

EXPLORER

NO FOLDER OPENED

OUTLINE

TIMELINE

RealEstateAU\_1000\_Samples.csv

C:\Users> laksh > Downloads > australiahousing.py > ...

```

1 import pandas as pd
2 from pandas_profiling import ProfileReport
3 df=pd.read_csv('RealEstateAU_1000_Samples.csv')
4 print(df)
5 profile=ProfileReport(df)
6 profile.to_file(output_file="housing.html")
7 profile=ProfileReport(df)
8 profile.to_file(output_file="housing.html")
9

```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL Python Debug Console

	index	TID	breadcrumb	...	bathroom_count	parking_count	RunDate
0	0	1350988	Buy>NT>DARWIN CITY	...	1.0	1.0	2022-05-27 15:54:05
1	1	1350989	Buy>NT>DARWIN CITY	...	2.0	2.0	2022-05-27 15:54:05
2	2	1350990	Buy>NT>DARWIN CITY	...	1.0	1.0	2022-05-27 15:54:05
3	3	1350991	Buy>NT>DARWIN CITY	...	1.0	0.0	2022-05-27 15:54:05
4	4	1350992	Buy>NT>DARWIN CITY	...	2.0	2.0	2022-05-27 15:54:05
...	...	...	...	...	...	...	...
995	995	1351983	Buy>NT>DARWIN	...	3.0	6.0	2022-05-27 15:54:05
996	996	1351984	Buy>NT>DARWIN	...	2.0	2.0	2022-05-27 15:54:05
997	997	1351985	Buy>NT>DARWIN	...	2.0	2.0	2022-05-27 15:54:05
998	998	1351986	Buy>NT>DARWIN	...	2.0	3.0	2022-05-27 15:54:05
999	999	1351987	Buy>NT>DARWIN	...	2.0	2.0	2022-05-27 15:54:05

[1000 rows x 27 columns]

Summarize dataset: 100% | 73/73 [00:10<00:00, 6.81it/s, Completed]

Generate report structure: 100% | 1/1 [00:05<00:00, 5.90s/it]

Render HTML: 0% | 0/1 [00:00<?, ?it/s]

## Pandas Profiling Report

Overview

Variables

Interactions

Correlations

Missing values

Sample

Overview

Alerts

49

Reproduction

## Dataset statistics

Number of variables	27
Number of observations	1000
Missing cells	4399
Missing cells (%)	16.3%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	211.1 KIB
Average record size in memory	216.1 B

## Variable types

Numeric	6
Categorical	19
Unsupported	2

## Variables

Select Columns ▾

# Variables

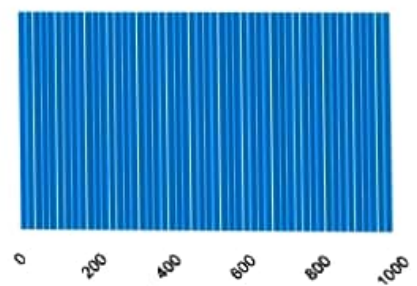
Select Columns ▾

df\_index

Real number (R)

HIGH CORRELATION UNIFORM UNIQUE

<b>Distinct</b>	1000	<b>Minimum</b>	0
<b>Distinct (%)</b>	100.0%	<b>Maximum</b>	999
<b>Missing</b>	0	<b>Zeros</b>	1
<b>Missing (%)</b>	0.0%	<b>Zeros (%)</b>	0.1%
<b>Infinite</b>	0	<b>Negative</b>	0
<b>Infinite (%)</b>	0.0%	<b>Negative (%)</b>	0.0%
<b>Mean</b>	499.5	<b>Memory size</b>	7.9 KiB



More details

## Pandas Profiling Report

Overview

Variables

Interactions

Correlations

Missing values

Sample

## product\_depth

Categorical

<b>Distinct</b>	4
<b>Distinct (%)</b>	0.4%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.9 KiB

[More details](#)

## bedroom\_count

Real number (R)

HIGH CORRELATION MISSING

<b>Distinct</b>	10	<b>Minimum</b>	0
<b>Distinct (%)</b>	1.0%	<b>Maximum</b>	9
<b>Missing</b>	33	<b>Zeros</b>	4

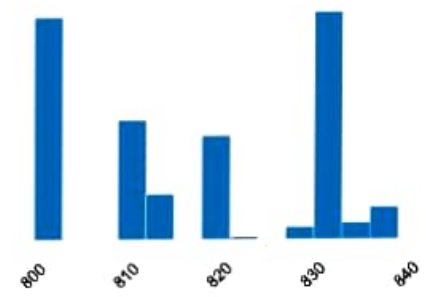


## Pandas Profiling Report

## zip\_code

Real number (E.)

<b>Distinct</b>	13	<b>Minimum</b>	800
<b>Distinct (%)</b>	1.3%	<b>Maximum</b>	839
<b>Missing</b>	0	<b>Zeros</b>	0
<b>Missing (%)</b>	0.0%	<b>Zeros (%)</b>	0.0%
<b>Infinite</b>	0	<b>Negative</b>	0
<b>Infinite (%)</b>	0.0%	<b>Negative (%)</b>	0.0%
<b>Mean</b>	816.646	<b>Memory size</b>	7.9 KiB

[More details](#)

## phone

Categorical

HIGH CARDINALITY
HIGH CORRELATION

<b>Distinct</b>	84	08 8943 3000	145
<b>Distinct (%)</b>	8.4%	08 8946 0500	62
		08 8931 5000	53

Pandas Profiling Report

Overview Variables Interactions Correlations Missing values Sample

HIGH CARDINALITY MISSING UNIFORM

Distinct	882	44 Woods S...	2
Distinct (%)	89.3%	309/3 Gardi...	2
Missing	12	9/144 Smith...	2
Missing (%)	1.2%	508/3 Gardi...	2
Memory size	7.9 KiB	1109/31 Wo...	2
		Other value...	978

More details

address\_1

Categorical

HIGH CARDINALITY MISSING UNIFORM

Distinct	882	44 Woods S...	2
Distinct (%)	89.3%	309/3 Gardi...	2
Missing	12	9/144 Smith...	2
Missing (%)	1.2%	508/3 Gardi...	2
Memory size	7.9 KiB	1109/31 Wo...	2
		Other value...	978

More details

## Pandas Profiling Report

Overview

Variables

Interactions

Correlations

Missing values

Sample

## bathroom\_count

Categorical

HIGH CORRELATION MISSING

<b>Distinct</b>	5
<b>Distinct (%)</b>	0.5%
<b>Missing</b>	33
<b>Missing (%)</b>	3.3%
<b>Memory size</b>	7.9 KiB

[More details](#)

## parking\_count

Real number (ℝ)

HIGH CORRELATION MISSING ZEROS

<b>Distinct</b>	11	<b>Minimum</b>	0
<b>Distinct (%)</b>	1.1%	<b>Maximum</b>	12
<b>Missing</b>	33	<b>Zeros</b>	54
<b>Missing (%)</b>	3.3%	<b>Zeros (%)</b>	5.4%

## Pandas Profiling Report

Overview

Variables

Interactions

Correlations

Missing values

Sample

## listing\_agency

Categorical

HIGH CARDINALITY HIGH CORRELATION

<b>Distinct</b>	85
<b>Distinct (%)</b>	8.5%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.9 KiB

[More details](#)

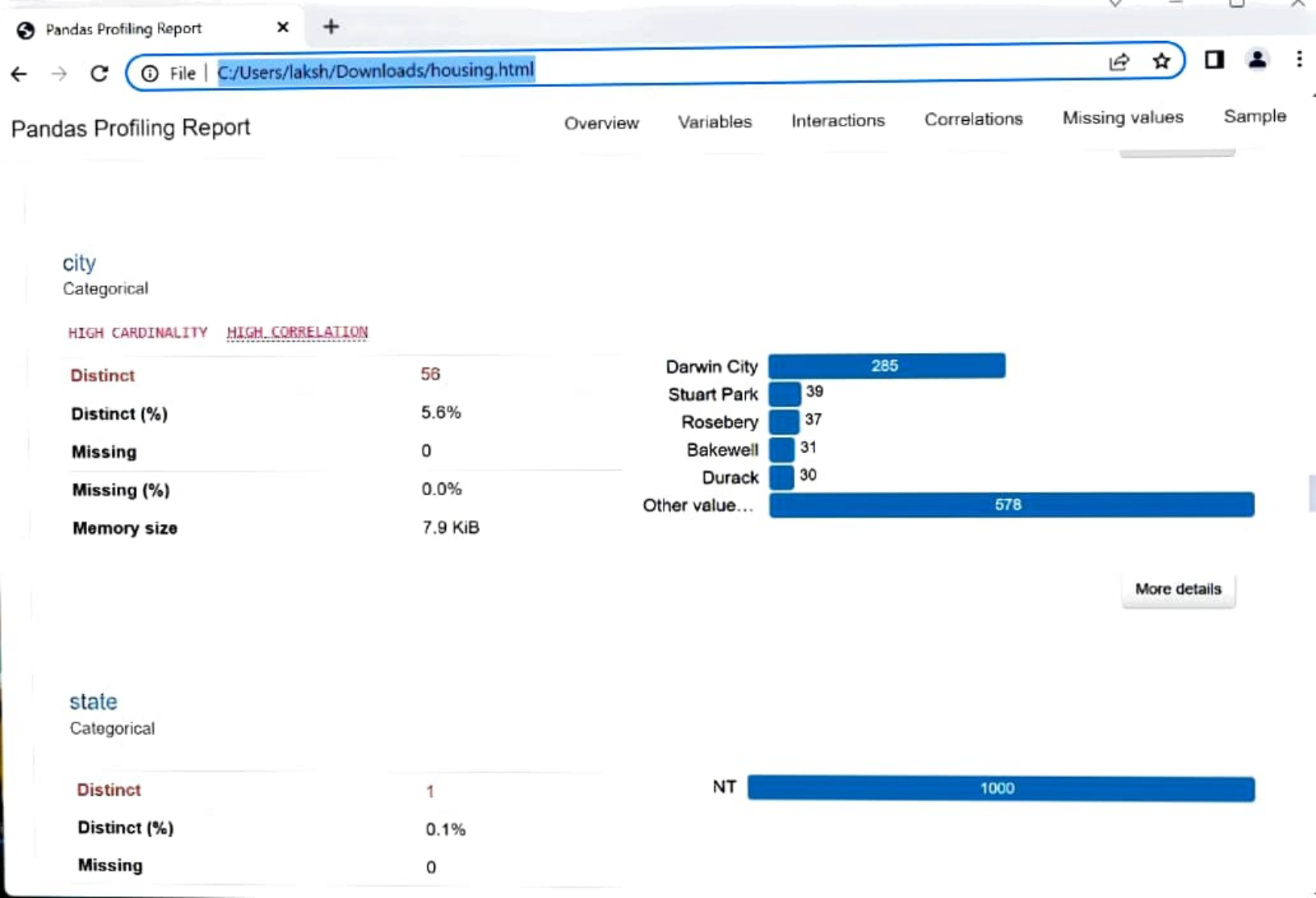
## price

Categorical

<b>Distinct</b>	494
<b>Distinct (%)</b>	49.4%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.9 KiB







### location\_type

Categorical

<b>Distinct</b>	1
<b>Distinct (%)</b>	0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.9 KiB



More details

### location\_name

Categorical

<b>Distinct</b>	494
<b>Distinct (%)</b>	49.4%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.9 KiB



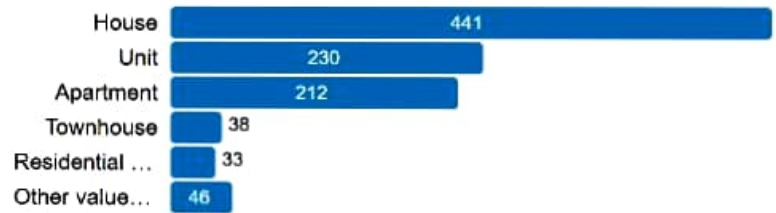
# Pandas Profiling Report

More details

## property\_type

Categorical

<b>Distinct</b>	13
<b>Distinct (%)</b>	1.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	7.9 KiB



More details

## building\_size

Categorical

HIGH CARDINALITY MISSING UNIFORM

<b>Distinct</b>	169
<b>Distinct (%)</b>	60.4%



More details

### preferred\_size

Categorical

HIGH CARDINALITY MISSING

Distinct	376
Distinct (%)	61.7%
Missing	391
Missing (%)	39.1%
Memory size	7.9 KiB



More details

### open\_date

Categorical

IMBALANCE MISSING

Distinct	15
Distinct (%)	5.0%



## Pandas Profiling Report

Overview

Variables

Interactions

Correlations

Missing values

Sample

## building\_size

Categorical

HIGH CARDINALITY MISSING UNIFORM

Distinct	169
Distinct (%)	60.4%
Missing	720
Missing (%)	72.0%
Memory size	7.9 KiB

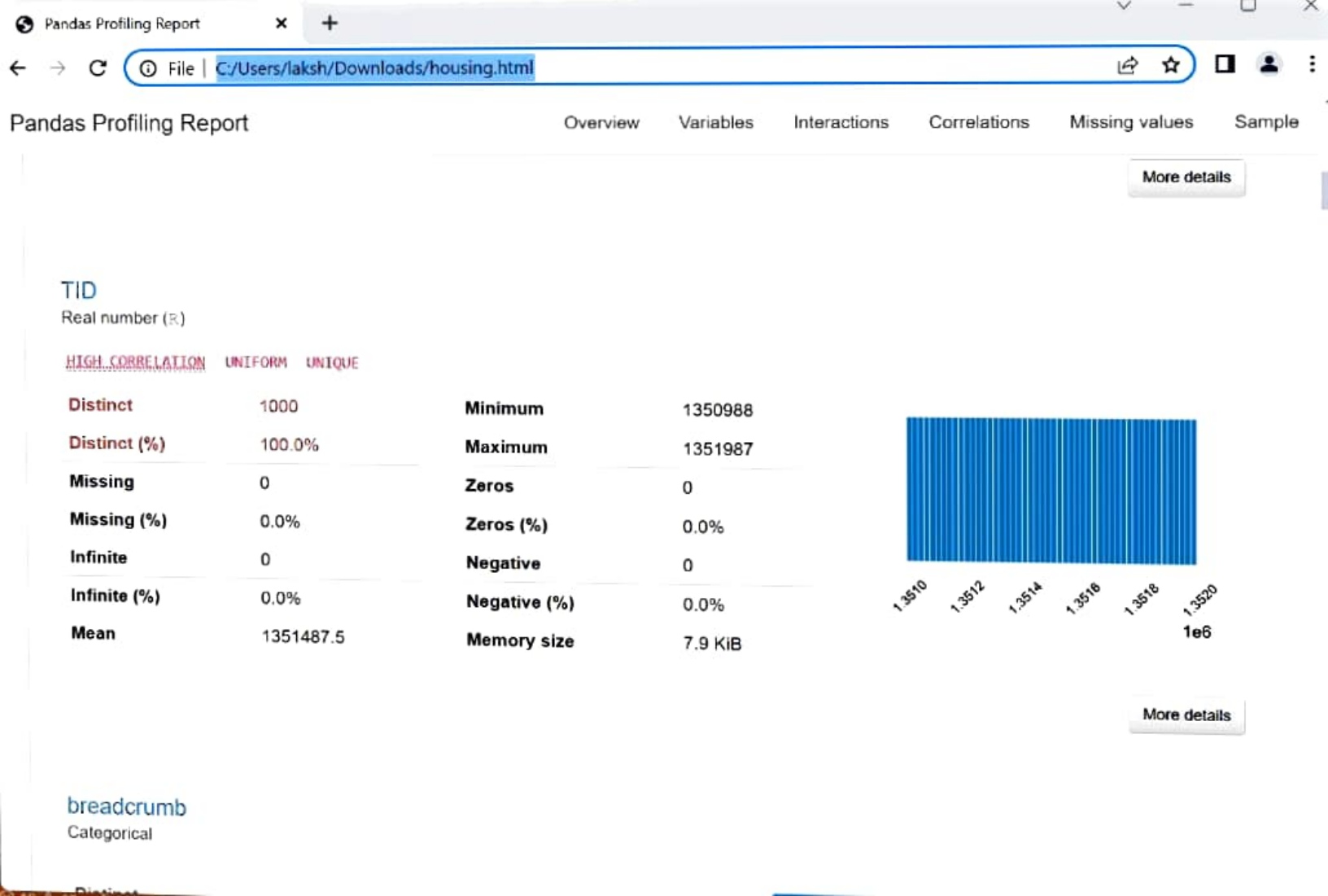
[More details](#)

## land\_size

Categorical

HIGH CARDINALITY MISSING

Distinct	346	2.02ha	14
Distinct (%)	64.9%	800m <sup>2</sup>	13
Missing	467	2ha	11
		450m <sup>2</sup>	10



More details

breadcrumb

Categorical

<b>Distinct</b>	2	Buy>NT>D...	816
<b>Distinct (%)</b>	0.2%	Buy>NT>D...	184
<b>Missing</b>	0		
<b>Missing (%)</b>	0.0%		
<b>Memory size</b>	7.9 KiB		

More details

category\_name

Categorical

<b>Distinct</b>	2	Real Estate ...	816
<b>Distinct (%)</b>	0.2%	Real Estate ...	184
<b>Missing</b>	0		

## Pandas Profiling Report

Overview

Variables

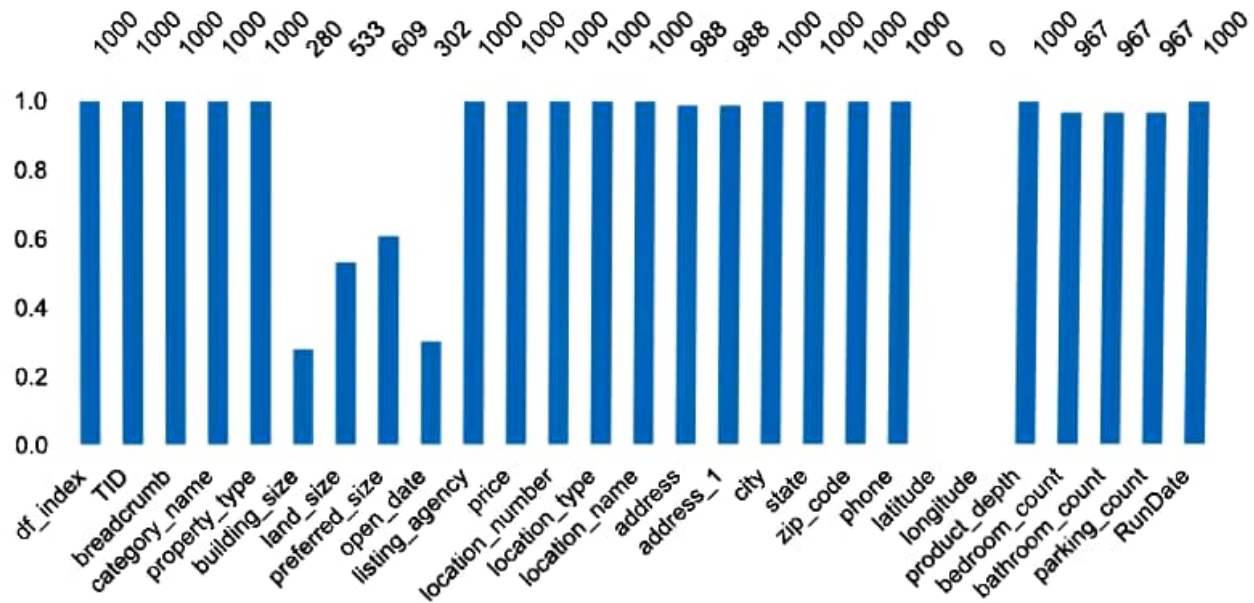
Interactions

Correlations

Missing values

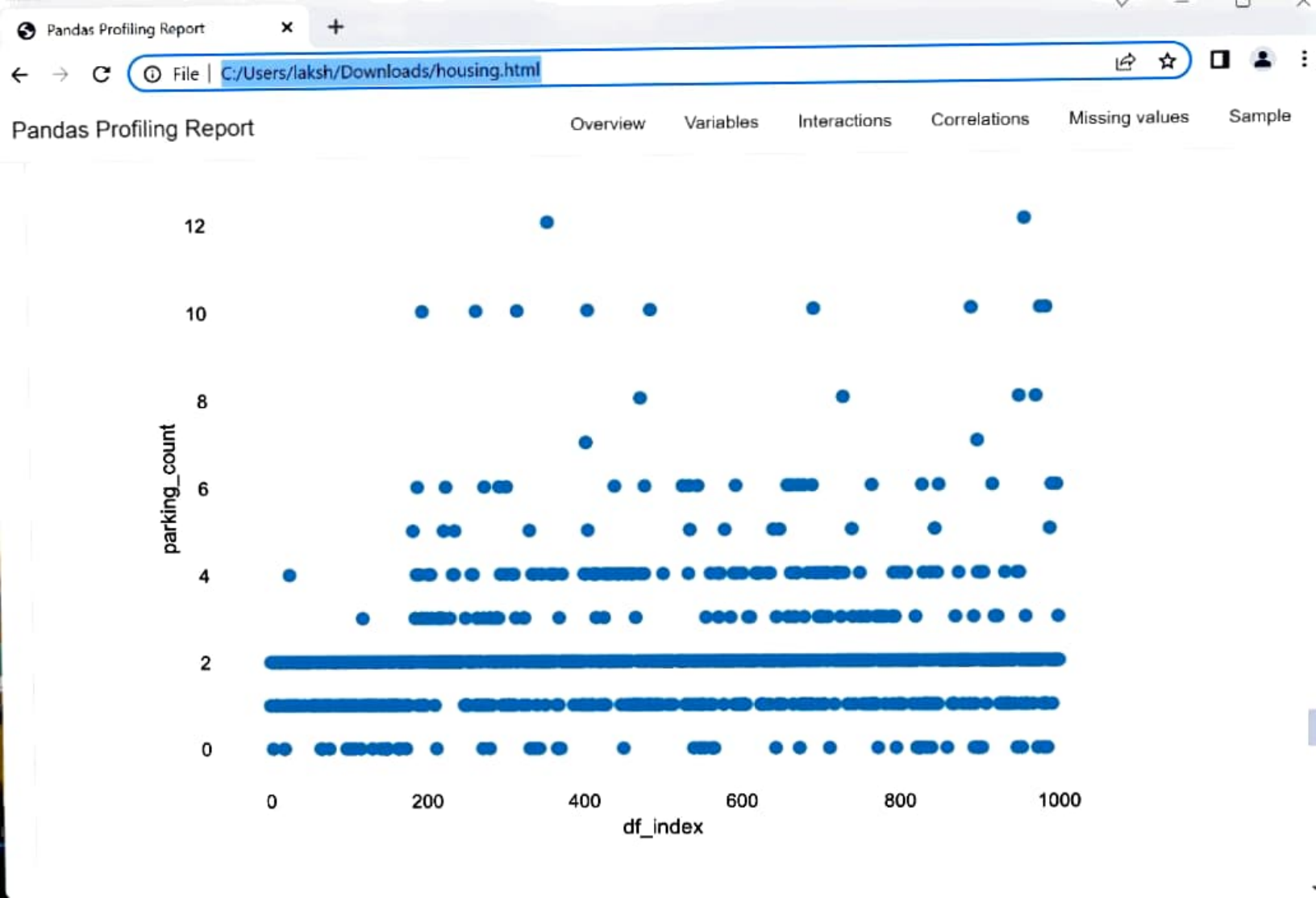
Sample

Count Matrix Heatmap



A simple visualization of nullity by column.





## Pandas Profiling Report

Overview

Variables

Interactions

Correlations

Missing values

Sample

First rows

Last rows

	df_index	TID	breadcrumb	category_name	property_type	building
0	0	1350988	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	House	NaN
1	1	1350989	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Apartment	171m <sup>2</sup>
2	2	1350990	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Unit	NaN
3	3	1350991	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	House	NaN
4	4	1350992	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Unit	201m <sup>2</sup>
5	5	1350993	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Apartment	218m <sup>2</sup>
6	6	1350994	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Unit	NaN
7	7	1350995	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Unit	NaN
8	8	1350996	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Unit	NaN
9	9	1350997	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY, NT 0800	Apartment	NaN

Report generated by YData.

