UNITED UNIVERSITY

SUBJECT NAME : DATA SCIENCE
PRACTICAL FILE
SESSION: 2025-26

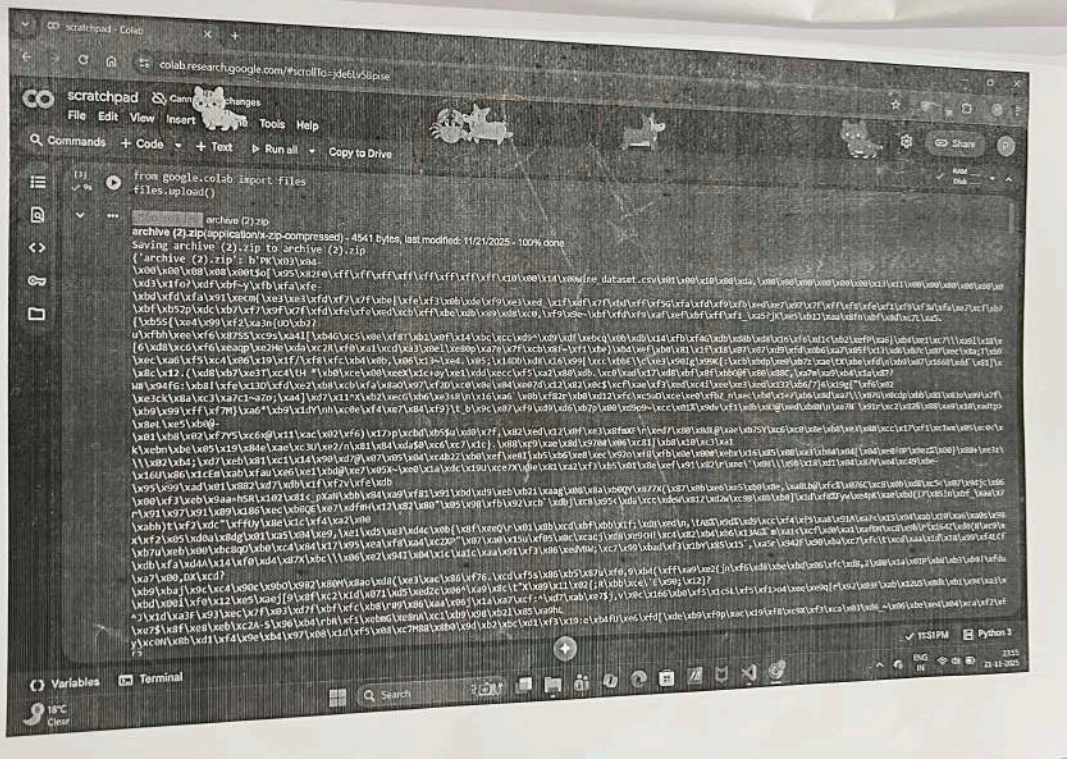
SUBMITTED BY:
STUDENT NAME: ANSHUMENDRA SINGH
(UNIVERSITY ROLL NO.):24201020021

SUBMITTED TO:

MR.SAMARTH SIR

COURSE: BCA IBM
SEMESTER: 3rd

UNITED UNIVERSITY
RAWATPUR, PRAYAGRAJ, UTTAR PRADESH- 211012



scratchpad - Colab

colab.research.google.com/#scrollTo=jde6Lv5Bpise

scratchpad — Cannot save changes

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▷ Run all Copy to Drive

Share

[7]
```
df.tail()
```

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	od280/od315_of_diluted_wi	
173	13.71	5.65	2.45	20.5	95	1.68	0.61	0.52	1.06	7.7	0.64		1
174	13.40	3.91	2.48	23.0	102	1.80	0.75	0.43	1.41	7.3	0.70		1
175	13.27	4.28	2.26	20.0	120	1.59	0.69	0.43	1.35	10.2	0.59		1
176	13.17	2.59	2.37	20.0	120	1.65	0.68	0.53	1.46	9.3	0.60		1
177	14.13	4.10	2.74	24.5	96	2.05	0.76	0.56	1.35	9.2	0.61		1

[9]
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.pyplot as plt
```

[11]
```
df = pd.read_csv('wine_dataset.csv')
print(df.head())
```

```
   alcohol  malic_acid   ash  alcalinity_of_ash  magnesium  total_phenols  \
0    14.23        1.71  2.43               15.6        127           2.80
1    13.20        1.78  2.14               11.2        100           2.65
2    13.16        2.36  2.67               18.6        101           2.80
3    14.37        1.95  2.50               16.8        113           3.85
4    13.24        2.59  2.87               21.0        118           2.80

   flavanoids  nonflavanoid_phenols  proanthocyanins  color_intensity   hue  \
0        3.06                  0.28             2.29             5.64  1.04
1        2.76                  0.26             1.28             4.38  1.05
2        3.24                  0.30             2.81             5.68  1.03
```

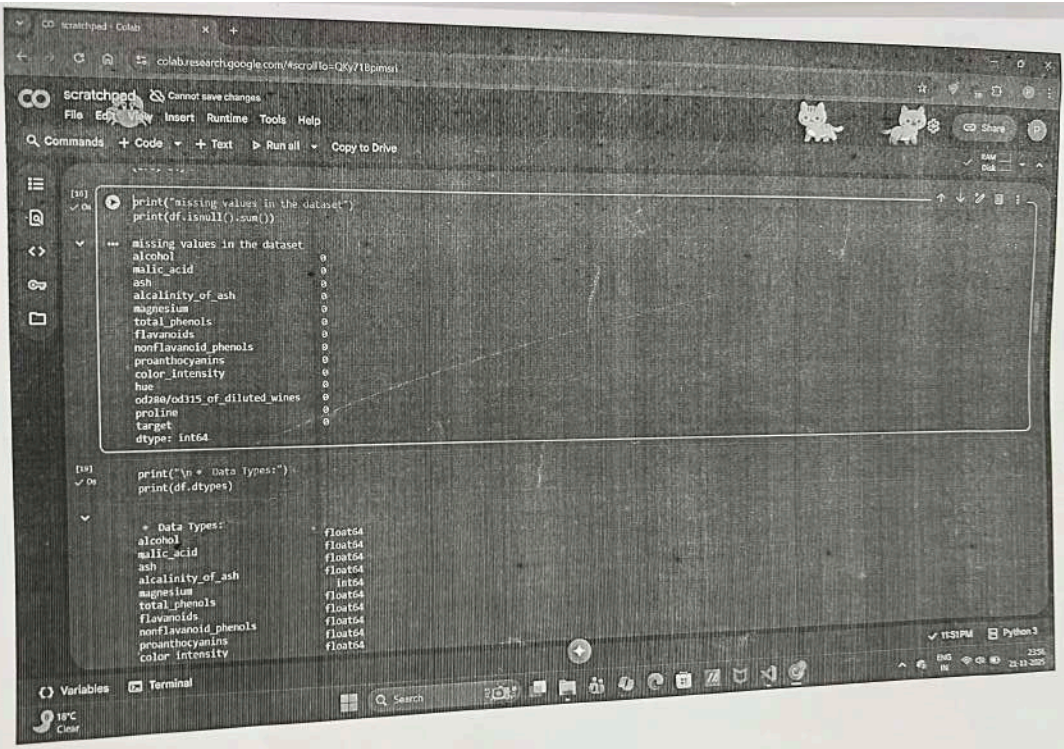
Variables Terminal

11:31PM Python 3

18°C Clear

Search

ENG IN 23:31 21-11-2025


scratchpad - Colab
colab.research.google.com/#scrollTo=QKy71Bpimsh
scratchpad
Cannot save changes
File Edit View Insert Runtime Tools Help
Commands + Code + Text Run all Copy to Drive
Share
[16]
print("missing values in the dataset")
print(df.isnull().sum())
missing values in the dataset
alcohol 0
malic_acid 0
ash 0
alcalinity_of_ash 0
magnesium 0
total_phenols 0
flavanoids 0
nonflavanoid_phenols 0
proanthocyanins 0
color_intensity 0
hue 0
od280/od315_of_diluted_wines 0
proline 0
target 0
dtype: int64
[18]
print("\n * Data Types:")
print(df.dtypes)
* Data Types:
alcohol float64
malic_acid float64
ash float64
alcalinity_of_ash float64
magnesium int64
total_phenols float64
flavanoids float64
nonflavanoid_phenols float64
proanthocyanins float64
color_intensity float64
11:51PM
Python 3
Variables
Terminal
Search
18°C
Clear

scratchpad - Colab

colab.research.google.com/#scrollTo=adWTjjVUn3rW

scratchpad

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all Copy to Drive

Share

```
print("\n📊 Summary Statistics:")
print(df.describe(include='all'))
```

```
📊 Summary Statistics:
          alcohol  malic_acid         ash  alcalinity_of_ash   magnesium  \
count  178.000000  178.000000  178.000000         178.000000  178.000000
mean    13.000618    2.336348    2.366517          19.494944   99.741573
std      0.811827    1.117146    0.274344           3.339564   14.282484
min     11.030000    0.740000    1.360000          10.600000   70.000000
25%     12.362500    1.602500    2.210000          17.200000   88.000000
50%     13.050000    1.865000    2.360000          19.500000   98.000000
75%     13.677500    3.082500    2.557500          21.500000  107.000000
max     14.830000    5.800000    3.230000          30.000000  162.000000

       total_phenols  flavanoids  nonflavanoid_phenols  proanthocyanins  \
count     178.000000  178.000000            178.000000       178.000000
mean        2.295112    2.029270              0.361854         1.590899
std         0.625851    0.998859              0.124453         0.572359
min         0.980000    0.340000              0.130000         0.410000
25%         1.742500    1.205000              0.270000         1.250000
50%         2.355000    2.135000              0.340000         1.555000
75%         2.800000    2.875000              0.437500         1.950000
max         3.880000    5.080000              0.660000         3.580000

       color_intensity         hue  od280/od315_of_diluted_wines      proline  \
count       178.000000  178.000000                    178.000000   178.000000
mean          5.058090    0.957449                      2.611685   746.893258
std           2.318286    0.228572                      0.709990   314.907474
min           1.280000    0.480000                      1.270000   278.000000
25%           3.220000    0.782500                      1.937500   500.500000
50%           4.690000    0.965000                      2.780000   673.500000
75%           6.200000    1.120000                      3.170000   985.000000
max          13.000000    1.710000                      4.000000  1680.000000

           target
count  178.000000
```

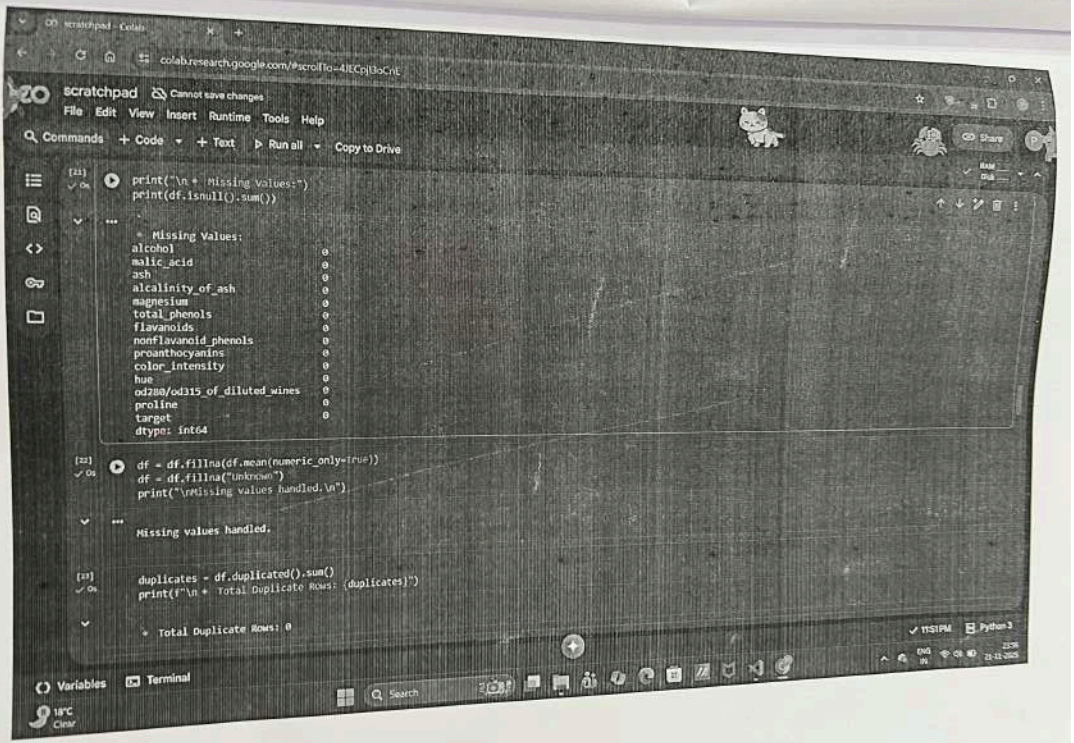
Variables Terminal

11:51PM Python 3

18°C Clear

Search

ENG IN 23:56 23-11-2025


scratchpad - Colab
colab.research.google.com/#scrollTo=4JECh13oCnE
scratchpad
Cannot save changes
File Edit View Insert Runtime Tools Help
Commands + Code + Text Run all Copy to Drive
Share
[21]
print("\n + Missing Values:")
print(df.isnull().sum())
+ Missing Values:
alcohol 0
malic_acid 0
ash 0
alcalinity_of_ash 0
magnesium 0
total_phenols 0
flavanoids 0
nonflavanoid_phenols 0
proanthocyanins 0
color_intensity 0
hue 0
od280/od315_of_diluted_wines 0
proline 0
target 0
dtype: int64
[22]
df = df.fillna(df.mean(numeric_only=True))
df = df.fillna("Unknown")
print("\nMissing values handled.\n")
Missing values handled.
[23]
duplicates = df.duplicated().sum()
print(f"\n + Total Duplicate Rows: {duplicates}")
+ Total Duplicate Rows: 0
11:51PM
Python 3
Variables
Terminal
18°C
Clear
Search

scratchpad - Colab

colab.research.google.com/#scrollTo=AJECpji3oCnE

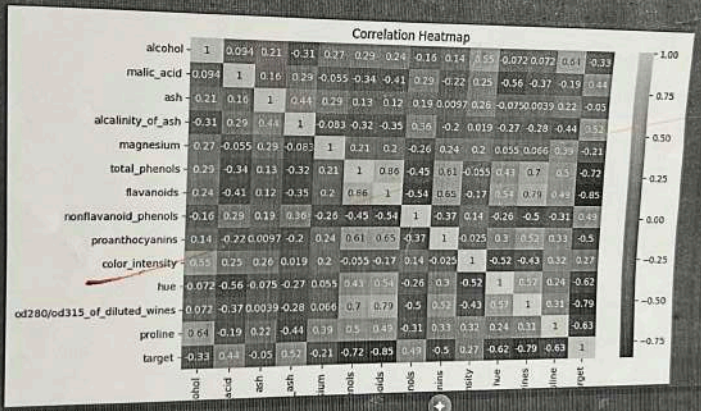
scratchpad | Cannot save changes

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▷ Run all Copy to Drive

Share

```
import seaborn as sns
plt.figure(figsize=(10,6))
sns.heatmap(df.corr(numeric_only=True), annot=True)
plt.title("Correlation Heatmap")
plt.show()
```



Variables Terminal

11:51PM Python 3

18°C Clear

Search

colab.research.google.com/#scrollTo=4UECpJ0oCnE

scratchpad Cannot save changes

Edit View Insert Runtime Tools Help

Commands + Code + Text ▷ Run all Copy to Drive

Share

```
[21] for col in df.select_dtypes(include="object"):
         print(f"Most common value in '{col}': {df[col].mode()[0]}")
```

```
[22] print("\nColumns with highest variance:")
     print(df.var(numeric_only=True).sort_values(ascending=False).head())
```

```
Columns with highest variance:
proline              99166.717355
magnesium              203.989335
alcalinity_of_ash       11.152686
color_intensity          5.374449
malic_acid               1.248015
dtype: float64
```

```
[37] num_columns = df.select_dtypes(include=np.number).columns

     if len(num_columns) > 0:
         sample_col = num_columns[0]
         print(f"\nTop 10 values in '{sample_col}':")
         print(df[sample_col].nlargest(10))
```

```
Top 10 values in 'alcohol':
8      14.83
13     14.75
6      14.39
14     14.38
46     14.38
3      14.37
158    14.34
16     14.30
0      14.23
```

11:51PM Python 3

Variables Terminal

18°C Clear

Search

ENG IN 22:57 25-11-2025

scratchpad - Colab

colab.research.google.com/#scrollTo=4JECpjlOCbE

scratchpad — Cannot save changes

File Edit View Insert Runtime Tools Help

Commands + Code + Text ▷ Run all Copy to Drive

Share

[40]
```
possible_groups = ["Category", "Region", "City", "Customer Segment", "Product"]
for g in possible_groups:
    if g in df.columns:
        print(f"\nAverage values grouped by {g}:")
        print(df.groupby(g).mean(numeric_only=True))
        break

print("\nSkewness of Numeric Columns:")
print(df.skew(numeric_only=True))
```

```
Skewness of Numeric Columns:
alcohol                         -0.051482
malic_acid                       1.039651
ash                             -0.176699
alcalinity_of_ash                0.213047
magnesium                        1.098191
total_phenols                    0.086639
flavanoids                       0.025344
nonflavanoid_phenols             0.450151
proanthocyanins                  0.517137
color_intensity                  0.868585
hue                              0.021091
od280/od315_of_diluted_wines    -0.307285
proline                          0.767822
target                           0.107431
dtype: float64
```

[42]
```
df[num_columns].hist(figsize=(12,6))
plt.show()


for col in num_columns:
    plt.figure(figsize=(5,3))
```

Variables Terminal

11:31 PM Python 3

18°C Clear

Search

scratchpad - Colab

colab.research.google.com/#scrollTo=4JECpJ3oCnE

scratchpad Cannot save changes

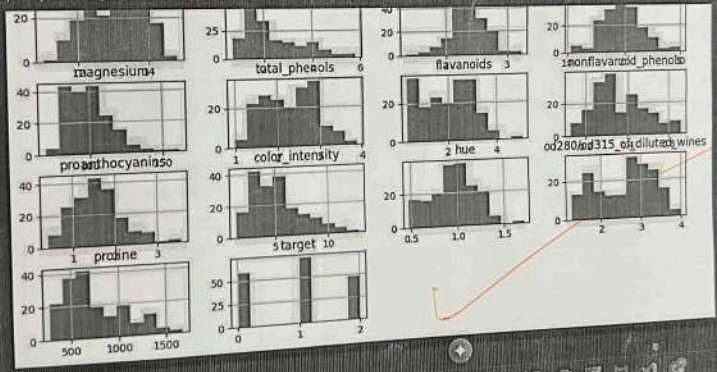
File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all Copy to Drive

Share

```
for col in num_columns:
    plt.figure(figsize=(5,3))
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()

print("\n✓ Analysis Completed Successfully!")
```



Variables Terminal

11:51PM Python 3

18°C Clear

Search

scratchpad - Colab

colab.research.google.com/#scrollTo=4JECpji3oCnE

scratchpad Cannot save changes

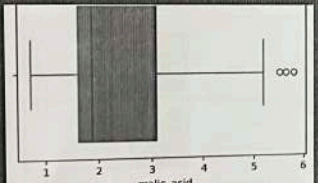
File Edit View Insert Runtime Tools Help

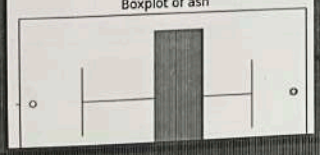
Commands + Code + Text Run all Copy to Drive

Share

```
for col in num_columns:
    plt.figure(figsize=(5,3))
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()

print("\n✓ Analysis Completed Successfully!")
```





Variables Terminal

11:51PM Python 3

18°C Clear

Search

ENG IN

23:57 22-11-2025

scratchpad - Colab

colab.research.google.com/#scrollTo=4JECpJ3oCnE

scratchpad | Cannot save changes

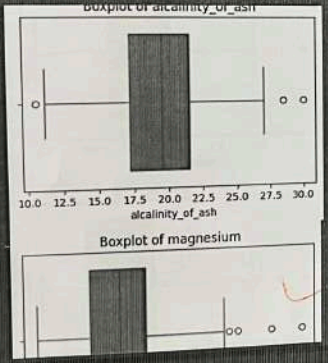
File Edit View Insert Runtime Tools Help

Commands + Code + Text | Run all | Copy to Drive

Share

```
for col in num_columns:
    plt.figure(figsize=(5,3))
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()

print("\n✓ Analysis Completed Successfully!")
```



Variables Terminal

11:51PM Python 3

18°C Clear

Search

scratchpad - Colab

colab.research.google.com/#scrollTo=4JECpjl3oCnE

scratchpad Cannot save changes

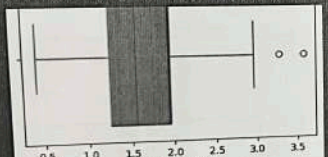
File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all Copy to Drive

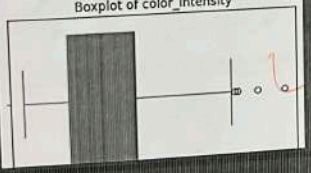
Share

```
for col in num_columns:
    plt.figure(figsize=(5,3))
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()

print("\n✓ Analysis Completed Successfully!")
```



Boxplot of color_intensity



Variables Terminal

11:51 PM Python 3

18°C Clear

Search

scratchpad - Colab

colab.research.google.com/#scrollTo=4JECpJ3oCnE

scratchpad Cannot save changes

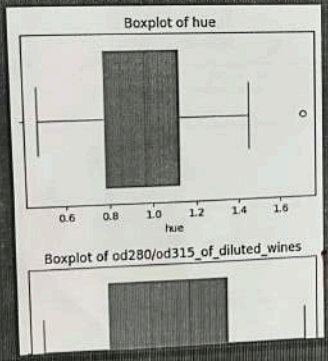
File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all Copy to Drive

Share

```
for col in num_columns:
    plt.figure(figsize=(5,3))
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()

print("\n✓ Analysis Completed Successfully!")
```

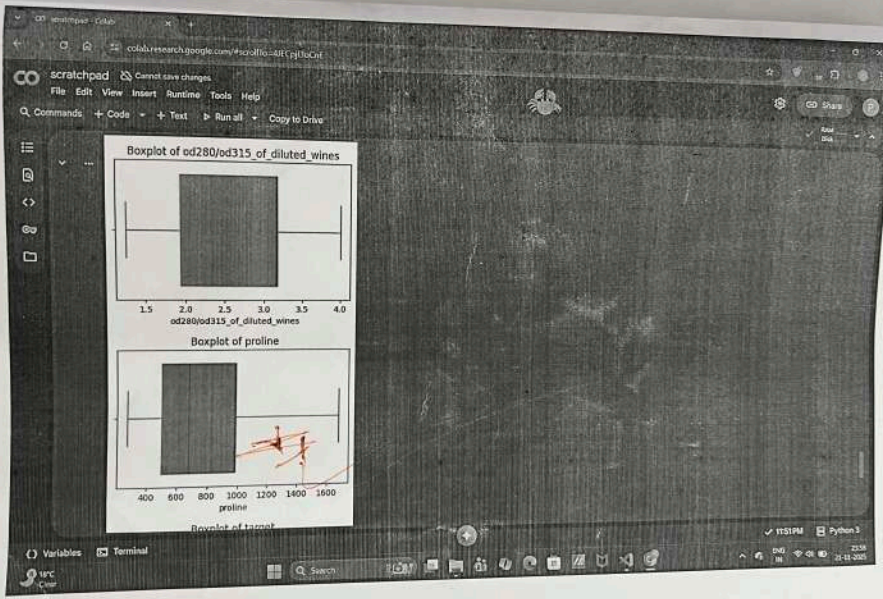


Variables Terminal

Python 3

18°C Clear

Search


scratchpad
Cannot save changes
File Edit View Insert Runtime Tools Help
Commands + Code + Text Run all Copy to Drive
Boxplot of od280/od315_of_diluted_wines
1.5 2.0 2.5 3.0 3.5 4.0
od280/od315_of_diluted_wines
Boxplot of proline
400 600 800 1000 1200 1400 1600
proline
Variables Terminal
Python 3