EDA AUTOMATION TOOL

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df=pd.read_csv("D:/Downloads/archive (20)/tested.csv")
```

In [4]:
```
df.head()
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	

ProfileReport

In [ ]:
```
from ydata_profiling import ProfileReport

ProfileReport(df, title="My Data Profile").to_file("report.html")
```

SWEETVIZ

In [ ]:
```
pip install sweetviz
```

In [ ]:
```
import sweetviz as sv
```

In [ ]:
```
df=pd.read_csv("D:/Downloads/archive (20)/tested.csv")
my_report = sv.analyze(df)
my_report.show_html()
```

```
Benefits of EDA Automation Tools:

    Efficiency: Automation tools can significantly reduce the time and
effort required to perform EDA tasks, especially for large and complex
datasets. They can quickly generate summary statistics, visualizations,
and insights, allowing analysts to focus on interpreting the results
rather than repetitive tasks.

    Consistency: Automated EDA processes ensure consistency in analysis
across different datasets and users. By following standardized procedures
and algorithms, these tools minimize human errors and biases that may
arise from manual analysis.

    Scalability: Automation tools are well-suited for handling large
volumes of data, making them scalable to diverse applications and
datasets. They can efficiently process extensive datasets that may be
impractical to analyze manually.

    Exploration: Automation tools can facilitate exploratory data analysis
by providing interactive visualizations and intuitive interfaces for data
exploration. They enable users to interactively explore the data, identify
patterns, and gain insights in real-time.

Challenges and Limitations:

    Overreliance: There's a risk of overreliance on automation tools,
leading to a lack of critical thinking and domain expertise in data
analysis. Users may become dependent on automated results without fully
understanding the underlying assumptions or nuances of the data.

    Black Box Nature: Some automation tools operate as "black boxes,"
meaning that users may not have full visibility into the algorithms and
methodologies used. This lack of transparency can make it challenging to
interpret the results accurately or troubleshoot issues effectively.

    Customization: Automation tools may lack flexibility or customization
options to accommodate specific analysis requirements or domain-specific
knowledge. Users may find it challenging to tailor the automated processes
to suit their unique needs or preferences.

    Quality Assurance: Automated EDA processes require rigorous quality
assurance to ensure the accuracy and reliability of the results. Users
must validate the outputs against manual analyses and domain knowledge to
verify their correctness.
```