

```
!pip install ydata-profiling
```

```
Requirement already satisfied: ydata-profiling in /usr/local/lib/python3.12/dist-packages (4.16.1)
Requirement already satisfied: scipy<1.16,>=1.4.1 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (1.15.3)
Requirement already satisfied: pandas!=1.4.0,<3.0,>1.1 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (2.2.2)
Requirement already satisfied: matplotlib<3.10,>=3.5 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (3.10.0)
Requirement already satisfied: pydantic>=2 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (2.11.7)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (6.0.2)
Requirement already satisfied: Jinja2<3.2,>=2.11.1 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (3.1.6)
Requirement already satisfied: visions<0.8.2,>=0.7.5 in /usr/local/lib/python3.12/dist-packages (from visions[type_image_path]<0.8.2)
Requirement already satisfied: numpy<2.2,>=1.16.0 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (2.0.2)
Requirement already satisfied: htmlmin==0.1.12 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (0.1.12)
Requirement already satisfied: phik<0.13,>=0.11.1 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (0.12.5)
Requirement already satisfied: requests<3,>=2.24.0 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (2.32.4)
Requirement already satisfied: tqdm<5,>=4.48.2 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (4.67.1)
Requirement already satisfied: seaborn<0.14,>=0.10.1 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (0.13.2)
Requirement already satisfied: multimethod<2,>=1.4 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (1.12)
Requirement already satisfied: statsmodels<1,>=0.13.2 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (0.14.5)
Requirement already satisfied: typeguard<5,>=3 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (4.4.4)
Requirement already satisfied: imagehash==4.3.1 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (4.3.1)
Requirement already satisfied: wordcloud==1.9.3 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (1.9.4)
Requirement already satisfied: dacite>=1.8 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (1.9.2)
Requirement already satisfied: numba<=0.61,>=0.56.0 in /usr/local/lib/python3.12/dist-packages (from ydata-profiling) (0.60.0)
Requirement already satisfied: PyWavelets in /usr/local/lib/python3.12/dist-packages (from imagehash==4.3.1->ydata-profiling) (1.9.0)
Requirement already satisfied: pillow in /usr/local/lib/python3.12/dist-packages (from imagehash==4.3.1->ydata-profiling) (11.3.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from Jinja2<3.2,>=2.11.1->ydata-profiling)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<3.10,>=3.5->ydata-profiling)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.12/dist-packages (from matplotlib<3.10,>=3.5->ydata-profiling)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib<3.10,>=3.5->ydata-profiling)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<3.10,>=3.5->ydata-profiling)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib<3.10,>=3.5->ydata-profiling)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib<3.10,>=3.5->ydata-profiling)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.12/dist-packages (from matplotlib<3.10,>=3.5->ydata-profiling)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.12/dist-packages (from numba<=0.61,>=0.56.0->ydata-profiling)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas!=1.4.0,<3.0,>1.1->ydata-profiling)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas!=1.4.0,<3.0,>1.1->ydata-profiling)
Requirement already satisfied: joblib>=0.14.1 in /usr/local/lib/python3.12/dist-packages (from phik<0.13,>=0.11.1->ydata-profiling)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->ydata-profiling)
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->ydata-profiling)
Requirement already satisfied: typing-extensions>=4.12.2 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->ydata-profiling)
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.12/dist-packages (from pydantic>=2->ydata-profiling)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24.0->ydata-profiling)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24.0->ydata-profiling)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24.0->ydata-profiling)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3,>=2.24.0->ydata-profiling)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.12/dist-packages (from statsmodels<1,>=0.13.2->ydata-profiling)
Requirement already satisfied: attr>=19.3.0 in /usr/local/lib/python3.12/dist-packages (from visions<0.8.2,>=0.7.5->visions[type_image_path])
Requirement already satisfied: networkx>=2.4 in /usr/local/lib/python3.12/dist-packages (from visions<0.8.2,>=0.7.5->visions[type_image_path])
Requirement already satisfied: puremagic in /usr/local/lib/python3.12/dist-packages (from visions<0.8.2,>=0.7.5->visions[type_image_path])
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.7->matplotlib<3.10,>=3.5)
```

```
import pandas as pd
from ydata_profiling import ProfileReport
```

```
df = pd.read_csv('retailsales.csv')
```

```
df.head()
```

	Transaction ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount
0	1	2023-11-24	CUST001	Male	34	Beauty	3	50	150
1	2	2023-02-27	CUST002	Female	26	Clothing	2	500	1000
2	3	2023-01-13	CUST003	Male	50	Electronics	1	30	30
3	4	2023-05-21	CUST004	Male	37	Clothing	1	500	500
4	5	2023-05-06	CUST005	Male	30	Beauty	2	50	100

```
profile = ProfileReport(df, title = 'Retail Sales EDA Report', explorative= True)
profile.to_file('retail.html')
```

```
Summarize dataset: 0%|          | 0/5 [00:00<?, ?it/s]
0%|          | 0/9 [00:00<?, ?it/s]
100%|██████████| 9/9 [00:00<00:00, 45.29it/s]
Generate report structure: 0%|          | 0/1 [00:00<?, ?it/s]
Render HTML: 0%|          | 0/1 [00:00<?, ?it/s]
Export report to file: 0%|          | 0/1 [00:00<?, ?it/s]
```

```
profile.to_notebook_iframe()
```



Retail Sales EDA Report

Overview Variables Interactions Correlations Missing values Sample

Dataset statistics		Variable types	
Number of variables	9	Numeric	3
Number of observations	1000	DateTime	1
Missing cells	0	Text	1
Missing cells (%)	0.0%	Categorical	4
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	260.3 KiB		
Average record size in memory	266.6 B		

Variables

Select Columns

Transaction ID

Real number (\mathbb{R})

Uniform Unique

Distinct	1000	Minimum	1
Distinct (%)	100.0%	Maximum	1000

```
# Get variable descriptions as dict, convert to DF, export

# Get the report data as a JSON string and convert to dictionary
import json
report_json = profile.to_json() # Get JSON representation
report_dict = json.loads(report_json) # Parse JSON string to dictionary

# Extract variables section
var_desc = report_dict.get('variables', {}) # Safely get 'variables' or empty dict if not found

# Convert to DataFrame and transpose
import pandas as pd
var_df = pd.DataFrame(var_desc).T # Transpose to make columns as rows

# Save to CSV
var_df.to_csv('retail_insights.csv') # Export (includes means, missing %, etc.)
```

Render JSON: 100% 1/1 [00:00<00:00, 13.16it/s]

```
# Get the report data
import json
report_json = profile.to_json()
report_dict = json.loads(report_json)

# Extract alerts section
alerts = report_dict.get('alerts', []) # Safely get 'alerts' or empty list

# Convert alerts to DataFrame
import pandas as pd
alerts_df = pd.DataFrame(alerts) # Use directly if alerts are dicts

# Save to CSV
alerts_df.to_csv('alerts_insights.csv')
```

