In [4]:
```
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import pandas_profiling as pp
```

In [3]:
```
#!pip3 install pandas_profiling
```

In [6]:
```
data_path = "C:/Users/kumarmanglam.thakur/OneDrive - IHS Markit/Study/AccelerateA
df = pd.read_csv(data_path)
df.head()
```

Out[6]:

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

In [7]:
```
pp.ProfileReport(df)
```

```
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]
Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]
Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]
```

Overview

Dataset statistics

Number of variables	9
Number of observations	13320
Missing cells	6201
Missing cells (%)	5.2%
Duplicate rows	267
Duplicate rows (%)	2.0%
Total size in memory	936.7 KiB
Average record size in memory	72.0 B

Variable types

Categorical	7
Numeric	2

Alerts

Dataset has 267 (2.0%) duplicate rows	**Duplicates**
`availability` has a high cardinality: 81 distinct values	**High cardinality**
`location` has a high cardinality: 1305 distinct values	**High cardinality**

Out[7]:

In [ ]: