

METROPOLITAN CITIES OF INDIA

INDUSTRY REVIEW

In India, the real estate sector is the second-highest employment generator, after the agriculture sector. Real estate sector in India is expected to reach 1 lakh crores by 2030. By 2025, it will contribute 13% to country's GDP. Emergence of nuclear families, rapid urbanisation and rising household income are likely to remain the key drivers for growth in all spheres of real estate, including residential, commercial, and retail. Rapid urbanisation in the country is pushing the growth of real estate. >70-75% of India's GDP will be contributed by urban areas by 2022.

LITERATURE SURVEY

To review the literature on the impacts of social, economic, environmental, political and technological aspects on shelter in relation to past and present developments, future indications along with a review on local and global contributions in these areas. The review will help in providing the knowledge to make decisions about price of housing properties in India. Despite the current surge in the construction of housing environments, the quest for predicting prices of housing properties of metropolitan areas in India remains elusive for a large segment of society. Decent affordable housing is emerging as a critical issue toward the social and economic well-being of the society. Demands for predicting prices of housing properties continue to far outstrip supply and projects that are currently built suffer from severe cost constraints, while lacking important qualities that relate to people needs and cultural aspirations. While recently there have been notable developments in housing design that demonstrate our increased understanding of social issues and construction techniques.

DATASET AND DOMAIN - DATA DICTIONARY

INDEX	DTYPE	DESCRIPTION
Price	int64	The price of the housing property or the value assessment in monetary terms
Area	int64	The constructed area of the housing property
City	object	This column describes the first-tier metropolitan cities where the housing property is located in
Location	object	The location of the housing property with respect to the cities
No. of Bedrooms	int64	Number of bedrooms available in a housing property
Resale	int64	The possibility of resale with respect to the housing property
Maintenance	int64	The availability of Maintenance in the housing property
Staff	int64	The availability of Staff in the housing property
Gymnasium	int64	The availability of Gymnasium in the housing property
Swimming Pool	int64	The availability of Swimming Pool in the housing property
Landscaped Gardens	int64	The availability of Landscaped Gardens in the housing property
Jogging Track	int64	The availability of Jogging Track in the housing property
Rain Water Harvesting	int64	The availability of Rain Water Harvesting in the housing property
Indoor Games	int64	The availability of Indoor Games in the housing property
Shopping Mall	int64	The availability of shopping mall in the housing property
intercom	int64	The availability of intercom in the housing property
Sports Facility	int64	The availability of Sports Facility in the housing property
ATM	int64	The availability of ATM in the housing property
Club House	int64	The availability of Club House in the housing property
School	int64	The availability of School in the housing property
24X7Security	int64	The availability of 24X7Security in the housing property
Power Backup	int64	The availability of Power Backup in the housing property
Car Parking	int64	The availability of Car Parking in the housing property
Staff Quarter	int64	The availability of Staff Quarter in the housing property
Cafeteria	int64	The availability of Cafeteria in the housing property
Multipurpose Room	int64	The availability of Multipurpose Room in the housing property
Hospital	int64	The availability of Hospital in the housing property
Washing Machine	int64	The availability of Washing Machine in the housing property
Gas connection	int64	The availability of Gas connection in the housing property
AC	int64	The availability of AC in the housing property
WIFI	int64	The availability of WIFI in the housing property
Children's play area	int64	The availability of Children's play area in the housing property
Lift Available	int64	The availability of Lift Available in the housing property
BED	int64	The availability of BED in the housing property
Vaastu Compliant	int64	Any Vaastu complaints regarding the housing property or none
Microwave	int64	The availability Microwave of in the housing property
Golf Course	int64	The availability of Golf Course in the housing property
TV	int64	The availability of TV in the housing property
Dining Table	int64	The availability of Dining Table in the housing property
Sofa	int64	The availability of Sofa in the housing property
Wardrobe	int64	The availability of Wardrobe in the housing property
Refrigerator	int64	The availability of Refrigerator in the housing property

VARIABLE CATEGORIZATION

CATEGORICAL COLUMNS:

3- (Price, No. of Bedrooms, Area)

NUMERICAL COLUMNS:

37- ('Location', 'No. of Bedrooms', 'Resale', 'Maintenance Staff', 'Gymnasium', 'Swimming Pool', 'Landscaped Gardens', 'Jogging Track', 'Rain-Water Harvesting', 'Indoor-Games', 'Shopping Mall', 'Intercom', 'Sports Facility', 'ATM', 'Club House', 'School', '24X7Security', 'Power Backup', 'Car Parking', 'Staff Quarter', 'Cafeteria', 'Multipurpose Room', 'Hospital', 'Washing Machine', 'Gas connection', 'AC', 'WIFI', 'Children's play area', 'Lift Available', 'BED', 'Vaastu Compliant', 'Microwave', 'Golf Course', 'TV', 'Dining Table', 'Sofa', 'Wardrobe', 'Refrigerator')

PRE-PROCESSING DATA ANALYSIS:

Shape of CHENNAI dataset: 5014 rows & 40 columns.

CHENNAI - NULL VALUE COUNT

Price	0
Area	0
Location	0
No. of Bedrooms	0
Resale	0
MaintenanceStaff	2781
Gymnasium	2781
SwimmingPool	2781
LandscapedGardens	2781
JoggingTrack	2781
RainWaterHarvesting	2781
IndoorGames	2781
ShoppingMall	2781
Intercom	2781
SportsFacility	2781
ATM	2781
ClubHouse	2781
School	2781
24X7Security	2781
PowerBackup	2781
CarParking	2781
StaffQuarter	2781
Cafeteria	2781
MultipurposeRoom	2781
Hospital	2781
Gasconnection	2781
AC	2781
Wifi	2781
Children'splayarea	2781
LiftAvailable	2781
VaastuCompliant	2781
GolfCourse	2781
Children'splayarea	2781

LiftAvailable	2781
BED	2781
VaastuCompliant	2781
Microwave	2781
GolfCourse	2781
TV	2781
DiningTable	2781
Sofa	2781
Wardrobe	2781
Refrigerator	2781

DROPPING REDUNDANT COLUMNS:

TV, Dining Table, Sofa, Wardrobe, Refrigerator, Washing Machine, BED, Microwave.

Shape of BANGALORE dataset: 6207 rows & 40 columns.

BANGALORE - NULL VALUE COUNT

Price	0
Area	0
Location	0
No. of Bedrooms	0
Resale	0
MaintenanceStaff	4256
Gymnasium	4256
SwimmingPool	4256
LandscapedGardens	4256
JoggingTrack	4256
RainWaterHarvesting	4256
IndoorGames	4256
ShoppingMall	4256
Intercom	4256
SportsFacility	4256
ATM	4256
ClubHouse	4256
School	4256
24X7Security	4256
PowerBackup	4256
CarParking	4256
StaffQuarter	4256
Cafeteria	4256
MultipurposeRoom	4256
Hospital	4256
WashingMachine	4256
Gasconnection	4256
AC	4256
Wifi	4256
Children'splayarea	4256
LiftAvailable	4256
BED	4256
VaastuCompliant	4256
Microwave	4256
GolfCourse	4256
TV	4256

DiningTable	4256
Sofa	4256
Wardrobe	4256
Refrigerator	4256

DROPPING REDUNDANT COLUMNS:

TV, Dining Table, Sofa, Wardrobe, Refrigerator, Washing Machine, BED, Microwave.

Shape of DELHI dataset: 4998 rows & 40 columns.

DELHI -NULL VALUE COUNT

Price	0
Area	0
Location	0
No. of Bedrooms	0
Resale	0
MaintenanceStaff	2996
Gymnasium	2996
SwimmingPool	2996
LandscapedGardens	2996
JoggingTrack	2996
RainWaterHarvesting	2996
IndoorGames	2996
ShoppingMall	2996
Intercom	2996
SportsFacility	2996
ATM	2996
ClubHouse	2996
School	2996
24X7Security	2996
PowerBackup	2996
CarParking	2996
StaffQuarter	2996
Cafeteria	2996
MultipurposeRoom	2996
Hospital	2996
WashingMachine	2996
Gasconnection	2996
AC	2996
Wifi	2996
Children'splayarea	2996
LiftAvailable	2996
BED	2996
VaastuCompliant	2996
Microwave	2996
GolfCourse	2996
TV	2996
DiningTable	2996
Sofa	2996
Wardrobe	2996
Refrigerator	2996

DROPPING REDUNDANT COLUMNS:

TV, Dining Table, Sofa, Wardrobe, Refrigerator, Washing Machine, BED, Microwave.

Shape of HYDERABAD dataset: 2518 rows & 32 columns.

HYDERABAD: (NULL VALUE COUNT)

Price	0
Area	0
Location	0
No. of Bedrooms	0
Resale	0
MaintenanceStaff	84
Gymnasium	84
SwimmingPool	84
LandscapedGardens	84
JoggingTrack	84
RainWaterHarvesting	84
IndoorGames	84
ShoppingMall	84
Intercom	84
SportsFacility	84
ATM	84
ClubHouse	84
School	84
24X7Security	84
PowerBackup	84
CarParking	84
StaffQuarter	84
Cafeteria	84
MultipurposeRoom	84
Hospital	84
WashingMachine	84
Gasconnection	84
AC	84
Wifi	84
Children'splayarea	84
LiftAvailable	84
BED	84
VaastuCompliant	84
Microwave	84
GolfCourse	84
TV	84
DiningTable	84
Sofa	84
Wardrobe	84
Refrigerator	84

DROPPING REDUNDANT COLUMNS:TV, Dining Table, Sofa, Wardrobe, Refrigerator, Washing Machine, BED, Microwave.

Shape of KOLKATA dataset: 6507 rows & 32 columns.

KOLKATA - NULL VALUE COUNT

No. of Bedrooms	1
MaintenanceStaff	6432
Gymnasium	6432
SwimmingPool	6432
LandscapedGardens	6432
JoggingTrack	6432
RainWaterHarvesting	6432
IndoorGames	6432
ShoppingMall	6432
Intercom	6432
SportsFacility	6432
ATM	6432
ClubHouse	6432
School	6432
24X7Security	6432
PowerBackup	6432
CarParking	6432
StaffQuarter	6432
Cafeteria	6432
MultipurposeRoom	6432
Hospital	6432
WashingMachine	6432
Gasconnection	6432
AC	6432
Wifi	6432
Children'splayarea	6432
LiftAvailable	6432
BED	6432
VaastuCompliant	6432
Microwave	6432
GolfCourse	6432
TV	6432
DiningTable	6432
Sofa	6432
Wardrobe	6432
Refrigerator	6432

DROPPING REDUNDANT COLUMNS:

TV, Dining Table, Sofa, Wardrobe, Refrigerator, Washing Machine, BED, Microwave.

Shape of MUMBAI dataset: 7719 rows& 40 columns.

MUMBAI: (NULL VALUE COUNT)

Price	0
Area	0
Location	0
No. of Bedrooms	0
Resale	0
MaintenanceStaff	6321
Gymnasium	6321
SwimmingPool	6321
LandscapedGardens	6321
JoggingTrack	6321
RainWaterHarvesting	6321
IndoorGames	6321
ShoppingMall	6321
Intercom	6321
SportsFacility	6321
ATM	6321
ClubHouse	6321
School	6321
24X7Security	6321
PowerBackup	6321
CarParking	6321
StaffQuarter	6321
Cafeteria	6321
MultipurposeRoom	6321
Hospital	6321
WashingMachine	6321
Gasconnection	6321
AC	6321
Wifi	6321
Children'splayarea	6321
LiftAvailable	6321
BED	6321
VaastuCompliant	6321
Microwave	6321
GolfCourse	6321
TV	6321
DiningTable	6321
Sofa	6321
Wardrobe	6321
Refrigerator	6321

DROPPING REDUNDANT COLUMNS

TV, Dining Table, Sofa, Wardrobe, Refrigerator, Washing Machine, BED, Microwave.

ATLERNATE SOURCE OF DATA

Source 1– <https://www.kaggle.com/vikrishnan/boston-house-prices>

Source 2 - <https://www.kaggle.com/gpandi007/usa-housing-dataset>

Source 3 - <https://www.kaggle.com/anthonyypino/melbourne-housing-market>

PROJECT JUSTIFICATION

Project Statement: This dataset contains the prices of housing properties across various tier one metropolitan cities containing the availability of amenities like furnishing, exercise facilities, electrical amenities etc. The objective concerning this project is to understand the relation between the factors that influence the pricing of housing properties.

Complexity involved: Categorical columns in the dataset had only binary values. Multicollinearity for several categorical columns were low irrespective of the city.

Project Outcome

Commercial: By completing this project it helps us understand the correlation between the factors that influence the pricing of properties thereby helping people involved in real estate or realtors who understand the science behind the pricing of housing properties thereby prices can be set more accurately thus helping them improve in monetary terms.