Name – Bhushan Yogesh Bhamare

Div – CS7

PRN – 202401110078

Roll No – CS7-69

Make Up Session Assignment

Take one real dataset. Find any 10 grains for the given dataset. Find solutions using pandas.

**Dataset Name –** White seeds dataset

Grain 1 - What is the shape of the dataset (rows, columns)?

df.shape

Output - (210, 8)

Grain 2 - How many grains belong to each variety?

df['variety'].value\_counts()

Output –

1 70

2 70

3 70

Grain 3 - What is the average area of grains for each variety?

df.groupby('variety')['area'].mean()

Output –

variety

1 15.478571

2 14.014714

3 15.048571

### Grain 4 - Which grain has the maximum length?

df[df['length'] == df['length'].max()]

area perimeter compactness length width asymmetry\_coef groove\_length variety

180 19.31 16.59 0.8687 6.579 3.814 0.545 5.762 3

Grain 5 - What is the correlation between area and perimeter?

df['area'].corr(df['perimeter'])

Output - 0.9946416242045072

### Grain 6 - **How many grains have compactness > 0.9?**

### df[df['compactness'] > 0.9].shape[0]

### Output – 54

### Grain 7 - **What is the mean groove length per variety?**

df.groupby('variety')['groove\_length'].mean()

Output –

variety

1 5.160857

2 4.933286

3 6.165714

Grain 8 - Are there any missing values in the dataset?

df.isnull().sum()

Output –

area 0

perimeter 0

compactness 0

length 0

width 0

asymmetry\_coef 0

groove\_length 0

variety 0

dtype: int64

Grain 9 - Class Distribution

df['Wheat Type'].value\_counts()

Output –

Kama 70

Rosa 70

Canadian 70

Name: Wheat Type, dtype: int64

### Grain 10 - What are the min/max values of each feature per seed type?

df.groupby('Wheat Type').agg(['min', 'max']).T

Output - Kama Rosa Canadian

min max min max min max

Area 10.59 19.18 11.23 18.71 6.812 19.24

Perimeter 12.41 15.68 12.92 15.59 13.43 15.56

Compactness 0.82 0.918 0.83 0.916 0.82 0.93