

Interactive-1

Interrupt

Clear All

Restart

Variables

Save

Export

Expand

Collapse

Python 3.11.7

import numpy as np

df = pd.read\_csv("G:/Pratik Profile PC/Download/California Lawyers/lawyers\_california.csv")

df

	address	infoSnippet	name	phone	rating	reviewSnippet	url
0	NaN	NaN	LemonBuyBack.com-Law Offices of Jon Jacobs	(916) 626-4164	5.0	NaN	https://www.yellowpages.com/nationwide/mip/lem...
1	NaN	NaN	Legal Aid Legal Services Corp	(844) 330-7981	NaN	NaN	https://www.yellowpages.com/nationwide/mip/leg... http://www.legalaidleg
2	NaN	NaN	LemonBuyBack.com-Law Offices of Jon Jacobs	(916) 626-4164	5.0	Bought a motor home in Fresno, CA on 7/21/18. ...	https://www.yellowpages.com/nationwide/mip/lem... http://lemon
3	2339 Gold Meadow Way, Gold River, CA 95670	At BPE Law Group, we specialize in addressing ...	BPE Law Group	(916) 546-4978	NaN	NaN	https://www.yellowpages.com/gold-river-ca/mip/... http://www
4	2344 Tulare St, Fresno, CA 93721	NaN	Cowin, William L	(559) 445-1234	NaN	NaN	https://www.yellowpages.com/fresno-ca/mip/cowi...
...	...	...	...	...	...	...	...
776	225 S Lake Ave Ste 300, Pasadena, CA 91101	NaN	The Law Offices of Gerald L. Marcus	(626) 396-1118	5.0	I had never been involved in an accident befor...	https://www.yellowpages.com/pasadena-ca/mip/th... http://www.geraldmarcuslaw.com/P
777	10 Keller St, Petaluma, CA 94952	You work hard to safeguard your family's best ...	Zimmerman Law Office	(707) 763-9525	NaN	NaN	https://www.yellowpages.com/petaluma-ca/mip/zi...
778	124 Howard St, Petaluma, CA 94952	1. Association Collaborative Council of the Re...	Martie H. Leys	(707) 789-0390	NaN	NaN	https://www.yellowpages.com/petaluma-ca/mip/ma...

Type 'python' code here and press Shift+Enter to run

Spaces: 4

Cell 3 of 22

Go Live

Codeium: [...]

01:11 AM  
05-03-2024

Interactive-1

Interrupt

Clear All

Restart

Variables

Save

Export

Expand

Collapse

Python 3.11.7

# Check for missing values  
print("\nMissing values in each column:")  
print(df.isnull().sum())

0.0s

Missing values in each column:  
address 43  
infoSnippet 387  
name 0  
phone 1  
rating 620  
reviewSnippet 627  
url 0  
website 159  
dtype: int64

# Data cleaning: Handling missing values  
# Fill missing values in 'address' column with a placeholder (e.g., "Unknown")  
df['address'].fillna("Unknown", inplace=True)

0.0s

# Fill missing values in 'infoSnippet' column with a placeholder (e.g., "No information available")  
df['infoSnippet'].fillna("No information available", inplace=True)

0.0s

# Since 'phone', 'name', and 'url' have only a few missing values, we might drop those rows.  
df.dropna(subset=['phone', 'name', 'url'], inplace=True)

0.0s

# For 'rating' column, you might use a specific value to indicate missing ratings (e.g., -1)  
df['rating'].fillna(-1, inplace=True)

0.0s

df['reviewSnippet'].fillna("No review available", inplace=True)

0.0s

df['website'].fillna("No website available", inplace=True)

0.0s

Type 'python' code here and press Shift+Enter to run

Spaces: 4

Cell 4 of 22

Go Live

Codeium: [...]

01:11 AM  
05-03-2024

Interactive-1 x

Interrupt | Clear All | Restart | Variables | Save | Export | Expand | Collapse | Python 3.11.7

```
print("\nMissing values after handling:")
print(df.isnull().sum())
```

[64] ✓ 0.0s

```
...
Missing values after handling:
address      0
infoSnippet  0
name         0
phone        0
rating       0
reviewSnippet 0
url          0
website      0
dtype: int64
```

```
# Generate descriptive statistics for numerical columns
numerical_stats = df['rating'].describe()
print("\nDescriptive statistics for ratings:")
print(numerical_stats)
```

[65] ✓ 0.0s

```
...
Descriptive statistics for ratings:
count    780.000000
mean     1.753846
std      1.535309
min      1.000000
25%      1.000000
50%      1.000000
75%      1.000000
max       5.000000
Name: rating, dtype: float64
```

```
df
```

[66] ✓ 0.0s

```
...

```

	address	infoSnippet	name	phone	rating	reviewSnippet	url	
0	Unknown	No information available	LemonBuyBack.com-Law Offices of Jon Jacobs	(916) 626-4164	5.0	No review available	https://www.yellowpages.com/nationwide/mip/lem...	No wet
1	Unknown	No information available	Legal Aid Legal Services Corp	(844) 330-7981	1.0	No review available	https://www.yellowpages.com/nationwide/mip/leg...	http://www.legalaidleg
2	Unknown	No information available	LemonBuyBack.com-Law Offices of Jon Jacobs	(916) 626-4164	5.0	Bought a motor home in Fresno, CA on	https://www.yellowpages.com/nationwide/mip/lem...	http://lemon

Type 'python' code here and press Shift+Enter to run

Spaces: 4 Cell 4 of 22 Go Live Codeium: [..] 01:12 AM 05-03-2024

Interactive-1 x

Interrupt | Clear All | Restart | Variables | Save | Export | Expand | Collapse | Python 3.11.7

```
# Drop irrelevant columns
columns_to_drop = ['address', 'infoSnippet', 'reviewSnippet', 'phone']
df.drop(columns=columns_to_drop, inplace=True)
```

[67] ✓ 0.0s

```
df
```

[68] ✓ 0.0s

```
...

```

	name	rating	url	website
0	LemonBuyBack.com-Law Offices of Jon Jacobs	5.0	https://www.yellowpages.com/nationwide/mip/lem...	No website available
1	Legal Aid Legal Services Corp	1.0	https://www.yellowpages.com/nationwide/mip/leg...	http://www.legalaidlegalservices.org
2	LemonBuyBack.com-Law Offices of Jon Jacobs	5.0	https://www.yellowpages.com/nationwide/mip/lem...	http://lemonbuyback.com
3	BPE Law Group	1.0	https://www.yellowpages.com/gold-river-ca/mip/...	http://www.bpelaw.com
4	Cowin, William L	1.0	https://www.yellowpages.com/fresno-ca/mip/cowi...	No website available
...	...	...	...	...
776	The Law Offices of Gerald L. Marcus	5.0	https://www.yellowpages.com/pasadena-ca/mip/th...	http://www.geraldmarcuslaw.com/Pasadena.aspx
777	Zimmerman Law Office	1.0	https://www.yellowpages.com/petaluma-ca/mip/zi...	No website available
778	Martie H. Leys	1.0	https://www.yellowpages.com/petaluma-ca/mip/ma...	No website available
779	Patrick H. Dwyer Attorney At Law	1.0	https://www.yellowpages.com/penn-valley-ca/mip/...	No website available
780	Norman J Ryker Attorney At Law	1.0	https://www.yellowpages.com/chico-ca/mip/norma...	No website available

780 rows x 4 columns

```
# Drop irrelevant columns if they exist in the DataFrame
columns_to_drop = ['address', 'infoSnippet', 'reviewSnippet', 'phone', 'url', 'website']
columns_to_drop = [col for col in columns_to_drop if col in df.columns] # Filter out columns that do not exist

# Drop the columns if they exist
if columns_to_drop:
    df.drop(columns_to_drop, axis=1, inplace=True)
    print("Columns dropped:", columns_to_drop)
else:
    print("Specified columns do not exist in the DataFrame.")

# Display the first few rows of the updated DataFrame
print("\nUpdated data after dropping irrelevant columns:")
print(df.head())
```

[71] ✓ 0.0s

```
...
Columns dropped: ['url', 'website']

Updated data after dropping irrelevant columns:

```

	name	rating
--	------	--------

Type 'python' code here and press Shift+Enter to run

Spaces: 4 Cell 4 of 22 Go Live Codeium: [..] 01:12 AM 05-03-2024



