A MINOR PROJECT

On

WARRANTY CLAIM FRAUD DETECTION USING DATA ANALYSIS

Dissertation submitted in the partial fulfillment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

By

DEPARTMENT OF INTERNSHIPS

TALARI TEJASWI **CSWDA133**

Under the esteemed Guidance of

Er. Y V D CHANDRA SEKHAR

Founder & Chief Executive Officer

CS CODENZ

LEARN HERE, LEAD ANYWHERE

CS CODENZ

GUDIVADA – 521 323, ANDHRA PRADESH., INDIA

2022-2023

A MINOR PROJECT

On

WARRANTY CLAIM FRAUD DETECTION USING DATA ANALYSIS

Dissertation submitted in the partial fulfillment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

By

DEPARTMENT OF INTERNSHIPS

TALARI TEJASWI **CSWDA133**

Under the esteemed Guidance of

Er. Y V D CHANDRA SEKHAR

Founder & Chief Executive Officer

CS CODENZ

LEARN HERE, LEAD ANYWHERE

CS CODENZ

GUDIVADA – 521 323, ANDHRA PRADESH., INDIA

2022-2023

CS CODENZ

CERTIFICATE

This is to certify that dissertation entitled **“Warranty Claim Fraud Detection Using Data Analysis”** submitted by **TALARI TEJASWI (CSWDA133)** in the partial fulfillment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** from **CS CODENZ** is a record of Bonafede work carried out by them under my guidance and supervision during the year 2022-2023. The result embodied in this dissertation have not been submitted by any other university or Institution for the award of any degree.

Signature of the Supervisor

Er. Y V D CHANDRA SEKHAR

Founder & CEO, CS CODENZ

DECLARATION

I **TALARI TEJASWI (CSWDA133)** declared that the dissertation report entitled **“Warranty Claim Fraud Detection Using Data Analysis”** is no more than 1,00,000 words in length including quotes and exclusive of tables, figures, bibliography, and references. This dissertation contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree ordiploma. Except where otherwise indicated this dissertation in our own work.

Roll No	**Name**	**Signature**
CSWDA133	TALARI TEJASWI	___________

Date:

Place:

COs, POs and PSOs Mapping

Subject Name	:	Major Project
Subject Code	:	PY42223
AcademicYear	:	2022 - 2023

Subject Code	Course Outcomes	
PR4204	CO1	Formulate solutions to computing problems using latest technologies and tools
	CO2	Work effectively in teams to design and implement solutions to computational problems and socially relevant issues
	CO3	Recognize the social and ethical responsibilities of a professional working in the discipline
	CO4	Apply advanced algorithmic and mathematical concepts to the design and analysis of software
	CO5	Devise a communication strategy (language, content and medium) to deliver messages according to the situation and need of the audience.
	CO6	Deliver effective presentations, extemporaneous or impromptu oral presentations. Setting up technical reports using technical tools.

CO-PO-PSOs Mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO 1	3	2	-	2	2	-	-	-	-	-	-	-	3	-	-
CO 2	2	3	-	2	2	-	-	-	-	-	-	-	3	-	-
CO 3	3	3	-	2	2	-	-	-	-	-	-	-	3	-	-
CO 4	3	3	-	2	2	-	-	-	-	-	-	-	3	-	-
CO 5	2	3	-	2	2	-	-	-	-	-	-	1	3	-	-
CO 6	2	3	2	2	3	-	-	-	2	2	2	2	3	-	-
Avg	**2.50**	**2.83**	**2.00**	**2.00**	**2.17**	**-**	**-**	**-**	**2.00**	**2.00**	**2.00**	**1.50**	**3.00**	**-**	**-**

Note: 1 – Good , 2 – Average, 3 - Excellent

Signature of Student with Date

Signature of Guide with Date

ACKNOWLEDGEMENT

This report dissertation could not have been written without the support of our guide **Er. Y V D Chandra Sekhar, Founder & CEO, CS CODENZ** who not only served as our superior but also encouraged and challenged us throughout our academic program our foremost thanks goes to his. Without his this dissertation would not have been possible. We appreciate him vast knowledge in many areas, and his insights, suggestions and guidance that helped to shapeour research skills

It is needed with a great sense of pleasure and immense sense of gratitude that we acknowledge the help of these individuals. We owe many thanks to many people who helped and supported us during the writing of this report

We are thankful to our project coordinator **Er. Y V D Chandra Sekhar,** Founder & CEO, CS CODENZ, for his continuous support

We express our sincere thanks to our respected for bet valuable suggestion and constant motivation that greatly helped us in successful completion of project We also take the privilegeto express our heartfelt gratitude to **Er. Y V D Chandra Sekhar, Founder & CEO,CS CODENZ**

We are thankful to all faculty members for extending their kind cooperation and assistance Finally,we are extremely thankful to our parents and friends for their constant helped moral support

Table Of Contents

ABSTRACT

The Primary objective of this "**WARRANTY CLAIM FRAUD DETECTION USING DATA ANALYSIS**" is to fulfill customer satisfaction and good trust along with challenges of the businesses. It includes data of historical warranty claims and real-time monitoring capabilities.

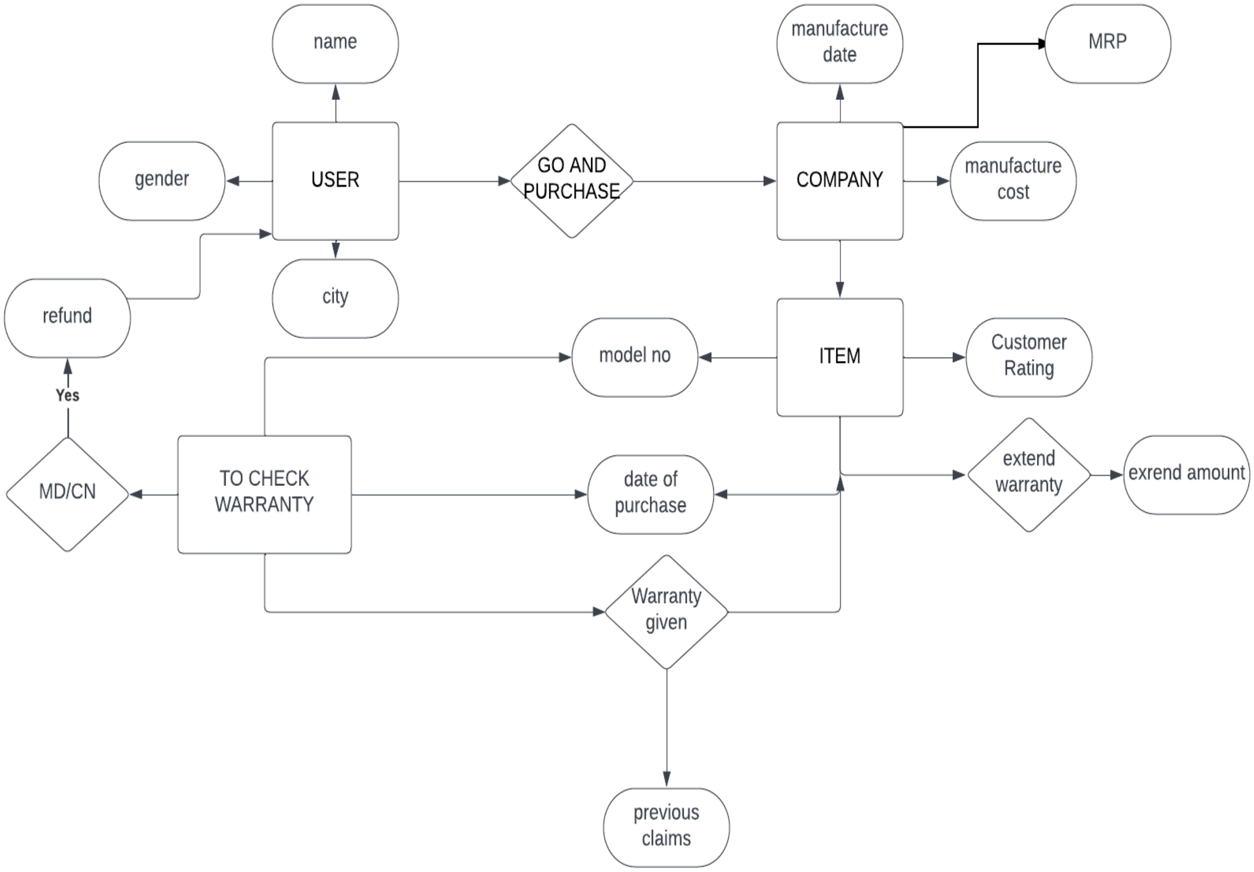
The project also focuses on achieving a high level of accuracy to find the frauds minimizing false positives. It will provide an interface for analysts to easily observe the frauds. It aims to provide a powerful tool to detect and combat warranty claim fraud. This is also help in reducing financial losses to the business providers.

PROBLEM STATEMENT

In today's competitive market, warranty claims are essential for maintaining customer trust and satisfaction. But warranty fraud remains a significant challenge for manufacturers and service providers, resulting in financial losses and eroding customer confidence.

The problem at hand is to develop an efficient and accurate warranty claim fraud detection system using data techniques.

E-R DIAGRAM


name
manufacture date
MRP
gender
USER
GO AND PURCHASE
COMPANY
manufacture cost
refund
city
model no
ITEM
Customer Rating
Yes
MD/CN
TO CHECK WARRANTY
date of purchase
extend warranty
exrend amount
Warranty given
previous claims

REQUIREMENTS

HARDWARE REQUIREMENTS

- Personal Computer / Laptop with minimum RAM (4 GB), ROM (128 GB) and Processor(i3)
- Good latency internet access

SOFTWARE REQUIREMENTS

- Basic Search Engine (Google)
- Google Colaboratory
- MICROSOFT WORD

FUNCTIONAL REQUIREMENTS

- Calculation
- Help in manipulating data and easy process.
- Graphical representation of Datasets

DESCRIPTION

Data Analytics is a process of scrutinizing the data to obtain accurate results. In data analytics the main purpose is extracting the original data from data. In this data analytics we need to perform the major that is data munging.

DATA MUNGING

It is a process of transferring unstructured data into structured format. The goal is to make the data more usable and valuable for analytics or other purposes.

STEPS OF DATA ANALYSIS

1. Defining the Question
2. Collecting the data
3. Cleaning the data
4. Analyzing the data
5. Sharing your results
6. Embracing your failures
7. Summary

CODE

1) DEFINING THE QUESTION

The first step in any data analysis process is to define your objective. In data analytics **jargon**, this is sometimes called the ‘problem statement’. The problem at hand is to develop an efficient and accurate warranty claim fraud detection system.

2) COLLECTING THE DATA

Pandas - Helps to create a dataset and it is also a library in python.

Pandas Package – It’s a group of Panal Data’s which are used to analyze the labelled data and relational data.

Series – A series is a method of pandas and labelled data. Series are nothing but columns in Excel sheet.

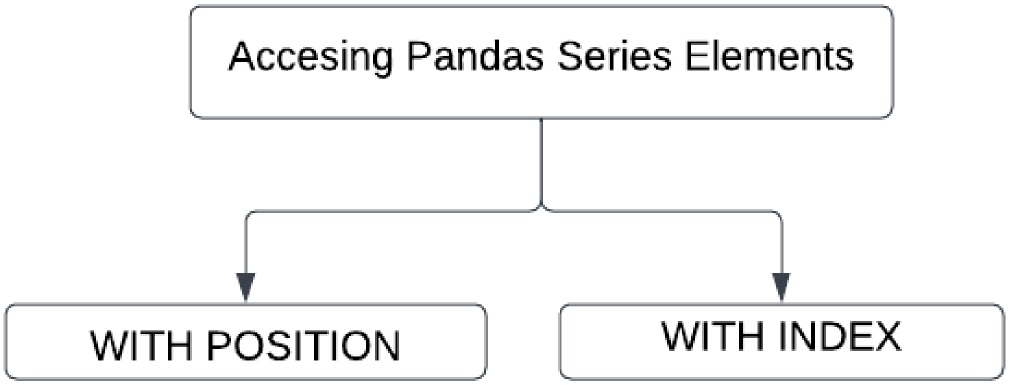
CREATING SERIES-

```
import pandas as pd
a=[10,20,30,'a',50]
b=pd.Series(a)
print(b)
```

```
0    10
1    20
2    30
3     a
4    50
dtype: object
```

```
import pandas as pd
a=[10,20,30,40,50]
b=pd.Series(a)
print(b)
```

```
0    10
1    20
2    30
3    40
4    50
dtype: int64
```



ACCESSING WITH POSITION-

```
#SLICING
import pandas as pd
a=[10,20,30,40,50]
b=pd.Series(a)
print(b[-2:])
```

```
3    40
4    50
dtype: int64
```

```
#SLICING
import pandas as pd
a=[10,20,30,40,50]
b=pd.Series(a)
print(b[-4:-2])
```

```
1    20
2    30
dtype: int64
```

```
import pandas as pd
a=[10,20,30,40,50]
b=pd.Series(a)
print(b[2:])
```

```
2    30
3    40
4    50
dtype: int64
```

ACCESSING WITH INDEX-

```
import pandas as pd
a=[10,20,'CS',40,50]
b=pd.Series(a,index=['!','@','#','$','%'])
print(b)
print("-------------------")
print(b['@'])
```

```
!    10
@    20
#    CS
$    40
%    50
dtype: object
-------------------
20
```

DATAFRAME

A data frame is a 2D data structure in which we store data in the form of tables. [rows x columns]

We can create a table via Data Frame i.e., known as DATASET.

CREATING A DATASET-

```
#creating empty data set
import pandas as pd
a=pd.DataFrame()
print(a)
```

```
Empty DataFrame
Columns: []
Index: []
```

Creating data set using list-

```
#creating dataframe by using list
import pandas as pd
a=[10,20,30,40,50]
b=pd.DataFrame(a)
print(b)
```

```
    0
0  10
1  20
2  30
3  40
4  50
```

Creating data set using Dict-

```
#creating using DICT
import pandas as pd
x=[{'a':10,'b':20,'c':30}]
y=pd.DataFrame(x)
print(y)
```

```
    a   b   c
0  10  20  30
```

```
#creating using DICT
import pandas as pd
x=[{'a':10,'b':20,'c':30}]
y=pd.DataFrame(x)
print(type(y))
```

<class 'pandas.core.frame.DataFrame'>

Creating dataset using Series

```
import pandas as pd
a=[10,20,30,40]
b=pd.Series(a)
c=pd.DataFrame(b)
print(c)
```

```
    0
0  10
1  20
2  30
3  40
```

CREATION OF LARGE DATASET-

```
import pandas as pd
a={'S.No':pd.Series([1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]),

    'Name':pd.Series(['vagesh','venkat','reddy','yashwanth','ram','kumar','abhi',
'bhavana','meghana','swapna','rani','keerthana','akash','siva','shankar','pradeep',
'sameena','samba','sai','vamsi']),

    'Gender':pd.Series(['m','m','m','m','m','m','f','f','f','f','f','f','m','m','m',
                        'm','f','m','m','m']),
    'City':pd.Series(['hyderabad','pune','nuzvid','vijayawada','mumbai','kolkata',
      'bangalore','chennai','delhi','vizag','tirupati','nellore','kadapa','kurnool',
      'kakinada','srikakulam','nizamabad','noida','khammam','nalgonda',]),

    'DateOfPurchase':pd.Series(["01 jun 2020","05 july 2020","28 aug 2020",
       "25 sep 2020","27 oct 2020","31 nov 2020","15 dec 2020","17 dec 2020",
       "04 jan 2021","17 jan 2021","21 jan 2021","14 feb 2021","16 feb 2021",
       "25 feb 2021","27 feb 2021","01 mar 2021","05 mar 2021","10 mar 2021",
       "14 mar 2021","18 mar 2021"]),

   'Customer_Rating(0-5)':([4,4,4,3,3,3,3,3,5,5,4,4,5,4,3,2,1,5,2,1]),
   'ManufactureDate':(["11 jun 2019","15 july 2019","18 aug 2019","15 sep 2019",
          "17 oct 2019","21 nov 2019","05 dec 2019","07 dec 2019","14 jan 2020",
          "07 jan 2020","11 jan 2019","04 feb 2019","06 feb 2019","15 feb 2019",
             "17 feb 2019","11 mar 2019","15 mar 2019","11 mar 2019",
              "04 mar 2019","28 mar2019"]),

   'Condition':(['used','used','used','new','new','used','new','new','new','used',
                'used','new','used','new','new','new','new','used','used','new']),

   'ModelNo':([101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,
               118,119,120]),
   'Mrp':([2380, 3472, 3127, 2704, 2575, 3210, 2999, 3866, 2750, 3665, 2455, 3766,
           3944, 2085, 3628, 2992, 3599, 2122, 2821,2200]),
   'SoldPrice':([2975, 2536, 2199, 3515, 3089, 2288, 2399, 3966, 3575, 3765, 2946,
                 1883, 3844, 2490, 2711, 3590, 2159, 2546, 3385,2100]),
   'Manufacturing_Cost':([230,340,310,270,250,320,290,380,270,360,240,370,390,200,
                          360,290,350,210,2800,220]),
   'Extend_Amount':([23,34,31,27,25,32,29,38,27,36,24,37,39,20,36,29,35,21,28,22]),
   'Warranty_Period':([1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]),
   'Warranty_periodLeft(in Months)':([0,2,4,0,6,8,9,4,7,5,8,0,0,6,5,4,2,2,1,1]),
   'MD/CN':([0,0,0,0,1,0,1,1,1,0,1,1,1,1,1,1,1,0,0,0]),
   'Previous_claims':([1,0,0,2,1,1,1,0,0,0,0,0,0,0,1,2,2,3,1,1]),
   'Refund_Duration(in days)':([3,3,4,3,5,7,2,2,2,2,3,3,3,3,4,4,5,5,7,7])
    }
b=pd.DataFrame(a)
print(b)
```

```
    S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5) \
0      1     vagesh      m   hyderabad    01 jun 2020                    4
1      2     venkat      m        pune   05 july 2020                    4
2      3      reddy      m      nuzvid    28 aug 2020                    4
3      4  yashwanth      m  vijayawada    25 sep 2020                    3
4      5        ram      m      mumbai    27 oct 2020                    3
5      6      kumar      m     kolkata    31 nov 2020                    3
6      7       abhi      f   bangalore    15 dec 2020                    3
7      8    bhavana      f     chennai    17 dec 2020                    3
8      9    meghana      f       delhi    04 jan 2021                    5
9     10     swapna      f       vizag    17 jan 2021                    5
10    11       rani      f    tirupati    21 jan 2021                    4
11    12  keerthana      f     nellore    14 feb 2021                    4
12    13      akash      m      kadapa    16 feb 2021                    5
13    14       siva      m     kurnool    25 feb 2021                    4
14    15    shankar      m    kakinada    27 feb 2021                    3
15    16    pradeep      m  srikakulam    01 mar 2021                    2
16    17    sameena      f   nizamabad    05 mar 2021                    1
17    18      samba      m       noida    10 mar 2021                    5
18    19        sai      m     khammam    14 mar 2021                    2
19    20      vamsi      m    nalgonda    18 mar 2021                    1

   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used      101  2380       2975                 230
1     15 july 2019      used      102  3472       2536                 340
2      18 aug 2019      used      103  3127       2199                 310
3      15 sep 2019       new      104  2704       3515                 270
4      17 oct 2019       new      105  2575       3089                 250
5      21 nov 2019      used      106  3210       2288                 320
6      05 dec 2019       new      107  2999       2399                 290
7      07 dec 2019       new      108  3866       3966                 380
8      14 jan 2020       new      109  2750       3575                 270
9      07 jan 2020      used      110  3665       3765                 360
10     11 jan 2019      used      111  2455       2946                 240
11     04 feb 2019       new      112  3766       1883                 370
12     06 feb 2019      used      113  3944       3844                 390
13     15 feb 2019       new      114  2085       2490                 200
14     17 feb 2019       new      115  3628       2711                 360
15     11 mar 2019       new      116  2992       3590                 290
16     15 mar 2019       new      117  3599       2159                 350
17     11 mar 2019      used      118  2122       2546                 210
18     04 mar 2019      used      119  2821       3385                2800
19     28 mar 2019       new      120  2200       2100                 220
```

```
    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0
1              34                1                               2      0
2              31                1                               4      0
3              27                1                               0      0
4              25                1                               6      1
5              32                1                               8      0
6              29                1                               9      1
7              38                1                               4      1
8              27                1                               7      1
9              36                1                               5      0
10             24                1                               8      1
11             37                1                               0      1
12             39                1                               0      1
13             20                1                               6      1
14             36                1                               5      1
15             29                1                               4      1
16             35                1                               2      1
17             21                1                               2      0
18             28                1                               1      0
19             22                1                               1      0

        Previous_claims  Refund_Duration(in days)
0                     1                         3
1                     0                         3
2                     0                         4
3                     2                         3
4                     1                         5
5                     1                         7
6                     1                         2
7                     0                         2
8                     0                         2
9                     0                         2
10                    0                         3
11                    0                         3
12                    0                         3
13                    0                         3
14                    1                         4
15                    2                         4
16                    2                         5
17                    3                         5
18                    1                         7
19                    1                         7
```

OPERATIONS ON DATASET

There are three operations we can perform on a dataset.

1. Row Operation
2. Column Operation
3. Selection Operation

ROW OPERATIONS-

- row selection

- row addition

- row deletion

ROW SELECTION-

```
#SELECTING ROW
x=b.iloc[0,]
print(x)
```

```
S.No                                           1

Name                                      vagesh
Gender                                         m
City                                   hyderabad
DateOfPurchase                       01 jun 2020
Customer_Rating(0-5)                           4
ManufactureDate                      11 jun 2019
Condition                                   used
ModelNo                                      101
Mrp                                         2380
SoldPrice                                   2975
Manufacturing_Cost                           230
Extend_Amount                                 23
Warranty_Period                                1
Warranty_periodLeft(in Months)                 0
MD/CN                                          0
Previous_claims                                1
Refund_Duration(in days)                       3
Profit                                       595
Profit_Percentage                           25.0
Name: 0, dtype: object
```

ROW ADDITION-

We can add the row for the dataset by using LOC() method

But the row data must be the same comparing to the other rows.

b.loc()

```
b.loc[20]=[1,'gut','m','vishakapatnam','01 jan 2023',5,'02 jan 2021',
           'used',1,2,3,4,5,6,7,8,9,10]
print(b)
```

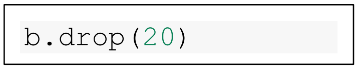
```
     S.No       Name   Gender          City   DateOfPurchase  \
0       1     vagesh        m     hyderabad      01 jun 2020
1       2     venkat        m          pune    05 july 2020
2       3      reddy        m        nuzvid      28 aug 2020
3       4  yashwanth        m     vijayawada     25 sep 2020
4       5        ram        m        mumbai      27 oct 2020
5       6      kumar        m       kolkata      31 nov 2020
6       7       abhi        f     bangalore      15 dec 2020
7       8    bhavana        f       chennai      17 dec 2020
8       9    meghana        f         delhi      04 jan 2021
9      10     swapna        f         vizag      17 jan 2021
10     11       rani        f      tirupati      21 jan 2021
11     12  keerthana        f       nellore      14 feb 2021
12     13      akash        m        kadapa      16 feb 2021
13     14       siva        m       kurnool      25 feb 2021
14     15    shankar        m      kakinada      27 feb 2021
15     16    pradeep        m    srikakulam      01 mar 2021
16     17    sameena        f     nizamabad      05 mar 2021
17     18      samba        m         noida      10 mar 2021
18     19        sai        m       khammam      14 mar 2021
19     20      vamsi        m      nalgonda      18 mar 2021
20      1        gut        m  vishakapatnam     01 jan 2023
```

```
  Customer_Rating(0-5) ManufactureDate Condition  ModelNo   Mrp  SoldPrice  \
0                     4     11 jun 2019      used      101  2380       2975
1                     4    15 july 2019      used      102  3472       2536
2                     4     18 aug 2019      used      103  3127       2199
3                     3     15 sep 2019       new      104  2704       3515
4                     3     17 oct 2019       new      105  2575       3089
5                     3     21 nov 2019      used      106  3210       2288
6                     3     05 dec 2019       new      107  2999       2399
7                     3     07 dec 2019       new      108  3866       3966
8                     5     14 jan 2020       new      109  2750       3575
9                     5     07 jan 2020      used      110  3665       3765
10                    4     11 jan 2019      used      111  2455       2946
11                    4     04 feb 2019       new      112  3766       1883
12                    5     06 feb 2019      used      113  3944       3844
13                    4     15 feb 2019       new      114  2085       2490
14                    3     17 feb 2019       new      115  3628       2711
15                    2     11 mar 2019       new      116  2992       3590
16                    1     15 mar 2019       new      117  3599       2159
17                    5     11 mar 2019      used      118  2122       2546
18                    2     04 mar 2019      used      119  2821       3385
19                    1     28 mar 2019       new      120  2200       2100
20                    5     02 jan 2021      used        1     2          3

    Manufacturing_Cost  Extend_Amount  Warranty_Period  /
0                  230             23                1
1                  340             34                1
2                  310             31                1
3                  270             27                1
4                  250             25                1
5                  320             32                1
6                  290             29                1
7                  380             38                1
8                  270             27                1
9                  360             36                1
10                 240             24                1
11                 370             37                1
12                 390             39                1
13                 200             20                1
14                 360             36                1
15                 290             29                1
16                 350             35                1
17                 210             21                1
18                2800             28                1
19                 220             22                1
20                   4              5                6
```

```
Warranty_periodLeft(in Months)  MD/CN  Previous_claims  \
0                                0      0                1
1                                2      0                0
2                                4      0                0
3                                0      0                2
4                                6      1                1
5                                8      0                1
6                                9      1                1
7                                4      1                0
8                                7      1                0
9                                5      0                0
10                               8      1                0
11                               0      1                0
12                               0      1                0
13                               6      1                0
14                               5      1                1
15                               4      1                2
16                               2      1                2
17                               2      0                3
18                               1      0                1
19                               1      0                1
20                               7      8                9

        Refund_Duration(in days)
0                          3
1                          3
2                          4
3                          3
4                          5
5                          7
6                          2
7                          2
8                          2
9                          2
10                         3
11                         3
12                         3
13                         3
14                         4
15                         4
16                         5
17                         5
18                         7
19                         7
20                        10
```

ROW DELETION-

By using drop() method we can delete the row.



```
b.drop(20)
```

```
   S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)  \
0      1     vagesh      m   hyderabad    01 jun 2020                     4
1      2     venkat      m        pune   05 july 2020                     4
2      3      reddy      m      nuzvid    28 aug 2020                     4
3      4  yashwanth      m  vijayawada    25 sep 2020                     3
4      5        ram      m      mumbai    27 oct 2020                     3
5      6      kumar      m     kolkata    31 nov 2020                     3
6      7       abhi      f   bangalore    15 dec 2020                     3
7      8    bhavana      f     chennai    17 dec 2020                     3
8      9    meghana      f       delhi    04 jan 2021                     5
9     10     swapna      f       vizag    17 jan 2021                     5
10    11       rani      f    tirupati    21 jan 2021                     4
11    12  keerthana      f     nellore    14 feb 2021                     4
12    13      akash      m      kadapa    16 feb 2021                     5
13    14       siva      m     kurnool    25 feb 2021                     4
14    15    shankar      m    kakinada    27 feb 2021                     3
15    16    pradeep      m  srikakulam    01 mar 2021                     2
16    17    sameena      f   nizamabad    05 mar 2021                     1
17    18      samba      m       noida    10 mar 2021                     5
18    19        sai      m     khammam    14 mar 2021                     2
19    20      vamsi      m    nalgonda    18 mar 2021                     1

   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used      101  2380       2975                 230
1     15 july 2019      used      102  3472       2536                 340
2      18 aug 2019      used      103  3127       2199                 310
3      15 sep 2019       new      104  2704       3515                 270
4      17 oct 2019       new      105  2575       3089                 250
5      21 nov 2019      used      106  3210       2288                 320
6      05 dec 2019       new      107  2999       2399                 290
7      07 dec 2019       new      108  3866       3966                 380
8      14 jan 2020       new      109  2750       3575                 270
9      07 jan 2020      used      110  3665       3765                 360
10     11 jan 2019      used      111  2455       2946                 240
11     04 feb 2019       new      112  3766       1883                 370
12     06 feb 2019      used      113  3944       3844                 390
13     15 feb 2019       new      114  2085       2490                 200
14     17 feb 2019       new      115  3628       2711                 360
15     11 mar 2019       new      116  2992       3590                 290
16     15 mar 2019       new      117  3599       2159                 350
17     11 mar 2019      used      118  2122       2546                 210
18     04 mar 2019      used      119  2821       3385                2800
19     28 mar 2019       new      120  2200       2100                 220
```

```
    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0
1              34                1                               2      0
2              31                1                               4      0
3              27                1                               0      0
4              25                1                               6      1
5              32                1                               8      0
6              29                1                               9      1
7              38                1                               4      1
8              27                1                               7      1
9              36                1                               5      0
10             24                1                               8      1
11             37                1                               0      1
12             39                1                               0      1
13             20                1                               6      1
14             36                1                               5      1
15             29                1                               4      1
16             35                1                               2      1
17             21                1                               2      0
18             28                1                               1      0
19             22                1                               1      0

     Previous_claims        Refund_Duration(in days)
0                  1                               3
1                  0                               3
2                  0                               4
3                  2                               3
4                  1                               5
5                  1                               7
6                  1                               2
7                  0                               2
8                  0                               2
9                  0                               2
10                 0                               3
11                 0                               3
12                 0                               3
13                 0                               3
14                 1                               4
15                 2                               4
16                 2                               5
17                 3                               5
18                 1                               7
19                 1                               7
```

COLUMN OPERATIONS

- Column Selection
- Column Addition
- Column deletion

COLUMN SELECTION-

We can select the column by using the column name and the data frame.

DataFrameObj.['column name']

```
print(b['Mrp'])
```

```
0     2380
1     3472
2     3127
3     2704
4     2575
5     3210
6     2999
7     3866
8     2750
9     3665
10    2455
11    3766
12    3944
13    2085
14    3628
15    2992
16    3599
17    2122
18    2821
19    2200
Name: Mrp, dtype: int64
```

COLUMN ADDITION

We can add columns by dataframe[' ']

I want to calculate profit of the data frame and add it to the existing columns.

PROFIT= SELLING PRICE – COST PRICE

Here in our DataFrame SoldPrice-Mrp=Profit

```
b['Profit']=b['SoldPrice']-b['Mrp']
print(b)
```

```
   S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)  \
0     1     vagesh      m   hyderabad    01 jun 2020                     4
1     2     venkat      m        pune   05 july 2020                     4
2     3      reddy      m      nuzvid    28 aug 2020                     4
3     4  yashwanth      m  vijayawada    25 sep 2020                     3
4     5        ram      m      mumbai    27 oct 2020                     3
5     6      kumar      m     kolkata    31 nov 2020                     3
6     7       abhi      f   bangalore    15 dec 2020                     3
7     8    bhavana      f     chennai    17 dec 2020                     3
8     9    meghana      f       delhi    04 jan 2021                     5
9    10     swapna      f       vizag    17 jan 2021                     5
10   11       rani      f    tirupati    21 jan 2021                     4
11   12  keerthana      f     nellore    14 feb 2021                     4
12   13      akash      m      kadapa    16 feb 2021                     5
13   14       siva      m     kurnool    25 feb 2021                     4
14   15    shankar      m    kakinada    27 feb 2021                     3
15   16    pradeep      m  srikakulam    01 mar 2021                     2
16   17    sameena      f   nizamabad    05 mar 2021                     1
17   18      samba      m       noida    10 mar 2021                     5
18   19        sai      m     khammam    14 mar 2021                     2
19   20      vamsi      m    nalgonda    18 mar 2021                     1


   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used      101  2380       2975                 230
1     15 july 2019      used      102  3472       2536                 340
2      18 aug 2019      used      103  3127       2199                 310
3      15 sep 2019       new      104  2704       3515                 270
4      17 oct 2019       new      105  2575       3089                 250
5      21 nov 2019      used      106  3210       2288                 320
6      05 dec 2019       new      107  2999       2399                 290
7      07 dec 2019       new      108  3866       3966                 380
8      14 jan 2020       new      109  2750       3575                 270
9      07 jan 2020      used      110  3665       3765                 360
10     11 jan 2019      used      111  2455       2946                 240
11     04 feb 2019       new      112  3766       1883                 370
12     06 feb 2019      used      113  3944       3844                 390
13     15 feb 2019       new      114  2085       2490                 200
14     17 feb 2019       new      115  3628       2711                 360
15     11 mar 2019       new      116  2992       3590                 290
16     15 mar 2019       new      117  3599       2159                 350
17     11 mar 2019      used      118  2122       2546                 210
18     04 mar 2019      used      119  2821       3385                2800
19     28 mar 2019       new      120  2200       2100                 220
```

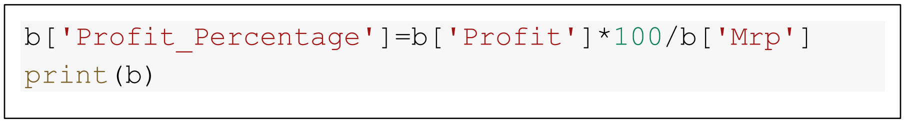
```
     Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0               23                1                               0      0
1               34                1                               2      0
2               31                1                               4      0
3               27                1                               0      0
4               25                1                               6      1
5               32                1                               8      0
6               29                1                               9      1
7               38                1                               4      1
8               27                1                               7      1
9               36                1                               5      0
10              24                1                               8      1
11              37                1                               0      1
12              39                1                               0      1
13              20                1                               6      1
14              36                1                               5      1
15              29                1                               4      1
16              35                1                               2      1
17              21                1                               2      0
18              28                1                               1      0
19              22                1                               1      0
```

```
      Previous_claims  Refund_Duration(in days)  Profit
0                  1                         3     595
1                  0                         3    -936
2                  0                         4    -928
3                  2                         3     811
4                  1                         5     514
5                  1                         7    -922
6                  1                         2    -600
7                  0                         2     100
8                  0                         2     825
9                  0                         2     100
10                 0                         3     491
11                 0                         3   -1883
12                 0                         3    -100
13                 0                         3     405
14                 1                         4    -917
15                 2                         4     598
16                 2                         5   -1440
17                 3                         5     424
18                 1                         7     564
19                 1                         7    -100
```

Similarly, I want to calculate profit percentage and add it to the columns list.

Profit %= profit*100 / Cost Price

According to our code



```
b['Profit_Percentage']=b['Profit']*100/b['Mrp']
print(b)
```

```
   S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)  \
0      1     vagesh      m   hyderabad    01 jun 2020                     4
1      2     venkat      m        pune   05 july 2020                     4
2      3      reddy      m      nuzvid    28 aug 2020                     4
3      4  yashwanth      m  vijayawada    25 sep 2020                     3
4      5        ram      m      mumbai    27 oct 2020                     3
5      6      kumar      m     kolkata    31 nov 2020                     3
6      7       abhi      f   bangalore    15 dec 2020                     3
7      8    bhavana      f     chennai    17 dec 2020                     3
8      9    meghana      f       delhi    04 jan 2021                     5
9     10     swapna      f       vizag    17 jan 2021                     5
10    11       rani      f    tirupati    21 jan 2021                     4
11    12  keerthana      f     nellore    14 feb 2021                     4
12    13      akash      m      kadapa    16 feb 2021                     5
13    14       siva      m     kurnool    25 feb 2021                     4
14    15    shankar      m    kakinada    27 feb 2021                     3
15    16    pradeep      m  srikakulam    01 mar 2021                     2
16    17    sameena      f   nizamabad    05 mar 2021                     1
17    18      samba      m       noida    10 mar 2021                     5
18    19        sai      m     khammam    14 mar 2021                     2
19    20      vamsi      m    nalgonda    18 mar 2021                     1
```

```
  ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0     11 jun 2019      used      101  2380       2975                 230
1    15 july 2019      used      102  3472       2536                 340
2     18 aug 2019      used      103  3127       2199                 310
3     15 sep 2019       new      104  2704       3515                 270
4     17 oct 2019       new      105  2575       3089                 250
5     21 nov 2019      used      106  3210       2288                 320
6     05 dec 2019       new      107  2999       2399                 290
7     07 dec 2019       new      108  3866       3966                 380
8     14 jan 2020       new      109  2750       3575                 270
9     07 jan 2020      used      110  3665       3765                 360
10    11 jan 2019      used      111  2455       2946                 240
11    04 feb 2019       new      112  3766       1883                 370
12    06 feb 2019      used      113  3944       3844                 390
13    15 feb 2019       new      114  2085       2490                 200
14    17 feb 2019       new      115  3628       2711                 360
15    11 mar 2019       new      116  2992       3590                 290
16    15 mar 2019       new      117  3599       2159                 350
17    11 mar 2019      used      118  2122       2546                 210
18    04 mar 2019      used      119  2821       3385                2800
19    28 mar 2019       new      120  2200       2100                 220

    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0
1              34                1                               2      0
2              31                1                               4      0
3              27                1                               0      0
4              25                1                               6      1
5              32                1                               8      0
6              29                1                               9      1
7              38                1                               4      1
8              27                1                               7      1
9              36                1                               5      0
10             24                1                               8      1
11             37                1                               0      1
12             39                1                               0      1
13             20                1                               6      1
14             36                1                               5      1
15             29                1                               4      1
16             35                1                               2      1
17             21                1                               2      0
18             28                1                               1      0
19             22                1                               1      0
```

```
       Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage
0                 1                         3     595          25.000000
1                 0                         3    -936         -26.958525
2                 0                         4    -928         -29.677007
3                 2                         3     811          29.992604
4                 1                         5     514          19.961165
5                 1                         7    -922         -28.722741
6                 1                         2    -600         -20.006669
7                 0                         2     100           2.586653
8                 0                         2     825          30.000000
9                 0                         2     100           2.728513
10                0                         3     491          20.000000
11                0                         3   -1883         -50.000000
12                0                         3    -100          -2.535497
13                0                         3     405          19.424460
14                1                         4    -917         -25.275634
15                2                         4     598          19.986631
16                2                         5   -1440         -40.011114
17                3                         5     424          19.981150
18                1                         7     564          19.992910
19                1                         7    -100          -4.545455
```

COLUMN DELETION –

To delete an entire column from a Pandas Series in Python, you can use the ’drop’ method or simply select the columns you want to keep.

```
del b['Mrp']
print(b)
```

The taken dataset ‘Mrp’ column is removed .

The output of the code is given below.

```
    S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)  \
0      1     vagesh      m   hyderabad    01 jun 2020                     4
1      2     venkat      m        pune   05 july 2020                     4
2      3      reddy      m      nuzvid    28 aug 2020                     4
3      4  yashwanth      m  vijayawada    25 sep 2020                     3
4      5        ram      m      mumbai    27 oct 2020                     3
5      6      kumar      m     kolkata    31 nov 2020                     3
6      7       abhi      f   bangalore    15 dec 2020                     3
7      8    bhavana      f     chennai    17 dec 2020                     3
8      9    meghana      f       delhi    04 jan 2021                     5
9     10     swapna      f       vizag    17 jan 2021                     5
10    11       rani      f    tirupati    21 jan 2021                     4
11    12  keerthana      f     nellore    14 feb 2021                     4
12    13      akash      m      kadapa    16 feb 2021                     5
13    14       siva      m     kurnool    25 feb 2021                     4
14    15    shankar      m    kakinada    27 feb 2021                     3
15    16    pradeep      m  srikakulam    01 mar 2021                     2
16    17    sameena      f   nizamabad    05 mar 2021                     1
17    18      samba      m       noida    10 mar 2021                     5
18    19        sai      m     khammam    14 mar 2021                     2
19    20      vamsi      m    nalgonda    18 mar 2021                     1
```

```
    ManufactureDate Condition  ModelNo  SoldPrice  Manufacturing_Cost  \
0       11 jun 2019      used      101       2975                 230
1      15 july 2019      used      102       2536                 340
2       18 aug 2019      used      103       2199                 310
3       15 sep 2019       new      104       3515                 270
4       17 oct 2019       new      105       3089                 250
5       21 nov 2019      used      106       2288                 320
6       05 dec 2019       new      107       2399                 290
7       07 dec 2019       new      108       3966                 380
8       14 jan 2020       new      109       3575                 270
9       07 jan 2020      used      110       3765                 360
10      11 jan 2019      used      111       2946                 240
11      04 feb 2019       new      112       1883                 370
12      06 feb 2019      used      113       3844                 390
13      15 feb 2019       new      114       2490                 200
14      17 feb 2019       new      115       2711                 360
15      11 mar 2019       new      116       3590                 290
16      15 mar 2019       new      117       2159                 350
17      11 mar 2019      used      118       2546                 210
18      04 mar 2019      used      119       3385                2800
19      28 mar 2019       new      120       2100                 220
```

```
    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0
1              34                1                               2      0
2              31                1                               4      0
3              27                1                               0      0
4              25                1                               6      1
5              32                1                               8      0
6              29                1                               9      1
7              38                1                               4      1
8              27                1                               7      1
9              36                1                               5      0
10             24                1                               8      1
11             37                1                               0      1
12             39                1                               0      1
13             20                1                               6      1
14             36                1                               5      1
15             29                1                               4      1
16             35                1                               2      1
17             21                1                               2      0
18             28                1                               1      0
19             22                1                               1      0


    Previous_claims  Refund_Duration(in days)
0                 1                         3
1                 0                         3
2                 0                         4
3                 2                         3
4                 1                         5
5                 1                         7
6                 1                         2
7                 0                         2
8                 0                         2
9                 0                         2
10                0                         3
11                0                         3
12                0                         3
13                0                         3
14                1                         4
15                2                         4
16                2                         5
17                3                         5
18                1                         7
19                1                         7
```

INFO( )

The info method provides a summary of the data including the data types of each column the number of non-null values.

```
print(b.info)
```

```
   S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)  \
0      1     vagesh      m   hyderabad    01 jun 2020                     4
1      2     venkat      m        pune   05 july 2020                     4
2      3      reddy      m      nuzvid    28 aug 2020                     4
3      4  yashwanth      m  vijayawada    25 sep 2020                     3
4      5        ram      m      mumbai    27 oct 2020                     3
5      6      kumar      m     kolkata    31 nov 2020                     3
6      7       abhi      f   bangalore    15 dec 2020                     3
7      8    bhavana      f     chennai    17 dec 2020                     3
8      9    meghana      f       delhi    04 jan 2021                     5
9     10     swapna      f       vizag    17 jan 2021                     5
10    11       rani      f    tirupati    21 jan 2021                     4
11    12  keerthana      f     nellore    14 feb 2021                     4
12    13      akash      m      kadapa    16 feb 2021                     5
13    14       siva      m     kurnool    25 feb 2021                     4
14    15    shankar      m    kakinada    27 feb 2021                     3
15    16    pradeep      m  srikakulam    01 mar 2021                     2
16    17    sameena      f   nizamabad    05 mar 2021                     1
17    18      samba      m       noida    10 mar 2021                     5
18    19        sai      m     khammam    14 mar 2021                     2
19    20      vamsi      m    nalgonda    18 mar 2021                     1
```

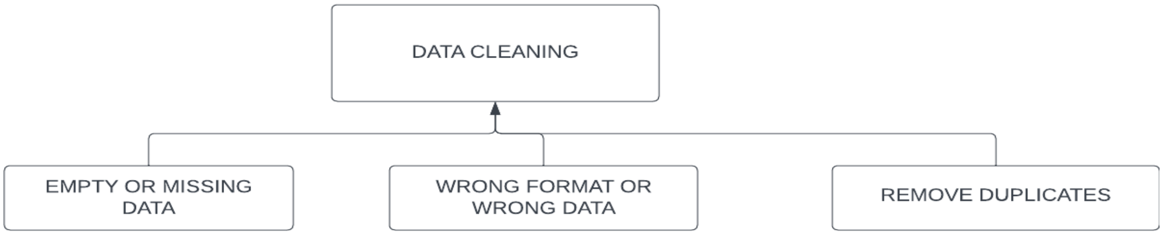
```
   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used      101  2380       2975                 230   
1     15 july 2019      used      102  3472       2536                 340   
2      18 aug 2019      used      103  3127       2199                 310   
3      15 sep 2019       new      104  2704       3515                 270   
4      17 oct 2019       new      105  2575       3089                 250   
5      21 nov 2019      used      106  3210       2288                 320   
6      05 dec 2019       new      107  2999       2399                 290   
7      07 dec 2019       new      108  3866       3966                 380   
8      14 jan 2020       new      109  2750       3575                 270   
9      07 jan 2020      used      110  3665       3765                 360   
10     11 jan 2019      used      111  2455       2946                 240   
11     04 feb 2019       new      112  3766       1883                 370   
12     06 feb 2019      used      113  3944       3844                 390   
13     15 feb 2019       new      114  2085       2490                 200   
14     17 feb 2019       new      115  3628       2711                 360   
15     11 mar 2019       new      116  2992       3590                 290   
16     15 mar 2019       new      117  3599       2159                 350   
17     11 mar 2019      used      118  2122       2546                 210   
18     04 mar 2019      used      119  2821       3385                2800   
19     28 mar 2019       new      120  2200       2100                 220   



    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0   
1              34                1                               2      0   
2              31                1                               4      0   
3              27                1                               0      0   
4              25                1                               6      1   
5              32                1                               8      0   
6              29                1                               9      1   
7              38                1                               4      1   
8              27                1                               7      1   
9              36                1                               5      0   
10             24                1                               8      1   
11             37                1                               0      1   
12             39                1                               0      1   
13             20                1                               6      1   
14             36                1                               5      1   
15             29                1                               4      1   
16             35                1                               2      1   
17             21                1                               2      0   
18             28                1                               1      0   
19             22                1                               1      0   
```

```
     Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage
0                  1                         3     595          25.000000
1                  0                         3    -936         -26.958525
2                  0                         4    -928         -29.677007
3                  2                         3     811          29.992604
4                  1                         5     514          19.961165
5                  1                         7    -922         -28.722741
6                  1                         2    -600         -20.006669
7                  0                         2     100           2.586653
8                  0                         2     825          30.000000
9                  0                         2     100           2.728513
10                 0                         3     491          20.000000
11                 0                         3   -1883         -50.000000
12                 0                         3    -100          -2.535497
13                 0                         3     405          19.424460
14                 1                         4    -917         -25.275634
15                 2                         4     598          19.986631
16                 2                         5   -1440         -40.011114
17                 3                         5     424          19.981150
18                 1                         7     564          19.992910
19                 1                         7    -100          -4.545455
```

3) CLEANING THE DATA

It is the process of remaining or replacing the NAN values.

NAN – Not A Null (or) Not A None



EMPTY OR MISSING DATA

We can solve empty cells by using two methods.

- isnull( )
- notnull( )

```
import pandas as pd
import numpy as np
a={'S.No':pd.Series([1,2,3,4,np.nan,6,7,8,9,10,11,12,np.nan,14,15,16,17,18,19,20]),
   'Name':pd.Series(['vagesh',np.nan,'reddy','yashwanth','ram','kumar','abhi',
   'bhavana','meghana','swapna','ram','keerthana',np.nan,'siva','shankar','pradeep',
   'sameena','samba','sai','vamsi']),

   'Gender':pd.Series(['m','m',np.nan,'m','m','m','f','f','f',np.nan,'f','f',np.nan,
                       'm','m','m','f',np.nan,'m',np.nan]),

   'City':pd.Series(['hyderabad','pune','mumbai','vijayawada','mumbai','kolkata',
       'bangalore',np.nan,'delhi','vizag','tirupati','nellore','kadapa','kurnool',
        np.nan,'srikakulam','nizamabad',np.nan,'khammam','nalgonda',]),

   'DateOfPurchase':pd.Series(["01 jun 2020","05 july 2020","28 aug 2020",
             "25 sep 2020","27 oct 2020","31 nov 2020","15 dec 2020","17 dec 2020",
             "04 jan 2021","17 jan 2021","21 jan 2021","14 feb 2021","16 feb 2021",
             "25 feb 2021","27 feb 2021","01 mar 2021","05 mar 2021","10 mar 2021",
             "14 mar 2021","18 mar 2021"]),
   'Customer_Rating(0-5)':([4,4,4,3,np.nan,3,3,3,5,5,4,np.nan,5,4,3,2,1,np.nan,2,1]),

   'ManufactureDate':(["11 jun 2019","15 july 2019","18 aug 2019","15 sep 2019",
         "17 oct 2019","21 nov 2019","05 dec 2019","07 dec 2019","14 jan 2020",
         "07 jan 2020","11 jan 2019","04 feb 2019","06 feb 2019","15 feb 2019",
         "17 feb 2019","11 mar 2019","15 mar 2019","11 mar 2019","04 mar 2019",
         "28 mar 2019"]),

   'Condition':(['used','used',np.nan,'new','new','used',np.nan,'new','new','used',
                'used','new',np.nan,'new','new',np.nan,'new','used','used','new']),

   'ModelNo':([101,102,103,103,105,106,107,108,109,110,111,111,np.nan,114,115,116,
               117,118,np.nan,120]),
   'Mrp':([2380, 3472, 3127, 2704, 2575, 3210, 2999, 3866, 2750, 3665, 2455, 3766,
           3944, 2085, 3628, 2992, 3599, 2122, 2821,2200]),
   'SoldPrice':([2975, 2536, 2199, 3515, 3089, 2288, 2399, 3966, 3575, 3765, 2946,
                 1883, 3844, 2490, 2711, 3590, 2159, 2546, 3385,2100]),
   'Manufacturing_Cost':([230,340,310,270,250,320,290,380,270,360,240,370,390,200,
                          360,290,350,210,2800,220]),
   'Extend_Amount':([23,34,31,27,25,32,29,38,27,36,24,37,39,20,36,29,35,21,28,22]),
   'Warranty_Period':([1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]),
   'Warranty_periodLeft(in Months)':([0,2,4,0,6,8,9,4,7,5,8,0,0,6,5,4,2,2,1,1]),
   'MD/CN':([0,0,0,0,1,0,1,1,1,0,1,1,1,1,1,1,1,0,0,0]),
   'Previous_claims':([1,0,0,2,1,1,1,0,0,0,0,0,0,0,1,2,2,3,1,1]),
   'Refund_Duration(in days)':([3,3,np.nan,3,5,7,2,2,2,2,3,3,np.nan,3,4,4,5,5,7,7])

   }
b=pd.DataFrame(a)
b['Profit']=b['SoldPrice']-b['Mrp']
b['Profit_Percentage']=b['Profit']*100/b['Mrp']
print(b)
```

```
   S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)  \
0    1.0     vagesh      m   hyderabad    01 jun 2020                   4.0   
1    2.0        NaN      m        pune   05 july 2020                   4.0   
2    3.0      reddy    NaN      mumbai    28 aug 2020                   4.0   
3    4.0  yashwanth      m  vijayawada    25 sep 2020                   3.0   
4    NaN        ram      m      mumbai    27 oct 2020                   NaN   
5    6.0      kumar      m     kolkata    31 nov 2020                   3.0   
6    7.0       abhi      f   bangalore    15 dec 2020                   3.0   
7    8.0    bhavana      f         NaN    17 dec 2020                   3.0   
8    9.0    meghana      f       delhi    04 jan 2021                   5.0   
9   10.0     swapna    NaN       vizag    17 jan 2021                   5.0   
10  11.0        ram      f    tirupati    21 jan 2021                   4.0   
11  12.0  keerthana      f     nellore    14 feb 2021                   NaN   
12   NaN        NaN    NaN      kadapa    16 feb 2021                   5.0   
13  14.0       siva      m     kurnool    25 feb 2021                   4.0   
14  15.0    shankar      m         NaN    27 feb 2021                   3.0   
15  16.0    pradeep      m  srikakulam    01 mar 2021                   2.0   
16  17.0    sameena      f   nizamabad    05 mar 2021                   1.0   
17  18.0      samba    NaN         NaN    10 mar 2021                   NaN   
18  19.0        sai      m     khammam    14 mar 2021                   2.0   
19  20.0      vamsi    NaN    nalgonda    18 mar 2021                   1.0   

   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used    101.0  2380       2975                 230   
1     15 july 2019      used    102.0  3472       2536                 340   
2      18 aug 2019       NaN    103.0  3127       2199                 310   
3      15 sep 2019       new    103.0  2704       3515                 270   
4      17 oct 2019       new    105.0  2575       3089                 250   
5      21 nov 2019      used    106.0  3210       2288                 320   
6      05 dec 2019       NaN    107.0  2999       2399                 290   
7      07 dec 2019       new    108.0  3866       3966                 380   
8      14 jan 2020       new    109.0  2750       3575                 270   
9      07 jan 2020      used    110.0  3665       3765                 360   
10     11 jan 2019      used    111.0  2455       2946                 240   
11     04 feb 2019       new    111.0  3766       1883                 370   
12     06 feb 2019       NaN      NaN  3944       3844                 390   
13     15 feb 2019       new    114.0  2085       2490                 200   
14     17 feb 2019       new    115.0  3628       2711                 360   
15     11 mar 2019       NaN    116.0  2992       3590                 290   
16     15 mar 2019       new    117.0  3599       2159                 350   
17     11 mar 2019      used    118.0  2122       2546                 210   
18     04 mar 2019      used      NaN  2821       3385                2800   
19     28 mar 2019       new    120.0  2200       2100                 220   
```

```
    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0   
1              34                1                               2      0   
2              31                1                               4      0   
3              27                1                               0      0   
4              25                1                               6      1   
5              32                1                               8      0   
6              29                1                               9      1   
7              38                1                               4      1   
8              27                1                               7      1   
9              36                1                               5      0   
10             24                1                               8      1   
11             37                1                               0      1   
12             39                1                               0      1   
13             20                1                               6      1   
14             36                1                               5      1   
15             29                1                               4      1   
16             35                1                               2      1   
17             21                1                               2      0   
18             28                1                               1      0   
19             22                1                               1      0   

    Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage  
0                 1                       3.0     595          25.000000  
1                 0                       3.0    -936         -26.958525  
2                 0                       NaN    -928         -29.677007  
3                 2                       3.0     811          29.992604  
4                 1                       5.0     514          19.961165  
5                 1                       7.0    -922         -28.722741  
6                 1                       2.0    -600         -20.006669  
7                 0                       2.0     100           2.586653  
8                 0                       2.0     825          30.000000  
9                 0                       2.0     100           2.728513  
10                0                       3.0     491          20.000000  
11                0                       3.0   -1883         -50.000000  
12                0                       NaN    -100          -2.535497  
13                0                       3.0     405          19.424460  
14                1                       4.0    -917         -25.275634  
15                2                       4.0     598          19.986631  
16                2                       5.0   -1440         -40.011114  
17                3                       5.0     424          19.981150  
18                1                       7.0     564          19.992910  
19                1                       7.0    -100          -4.545455  
```

```
print(b.isnull())
```

```
     S.No   Name  Gender   City  DateOfPurchase  Customer_Rating(0-5)  \
0   False  False   False  False           False                 False
1   False   True   False  False           False                 False
2   False  False    True  False           False                 False
3   False  False   False  False           False                 False
4    True  False   False  False           False                  True
5   False  False   False  False           False                 False
6   False  False   False  False           False                 False
7   False  False   False   True           False                 False
8   False  False   False  False           False                 False
9   False  False    True  False           False                 False
10  False  False   False  False           False                 False
11  False  False   False  False           False                  True
12   True   True    True  False           False                 False
13  False  False   False  False           False                 False
14  False  False   False   True           False                 False
15  False  False   False  False           False                 False
16  False  False   False  False           False                 False
17  False  False    True   True           False                  True
18  False  False   False  False           False                 False
19  False  False    True  False           False                 False

 ManufactureDate  Condition  ModelNo    Mrp  SoldPrice  Manufacturing_Cost  \
0             False      False    False  False      False               False
1             False      False    False  False      False               False
2             False       True    False  False      False               False
3             False      False    False  False      False               False
4             False      False    False  False      False               False
5             False      False    False  False      False               False
6             False       True    False  False      False               False
7             False      False    False  False      False               False
8             False      False    False  False      False               False
9             False      False    False  False      False               False
10            False      False    False  False      False               False
11            False      False    False  False      False               False
12            False       True     True  False      False               False
13            False      False    False  False      False               False
14            False      False    False  False      False               False
15            False       True    False  False      False               False
16            False      False    False  False      False               False
17            False      False    False  False      False               False
18            False      False     True  False      False               False
19            False      False    False  False      False               False
```

```
    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0           False            False                           False  False   
1           False            False                           False  False   
2           False            False                           False  False   
3           False            False                           False  False   
4           False            False                           False  False   
5           False            False                           False  False   
6           False            False                           False  False   
7           False            False                           False  False   
8           False            False                           False  False   
9           False            False                           False  False   
10          False            False                           False  False   
11          False            False                           False  False   
12          False            False                           False  False   
13          False            False                           False  False   
14          False            False                           False  False   
15          False            False                           False  False   
16          False            False                           False  False   
17          False            False                           False  False   
18          False            False                           False  False   
19          False            False                           False  False   

    Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage  
0             False                     False   False              False  
1             False                     False   False              False  
2             False                      True   False              False  
3             False                     False   False              False  
4             False                     False   False              False  
5             False                     False   False              False  
6             False                     False   False              False  
7             False                     False   False              False  
8             False                     False   False              False  
9             False                     False   False              False  
10            False                     False   False              False  
11            False                     False   False              False  
12            False                      True   False              False  
13            False                     False   False              False  
14            False                     False   False              False  
15            False                     False   False              False  
16            False                     False   False              False  
17            False                     False   False              False  
18            False                     False   False              False  
19            False                     False   False              False  
```

```
print(b.notnull())
```

```
     S.No   Name  Gender   City  DateOfPurchase  Customer_Rating(0-5)  \
0    True   True    True   True            True                  True
1    True  False    True   True            True                  True
2    True   True   False   True            True                  True
3    True   True    True   True            True                  True
4   False   True    True   True            True                 False
5    True   True    True   True            True                  True
6    True   True    True   True            True                  True
7    True   True    True  False            True                  True
8    True   True    True   True            True                  True
9    True   True   False   True            True                  True
10   True   True    True   True            True                  True
11   True   True    True   True            True                 False
12  False  False   False   True            True                  True
13   True   True    True   True            True                  True
14   True   True    True  False            True                  True
15   True   True    True   True            True                  True
16   True   True    True   True            True                  True
17   True   True   False  False            True                 False
18   True   True    True   True            True                  True
19   True   True   False   True            True                  True

  ManufactureDate  Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0              True       True     True  True       True                True
1              True       True     True  True       True                True
2              True      False     True  True       True                True
3              True       True     True  True       True                True
4              True       True     True  True       True                True
5              True       True     True  True       True                True
6              True      False     True  True       True                True
7              True       True     True  True       True                True
8              True       True     True  True       True                True
9              True       True     True  True       True                True
10             True       True     True  True       True                True
11             True       True     True  True       True                True
12             True      False    False  True       True                True
13             True       True     True  True       True                True
14             True       True     True  True       True                True
15             True      False     True  True       True                True
16             True       True     True  True       True                True
17             True       True     True  True       True                True
18             True       True    False  True       True                True
19             True       True     True  True       True                True
```

```
    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0            True             True                            True   True   
1            True             True                            True   True   
2            True             True                            True   True   
3            True             True                            True   True   
4            True             True                            True   True   
5            True             True                            True   True   
6            True             True                            True   True   
7            True             True                            True   True   
8            True             True                            True   True   
9            True             True                            True   True   
10           True             True                            True   True   
11           True             True                            True   True   
12           True             True                            True   True   
13           True             True                            True   True   
14           True             True                            True   True   
15           True             True                            True   True   
16           True             True                            True   True   
17           True             True                            True   True   
18           True             True                            True   True   
19           True             True                            True   True   


    Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage  
0              True                      True    True               True  
1              True                      True    True               True  
2              True                     False    True               True  
3              True                      True    True               True  
4              True                      True    True               True  
5              True                      True    True               True  
6              True                      True    True               True  
7              True                      True    True               True  
8              True                      True    True               True  
9              True                      True    True               True  
10             True                      True    True               True  
11             True                      True    True               True  
12             True                     False    True               True  
13             True                      True    True               True  
14             True                      True    True               True  
15             True                      True    True               True  
16             True                      True    True               True  
17             True                      True    True               True  
18             True                      True    True               True  
19             True                      True    True               True  
```

WRONG FORMAT OR WRONG DATA

It means an individual column contains same datatype values , if not it doesn't perform any statistical operations.

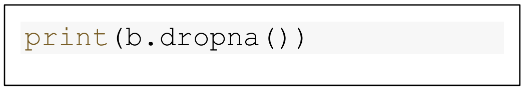
There are two methods to handle the data .

- dropna( ) – it drops all the NAN values
- fillna( ) – it is used to fill in the NAN values

Fillna() is again divided into two types

Fillna(method='pad') – replaces NAN values with the previous values.

Fillna(method='bfill') – replaces NAN values with the next values.



```
print(b.dropna())
```

```
   S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)  \
0    1.0     vagesh      m   hyderabad    01 jun 2020                   4.0
3    4.0  yashwanth      m  vijayawada    25 sep 2020                   3.0
5    6.0      kumar      m     kolkata    31 nov 2020                   3.0
8    9.0    meghana      f       delhi    04 jan 2021                   5.0
10  11.0        ram      f    tirupati    21 jan 2021                   4.0
13  14.0       siva      m     kurnool    25 feb 2021                   4.0
16  17.0    sameena      f   nizamabad    05 mar 2021                   1.0

   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used    101.0  2380       2975                 230
3      15 sep 2019       new    103.0  2704       3515                 270
5      21 nov 2019      used    106.0  3210       2288                 320
8      14 jan 2020       new    109.0  2750       3575                 270
10     11 jan 2019      used    111.0  2455       2946                 240
13     15 feb 2019       new    114.0  2085       2490                 200
16     15 mar 2019       new    117.0  3599       2159                 350


    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0
3              27                1                               0      0
5              32                1                               8      0
8              27                1                               7      1
10             24                1                               8      1
13             20                1                               6      1
16             35                1                               2      1
```

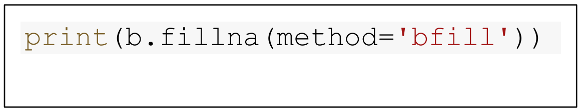
```
    Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage
0                 1                       3.0     595          25.000000
3                 2                       3.0     811          29.992604
5                 1                       7.0    -922         -28.722741
8                 0                       2.0     825          30.000000
10                0                       3.0     491          20.000000
13                0                       3.0     405          19.424460
16                2                       5.0   -1440         -40.011114
```

```python
print(b.fillna(method='pad'))
```

```
    S.No       Name Gender         City DateOfPurchase  Customer_Rating(0-5)  \
0    1.0     vagesh      m    hyderabad    01 jun 2020                   4.0
1    2.0     vagesh      m         pune   05 july 2020                   4.0
2    3.0      reddy      m       mumbai    28 aug 2020                   4.0
3    4.0  yashwanth      m   vijayawada    25 sep 2020                   3.0
4    4.0        ram      m       mumbai    27 oct 2020                   3.0
5    6.0      kumar      m      kolkata    31 nov 2020                   3.0
6    7.0       abhi      f    bangalore    15 dec 2020                   3.0
7    8.0    bhavana      f    bangalore    17 dec 2020                   3.0
8    9.0    meghana      f        delhi    04 jan 2021                   5.0
9   10.0     swapna      f        vizag    17 jan 2021                   5.0
10  11.0        ram      f     tirupati    21 jan 2021                   4.0
11  12.0  keerthana      f      nellore    14 feb 2021                   4.0
12  12.0  keerthana      f       kadapa    16 feb 2021                   5.0
13  14.0       siva      m      kurnool    25 feb 2021                   4.0
14  15.0    shankar      m      kurnool    27 feb 2021                   3.0
15  16.0    pradeep      m   srikakulam    01 mar 2021                   2.0
16  17.0    sameena      f    nizamabad    05 mar 2021                   1.0
17  18.0      samba      f    nizamabad    10 mar 2021                   1.0
18  19.0        sai      m      khammam    14 mar 2021                   2.0
19  20.0      vamsi      m     nalgonda    18 mar 2021                   1.0
```

```
   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used    101.0  2380       2975                 230
1     15 july 2019      used    102.0  3472       2536                 340
2      18 aug 2019      used    103.0  3127       2199                 310
3      15 sep 2019       new    103.0  2704       3515                 270
4      17 oct 2019       new    105.0  2575       3089                 250
5      21 nov 2019      used    106.0  3210       2288                 320
6      05 dec 2019      used    107.0  2999       2399                 290
7      07 dec 2019       new    108.0  3866       3966                 380
8      14 jan 2020       new    109.0  2750       3575                 270
9      07 jan 2020      used    110.0  3665       3765                 360
10     11 jan 2019      used    111.0  2455       2946                 240
11     04 feb 2019       new    111.0  3766       1883                 370
12     06 feb 2019       new    111.0  3944       3844                 390
13     15 feb 2019       new    114.0  2085       2490                 200
14     17 feb 2019       new    115.0  3628       2711                 360
15     11 mar 2019       new    116.0  2992       3590                 290
16     15 mar 2019       new    117.0  3599       2159                 350
17     11 mar 2019      used    118.0  2122       2546                 210
18     04 mar 2019      used    118.0  2821       3385                2800
19     28 mar 2019       new    120.0  2200       2100                 220


    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0
1              34                1                               2      0
2              31                1                               4      0
3              27                1                               0      0
4              25                1                               6      1
5              32                1                               8      0
6              29                1                               9      1
7              38                1                               4      1
8              27                1                               7      1
9              36                1                               5      0
10             24                1                               8      1
11             37                1                               0      1
12             39                1                               0      1
13             20                1                               6      1
14             36                1                               5      1
15             29                1                               4      1
16             35                1                               2      1
17             21                1                               2      0
18             28                1                               1      0
19             22                1                               1      0
```

```
      Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage
0                 1                       3.0     595          25.000000
1                 0                       3.0    -936         -26.958525
2                 0                       3.0    -928         -29.677007
3                 2                       3.0     811          29.992604
4                 1                       5.0     514          19.961165
5                 1                       7.0    -922         -28.722741
6                 1                       2.0    -600         -20.006669
7                 0                       2.0     100           2.586653
8                 0                       2.0     825          30.000000
9                 0                       2.0     100           2.728513
10                0                       3.0     491          20.000000
11                0                       3.0   -1883         -50.000000
12                0                       3.0    -100          -2.535497
13                0                       3.0     405          19.424460
14                1                       4.0    -917         -25.275634
15                2                       4.0     598          19.986631
16                2                       5.0   -1440         -40.011114
17                3                       5.0     424          19.981150
18                1                       7.0     564          19.992910
19                1                       7.0    -100          -4.545455
```



```
print(b.fillna(method='bfill'))
```

```
   S.No       Name Gender        City DateOfPurchase  Customer_Rating(0-5)   \
0    1.0     vagesh      m   hyderabad    01 jun 2020                    4.0
1    2.0      reddy      m        pune   05 july 2020                    4.0
2    3.0      reddy      m      mumbai    28 aug 2020                    4.0
3    4.0  yashwanth      m  vijayawada    25 sep 2020                    3.0
4    6.0        ram      m      mumbai    27 oct 2020                    3.0
5    6.0      kumar      m     kolkata    31 nov 2020                    3.0
6    7.0       abhi      f   bangalore    15 dec 2020                    3.0
7    8.0    bhavana      f       delhi    17 dec 2020                    3.0
8    9.0    meghana      f       delhi    04 jan 2021                    5.0
9   10.0     swapna      f       vizag    17 jan 2021                    5.0
10  11.0        ram      f    tirupati    21 jan 2021                    4.0
11  12.0  keerthana      f     nellore    14 feb 2021                    5.0
12  14.0       siva      m      kadapa    16 feb 2021                    5.0
13  14.0       siva      m     kurnool    25 feb 2021                    4.0
14  15.0    shankar      m  srikakulam    27 feb 2021                    3.0
15  16.0    pradeep      m  srikakulam    01 mar 2021                    2.0
16  17.0    sameena      f   nizamabad    05 mar 2021                    1.0
17  18.0      samba      m     khammam    10 mar 2021                    2.0
18  19.0        sai      m     khammam    14 mar 2021                    2.0
19  20.0      vamsi    NaN    nalgonda    18 mar 2021                    1.0
```

```
   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used    101.0  2380       2975                 230   
1     15 july 2019      used    102.0  3472       2536                 340   
2      18 aug 2019       new    103.0  3127       2199                 310   
3      15 sep 2019       new    103.0  2704       3515                 270   
4      17 oct 2019       new    105.0  2575       3089                 250   
5      21 nov 2019      used    106.0  3210       2288                 320   
6      05 dec 2019       new    107.0  2999       2399                 290   
7      07 dec 2019       new    108.0  3866       3966                 380   
8      14 jan 2020       new    109.0  2750       3575                 270   
9      07 jan 2020      used    110.0  3665       3765                 360   
10     11 jan 2019      used    111.0  2455       2946                 240   
11     04 feb 2019       new    111.0  3766       1883                 370   
12     06 feb 2019       new    114.0  3944       3844                 390   
13     15 feb 2019       new    114.0  2085       2490                 200   
14     17 feb 2019       new    115.0  3628       2711                 360   
15     11 mar 2019       new    116.0  2992       3590                 290   
16     15 mar 2019       new    117.0  3599       2159                 350   
17     11 mar 2019      used    118.0  2122       2546                 210   
18     04 mar 2019      used    120.0  2821       3385                2800   
19     28 mar 2019       new    120.0  2200       2100                 220   

    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0   
1              34                1                               2      0   
2              31                1                               4      0   
3              27                1                               0      0   
4              25                1                               6      1   
5              32                1                               8      0   
6              29                1                               9      1   
7              38                1                               4      1   
8              27                1                               7      1   
9              36                1                               5      0   
10             24                1                               8      1   
11             37                1                               0      1   
12             39                1                               0      1   
13             20                1                               6      1   
14             36                1                               5      1   
15             29                1                               4      1   
16             35                1                               2      1   
17             21                1                               2      0   
18             28                1                               1      0   
19             22                1                               1      0   
```

```
    Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage
0                 1                       3.0     595          25.000000
1                 0                       3.0    -936         -26.958525
2                 0                       3.0    -928         -29.677007
3                 2                       3.0     811          29.992604
4                 1                       5.0     514          19.961165
5                 1                       7.0    -922         -28.722741
6                 1                       2.0    -600         -20.006669
7                 0                       2.0     100           2.586653
8                 0                       2.0     825          30.000000
9                 0                       2.0     100           2.728513
10                0                       3.0     491          20.000000
11                0                       3.0   -1883         -50.000000
12                0                       3.0    -100          -2.535497
13                0                       3.0     405          19.424460
14                1                       4.0    -917         -25.275634
15                2                       4.0     598          19.986631
16                2                       5.0   -1440         -40.011114
17                3                       5.0     424          19.981150
18                1                       7.0     564          19.992910
19                1                       7.0    -100          -4.545455
```

REMOVE DUPLICATES

```
print(b.drop_duplicates)
```

```
    S.No       Name Gender       City DateOfPurchase  Customer_Rating(0-5)  \
0    1.0     vagesh      m  hyderabad    01 jun 2020                   4.0
1    2.0        NaN      m       pune   05 july 2020                   4.0
2    3.0      reddy    NaN     mumbai    28 aug 2020                   4.0
3    4.0  yashwanth      m vijayawada    25 sep 2020                   3.0
4    NaN        ram      m     mumbai    27 oct 2020                   NaN
5    6.0      kumar      m    kolkata    31 nov 2020                   3.0
6    7.0       abhi      f  bangalore    15 dec 2020                   3.0
7    8.0    bhavana      f        NaN    17 dec 2020                   3.0
8    9.0    meghana      f      delhi    04 jan 2021                   5.0
9   10.0     swapna    NaN      vizag    17 jan 2021                   5.0
10  11.0        ram      f   tirupati    21 jan 2021                   4.0
11  12.0  keerthana      f    nellore    14 feb 2021                   NaN
12   NaN        NaN    NaN     kadapa    16 feb 2021                   5.0
13  14.0       siva      m    kurnool    25 feb 2021                   4.0
14  15.0    shankar      m        NaN    27 feb 2021                   3.0
15  16.0    pradeep      m srikakulam    01 mar 2021                   2.0
16  17.0    sameena      f  nizamabad    05 mar 2021                   1.0
17  18.0      samba    NaN        NaN    10 mar 2021                   NaN
18  19.0        sai      m    khammam    14 mar 2021                   2.0
19  20.0      vamsi    NaN    nalgonda    18 mar 2021                 1.0
```

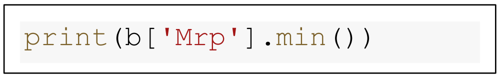
```
   ManufactureDate Condition  ModelNo   Mrp  SoldPrice  Manufacturing_Cost  \
0      11 jun 2019      used    101.0  2380       2975                 230
1     15 july 2019      used    102.0  3472       2536                 340
2      18 aug 2019       NaN    103.0  3127       2199                 310
3      15 sep 2019       new    103.0  2704       3515                 270
4      17 oct 2019       new    105.0  2575       3089                 250
5      21 nov 2019      used    106.0  3210       2288                 320
6      05 dec 2019       NaN    107.0  2999       2399                 290
7      07 dec 2019       new    108.0  3866       3966                 380
8      14 jan 2020       new    109.0  2750       3575                 270
9      07 jan 2020      used    110.0  3665       3765                 360
10     11 jan 2019      used    111.0  2455       2946                 240
11     04 feb 2019       new    111.0  3766       1883                 370
12     06 feb 2019       NaN      NaN  3944       3844                 390
13     15 feb 2019       new    114.0  2085       2490                 200
14     17 feb 2019       new    115.0  3628       2711                 360
15     11 mar 2019       NaN    116.0  2992       3590                 290
16     15 mar 2019       new    117.0  3599       2159                 350
17     11 mar 2019      used    118.0  2122       2546                 210
18     04 mar 2019      used      NaN  2821       3385                2800
19     28 mar 2019       new   120.0  2200       2100                 220


    Extend_Amount  Warranty_Period  Warranty_periodLeft(in Months)  MD/CN  \
0              23                1                               0      0
1              34                1                               2      0
2              31                1                               4      0
3              27                1                               0      0
4              25                1                               6      1
5              32                1                               8      0
6              29                1                               9      1
7              38                1                               4      1
8              27                1                               7      1
9              36                1                               5      0
10             24                1                               8      1
11             37                1                               0      1
12             39                1                               0      1
13             20                1                               6      1
14             36                1                               5      1
15             29                1                               4      1
16             35                1                               2      1
17             21                1                               2      0
18             28                1                               1      0
19             22                1                               1      0
```

```
        Previous_claims  Refund_Duration(in days)  Profit  Profit_Percentage
0                     1                       3.0     595          25.000000
1                     0                       3.0    -936         -26.958525
2                     0                       NaN    -928         -29.677007
3                     2                       3.0     811          29.992604
4                     1                       5.0     514          19.961165
5                     1                       7.0    -922         -28.722741
6                     1                       2.0    -600         -20.006669
7                     0                       2.0     100           2.586653
8                     0                       2.0     825          30.000000
9                     0                       2.0     100           2.728513
10                    0                       3.0     491          20.000000
11                    0                       3.0   -1883         -50.000000
12                    0                       NaN    -100          -2.535497
13                    0                       3.0     405          19.424460
14                    1                       4.0    -917         -25.275634
15                    2                       4.0     598          19.986631
16                    2                       5.0   -1440         -40.011114
17                    3                       5.0     424          19.981150
18                    1                       7.0     564          19.992910
19                    1                       7.0    -100          -4.545455
```

4) ANALYZING THE DATA

In this section we need to perform or manipulate statistical operations. This is the major step in the entire data analysis part in which we can perform different statistical tool operations. These are known as Computational Tools.

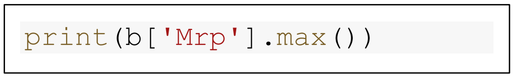
Min( )
Used to find the minimum value in the entire column.



```
print(b['Mrp'].min())
```

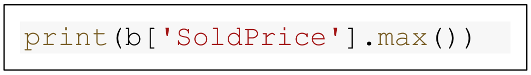
2085

Max( )
Used to find the maximum value in the entire column.



```
print(b['Mrp'].max())
```

3944



```
print(b['SoldPrice'].max())
```

3966

Rank( )

Gives a rank to the columns or entire data frame according to the ascending order or descending order for Numerical values , for alphabets it follows alphabetical order.

```
print(b['Mrp'].rank())
```

```
0      4.0
1     14.0
2     12.0
3      7.0
4      6.0
5     13.0
6     11.0
7     19.0
8      8.0
9     17.0
10     5.0
11    18.0
12    20.0
13     1.0
14    16.0
15    10.0
16    15.0
17     2.0
18     9.0
19     3.0
Name: Mrp, dtype: float64
```

CORRELATION-

- It is a relation between two data column data members.
- We use method called corr( )
- It is scaled form of a covariance.
- Correlation values lies between (-1 to +1)

Attribute1.corr(attribute2)

Types of correlations:

we have three types of correlation.

- Positive – (0 to1)
- Negative – (0 to -1)
- No correlation – (0)

```
print(b['Extend_Amount'].corr(b['MD/CN']))
```

0.22010645891035724

```
print(b['Mrp'].corr(b['Extend_Amount']))
```

0.9986407965884486

```
print(b['SoldPrice'].corr(b['Extend_Amount']))
```

0.1836663464733947

```
print(b['Previous_claims'].corr(b['MD/CN']))
```

-0.207514339159822 34

```
print(b['SoldPrice'].corr(b['MD/CN']))
```

0.12165190710678243

COVARIANCE-

The covariance is the relation between two data members of two different columns.

- It is the measure of a correlation.
- It lies between (-∞ to ∞).

```
print(b['Mrp'].cov(b['SoldPrice']))
```

73196.52631578947

```
print(b['Mrp'].cov(b['Extend_Amount']))
```

3646.9473684210525

```
print(b['SoldPrice'].cov(b['Extend_Amount']))
```

724.8605263157896

```
print(b['Previous_claims'].cov(b['MD/CN']))
```

-0.09473684210526313

```
print(b['SoldPrice'].cov(b['MD/CN']))
```

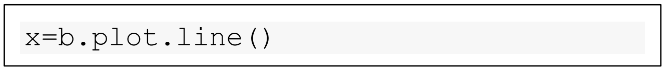
40.7078947368421

5)VISUALIZING THE DATA & SHARING THE RESULT

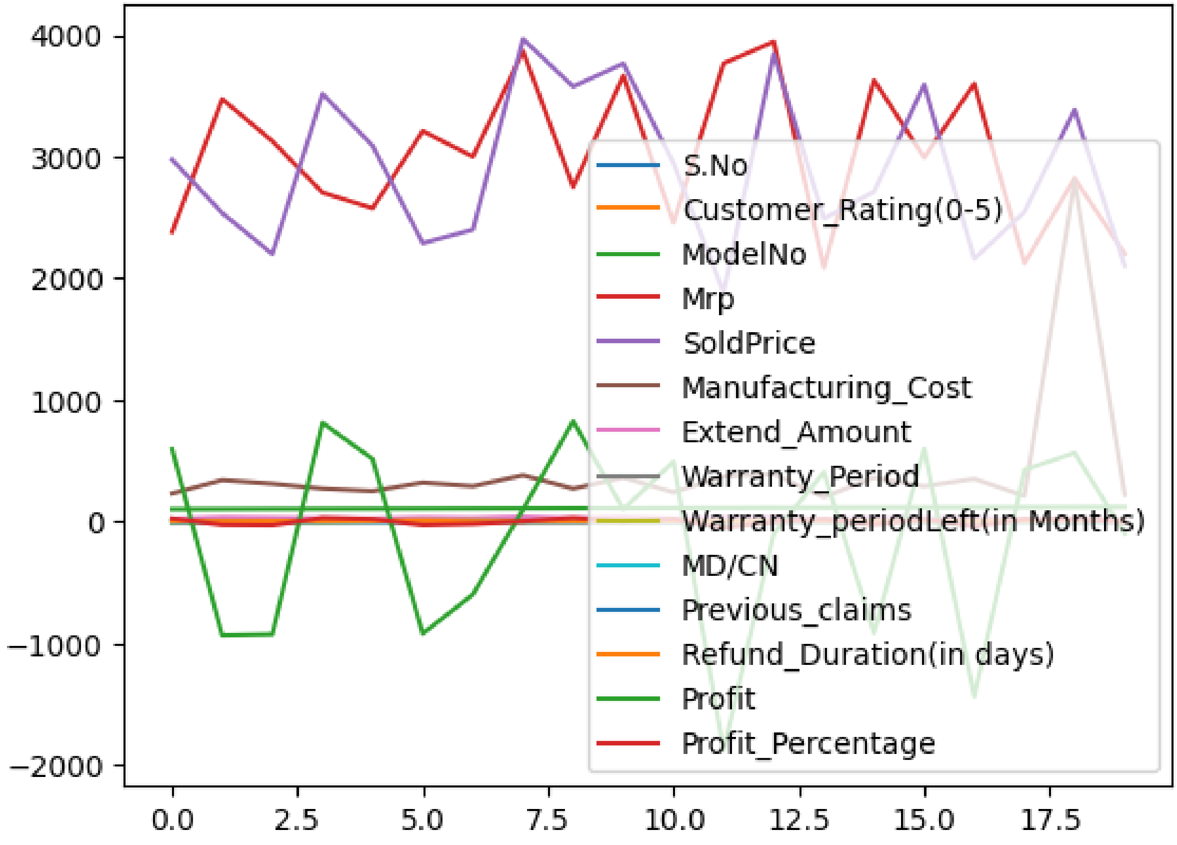
Data visualization is a process of representing data in a graphical way. Here we can represent the below graph formats.

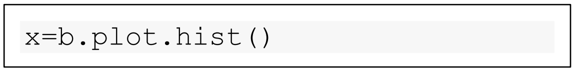
1. Line Graph
2. Bar Graph
3. Box Graph
4. KDE Graph
5. Area Graph
6. Histogram Graph

Here we need to use matplotlib module to represent graphs using the code.

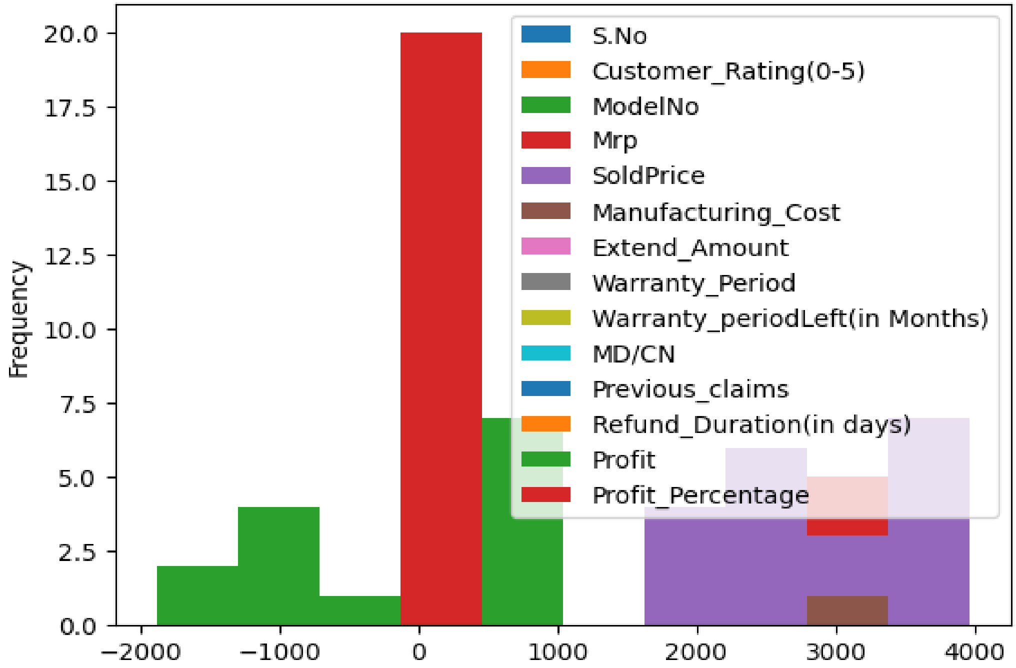


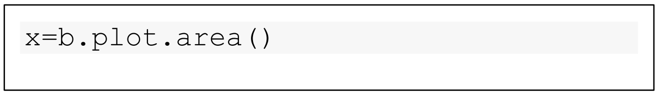
```
x=b.plot.line()
```



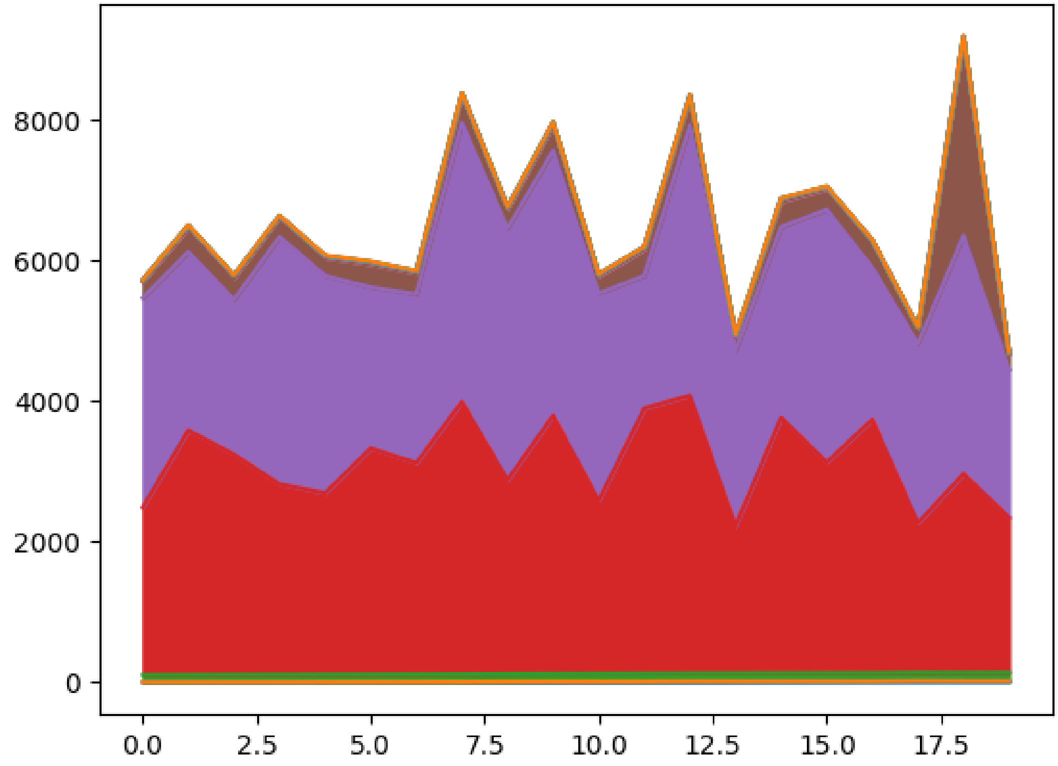


```
x=b.plot.hist()
```



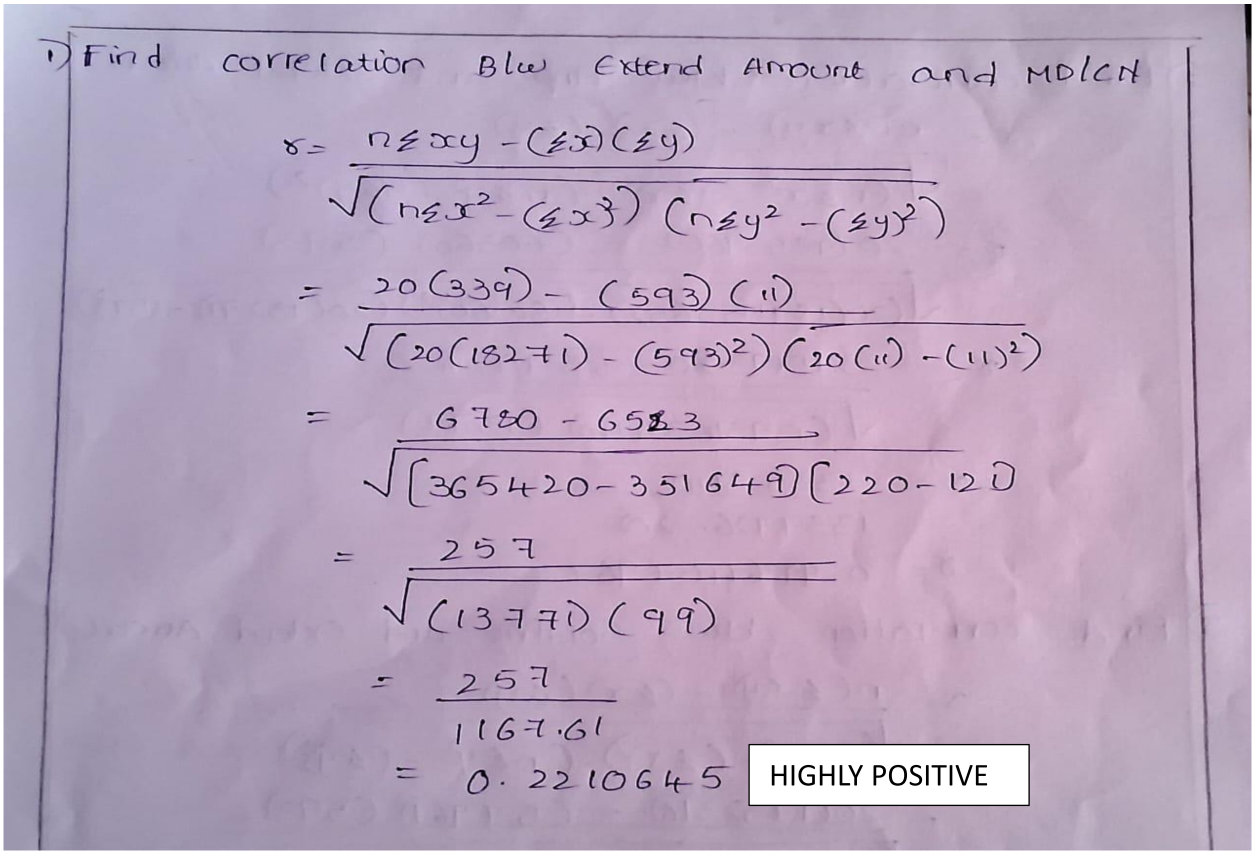


```
x=b.plot.area()
```



OUTPUT

CORELATION AND CO-VARIANCE ARE MANUALLY CALCULATED



1) Find correlation B/w Extend Amount and MD/CH

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{(n\Sigma x^2 - (\Sigma x)^2)(n\Sigma y^2 - (\Sigma y)^2)}}$$

$$= \frac{20(339) - (593)(11)}{\sqrt{(20(18271) - (593)^2)(20(11) - (11)^2)}}$$

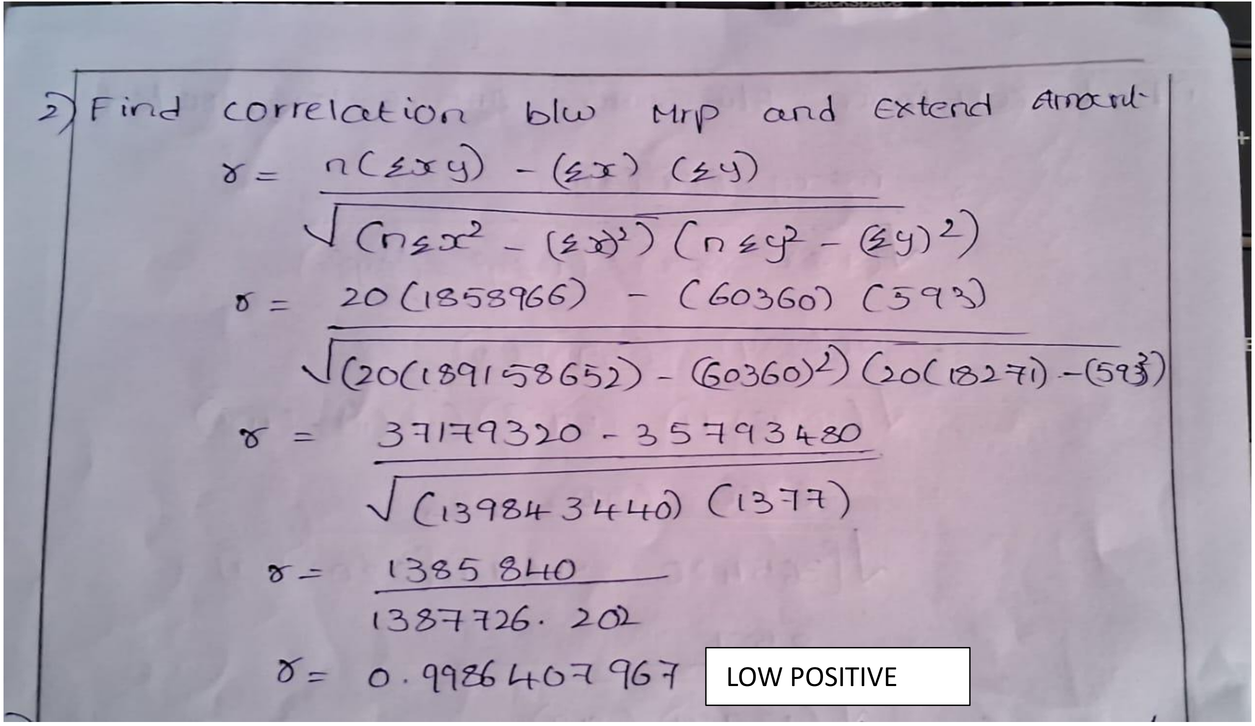
$$= \frac{6780 - 6523}{\sqrt{[365420 - 351649][220 - 121]}}$$

$$= \frac{257}{\sqrt{(13771)(99)}}$$

$$= \frac{257}{1167.61}$$

$$= 0.2210645$$

HIGHLY POSITIVE



2) Find correlation b/w Mrp and Extend Amount

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{(n\Sigma x^2 - (\Sigma x)^2)(n\Sigma y^2 - (\Sigma y)^2)}}$$

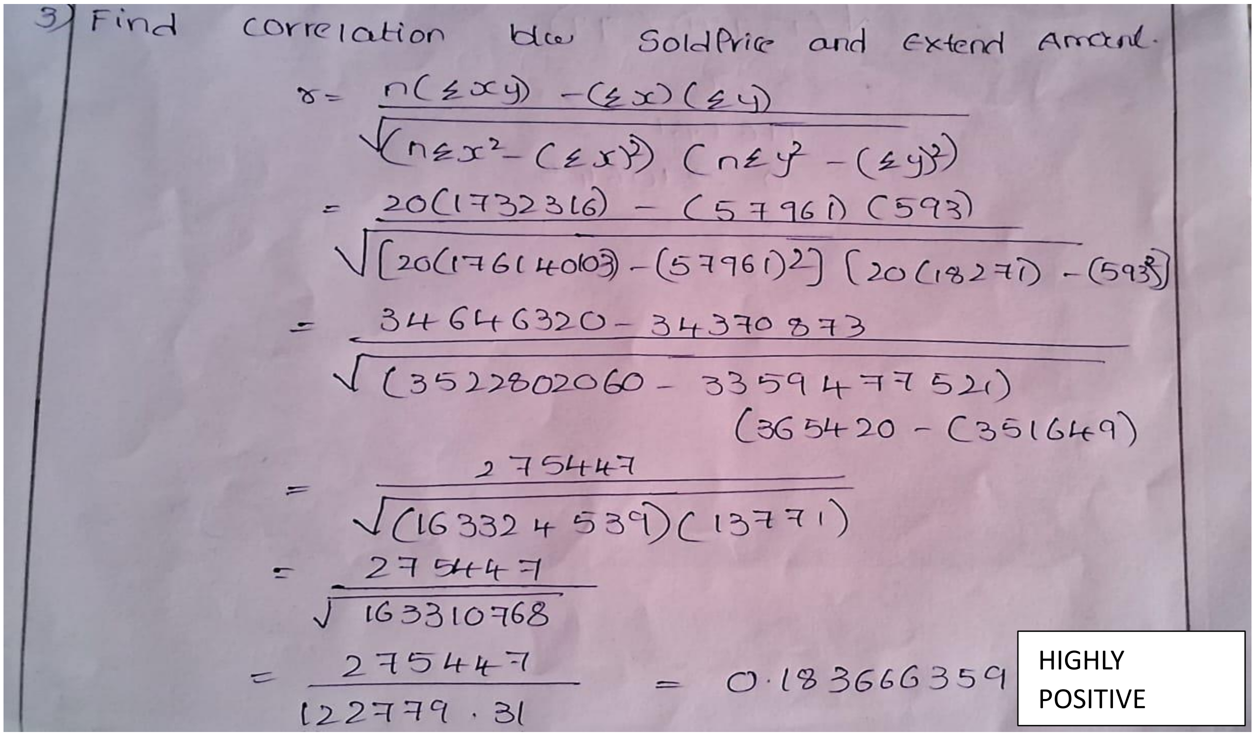
$$r = \frac{20(1858966) - (60360)(593)}{\sqrt{(20(189158652) - (60360)^2)(20(18271) - (593)^2)}}$$

$$r = \frac{37179320 - 35793480}{\sqrt{(139843440)(1377)}}$$

$$r = \frac{1385840}{1387726.202}$$

$$r = 0.9986407967$$

LOW POSITIVE



3) Find correlation b/w SoldPrice and Extend Amount.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)\ (n\sum y^2 - (\sum y)^2)}}$$

$$= \frac{20(1732316) - (57961)(593)}{\sqrt{[20(176140103) - (57961)^2]\ [20(18271) - (593)^2]}}$$

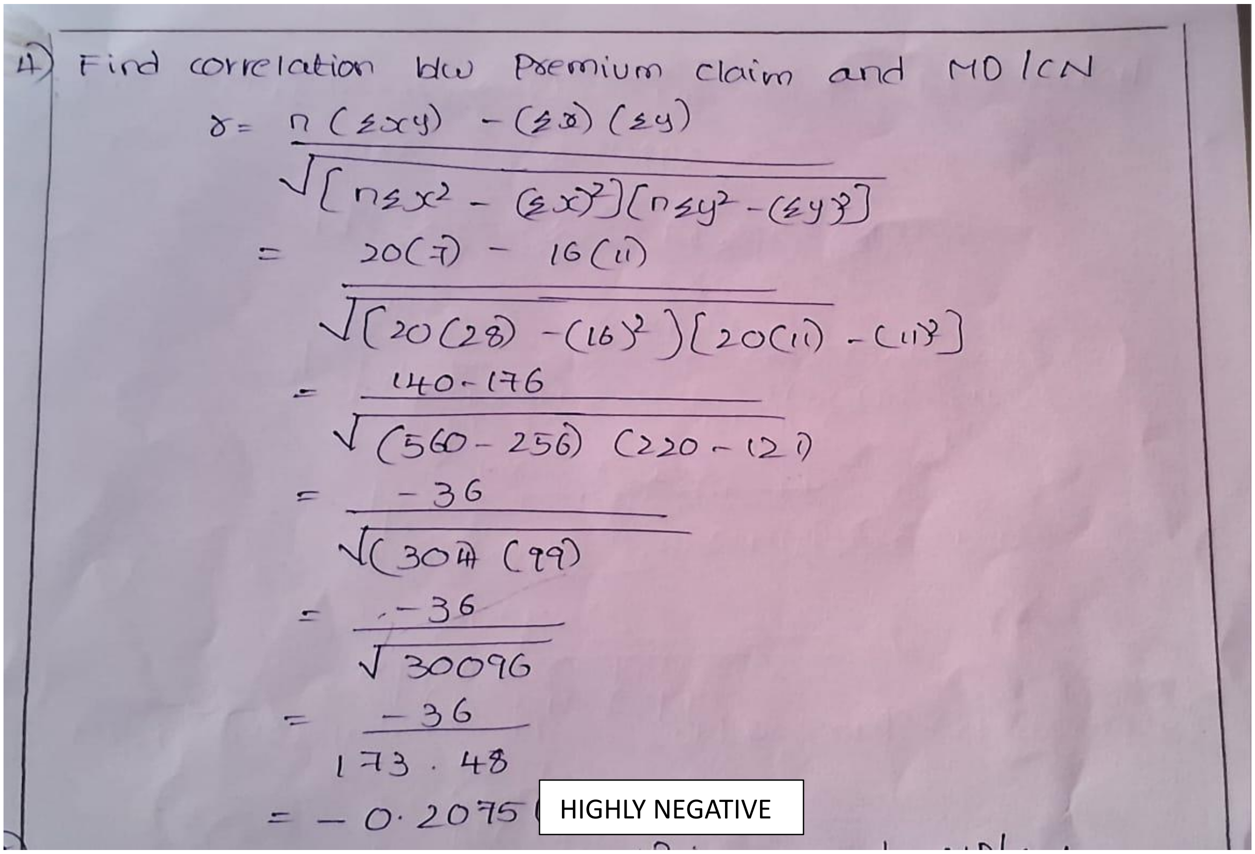
$$= \frac{34646320 - 34370873}{\sqrt{(3522802060 - 3359477521)\ (365420 - (351649)}}$$

$$= \frac{275447}{\sqrt{(163324539)(13771)}}$$

$$= \frac{275447}{\sqrt{163310768}}$$

$$= \frac{275447}{122779.31} = 0.183666359$$

HIGHLY POSITIVE



4) Find correlation b/w Premium claim and MD/CN

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$= \frac{20(7) - 16(11)}{\sqrt{[20(28) - (16)^2)[20(11) - (11)^2]}}$$

$$= \frac{140 - 176}{\sqrt{(560 - 256)\ (220 - 121)}}$$

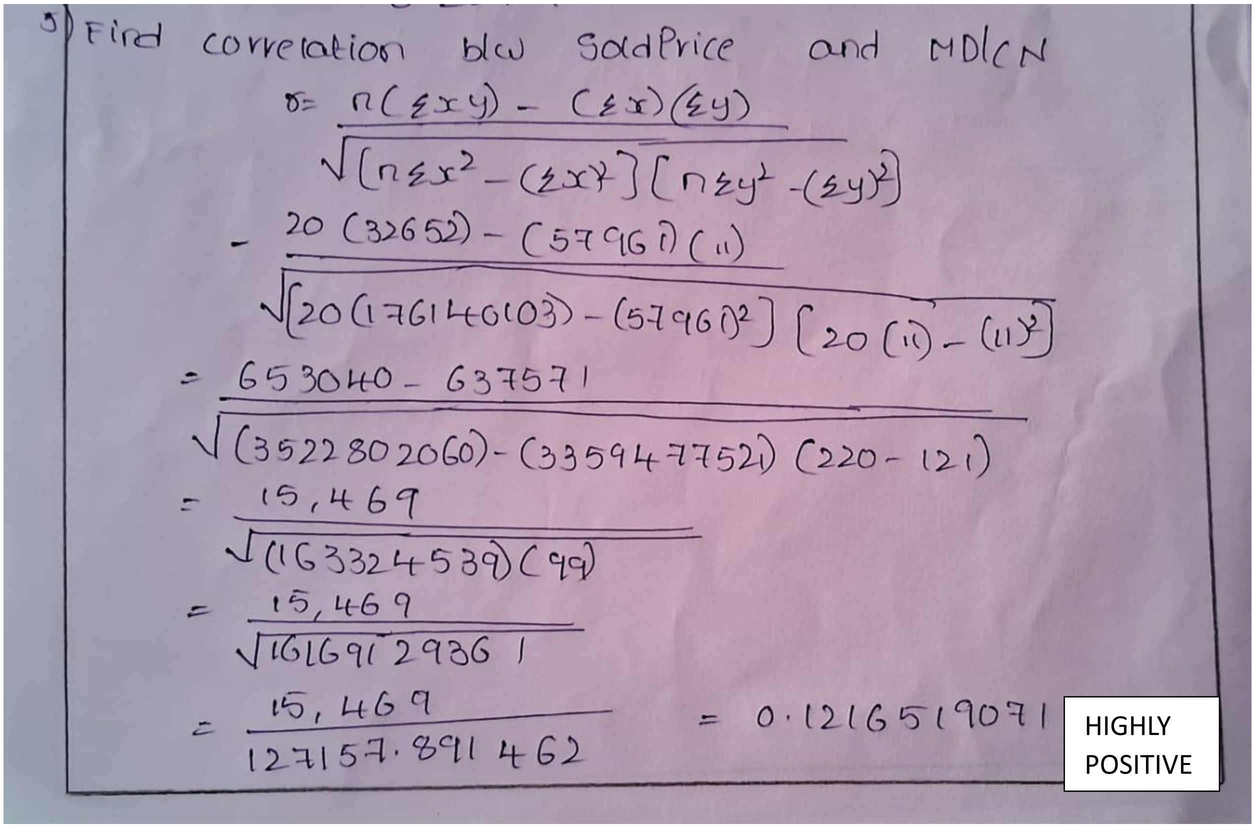
$$= \frac{-36}{\sqrt{(304)\ (99)}}$$

$$= \frac{-36}{\sqrt{30096}}$$

$$= \frac{-36}{173.48}$$

$$= -0.2075$$

HIGHLY NEGATIVE



5) Find correlation b/w SoldPrice and MD/CN

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{20(32652) - (57961)(11)}{\sqrt{[20(176140103) - (57961)^2][20(11) - (11)^2]}}$$

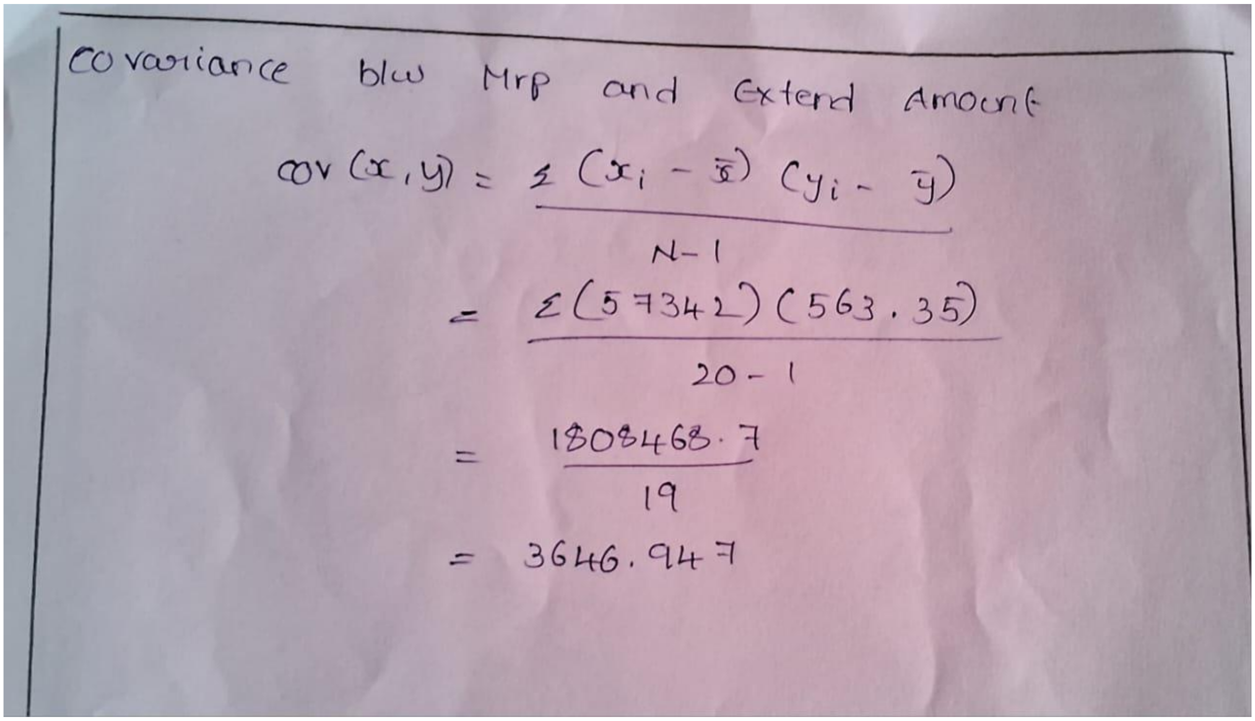
$$= \frac{653040 - 637571}{\sqrt{(3522802060) - (3359477521)(220 - 121)}}$$

$$= \frac{15,469}{\sqrt{(163324539)(99)}}$$

$$= \frac{15,469}{\sqrt{16169129361}}$$

$$= \frac{15,469}{127157.891462} = 0.1216519071$$

HIGHLY POSITIVE



Covariance b/w Mrp and Extend Amount

$$\text{cov}(x, y) = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

$$= \frac{\Sigma(57342)(563.35)}{20 - 1}$$

$$= \frac{1808468.7}{19}$$

$$= 3646.947$$

SUMMARY

In this project, we aimed to develop a warranty claim fraud detection system using data analytics, with a focus on time series analysis techniques implemented in Python.

Warranty claim fraud is a significant issue for many businesses, and detecting fraudulent claims is essential to minimize financial losses and maintain trust with customers.

Benefits:

- Improved fraud detection capabilities, leading to cost savings for the organization.
- Enhanced customer trust and satisfaction by reducing fraudulent claims.
- Data-driven decision-making for warranty claim management.