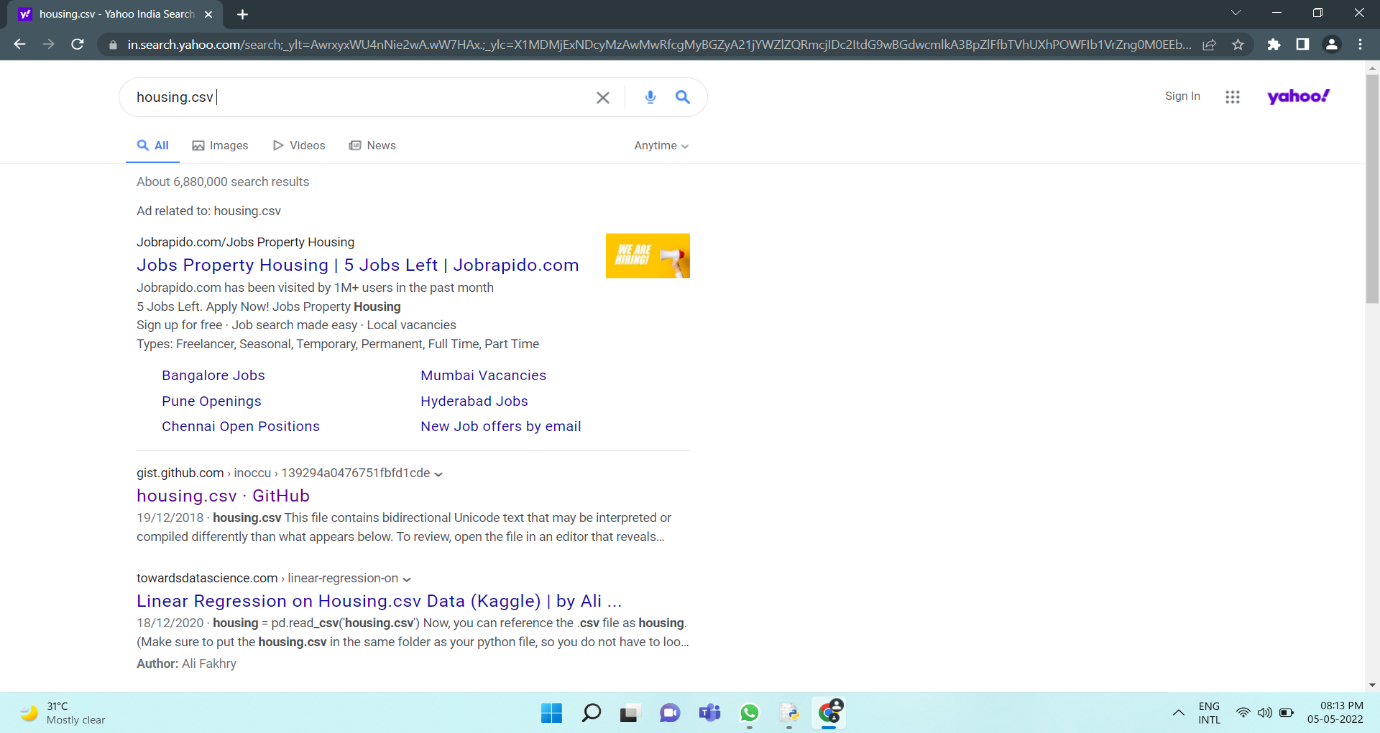
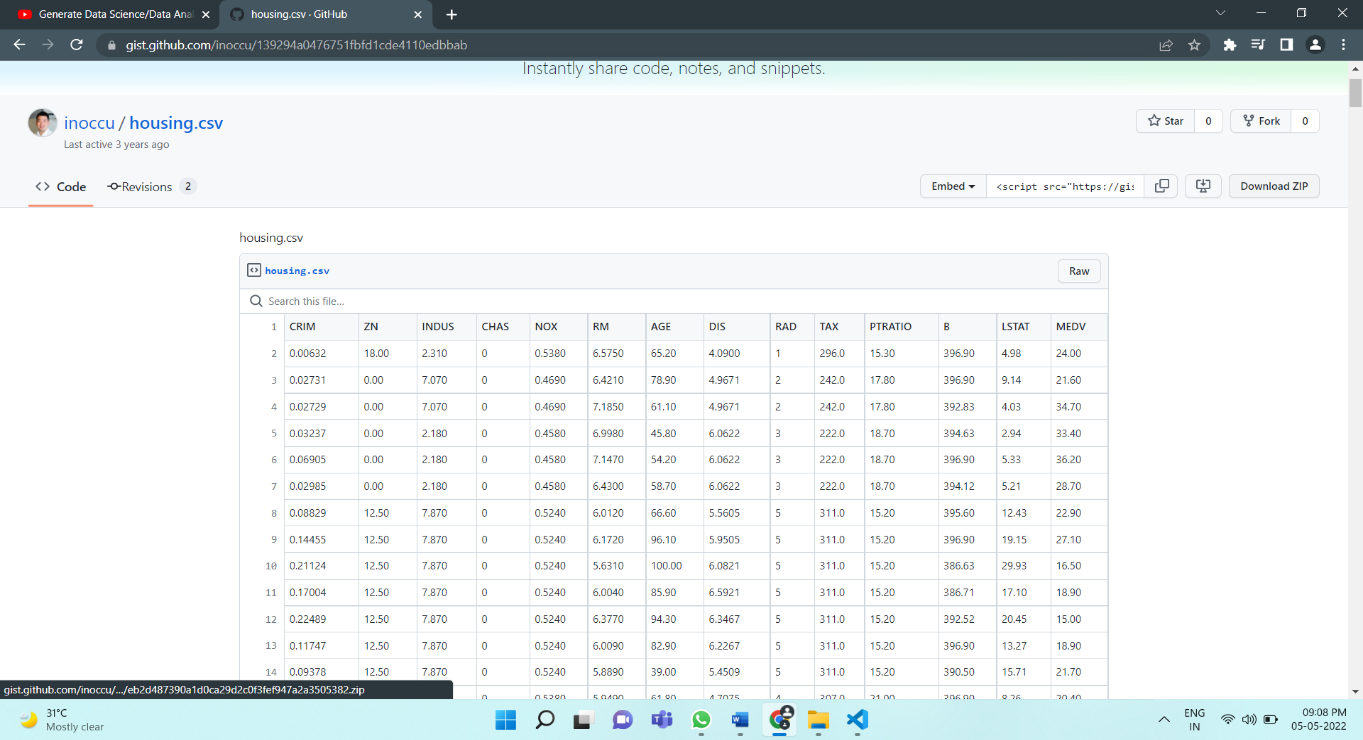
Major project

Step-1: Downloading any data set from Browser.

\*Open your browser like google chrome and search “housing.csv”.

\*Then download data set of housing.





Step-2: Import pandas package.{‘pip install pandas-profiling’}

Step-3: Enter the below code….

#Importing pandas package

import pandas as pd

#Importing pandas-profiling package

from pandas\_profiling import ProfileReport

#Reading csv file (Data set file)

df = pd.read\_csv('housing.csv')

#printing csv file (Data set file)

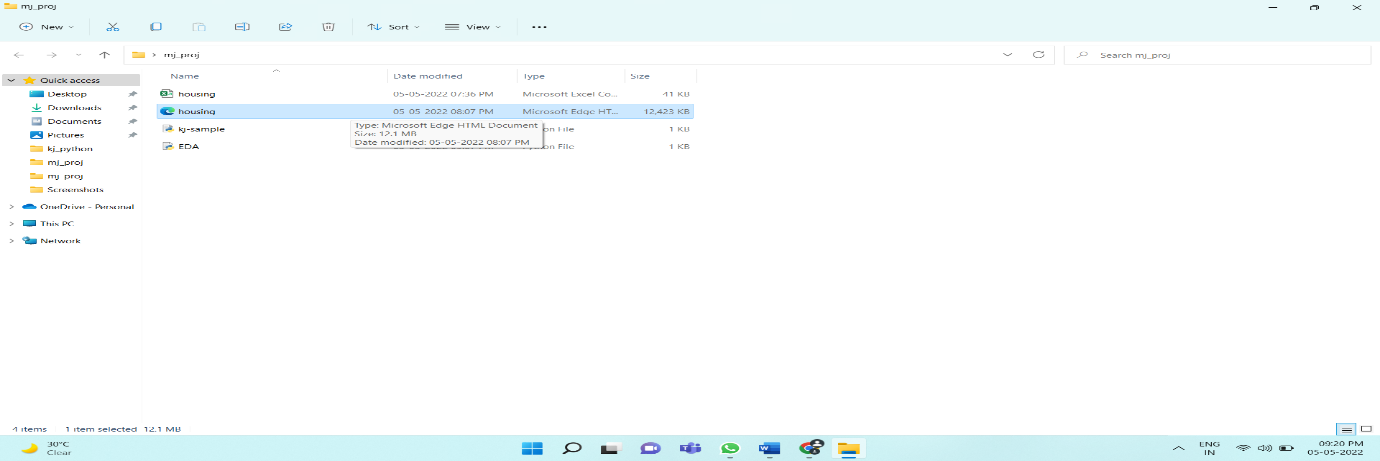
print(df)

#Analaysis in html file

profile = ProfileReport(df)

profile.to\_file(output\_file="housing.html")

A html file of “housing” is generated….



Open the housing html file which is our required “EDA”.

Here is the file….



The above file is the analysis of dataset {Housing}which we have downloaded from browser.

[Pandas Profiling Report](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#top)

* [Overview](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#overview)
* [Variables](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#variables)
* [Interactions](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#interactions)
* [Correlations](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#correlations)
* [Missing values](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#missing)
* [Sample](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#sample)

# Overview

* [Overview](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#overview-dataset_overview)
* [Alerts **48**](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#overview-alerts)
* [Reproduction](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#overview-reproduction)

Dataset statistics

|  |  |
| --- | --- |
| **Number of variables** | 14 |
| **Number of observations** | 506 |
| **Missing cells** | 0 |
| **Missing cells (%)** | 0.0% |
| **Duplicate rows** | 0 |
| **Duplicate rows (%)** | 0.0% |
| **Total size in memory** | 55.5 KiB |
| **Average record size in memory** | 112.3 B |

Variable types

|  |  |
| --- | --- |
| **Numeric** | 13 |
| **Categorical** | 1 |

# Variables

[CRIM](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_-3773152339091776607)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 504 |
| **Distinct (%)** | 99.6% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 3.613523557 |
| **Minimum** | 0.00632 |
| **Maximum** | 88.9762 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[ZN](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_-4486504895041887215)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
ZEROS

|  |  |
| --- | --- |
| **Distinct** | 26 |
| **Distinct (%)** | 5.1% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 11.36363636 |
| **Minimum** | 0 |
| **Maximum** | 100 |
| **Zeros** | 372 |
| **Zeros (%)** | 73.5% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[INDUS](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_7216159658809976900)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 76 |
| **Distinct (%)** | 15.0% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 11.13677866 |
| **Minimum** | 0.46 |
| **Maximum** | 27.74 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[CHAS](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_5930626085856088489)  
Categorical

|  |  |
| --- | --- |
| **Distinct** | 2 |
| **Distinct (%)** | 0.4% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Memory size** | 4.1 KiB |
| 0 | | 471 |
| 1 | | 35 |

Toggle details

[NOX](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_5153280459864537968)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 81 |
| **Distinct (%)** | 16.0% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 0.5546950593 |
| **Minimum** | 0.385 |
| **Maximum** | 0.871 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[RM](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_-3589895829841017829)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 446 |
| **Distinct (%)** | 88.1% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 6.284634387 |
| **Minimum** | 3.561 |
| **Maximum** | 8.78 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[AGE](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_6261799268731839478)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 356 |
| **Distinct (%)** | 70.4% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 68.57490119 |
| **Minimum** | 2.9 |
| **Maximum** | 100 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[DIS](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_-2468286617200742458)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 412 |
| **Distinct (%)** | 81.4% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 3.795042688 |
| **Minimum** | 1.1296 |
| **Maximum** | 12.1265 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[RAD](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_194995499330958695)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 9 |
| **Distinct (%)** | 1.8% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 9.549407115 |
| **Minimum** | 1 |
| **Maximum** | 24 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[TAX](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_6174113874344365254)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 66 |
| **Distinct (%)** | 13.0% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 408.2371542 |
| **Minimum** | 187 |
| **Maximum** | 711 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[PTRATIO](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_-3282920435940328703)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 46 |
| **Distinct (%)** | 9.1% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 18.4555336 |
| **Minimum** | 12.6 |
| **Maximum** | 22 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[B](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_3080703646425874869)  
Real number (ℝ≥0)

HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 357 |
| **Distinct (%)** | 70.6% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 356.6740316 |
| **Minimum** | 0.32 |
| **Maximum** | 396.9 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[LSTAT](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_-6592746111590702744)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 455 |
| **Distinct (%)** | 89.9% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 12.65306324 |
| **Minimum** | 1.73 |
| **Maximum** | 37.97 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

[MEDV](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#pp_var_3063150975125133597)  
Real number (ℝ≥0)

HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION  
HIGH CORRELATION

|  |  |
| --- | --- |
| **Distinct** | 229 |
| **Distinct (%)** | 45.3% |
| **Missing** | 0 |
| **Missing (%)** | 0.0% |
| **Infinite** | 0 |
| **Infinite (%)** | 0.0% |
| **Mean** | 22.53280632 |
| **Minimum** | 5 |
| **Maximum** | 50 |
| **Zeros** | 0 |
| **Zeros (%)** | 0.0% |
| **Negative** | 0 |
| **Negative (%)** | 0.0% |
| **Memory size** | 4.1 KiB |

Toggle details

# Interactions





# Correlations

Toggle correlation descriptions

* [Spearman's ρ](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#correlations_tab-spearman)
* [Pearson's r](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#correlations_tab-pearson)
* [Kendall's τ](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#correlations_tab-kendall)
* [Phik (φk)](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#correlations_tab-phi_k)

# Missing values

* [Count](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#missing-bar)
* [Matrix](file:///C:\Users\SHAIKR~1\AppData\Local\Temp\%7BBDC325D4-CF19-4479-8A18-4B1DCC9F239A%7D\housing.html#missing-matrix)

A simple visualization of nullity by column.

# Sample

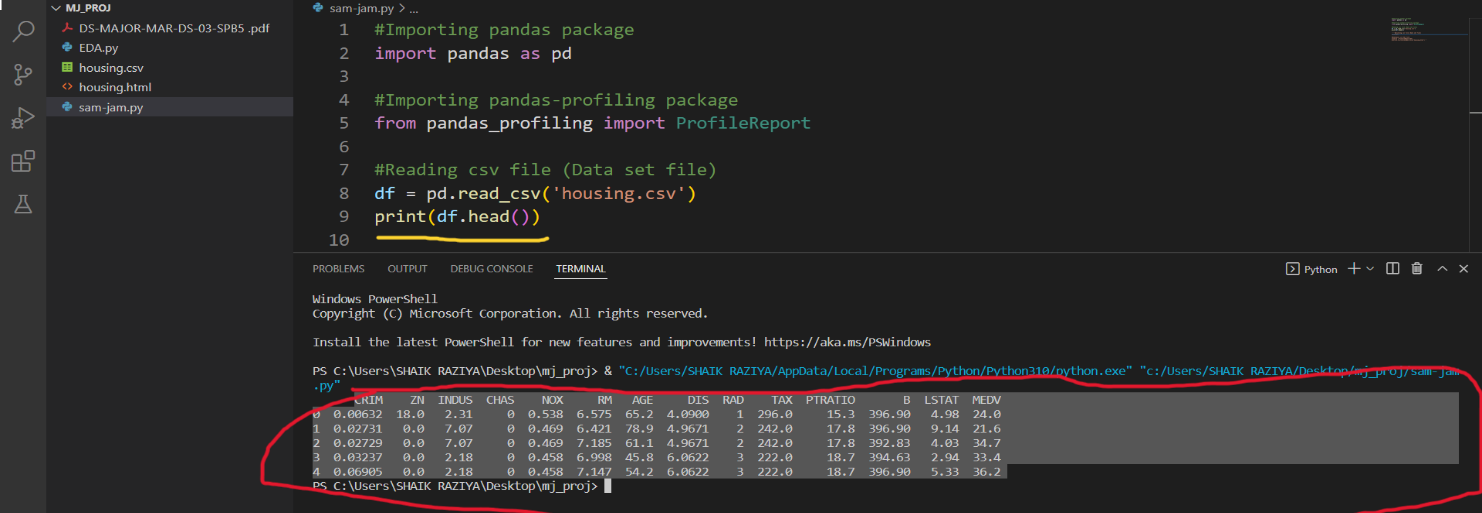
## First rows

|  | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LSTAT** | **MEDV** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 0.00632 | 18.0 | 2.31 | 0 | 0.538 | 6.575 | 65.2 | 4.0900 | 1 | 296.0 | 15.3 | 396.90 | 4.98 | 24.0 |
| **1** | 0.02731 | 0.0 | 7.07 | 0 | 0.469 | 6.421 | 78.9 | 4.9671 | 2 | 242.0 | 17.8 | 396.90 | 9.14 | 21.6 |
| **2** | 0.02729 | 0.0 | 7.07 | 0 | 0.469 | 7.185 | 61.1 | 4.9671 | 2 | 242.0 | 17.8 | 392.83 | 4.03 | 34.7 |
| **3** | 0.03237 | 0.0 | 2.18 | 0 | 0.458 | 6.998 | 45.8 | 6.0622 | 3 | 222.0 | 18.7 | 394.63 | 2.94 | 33.4 |
| **4** | 0.06905 | 0.0 | 2.18 | 0 | 0.458 | 7.147 | 54.2 | 6.0622 | 3 | 222.0 | 18.7 | 396.90 | 5.33 | 36.2 |
| **5** | 0.02985 | 0.0 | 2.18 | 0 | 0.458 | 6.430 | 58.7 | 6.0622 | 3 | 222.0 | 18.7 | 394.12 | 5.21 | 28.7 |
| **6** | 0.08829 | 12.5 | 7.87 | 0 | 0.524 | 6.012 | 66.6 | 5.5605 | 5 | 311.0 | 15.2 | 395.60 | 12.43 | 22.9 |
| **7** | 0.14455 | 12.5 | 7.87 | 0 | 0.524 | 6.172 | 96.1 | 5.9505 | 5 | 311.0 | 15.2 | 396.90 | 19.15 | 27.1 |
| **8** | 0.21124 | 12.5 | 7.87 | 0 | 0.524 | 5.631 | 100.0 | 6.0821 | 5 | 311.0 | 15.2 | 386.63 | 29.93 | 16.5 |
| **9** | 0.17004 | 12.5 | 7.87 | 0 | 0.524 | 6.004 | 85.9 | 6.5921 | 5 | 311.0 | 15.2 | 386.71 | 17.10 | 18.9 |

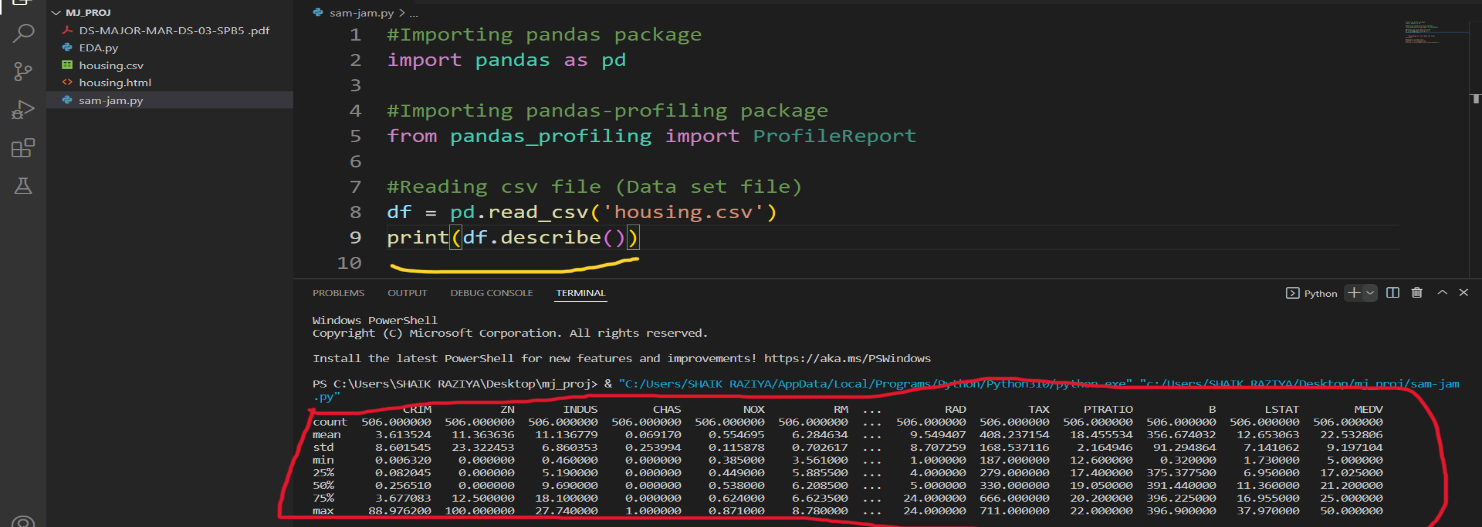
## Last rows

|  | **CRIM** | **ZN** | **INDUS** | **CHAS** | **NOX** | **RM** | **AGE** | **DIS** | **RAD** | **TAX** | **PTRATIO** | **B** | **LSTAT** | **MEDV** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **496** | 0.28960 | 0.0 | 9.69 | 0 | 0.585 | 5.390 | 72.9 | 2.7986 | 6 | 391.0 | 19.2 | 396.90 | 21.14 | 19.7 |
| **497** | 0.26838 | 0.0 | 9.69 | 0 | 0.585 | 5.794 | 70.6 | 2.8927 | 6 | 391.0 | 19.2 | 396.90 | 14.10 | 18.3 |
| **498** | 0.23912 | 0.0 | 9.69 | 0 | 0.585 | 6.019 | 65.3 | 2.4091 | 6 | 391.0 | 19.2 | 396.90 | 12.92 | 21.2 |
| **499** | 0.17783 | 0.0 | 9.69 | 0 | 0.585 | 5.569 | 73.5 | 2.3999 | 6 | 391.0 | 19.2 | 395.77 | 15.10 | 17.5 |
| **500** | 0.22438 | 0.0 | 9.69 | 0 | 0.585 | 6.027 | 79.7 | 2.4982 | 6 | 391.0 | 19.2 | 396.90 | 14.33 | 16.8 |
| **501** | 0.06263 | 0.0 | 11.93 | 0 | 0.573 | 6.593 | 69.1 | 2.4786 | 1 | 273.0 | 21.0 | 391.99 | 9.67 | 22.4 |
| **502** | 0.04527 | 0.0 | 11.93 | 0 | 0.573 | 6.120 | 76.7 | 2.2875 | 1 | 273.0 | 21.0 | 396.90 | 9.08 | 20.6 |
| **503** | 0.06076 | 0.0 | 11.93 | 0 | 0.573 | 6.976 | 91.0 | 2.1675 | 1 | 273.0 | 21.0 | 396.90 | 5.64 | 23.9 |
| **504** | 0.10959 | 0.0 | 11.93 | 0 | 0.573 | 6.794 | 89.3 | 2.3889 | 1 | 273.0 | 21.0 | 393.45 | 6.48 | 22.0 |
| **505** | 0.04741 | 0.0 | 11.93 | 0 | 0.573 | 6.030 | 80.8 | 2.5050 | 1 | 273.0 | 21.0 | 396.90 | 7.88 | 11.9 |

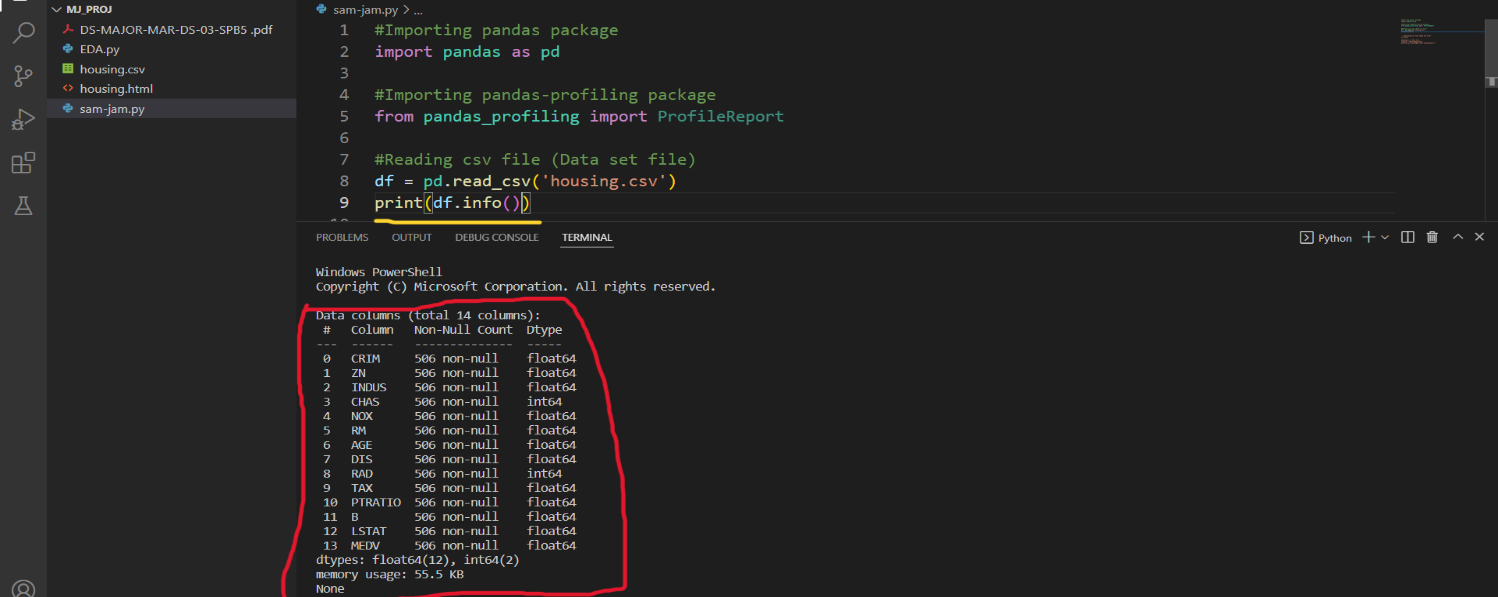
Report generated with [pandas-profiling](https://github.com/pandas-profiling/pandas-profiling).

#df.head()

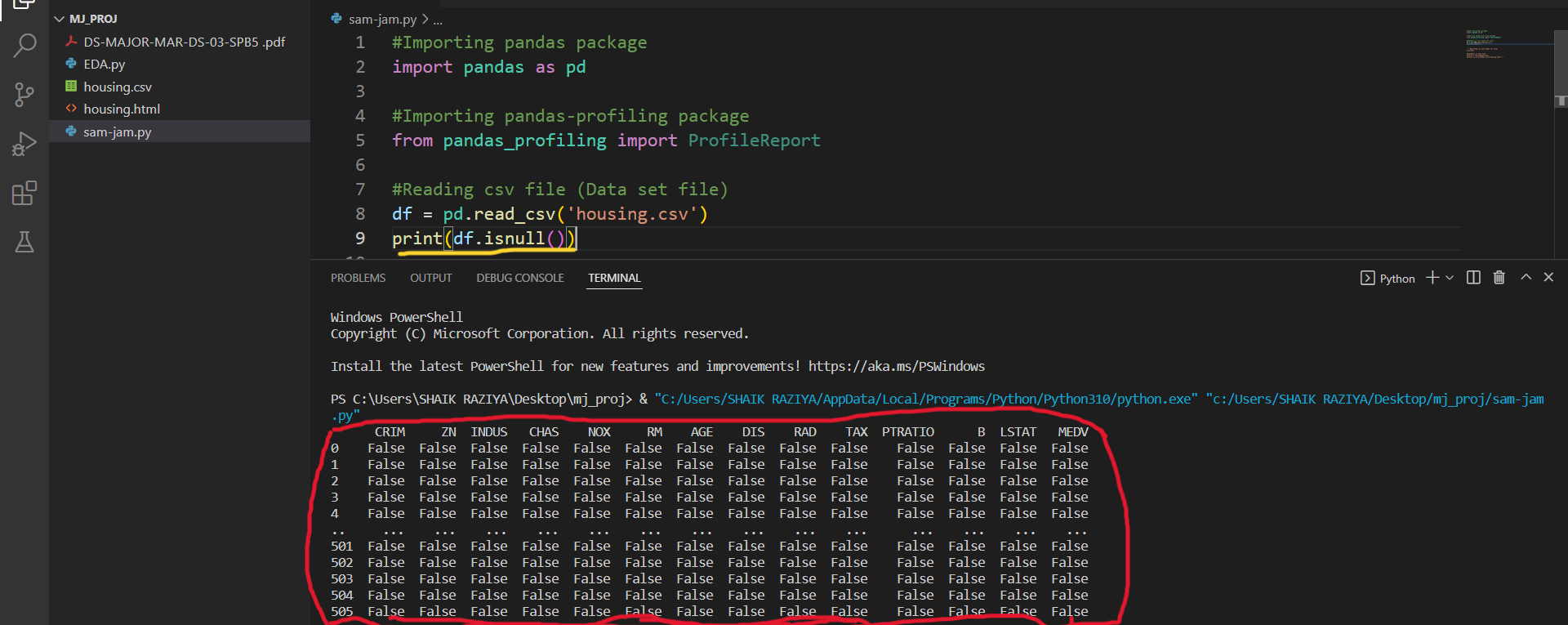
#df.describe()



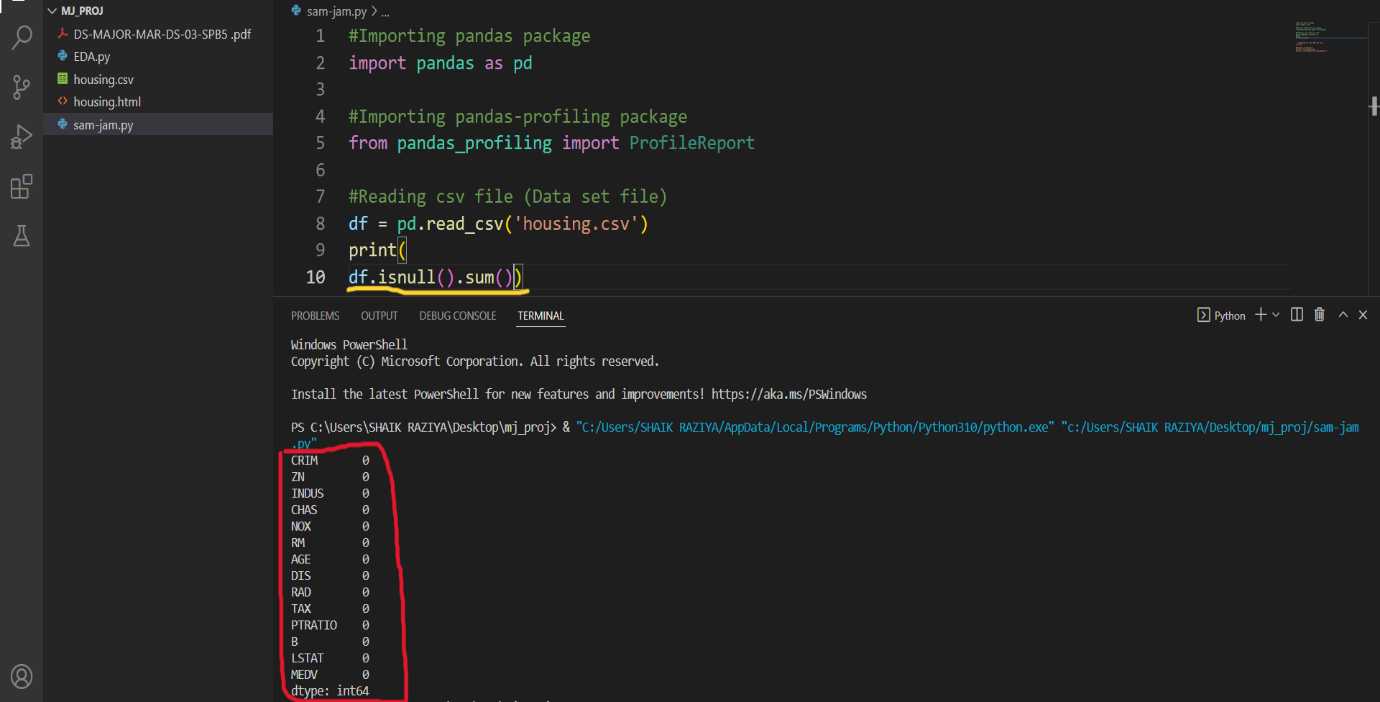
#df.info()



df.isnull()



df.isnull().sum()



PROJECT SUBMITTED TO:

[event@teachnook.com](mailto:event@teachnook.com)

GOOGLE CLASSROOM NAME:

DS-03-SPB1

PROJECT SUBMITTED BY:

SHAIK.KHAJA MOHIDDIN

SHAIK.REHANA

MOBILE NO:

6303709931

9392479176

EMAIL ID:

[shaikkhajamohiddin970@gmail.com](mailto:shaikkhajamohiddin970@gmail.com)

[shaikmysoorvali@gmail.com](mailto:shaikmysoorvali@gmail.com)