```
In [ ]:   #Name: Nupur Kuhite
          #Roll no.: AD22026
          #DA Post Lab
          #Aim: Profiling Large Datasets with Pandas Profiling:
          #     Automating data profiling for large datasets using Pandas Profiling.
```

```
In [5]:   import pandas as pd
          import seaborn as sns
          from ydata_profiling import ProfileReport
```

```
In [12]:  df = sns.load_dataset('titanic')
          df
```

Out[12]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who
0	0	3	male	22.0	1	0	7.2500	S	Third	man
1	1	1	female	38.0	1	0	71.2833	C	First	woman
2	1	3	female	26.0	0	0	7.9250	S	Third	woman
3	1	1	female	35.0	1	0	53.1000	S	First	woman
4	0	3	male	35.0	0	0	8.0500	S	Third	man
...	...	...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S	Second	man
887	1	1	female	19.0	0	0	30.0000	S	First	woman
888	0	3	female	NaN	1	2	23.4500	S	Third	woman
889	1	1	male	26.0	0	0	30.0000	C	First	man
890	0	3	male	32.0	0	0	7.7500	Q	Third	man

891 rows × 15 columns

```
In [13]:  profile = ProfileReport(df, title="Titanic Dataset Profiling", explorative=True)
```

```
In [11]:  profile.to_file("titanic_dataset_profile.html")
```

```
Summarize dataset:  77%|████████████████████          | 17/22 [00:01<00:00,
6.16it/s, Calculate auto correlation]C:\Users\nupur\AppData\Local\Packages\Python
SoftwareFoundation.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311
\site-packages\ydata_profiling\model\correlations.py:66: UserWarning: There was a
n attempt to calculate the auto correlation, but this failed.
To hide this warning, disable the calculation
(using `df.profile_report(correlations={"auto": {"calculate": False}})`
If this is problematic for your use case, please report this as an issue:
https://github.com/ydataai/ydata-profiling/issues
(include the error message: 'Function <code object pandas_auto_compute at 0x00000
15F3A696630, file "C:\Users\nupur\AppData\Local\Packages\PythonSoftwareFoundatio
n.Python.3.11_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site-packages\yda
ta_profiling\model\pandas\correlations_pandas.py", line 167>')
  warnings.warn(
Summarize dataset: 100%|███████████████████████████████| 41/
41 [00:07<00:00,  5.31it/s, Completed]
Generate report structure: 100%|███████████████████████
███████| 1/1 [00:09<00:00,  9.24s/it]
Render HTML: 100%|█████████████████████████████████████
███████| 1/1 [00:02<00:00,  2.91s/it]
Export report to file: 100%|███████████████████████████
███████| 1/1 [00:00<00:00, 46.44it/s]
```