

## ▼ Transforming Raw cpdata to mergable data

### ▼ Filter cpdata.csv to MergeFileCrop.csv

#### File fertilizer.csv to MergeFileFert.csv

```
import pandas as pd
```

```
crop_data_path = '/content/cpdata.csv'
fertilizer_data_path = '/content/Fertilizer.csv'
```

```
crop = pd.read_csv(crop_data_path)
fert = pd.read_csv(fertilizer_data_path)
```

```
crop.head()
```

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice

```
fert.head()
```

	Unnamed: 0	Crop	N	P	K	pH
0	0	Rice	80	40	40	5.5
1	1	Jowar(Sorghum)	80	40	40	5.5
2	2	Barley(JAV)	70	40	45	5.5
3	3	Maize	80	40	20	5.5
4	4	Ragi( naachnnii)	50	40	20	5.5

```
# Function for lowering the cases
def change_case(i):
```

```
i = i.replace(" ", "")
i = i.lower()
return i
```

```
fert['Crop'] = fert['Crop'].apply(change_case)
crop['label'] = crop['label'].apply(change_case)
```

```
#make some changes in ferttilizer dataset
```

```
fert['Crop'] = fert['Crop'].replace('mungbeans','mungbean')
fert['Crop'] = fert['Crop'].replace('lentils(masoordal)','lentil')
fert['Crop'] = fert['Crop'].replace('pigeonpeas(toordal)','pigeonpeas')
fert['Crop'] = fert['Crop'].replace('mothbean(matki)','mothbeans')
fert['Crop'] = fert['Crop'].replace('chickpeas(channa)','chickpea')
```

```
crop.head()
```

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice

```
crop.tail()
```

	temperature	humidity	ph	rainfall	label
3095	25.287846	89.636679	6.765095	58.286977	watermelon
3096	26.638386	84.695469	6.189214	48.324286	watermelon
3097	25.331045	84.305338	6.904242	41.532187	watermelon
3098	26.897502	83.892415	6.463271	43.971937	watermelon
3099	26.986037	89.413849	6.260839	58.548767	watermelon

```
crop_names = crop['label'].unique()
crop_names
```

```
array(['rice', 'wheat', 'mungbean', 'tea', 'millet', 'maize', 'lentil',
       'jute', 'coffee', 'cotton', 'groundnut', 'peas', 'rubber',
       'sugarcane', 'tobacco', 'kidneybeans', 'mothbeans', 'coconut',
       'blackgram', 'adzukibeans', 'pigeonpeas', 'chickpea', 'banana',
```

```
'grapes', 'apple', 'mango', 'muskmelon', 'orange', 'papaya',
'pomegranate', 'watermelon'], dtype=object)
```

```
fert.head()
```

	Unnamed: 0	Crop	N	P	K	pH
0	0	rice	80	40	40	5.5
1	1	jowar(sorghum)	80	40	40	5.5
2	2	barley(jav)	70	40	45	5.5
3	3	maize	80	40	20	5.5
4	4	ragi(naachnnii)	50	40	20	5.5

```
del fert['Unnamed: 0']
```

```
crop_names_from_fert = fert['Crop'].unique()
crop_names_from_fert
```

```
array(['rice', 'jowar(sorghum)', 'barley(jav)', 'maize',
'ragi(naachnnii)', 'chickpea', 'frenchbeans(farasbi)',
'favabeans(papdi-val)', 'limabeans(pavta)', 'clusterbeans(gavar)',
'soyabean', 'blackeyedbeans(chawli)', 'kidneybeans', 'pigeonpeas',
'mothbeans', 'mungbean', 'greenpeas', 'horsegram(kulthi)',
'blackgram', 'rapeseed(mohri)', 'corianderseeds', 'mustardseeds',
'sesameseed', 'cuminseeds', 'lentil', 'brinjal', 'beetroot',
'bittergourd', 'bottlegourd', 'capsicum', 'cabbage', 'carrot',
'cauliflower', 'cucumber', 'corianderleaves', 'curryleaves',
'drumstick-moringa', 'chili', 'ladyfinger', 'mushroom', 'onion',
'potato', 'pumpkin', 'radish', 'olive', 'sweetpotato',
'fenugreekleaf(methi)', 'spinach', 'ridgegourd',
'gooseberry(amla)', 'jambun(syzygiumcumini)',
'ziziphusmauritiana(bor)', 'garciniaindica(kokam)', 'tamarind',
'tapioca(suran)', 'garlic', 'lemon', 'tomato', 'ashgourd',
'pineapple', 'pomegranate', 'banana', 'mango', 'grapes',
'jackfruit', 'guava', 'watermelon', 'muskmelon', 'apricot',
'apple', 'chickoo', 'custardapple', 'dates', 'figs', 'orange',
'papaya', 'aniseed', 'asafoetida', 'bayleaf', 'blackpepper',
'cardamom', 'cinnamon', 'cloves', 'jaiphal(nutmeg)', 'ginger',
'turmeric', 'cashewnuts', 'raisins', 'coconut', 'almondnut',
'arecanut', 'pistachionut', 'lemongrass', 'cotton', 'jute',
'coffee', 'sunflower'], dtype=object)
```

```
for i in crop_names_from_fert:
    print(crop[crop['label'] == i])
```

```
temperature    humidity      ph    rainfall label
0      20.879744   82.002744  6.502985   202.935536  rice
```

1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice
..	...	...	...	...	...
95	22.683191	83.463583	6.604993	194.265172	rice
96	21.533463	82.140041	6.500343	295.924880	rice
97	21.408658	83.329319	5.935745	287.576693	rice
98	26.543481	84.673536	7.072656	183.622266	rice
99	23.359054	83.595123	5.333323	188.413665	rice

[100 rows x 5 columns]

Empty DataFrame

Columns: [temperature, humidity, ph, rainfall, label]

Index: []

Empty DataFrame

Columns: [temperature, humidity, ph, rainfall, label]

Index: []

	temperature	humidity	ph	rainfall	label
500	22.613600	63.690706	5.749914	87.759539	maize
501	26.100184	71.574769	6.931757	102.266244	maize
502	23.558821	71.593514	6.657965	66.719955	maize
503	19.972160	57.682729	6.596061	60.651715	maize
504	18.478913	62.695039	5.970458	65.438354	maize
..	...	...	...	...	...
595	18.928519	72.800861	6.158860	82.341629	maize
596	23.305468	63.246480	6.385684	108.760300	maize
597	18.748267	62.498785	6.417820	70.234016	maize
598	19.742133	59.662631	6.381202	65.508614	maize
599	25.730444	70.747393	6.877869	98.737713	maize

[100 rows x 5 columns]

Empty DataFrame

Columns: [temperature, humidity, ph, rainfall, label]

Index: []

	temperature	humidity	ph	rainfall	label
2100	17.024985	16.988612	7.485996	88.551231	chickpea
2101	19.020613	17.131591	6.920251	79.926981	chickpea
2102	17.887765	15.405897	5.996932	68.549329	chickpea
2103	18.868056	15.658092	6.391174	88.510490	chickpea
2104	18.369526	19.563810	7.152811	79.263577	chickpea
...	...	...	...	...	...
2195	17.341502	18.756263	8.861480	67.954543	chickpea
2196	17.437327	14.338474	7.861128	73.092670	chickpea
2197	18.897802	19.761829	7.452671	69.095125	chickpea
2198	18.591908	14.779596	7.168096	89.609825	chickpea
2199	18.315615	15.361435	7.263119	81.787105	chickpea

[100 rows x 5 columns]

Empty DataFrame

Columns: [temperature, humidity, ph, rainfall, label]

Index: []

Empty DataFrame

Columns: [temperature, humidity, ph, rainfall, label]

Index: []

Empty DataFrame

```
crop['label']
```

0	rice
1	rice
2	rice
3	rice
4	rice
...	
3095	watermelon
3096	watermelon
3097	watermelon
3098	watermelon
3099	watermelon

Name: label, Length: 3100, dtype: object

```
extract_labels = []
for i in crop_names_from_fert:
    if i in crop_names:
        extract_labels.append(i)
```

```
# using extract labels on crop to get all the data related to those labels
new_crop = pd.DataFrame(columns = crop.columns)
new_fert = pd.DataFrame(columns = fert.columns)
```

```
for label in extract_labels:
    new_crop = new_crop.append(crop[crop['label'] == label])
```

```
for label in extract_labels:
    new_fert = new_fert.append(fert[fert['Crop'] == label].iloc[0])
```

```
new_crop
```

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice

new\_fert

	Crop	N	P	K	pH
0	rice	80	40	40	5.5
3	maize	80	40	20	5.5
5	chickpea	40	60	80	5.5
12	kidneybeans	20	60	20	5.5
13	pigeonpeas	20	60	20	5.5
14	mothbeans	20	40	20	5.5
15	mungbean	20	40	20	5.5
18	blackgram	40	60	20	5.0
24	lentil	20	60	20	5.5
60	pomegranate	20	10	40	5.5
61	banana	100	75	50	6.5
62	mango	20	20	30	5.0
63	grapes	20	125	200	4.0
66	watermelon	100	10	50	5.5
67	muskmelon	100	10	50	5.5
69	apple	20	125	200	6.5
74	orange	20	10	10	4.0
75	papaya	50	50	50	6.0
88	coconut	20	10	30	5.0
93	cotton	120	40	20	5.5
94	jute	80	40	40	5.5
95	coffee	100	20	30	5.5

```
new_crop.to_csv('MergeFileCrop.csv')
new_fert.to_csv('MergeFileFert.csv')
```

[Colab paid products](#) - [Cancel contracts here](#)

