

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

1. Downloading the data set from Kaggle
2. Loading the Dataset

```
House_price = pd.read_csv('/content/House Price India.csv')
House_price.head()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	grade of the house
0	6762810145	42491	5	2.50	3650	9050	2.0				
1	6762810635	42491	4	2.50	2920	4000	1.5				
2	6762810998	42491	5	2.75	2910	9480	1.5				
3	6762812605	42491	4	2.50	3310	42998	2.0				
4	6762812919	42491	3	2.00	2710	4500	1.5				

5 rows × 23 columns

```
House_price.shape
```

(14620, 23)

```
House_price.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14620 entries, 0 to 14619
Data columns (total 23 columns):
 #   Column          Non-Null Count  Dtype  
--- 
 0   id              14620 non-null   int64  
 1   Date            14620 non-null   int64  
 2   number of bedrooms  14620 non-null   int64  
 3   number of bathrooms 14620 non-null   float64 
 4   living area     14620 non-null   int64  
 5   lot area         14620 non-null   int64  
 6   number of floors 14620 non-null   float64 
 7   waterfront present 14620 non-null   int64  
 8   number of views   14620 non-null   int64  
 9   condition of the house 14620 non-null   int64  
 10  grade of the house 14620 non-null   int64
```

```

11 Area of the house(excluding basement) 14620 non-null int64
12 Area of the basement 14620 non-null int64
13 Built Year 14620 non-null int64
14 Renovation Year 14620 non-null int64
15 Postal Code 14620 non-null int64
16 Latitude 14620 non-null float64
17 Longitude 14620 non-null float64
18 living_area_renov 14620 non-null int64
19 lot_area_renov 14620 non-null int64
20 Number of schools nearby 14620 non-null int64
21 Distance from the airport 14620 non-null int64
22 Price 14620 non-null int64
dtypes: float64(4), int64(19)
memory usage: 2.6 MB

```

`House_price.isnull().any()`

id	False
Date	False
number of bedrooms	False
number of bathrooms	False
living area	False
lot area	False
number of floors	False
waterfront present	False
number of views	False
condition of the house	False
grade of the house	False
Area of the house(excluding basement)	False
Area of the basement	False
Built Year	False
Renovation Year	False
Postal Code	False
Latitude	False
Longitude	False
living_area_renov	False
lot_area_renov	False
Number of schools nearby	False
Distance from the airport	False
Price	False

`dtype: bool`

`House_price.isnull().sum()`

id	0
Date	0
number of bedrooms	0
number of bathrooms	0
living area	0
lot area	0
number of floors	0
waterfront present	0
number of views	0
condition of the house	0
grade of the house	0

```
Area of the house(excluding basement)      0
Area of the basement                      0
Built Year                                0
Renovation Year                           0
Postal Code                               0
Latitude                                  0
Longitude                                 0
living_area_renov                         0
lot_area_renov                            0
Number of schools nearby                  0
Distance from the airport                 0
Price                                     0
dtype: int64
```

```
House_price.describe()
```

	id	Date	number of bedrooms	number of bathrooms	livin are
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.26299
std	6.237575e+03	67.347991	0.938719	0.769934	928.27572
min	6.762810e+09	42491.000000	1.000000	0.500000	370.00000
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.00000
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.00000
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.00000
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.00000

8 rows × 23 columns

3. Performing the Visualizations

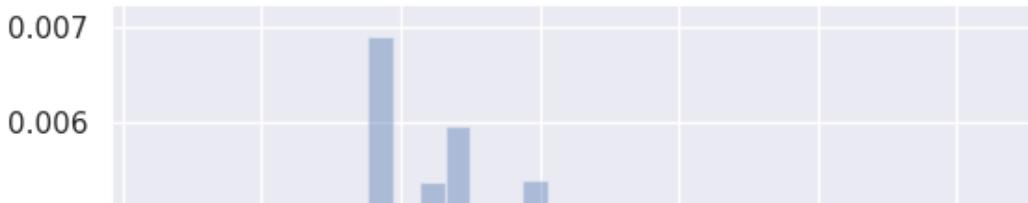
Univariant Analysis

```
sns.set_theme()
```

```
sns.distplot(House_price.Date)
```

```
<ipython-input-21-9724c665c443>:1: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn  
Please adapt your code to use either `displot` (a figure-level fun  
similar flexibility) or `histplot` (an axes-level function for his  
For a guide to updating your code to use the new functions, please  
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

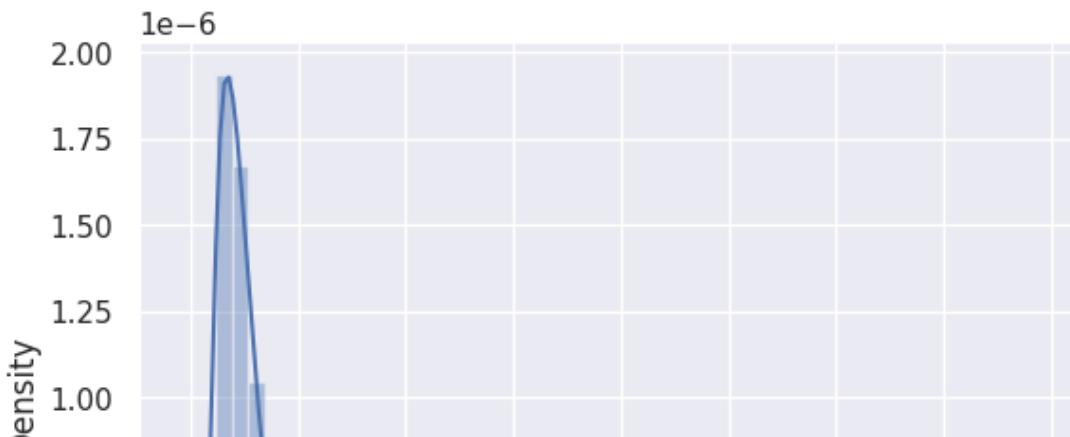
```
    sns.distplot(House_price.Date)  
<Axes: xlabel='Date', ylabel='Density'>
```



```
sns.distplot(House_price.Price)
```

```
<ipython-input-22-a54c73afe3f8>:1: UserWarning:  
`distplot` is a deprecated function and will be removed in seaborn  
Please adapt your code to use either `displot` (a figure-level fun  
similar flexibility) or `histplot` (an axes-level function for his  
For a guide to updating your code to use the new functions, please  
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
    sns.distplot(House_price.Price)  
<Axes: xlabel='Price', ylabel='Density'>
```



```
sns.scatterplot(House_price.living_area_renov,color='g')
```

```
<Axes: ylabel='living_area_renov'>
```

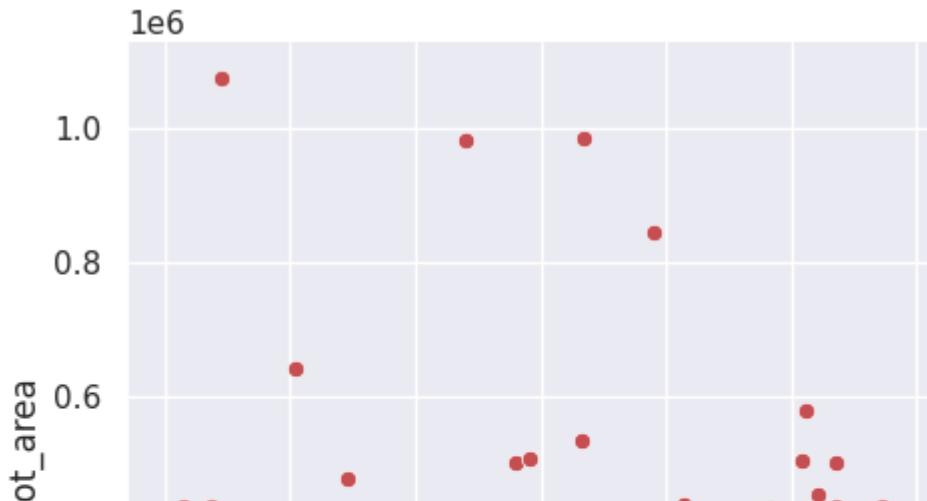


```
#Changing the column name
```

```
House_price.rename(columns= {'number of bedrooms' : 'number_of_bedrooms','lot area':'lot_a
```

```
sns.relplot(House_price.lot_area,color='r')
```

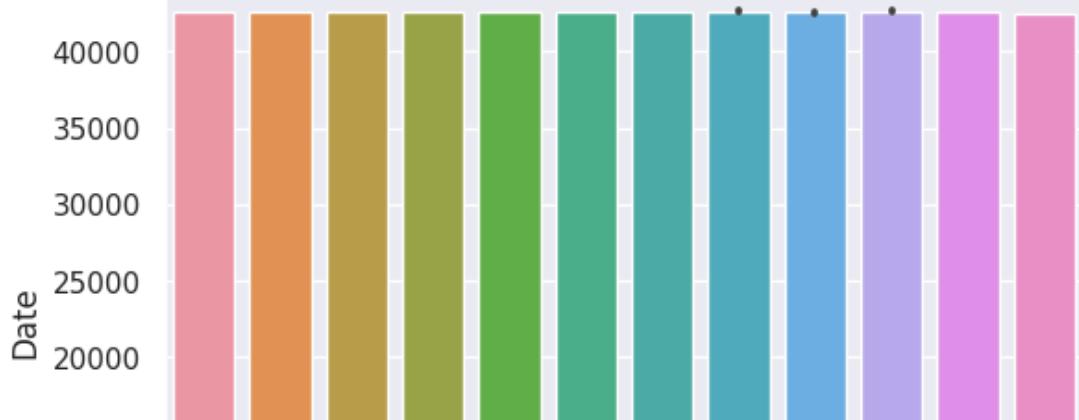
```
<seaborn.axisgrid.FacetGrid at 0x78300778f730>
```



```
House_price.rename(columns={'Built Year' : 'Built_Year','number of views' : 'number_of_view
```

```
sns.barplot(x='number_of_bedrooms',y='Date',data=House_price)
```

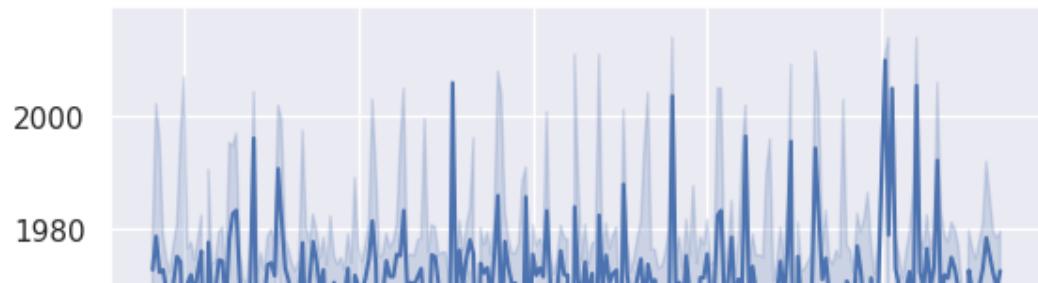
```
<Axes: xlabel='number_of_bedrooms', ylabel='Date'>
```



Bi- Variant Analysis

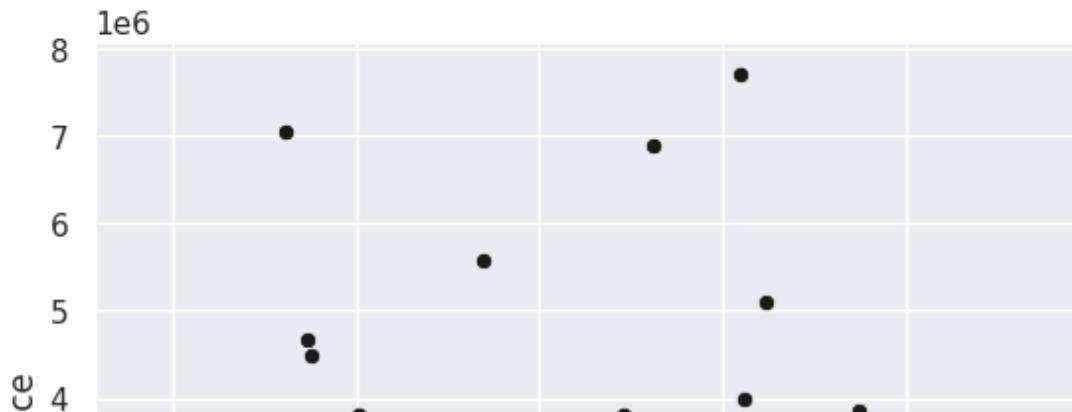
```
sns.lineplot(x=House_price.Date, y=House_price.Built_Year)
```

```
<Axes: xlabel='Date', ylabel='Built_Year'>
```



```
sns.scatterplot(x=House_price.Date,y=House_price.Price,color='k')
```

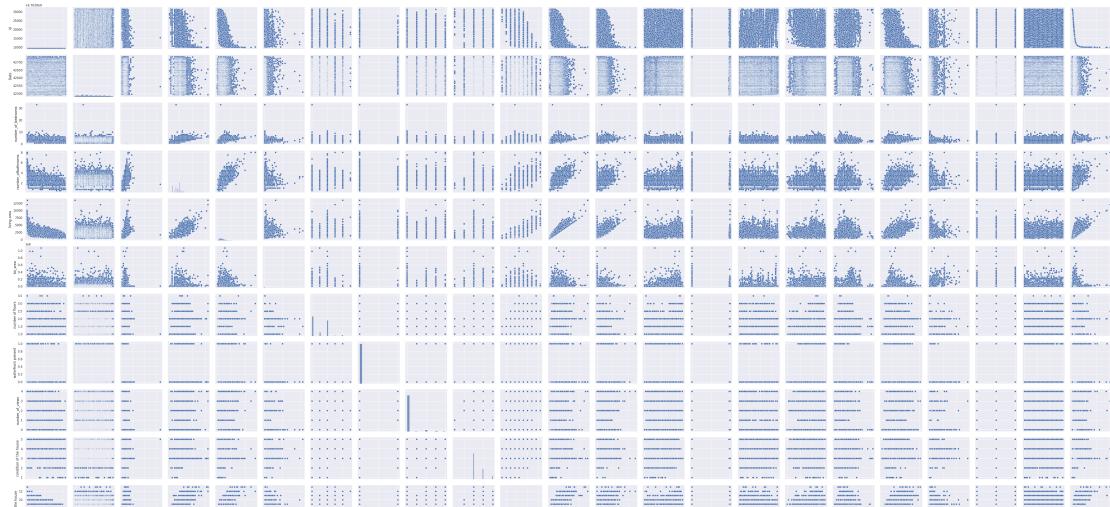
```
<Axes: xlabel='Date', ylabel='Price'>
```



Multivariant Analysis

```
sns.pairplot(House_price)
```

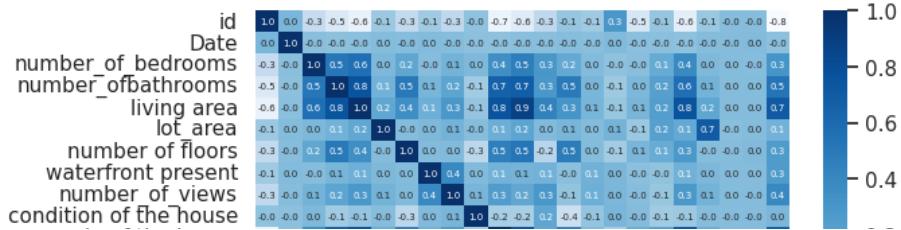
```
<seaborn.axisgrid.PairGrid at 0x782f8e73beb0>
```



```
correlation =House_price.corr()
```

```
sns.heatmap(House_price.corr(),annot=True,fmt='.1f',annot_kws={'size':5},cmap='Blues')
```

<Axes: >



4. Perform descriptive statistics on the dataset

House_price.count()

id	14620
Date	14620
number_of_bedrooms	14620
number_ofbathrooms	14620
living area	14620
lot_area	14620
number of floors	14620
waterfront present	14620
number_of_views	14620
condition of the house	14620
grade of the house	14620
Area of the house(excluding basement)	14620
Area of the basement	14620
Built_Year	14620
Renovation Year	14620
Postal Code	14620
Latitude	14620
Longitude	14620
living_area_renov	14620
lot_area_renov	14620
Number of schools nearby	14620
Distance from the airport	14620
Price	14620
dtype: int64	

House_price.mean()

id	6.762821e+09
Date	4.260454e+04
number_of_bedrooms	3.379343e+00
number_ofbathrooms	2.129583e+00
living area	2.098263e+03
lot_area	1.509328e+04
number of floors	1.502360e+00
waterfront present	7.660739e-03
number_of_views	2.331053e-01
condition of the house	3.430506e+00
grade of the house	7.682421e+00
Area of the house(excluding basement)	1.801784e+03
Area of the basement	2.964791e+02
Built_Year	1.970926e+03

```

Renovation Year          9.092401e+01
Postal Code              1.220331e+05
Latitude                 5.279285e+01
Longitude                -1.144040e+02
living_area_renov        1.996702e+03
lot_area_renov           1.275350e+04
Number of schools nearby 2.012244e+00
Distance from the airport 6.495096e+01
Price                     5.389322e+05
dtype: float64

```

`House_price.median()`

```

id                         6.762821e+09
Date                       4.260000e+04
number_of_bedrooms         3.000000e+00
number_ofbathrooms         2.250000e+00
living area                1.930000e+03
lot_area                   7.620000e+03
number of floors            1.500000e+00
waterfront present         0.000000e+00
number_of_views             0.000000e+00
condition of the house    3.000000e+00
grade of the house          7.000000e+00
Area of the house(excluding basement) 1.580000e+03
Area of the basement       0.000000e+00
Built_Year                  1.975000e+03
Renovation Year             0.000000e+00
Postal Code                 1.220320e+05
Latitude                    5.280640e+01
Longitude                  -1.144210e+02
living_area_renov           1.850000e+03
lot_area_renov              7.620000e+03
Number of schools nearby   2.000000e+00
Distance from the airport  6.500000e+01
Price                       4.500000e+05
dtype: float64

```

`House_price.std()`

```

id                         6237.574799
Date                       67.347991
number_of_bedrooms         0.938719
number_ofbathrooms         0.769934
living area                928.275721
lot_area                   37919.621304
number of floors            0.540239
waterfront present         0.087193
number_of_views             0.766259
condition of the house    0.664151
grade of the house          1.175033
Area of the house(excluding basement) 833.809963
Area of the basement       448.551409
Built_Year                  29.493625
Renovation Year             416.216661

```

```
Postal Code          19.082418
Latitude            0.137522
Longitude           0.141326
living_area_renov   691.093366
lot_area_renov      26058.414467
Number of schools nearby   0.817284
Distance from the airport   8.936008
Price                367532.380804
dtype: float64
```

```
House_price.min()
```

```
id                  6.762810e+09
Date                4.249100e+04
number_of_bedrooms  1.000000e+00
number_ofbathrooms  5.000000e-01
living area         3.700000e+02
lot_area            5.200000e+02
number of floors    1.000000e+00
waterfront present 0.000000e+00
number_of_views     0.000000e+00
condition of the house  1.000000e+00
grade of the house  4.000000e+00
Area of the house(excluding basement) 3.700000e+02
Area of the basement 0.000000e+00
Built_Year          1.900000e+03
Renovation Year    0.000000e+00
Postal Code         1.220030e+05
Latitude            5.238590e+01
Longitude           -1.147090e+02
living_area_renov   4.600000e+02
lot_area_renov      6.510000e+02
Number of schools nearby   1.000000e+00
Distance from the airport   5.000000e+01
Price                7.800000e+04
dtype: float64
```

```
House_price.max()
```

```
id                  6.762832e+09
Date                4.273400e+04
number_of_bedrooms  3.300000e+01
number_ofbathrooms  8.000000e+00
living area         1.354000e+04
lot_area            1.074218e+06
number of floors    3.500000e+00
waterfront present 1.000000e+00
number_of_views     4.000000e+00
condition of the house  5.000000e+00
grade of the house  1.300000e+01
Area of the house(excluding basement) 9.410000e+03
Area of the basement 4.820000e+03
Built_Year          2.015000e+03
Renovation Year    2.015000e+03
Postal Code         1.220720e+05
```

```

Latitude           5.300760e+01
Longitude          -1.135050e+02
living_area_renov 6.110000e+03
lot_area_renov    5.606170e+05
Number of schools nearby 3.000000e+00
Distance from the airport 8.000000e+01
Price              7.700000e+06
dtype: float64

```

```
House_price.describe()
```

	id	Date	number_of_bedrooms	number_ofbathro
count	1.462000e+04	14620.000000	14620.000000	14620.000000
mean	6.762821e+09	42604.538646	3.379343	2.129
std	6.237575e+03	67.347991	0.938719	0.769
min	6.762810e+09	42491.000000	1.000000	0.500
25%	6.762815e+09	42546.000000	3.000000	1.750
50%	6.762821e+09	42600.000000	3.000000	2.250
75%	6.762826e+09	42662.000000	4.000000	2.500
max	6.762832e+09	42734.000000	33.000000	8.000

8 rows × 23 columns

5.Handle the missing values

```
df= pd.read_csv('/content/Placement_Dataset.csv')
```

```
df.isnull().any()
```

s1_no	False
gender	False
ssc_p	False
ssc_b	False
hsc_p	False
hsc_b	False
hsc_s	False
degree_p	False
degree_t	False
workex	False
etest_p	False
specialisation	False
mba_p	False

```
status          False
salary          True
dtype: bool
```

```
df.isnull().sum()
```

```
sl_no           0
gender          0
ssc_p           0
ssc_b           0
hsc_p           0
hsc_b           0
hsc_s           0
degree_p        0
degree_t        0
workex          0
etest_p         0
specialisation  0
mba_p           0
status          0
salary          67
dtype: int64
```

```
#Replacing with median value
```

```
df['salary'].fillna(df['salary'].median(), inplace=True)
df.isnull().sum()
```

```
sl_no           0
gender          0
ssc_p           0
ssc_b           0
hsc_p           0
hsc_b           0
hsc_s           0
degree_p        0
degree_t        0
workex          0
etest_p         0
specialisation  0
mba_p           0
status          0
salary          0
dtype: int64
```

```
df.isnull().any()
```

```
sl_no          False
gender         False
ssc_p          False
ssc_b          False
hsc_p          False
hsc_b          False
hsc_s          False
```

```
degree_p      False
degree_t      False
workex        False
etest_p       False
specialisation False
mba_p         False
status         False
salary         False
dtype: bool
```

