

EDA by pandas_profiling

In [1]:

```
import pandas as pd
from pandas_profiling import ProfileReport
```

C:\Users\SAMEER\AppData\Local\Temp\ipykernel_19244\3034448462.py:2: DeprecationWarning: `import pandas_profiling` is going to be deprecated by April 1st. Please use `import ydata_profiling` instead.
from pandas_profiling import ProfileReport

In [2]:

```
data = pd.read_csv(r"D:\Sabina\ANN\30th,31st\ANN_ 1st\Churn_Modelling.csv")
```

In [3]:

data

Out[3]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure
0	1	15634602	Hargrave	619	France	Female	42	2
1	2	15647311	Hill	608	Spain	Female	41	1
2	3	15619304	Onio	502	France	Female	42	8
3	4	15701354	Boni	699	France	Female	39	1
4	5	15737888	Mitchell	850	Spain	Female	43	2
...
9995	9996	15606229	Obijiaku	771	France	Male	39	5
9996	9997	15569892	Johnstone	516	France	Male	35	10
9997	9998	15584532	Liu	709	France	Female	36	7
9998	9999	15682355	Sabbatini	772	Germany	Male	42	3
9999	10000	15628319	Walker	792	France	Female	28	4

10000 rows × 14 columns



In [6]:

```
profile.to_file(output_file="Discharge_profiling.html")
profile
```

Export report to file:

1/1 [00:00<00:00,
100%66.69it/s]

Overview

Dataset statistics

Number of variables	14
Number of observations	10000
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	2.6 MiB
Average record size in memory	276.4 B

Variable types

Numeric	7
Text	1
Categorical	6

Alerts

RowNumber is uniformly distributed	Uniform
RowNumber has unique values	Unique

Out[6]:

In []: