Logo

Exploratory Data Analysis Report

Developed by: Mayank Sharma & Garvit Kedia

Table of Contents

Dataset Overview

Statistic	Value
Shape (Rows, Columns)	(483, 28)

Data Types	['int64', 'float64']

Missing Values

Column Name	Total Missing Values	Percentage Missing
Well	0	0.0%
BHP	0	0.0%
RP	0	0.0%
LFR	0	0.0%
TS	1	0.2%
GF	0	0.0%
SP	7	1.4%
WC	7	1.4%
Po	7	1.4%
PermR	7	1.4%
PI	7	1.4%
BC	0	0.0%
ER	0	0.0%
DT	0	0.0%
NWBZ	0	0.0%
FZ,	0	0.0%
PiezC	0	0.0%
FD	1	0.2%
PipDT	0	0.0%
BP	0	0.0%
OV	0	0.0%
OCF	0	0.0%

VC	0	0.0%
THA	0	0.0%
FR	0	0.0%
WR	0	0.0%
DOP	0	0.0%
Permeability	0	0.0%

Summary Statistics for Numerical Features

Statistic	FD	PipDT	BP
Count	482	483	483
Mean	0.86	1973.42	3.52
Standard Deviation	0.12	183.51	0.87
Minimum	0.68	1486.00	1.78
Q1	0.76	1835.00	2.68
Median (Q2)	0.82	2000.00	3.20
Q3	0.97	2124.00	4.40
Maximum	1.18	2457.00	4.60

Correlation Matrix

	PermR	GF	BHP
PermR	1.00	-0.02	0.07
GF	-0.02	1.00	0.35
BHP	0.07	0.35	1.00

Outliers Summary

Column Name	Outlier Detection Method	Number of Outliers
Well	Z-score	0
Well	IQR Method	14
BHP	Z-score	7
BHP	IQR Method	54
RP	Z-score	3
RP	IQR Method	44
LFR	Z-score	3
LFR	IQR Method	12
TS	Z-score	4
TS	IQR Method	4
GF	Z-score	4
GF	IQR Method	12
SP	Z-score	1
SP	IQR Method	2
WC	Z-score	0
WC	IQR Method	0
Po	Z-score	6
Po	IQR Method	22
PermR	Z-score	8
PermR	IQR Method	68
PI	Z-score	13
PI	IQR Method	65
BC	Z-score	0
BC	IQR Method	0
ER	Z-score	11

ER	IQR Method	83
DT	Z-score	1
DT	IQR Method	1
NWBZ	Z-score	10
NWBZ	IQR Method	60
FZ,	Z-score	23
FZ,	IQR Method	55
PiezC	Z-score	11
PiezC	IQR Method	70
FD	Z-score	0
FD	IQR Method	0
PipDT	Z-score	0
PipDT	IQR Method	0
BP	Z-score	0
BP	IQR Method	0
OV	Z-score	29
OV	IQR Method	189
OCF	Z-score	0
OCF	IQR Method	69
VC	Z-score	0
VC	IQR Method	0
THA	Z-score	4
THA	IQR Method	4
FR	Z-score	9
FR	IQR Method	59
WR	Z-score	7
WR	IQR Method	0
DOP	Z-score	14
DOP	IQR Method	49
Permeability	Z-score	8

Permeability	IQR Method	58

Duplicates

Column Name	Total Duplicates	Percentage of Duplicates
Well	441	91.30%
BHP	209	43.27%
RP	198	40.99%
LFR	197	40.79%
TS	459	95.03%
GF	469	97.10%
SP	472	97.72%
WC	303	62.73%
Po	463	95.86%
PermR	166	34.37%
PI	118	24.43%
BC	476	98.55%
ER	114	23.60%
DT	397	82.19%
NWBZ	90	18.63%
FZ,	76	15.73%
PiezC	56	11.59%
FD	347	71.84%
PipDT	398	82.40%
BP	478	98.96%
OV	473	97.93%
OCF	464	96.07%

VC	468	96.89%
THA	463	95.86%
FR	259	53.62%
WR	483	100.00%
DOP	424	87.78%
Permeability	168	34.78%