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

In [3]:
```
data = pd.read_csv('../data/training_data_clean.csv')
```

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

Out[4]:

	Facies	Formation	WellName	Depth	GR	ILD_log10	DeltaPHI	PHIND	PE	NM_M	RELPOS
0	3	A1 SH	SHRIMPLIN	2793.0	77.45	0.664	9.9	11.915	4.6	1	1.000
1	3	A1 SH	SHRIMPLIN	2793.5	78.26	0.661	14.2	12.565	4.1	1	0.979
2	3	A1 SH	SHRIMPLIN	2794.0	79.05	0.658	14.8	13.050	3.6	1	0.957
3	3	A1 SH	SHRIMPLIN	2794.5	86.10	0.655	13.9	13.115	3.5	1	0.936
4	3	A1 SH	SHRIMPLIN	2795.0	74.58	0.647	13.5	13.300	3.4	1	0.915

In [5]:
```
data.shape
```

Out[5]:
```
(3232, 11)
```

In [6]:
```
data.profile_report(style={'full_width':True})
```

Overview

Dataset info

Number of variables	11
Number of observations	3232
Missing cells	0 (0.0%)
Duplicate rows	1 (< 0.1%)
Total size in memory	277.8 KiB
Average record size in memory	88.0 B

Variables types

Numeric	8
Categorical	3
Boolean	0
Date	0
URL	0
Text (Unique)	0
Rejected	0
Unsupported	0

Warnings

Dataset has 1 (< 0.1%) duplicate rows	**Warning**

Out[6]:

In [6]:
```
profile = data.profile_report()
rejected_variables = profile.get_rejected_variables(threshold=0.9)
```

In [ ]:
```

```