Advance Software Engineering (SE-406)

LAB A1-G3

Laboratory Manual

Department of Software Engineering

DELHI TECHNOLOGICAL UNIVERSITY(DTU)

Shahbad Daulatpur, Bawana Road, Delhi-110042

Submitted to: -

Ms. Parul Sharma

Submitted by:-

Name: ASHISH KUMAR

Roll number: 2K18/SE/041

EXPERIMENT 6

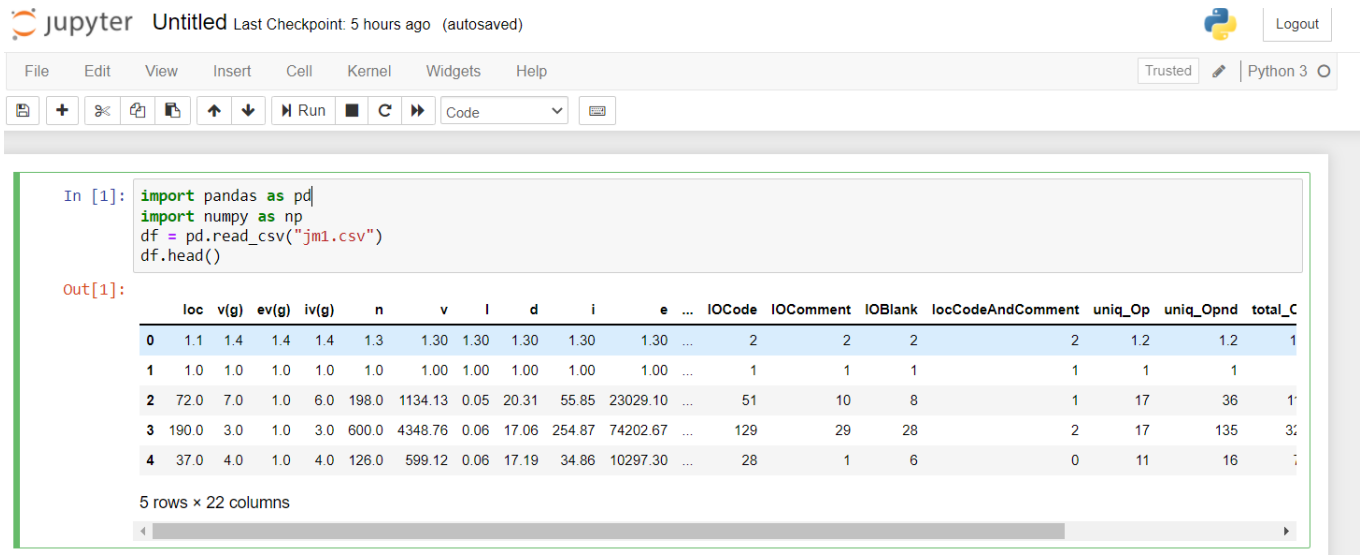
- ASHISH KUMAR
- 2K18/SE/041

Aim:- Preprocessing and cleaning of data collected in experiment 5.

Introduction:- Data Preprocessing is the most important step when we are building our model. In **Data Preprocessing** step, the data is transformed into a form where it becomes suitable for model ingestion. **Data Cleaning** is the process of analyzing data for finding incorrect, corrupt, and missing values and ablating it to make it suitable for input to data analytics and various machine learning algorithms.

Code & Output:-

```
import pandas as pd
import numpy as np
df = pd.read_csv("jm1.csv")
df.head()
```

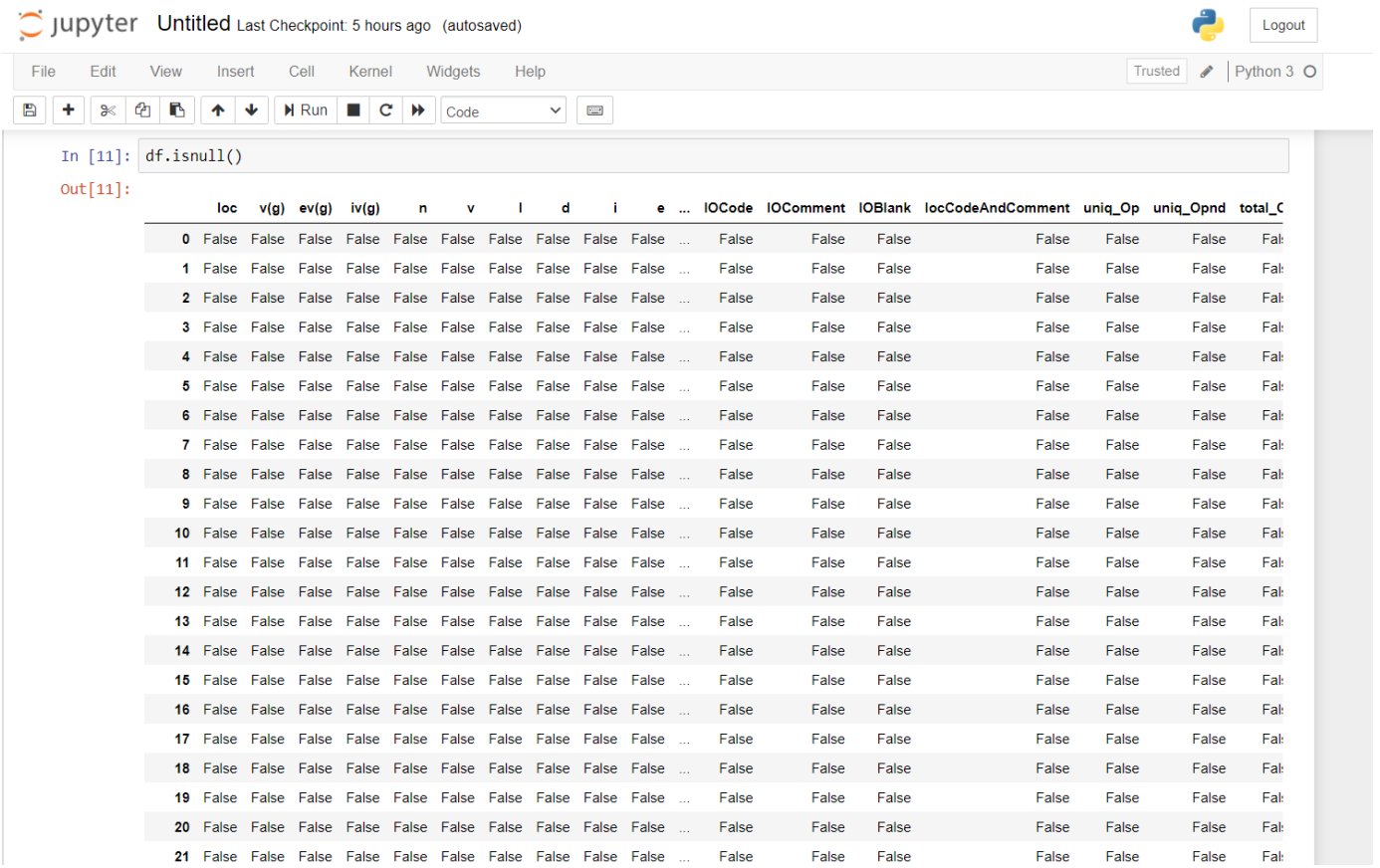


	loc	v(g)	ev(g)	iv(g)	n	v	l	d	i	e	...	lOCode	lOComment	lOBlank	locCodeAndComment	uniq_Op	uniq_Opnd	total_C
0	1.1	1.4	1.4	1.4	1.3	1.30	1.30	1.30	1.30	1.30	...	2	2	2	2	1.2	1.2	1
1	1.0	1.0	1.0	1.0	1.0	1.00	1.00	1.00	1.00	1.00	...	1	1	1	1	1	1	
2	72.0	7.0	1.0	6.0	198.0	1134.13	0.05	20.31	55.85	23029.10	...	51	10	8	1	17	36	1
3	190.0	3.0	1.0	3.0	600.0	4348.76	0.06	17.06	254.87	74202.67	...	129	29	28	2	17	135	3
4	37.0	4.0	1.0	4.0	126.0	599.12	0.06	17.19	34.86	10297.30	...	28	1	6	0	11	16	

5 rows × 22 columns

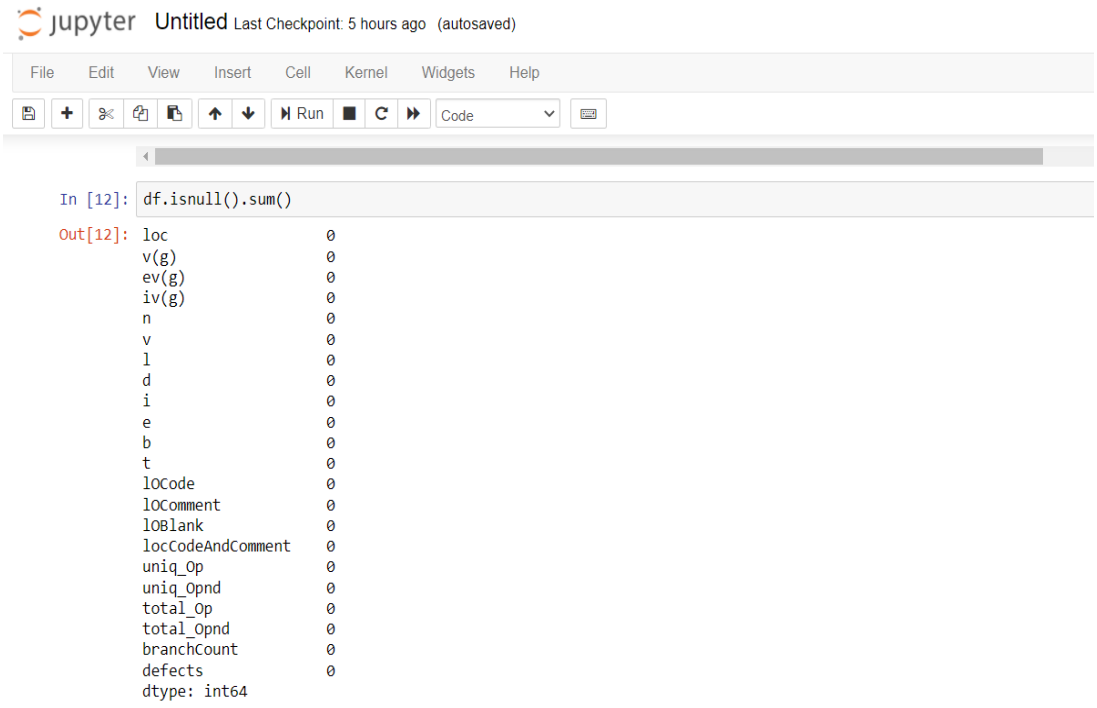
Note: 'isnull' helps us find where in our dataset there are missing values. This is useful information as this is what we need to correct while data cleaning.

df.isnull()



	loc	v(g)	ev(g)	iv(g)	n	v	l	d	i	e	...	lOCode	lOComment	lOBlank	locCodeAndComment	uniq_Op	uniq_Opnd	total_C
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
5	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
6	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
7	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
8	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
9	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
10	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
11	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
12	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
13	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
14	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
15	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
16	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
17	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
18	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
19	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
20	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal
21	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	Fal

df.isnull().sum() #shows how many of the null values
#it shows zero missing values.



Result:- Since there are no missing value, no **data cleaning needed** because the data is all important.

Learning from experiment:- We are successful in Preprocessing and cleaning of a dataset collected in experiment 5 i.e. NASA Metrics Data Program.