

# Code OG

## **Problem Statement:**

To make an AI-powered data quality enhancer.

## **Tasks to perform:**

- 1) Remove inconsistencies
- 2) Deal with anomalies
- 3) Duplicate detection and handling
- 4) Reducing pre-processing time

## **Additional agendas:**

- 1) Selecting methods based on dataset context
- 2) Creating a good user interface
- 3) Deploying on cloud

## **Tasks Achieved:**

- 1) Duplicate Removal
- 2) Filling missing values
- 3) Outlier detection

## **Sample outputs:**

The sample outputs are provided using a sample dataset and the outputs are taken directly from the compiler. This document is taken based on the first phase progress as of 07/02/2025 11:00 pm IST.

Data before:

Home | dataSet\_1corrected.csv | Untitled | dataSet\_1.csv

localhost:8888/edit/archive%20(1)/dataSet\_1.csv

Jupyter | dataSet\_1.csv Last Checkpoint: 3 days ago

File Edit View Settings Help

Delimiter: ,

	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
1	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
2	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
3	No	No	Month-to-month	Yes	Electronic check	53.85	108.15	Yes
4	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No
5	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes
6	Yes	Yes	Month-to-month	Yes	Electronic check	99.65		Yes
7	Yes	No	Month-to-month	Yes	Credit card (automatic)	89.1		No
8	No	No	Month-to-month	No	Mailed check	29.75	301.9	No
9	Yes	Yes	Month-to-month	Yes	Electronic check	abod\$#	3046.05	Yes
10	No	No	One year	No	Bank transfer (automatic)	56.15	3487.95	No
11	No	No	Month-to-month	Yes	Mailed check	49.95	587.45	No
12	No internet service	No internet service	Two year	No	Credit card (automatic)	18.95	326.8	No
13	Yes	Yes	One year	No	Credit card (automatic)	100.35	5681.1	No
14	Yes	Yes	Month-to-month	Yes	Bank transfer (automatic)	103.7	5036.3	Yes
15	Yes	Yes	Month-to-month	Yes	Electronic check	105.5	2686.05	No
16	Yes	Yes	Two year	No	Credit card (automatic)	113.25	7895.15	No
17	No internet service	No internet service	One year	No	Mailed check	20.65	1022.95	No
18	Yes	Yes	Two year	No	Bank transfer (automatic)	106.7	7382.25	No
19	No	No	Month-to-month	No	Credit card (automatic)	55.2	528.35	Yes
20	No	Yes	Month-to-month	Yes	Electronic check	90.05	1862.9	No
21	No	Yes	Month-to-month	Yes	Electronic check	39.65	39.65	Yes
22	No internet service	No internet service	One year	No	Bank transfer (automatic)	19.8	202.25	No
23	No internet service	No internet service	Month-to-month	No	Mailed check	20.15	20.15	Yes
24	No	No	Two year	Yes	Credit card (automatic)	59.9	3505.1	No
25	No	No	Month-to-month	No	Credit card (automatic)	59.6	2970.3	No
26	No	No	Month-to-month	Yes	Bank transfer (automatic)	55.3	1530.6	No
27	Yes	Yes	Month-to-month	Yes	Electronic check	99.14	4749.14	Yes

Type here to search | 24°C Mostly cloudy | 11:07 PM 2/7/2025

Home

dataSet\_1corrected.csv

Untitled

dataSet\_1.csv

<

Home

dataSet\_1corrected.csv

Untitled

dataSet\_1.csv

localhost:8888/notebooks/Untitled.ipynb

Jupyter

Untitled Last Checkpoint: 1 hour ago

File Edit View Run Kernel Settings Help

JupyterLab Python 3 (ipykernel)

```
def main(inputsize_path):
    df = pd.read_csv(file_path)
    Enhancer(inp,df)

def Main_1(df):
    a= clean_dataset(df, method="mean")
    b= Fill_missing_values(a)
    c=detect_outliers_iforest(b, contamination=0.05, n_estimators=100, random_state=42)
    print(c)
    c.to_csv('archive (1)/dataSet_1corrected.csv',index=False)
    Main ('Medical','archive (1)/dataSet_1.csv')
```

Cleaned Dataset:

Original Data with Missing Values:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure
0	7590-VHVEG	Female	0.000000	Yes	No	1.000000
1	7590-VHVEG	Female	0.000000	Yes	No	1.000000
2	3668-QPYBK	Male	0.162147	No	No	2.000000
3	7795-CFDCM	Male	0.162147	No	No	45.000000
4	9237-HQITU	Male	0.162147	No	No	32.366463
...	...	...	...	...	...	...
7038	6840-RESVB	Male	0.162147	No	No	32.366463
7039	2234-XADUM	Male	0.162147	No	No	32.366463
7040	4801-JJAZL	Male	0.162147	No	No	32.366463
7041	8361-LTRMD	Male	0.162147	No	No	32.366463
7042	3186-AJIEK	Male	0.162147	No	No	32.366463

	PhoneService	MultipleLines	InternetService	OnlineSecurity
0	No	No phone service	DSL	No
1	No	No phone service	DSL	No
2	Yes	No	Fiber optic	Yes
3	Yes	No	Fiber optic	No
4	Yes	No	Fiber optic	No
...	...	...	...	...
7038	Yes	No	Fiber optic	No
7039	Yes	No	Fiber optic	No
7040	Yes	No	Fiber optic	No
7041	Yes	No	Fiber optic	No
7042	Yes	No	Fiber optic	No

DeviceProtection TechSupport StreamingTV StreamingMovies Contract