this change is the pandemic. However, an in-depth analysis is required to expand our knowledge in this domain.

(iii) **Correlation**

A 32% correlation was found between the parameters Price and Area before the pandemic as it can be seen from the graph (Fig. 4).

Almost 21% correlation has been seen after the pandemic hit. This suggests that the relationship between these parameters have changed significantly roughly giving us a numerical idea of the impact. However, further analysis is needed to be able to draw conclusions (Fig. 5).

Results and Discussion

(i) **Linear Regression Model**

We are applying this model to prove our hypothesis obtained from correlation that the real estate market was on a stagnant state
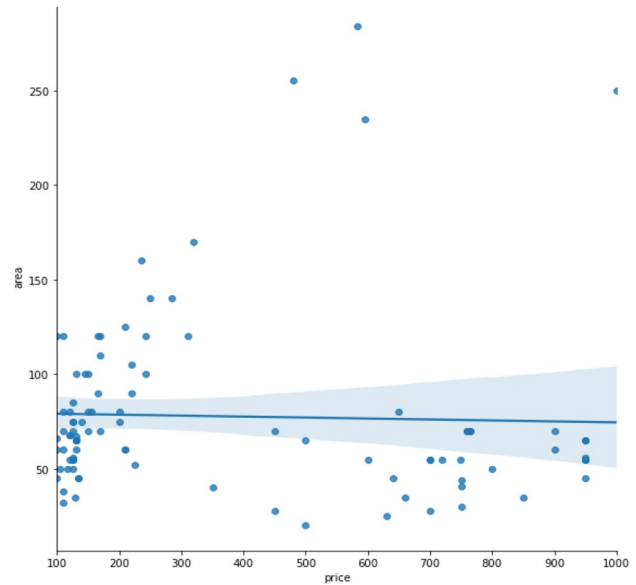

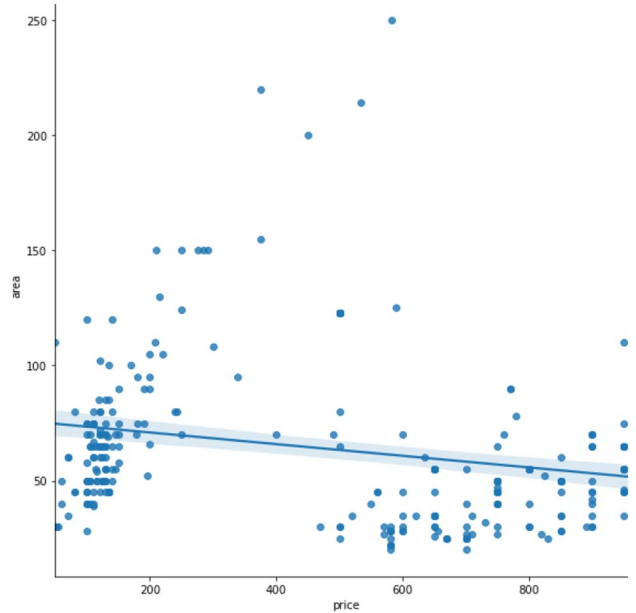**Fig. 4** Correlation between price and area before pandemic hit

Fig. 5 Correlation between price and area after pandemic hit

before the pandemic but after the pandemic, the real estate market saw an unusual inflation in the market but that did not last long resulting in the drop in prices across the entire city (Fig. 6).

As seen, our prediction seems to be proven by the linear regression as most of the real estate have seen a major drop in their values excluding some posh areas.

(ii) *K-Means Clustering*

The main role of this model is to make clusters based on similar properties. In this case, we will make use of 4 ($K = 4$) clusters and then club them according to their location.
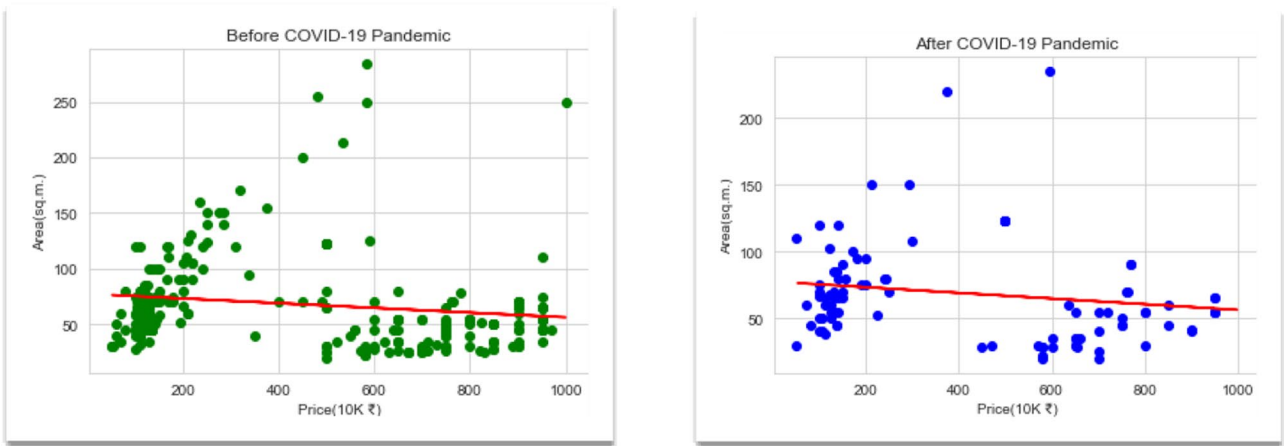
Fig. 6 Predicting values before and after the pandemic

Before the Pandemic

The four clusters formed are represented by four categories: Trivial estate with rational price, small estate with very exorbitant value, huge estate with high variety of prices and big flats with restrained to exorbitant price (Fig. 7).
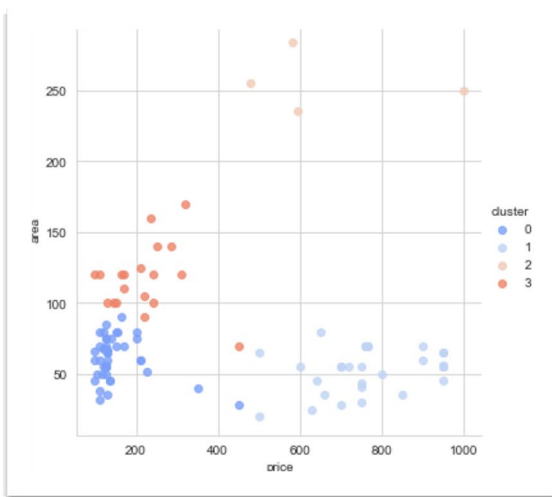
After the Pandemic

As seen from the graph, the range of prices have decreased except for cluster 0 in which the prices have increased significantly. In other words, the high class/developed areas of Mumbai saw next to no change or even an increase, while the remaining areas saw a high drop in the prices because of this pandemic (Fig. 8).

(iii) **Ways to improve the real estate market**
(a) Taking account of the governments steps towards reducing inflation in the Indian economy resulting in the betterment of the real estate sector.
(b) Attract people with extravagant discounts and offers to increase sale in the real estate domain.
(c) Bring the migrant workers back to the city so that construct work can resume in full stride.

(a)
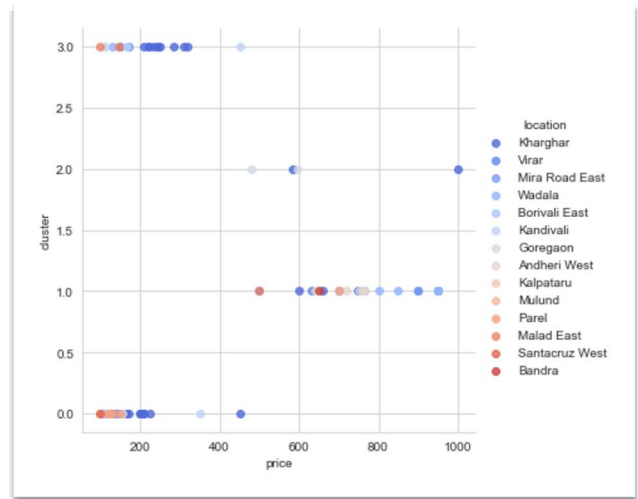
(b)

Fig. 7 **a** Visualization of clusters **b** Clusters based on location before the pandemic
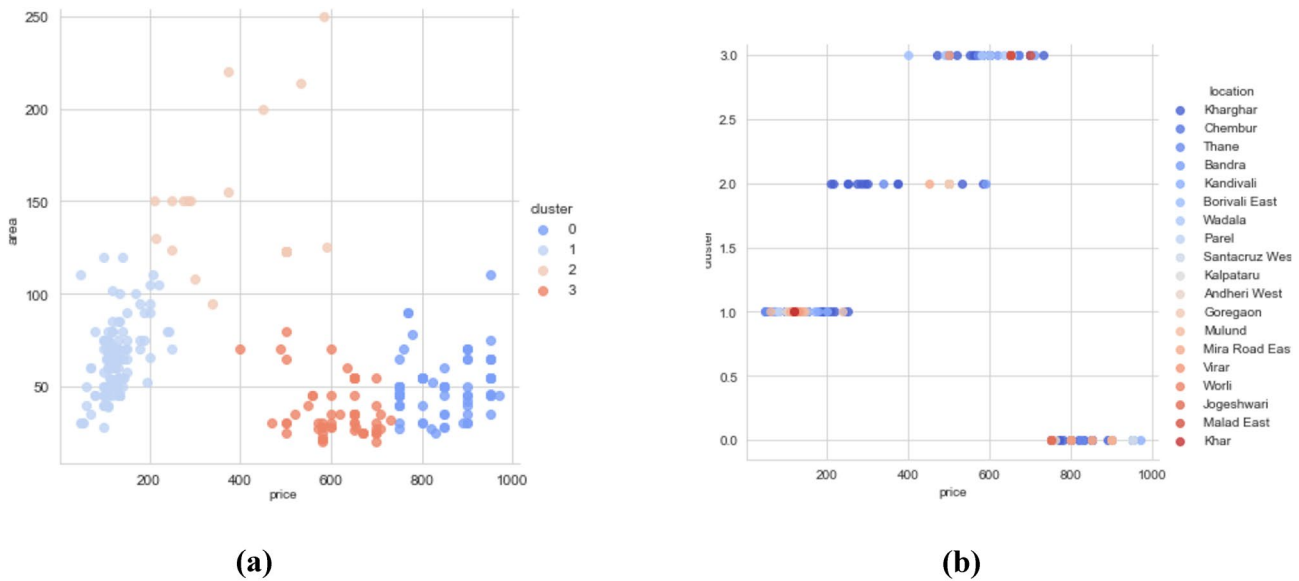
(a)   **(b)**

Fig. 8 **a** Visualization of clusters **b** Clusters based on location after the pandemic

(d) Inculcating various payment methods such as down-payment, multi-level payment plan and lucrative loans as it will attract more home-buyers.

(e) Accumulation of government led incentives such as tax reimbursements, comfort on loans will be welcomed with open arms by the general public.

Conclusion

The main conclusion of this project that no matter what area it is, covid has had its effect on the real estate either in a positive way or in a negative way. No place has is left unaffected by this pandemic. Prices of few posh areas have gone to a much greater height, while prices of few areas have seen a downtrend. Covid has instilled in the minds of people, a sense of fear, a fear of safety, so they have started to maintain a social distance, avoid going to crowded places, and also looking for sanitation where ever they go, so they prefer an area with good neighborhood sanitation, less crowded areas and are willing to pay a higher price for such areas. Such places in Mumbai include Malabar Hills, Cuffe Parade, etc., these are some of the posh areas of Mumbai with good sanitation. On the other hand, real estate prices of more crowded places have gone down as people are wanting to sell their properties in that area and there are less buyers for that area and people are considering the fact that this might happen again in future. So emotional thinking is one of the main reasons for this much amount of movement in the prices of real estate.

This change in the real estate price is also due to the inflation, one of the main reasons of which is the fall of economy which is on its path back as before. To counter it, either we can start selling contracts of properties like we do in stock market futures buying and selling to protect the interest of both, the buyer and the seller or counter inflation as Governments can employ a contractionary monetary policy that reduces the money supply within an economy via decrease in bond prices and increase in interest rates. But now the real estate industry has to learn towards adapting to these strange circumstances.

Real estate pricing took a major hit due to the COVID-19 pandemic as well as the emotions of people, as they have finally begun to understand the real situation or coming to terms with it. Thus, we try to crack the trend set by the pandemic for real estate. Analysis and prediction through machine learning models in order to visualize the outcome to help the industry grow more.

Declarations

Conflict of interest The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Consent to publication This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

References

1. D'Lima W, Lopez LA, Pradhan A (2020) Covid-19 and housing market effects. SSRN Electron J. No. 3647252
2. Segnon M, Gupta R, Lesame K, Wohar ME. High-frequency volatility forecasting of US housing markets. J Real Estate Financ Econ. 2020;62(2):283–317. https://doi.org/10.1007/s11146-020-09745-w.
3. Chauhan AS, Kapoor S (2021) Impact of coronavirus on Indian real estate. India Corporate Law. A Cyril Amarchand Mangaldas Blog. https://corporate.cyrilamarchandblogs.com
4. Sharifi A, Khavarian-Garmsir AR. The COVID-19 pandemic: Impacts on cities and major lessons for urbanplanning, design, and management. Sci Total Environ. 2020;749(2020): 142391. https://doi.org/10.1016/j.scitotenv.2020.142391.
5. Tupenaite L, Lill I, Geipele I, Naimaviciene J. Ranking of sustainability indicators for assessment of the new housing development projects: case of the Baltic states. Sustain Indic Environ Manag. 2017;6(4):55. https://doi.org/10.3390/resources6040055.
6. Pai PF, Wang WC. Using machine learning models and actual transaction data for predicting real estate prices. Appl Sci. 2020;10(5832):1–10. https://doi.org/10.3390/app10175832.
7. Del Giudice V, De Paola P, Del Giudice FP. COVID-19 infects real estate markets: short and mid-run effects on housing prices in Campania region (Italy). Soc Sci. 2020;9(7):114. https://doi.org/10.3390/socsci9070114.
8. Xue H (2015) The prediction on residential real estate price based on BPNN. In: 8th International conference on intelligent computation technology and automation (ICICTA), p 1008–1013. https://doi.org/10.1109/ICICTA.2015.256
9. Al-Gbury O, Kurnaz S (2020) Real estate price range prediction using artificial neural network and grey wolf optimizer. In: 4th International symposium on multidisciplinary studies and innovative technologies (ISMSIT), p 15. https://doi.org/10.1109/ISMSIT50672.2020.9254972
10. Phan, T.D (2018). Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia. *International Conference on Machine Learning and Data Engineering (ICMLDE)*,35–42. https://doi.org/10.1109/iCMLDE.2018.00017.
11. Ahtesham, M., Bawany, N.Z., & Fatima. F (2020). House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan. *21st International Arab Conference on Information Technology (ACIT),* 1–5. https://doi.org/10.1109/ACIT50332.2020.9300074.
12. De Nadai M, Lepri B (2018) The economic value of neighborhoods: predicting real estate prices from the urban environment. In: IEEE 5th international conference on data science and advanced analytics (DSAA), p 323–330. /https://doi.org/10.1109/DSAA.2018.00043
13. Salnikov VA, Mikheeva OM. Models for predicting prices in the moscow residential real estate market. Stud Russ Econ Dev. 2018;29(1):94–101. https://doi.org/10.1134/S1075700718010136.
14. Li L, Chu K (2017) Prediction of real estate price variation based on economic parameters. In: International conference on applied system innovation (ICASI), p 87–90. https://doi.org/10.1109/ICASI.2017.7988353
15. Lu S, Li Z, Qin Z, Yang X, Goh RSM (2017) A hybrid regression technique for house prices prediction. In: IEEE international conference on industrial engineering and engineering management (IEEM), p 319–323. https://doi.org/10.1109/IEEM.2017.8289904
16. Palma M, Cappello C, De Iaco S, et al. The residential real estate market in Italy: a spatio-temporal analysis. Qual Quant Int J Methodol. 2019;53(5):2451–72. https://doi.org/10.1007/s11135-018-0768-8.