

Dataset

March 8, 2023

```
[1]: import pandas as pd

#grafting/ acreage table
graft_df = pd.read_csv("graft.csv")

graft_df
```

```
[1]:
```

	ranch_no	ranch_sub	var_cd	variety_desc	grafted	acres	\
0	3	NaN	300	ALLISON - SHEEGENE 20	True	72.8	
1	14	NaN	300	ALLISON - SHEEGENE 20	True	7.5	
2	14	NaN	300	ALLISON - SHEEGENE 20	True	36.8	
3	16	N	300	ALLISON - SHEEGENE 20	True	21.0	
4	17	N	300	ALLISON - SHEEGENE 20	True	76.0	
..	
156	20	NN	314	SWEET GLOBE	True	34.1	
157	37	E	94	SWEET SCARLET	True	51.6	
158	37	M	94	SWEET SCARLET	True	47.8	
159	37	W	94	SWEET SCARLET	True	50.4	
160	29	SW	2	THOMPSON SEEDLESS	False	134.8	

```
year_planted_grafted
0          2019
1          2017
2          2020
3          2019
4          2017
..          ...
156         2020
157         2017
158         2017
159         2017
160         1990
```

[161 rows x 7 columns]

```
[2]: #production table
prod_df = pd.read_csv("production.csv")
```

```
prod_df
```

```
[2]:      prod_wk  ranch_no ranch_sub  var_cd      variety_desc  var_boxes  \
0      7/13/2015         1      NaN    178  ARRA 15 - SWEETIES      3905
1      7/20/2015         1      NaN    178  ARRA 15 - SWEETIES      1075
2      7/27/2015         1      NaN    178  ARRA 15 - SWEETIES      2047
3       8/3/2015         1      NaN    178  ARRA 15 - SWEETIES      6121
4      8/10/2015         1      NaN    178  ARRA 15 - SWEETIES       724
...
4123  7/18/2022         50      NaN    307  IVORY - SHEEGENE 21      9836
4124  7/25/2022         50      NaN    307  IVORY - SHEEGENE 21     27029
4125   8/1/2022         50      NaN    307  IVORY - SHEEGENE 21    24363
4126   8/8/2022         50      NaN    307  IVORY - SHEEGENE 21    11359
4127  8/15/2022         50      NaN    307  IVORY - SHEEGENE 21     2189
```

```
      color  prod_date_cnt  prod_wk_no  prod_yr  min_date  max_date
0    GREEN              4           29    2015  7/15/2015  7/18/2015
1    GREEN              1           30    2015  7/21/2015  7/21/2015
2    GREEN              2           31    2015  7/31/2015  8/1/2015
3    GREEN              5           32    2015   8/3/2015  8/7/2015
4    GREEN              1           33    2015  8/15/2015  8/15/2015
...
4123  GREEN              3           29    2022  7/20/2022  7/22/2022
4124  GREEN              5           30    2022  7/25/2022  7/29/2022
4125  GREEN              5           31    2022   8/1/2022  8/5/2022
4126  GREEN              5           32    2022   8/8/2022  8/12/2022
4127  GREEN              3           33    2022  8/15/2022  8/17/2022
```

```
[4128 rows x 12 columns]
```

```
[3]: #merge graft and production datasets
df = prod_df.merge(graft_df)
```

```
df
```

```
[3]:      prod_wk  ranch_no ranch_sub  var_cd      variety_desc  var_boxes  \
0      7/13/2015         1      NaN    178  ARRA 15 - SWEETIES      3905
1      7/20/2015         1      NaN    178  ARRA 15 - SWEETIES      1075
2      7/27/2015         1      NaN    178  ARRA 15 - SWEETIES      2047
3       8/3/2015         1      NaN    178  ARRA 15 - SWEETIES      6121
4      8/10/2015         1      NaN    178  ARRA 15 - SWEETIES       724
...
4266  8/15/2022         50      NaN    307  IVORY - SHEEGENE 21      2189
4267  8/15/2022         50      NaN    307  IVORY - SHEEGENE 21      2189
4268  8/15/2022         50      NaN    307  IVORY - SHEEGENE 21      2189
4269  8/15/2022         50      NaN    307  IVORY - SHEEGENE 21      2189
```

4270 8/15/2022 50 NaN 307 IVORY - SHEEGENE 21 2189

	color	prod_date_cnt	prod_wk_no	prod_yr	min_date	max_date	\
0	GREEN	4	29	2015	7/15/2015	7/18/2015	
1	GREEN	1	30	2015	7/21/2015	7/21/2015	
2	GREEN	2	31	2015	7/31/2015	8/1/2015	
3	GREEN	5	32	2015	8/3/2015	8/7/2015	
4	GREEN	1	33	2015	8/15/2015	8/15/2015	
...	
4266	GREEN	3	33	2022	8/15/2022	8/17/2022	
4267	GREEN	3	33	2022	8/15/2022	8/17/2022	
4268	GREEN	3	33	2022	8/15/2022	8/17/2022	
4269	GREEN	3	33	2022	8/15/2022	8/17/2022	
4270	GREEN	3	33	2022	8/15/2022	8/17/2022	

	grafted	acres	year_planted_grafted
0	True	36.00	2009
1	True	36.00	2009
2	True	36.00	2009
3	True	36.00	2009
4	True	36.00	2009
...
4266	False	27.70	2018
4267	False	24.50	2019
4268	True	18.92	2019
4269	True	114.36	2020
4270	True	14.58	2021

[4271 rows x 15 columns]

```
[4]: #Make prod_wk datetime
df['prod_wk'] = pd.to_datetime(df['prod_wk'])

#Remove unneeded columns
df = df.drop(['prod_date_cnt', 'prod_wk_no', 'min_date', 'max_date' ], axis=1)

boxes = df['var_boxes']
acres = df['acres']

#create a new column for production yield (boxes per acre)
df['yield'] = boxes/acres

#create a new column for age of the variety
prod_yr = df['prod_yr']
graft_year = df['year_planted_grafted']
df['age'] = prod_yr - graft_year
```

```
#create a new column for yield classification:
# High(2) : yield > 170 | Medium(1) : 170 > yield > 55 | Low(0) : yield < 55
df['yield_class'] = df['yield'].apply(lambda x: 'high' if x > 175 else
    ↪('medium' if x >= 55 else 'low'))

df
```

```
[4]:
```

	prod_wk	ranch_no	ranch_sub	var_cd	variety_desc	var_boxes	\
0	2015-07-13	1	NaN	178	ARRA 15 - SWEETIES	3905	
1	2015-07-20	1	NaN	178	ARRA 15 - SWEETIES	1075	
2	2015-07-27	1	NaN	178	ARRA 15 - SWEETIES	2047	
3	2015-08-03	1	NaN	178	ARRA 15 - SWEETIES	6121	
4	2015-08-10	1	NaN	178	ARRA 15 - SWEETIES	724	
...	
4266	2022-08-15	50	NaN	307	IVORY - SHEEGENE 21	2189	
4267	2022-08-15	50	NaN	307	IVORY - SHEEGENE 21	2189	
4268	2022-08-15	50	NaN	307	IVORY - SHEEGENE 21	2189	
4269	2022-08-15	50	NaN	307	IVORY - SHEEGENE 21	2189	
4270	2022-08-15	50	NaN	307	IVORY - SHEEGENE 21	2189	

	color	prod_yr	grafted	acres	year_planted_grafted	yield	age	\
0	GREEN	2015	True	36.00	2009	108.472222	6	
1	GREEN	2015	True	36.00	2009	29.861111	6	
2	GREEN	2015	True	36.00	2009	56.861111	6	
3	GREEN	2015	True	36.00	2009	170.027778	6	
4	GREEN	2015	True	36.00	2009	20.111111	6	
...	
4266	GREEN	2022	False	27.70	2018	79.025271	4	
4267	GREEN	2022	False	24.50	2019	89.346939	3	
4268	GREEN	2022	True	18.92	2019	115.697674	3	
4269	GREEN	2022	True	114.36	2020	19.141308	2	
4270	GREEN	2022	True	14.58	2021	150.137174	1	

	yield_class
0	medium
1	low
2	medium
3	medium
4	low
...	...
4266	medium
4267	medium
4268	medium
4269	low
4270	medium

[4271 rows x 14 columns]

```
[5]: df = df.rename(columns={'prod_wk': 'Date'})

#import weather and precipitation dataset
temps_df = pd.read_csv("temps.csv")

#drop unneeded columns
temps_df = temps_df.drop(['Unnamed: 4', 'Unnamed: 5'], axis=1)

# convert the 'Date' column to datetime format
temps_df['Date'] = pd.to_datetime(temps_df['Date'])

# set the 'Date' column as the index
temps_df.set_index('Date', inplace=True)

# resample the data by week and calculate the mean of 'Daily_High'
weekly_avg = temps_df['Daily_High'].resample('W').mean()

#Make weekly_avg into a dataframe
weekly_avg = pd.DataFrame(weekly_avg)

#Add weekly_low_avg to weekly_avg
weekly_avg['weekly_low_avg'] = temps_df['Daily_Low'].resample('W').mean()
#Add weekly_precipitation_avg to weekly_avg
weekly_avg['weekly_precipitation_avg'] = temps_df['Precipitation '].
    ↪resample('W').mean()
#Rename 'Daily_High' to 'weekly_high_avg'
weekly_avg = weekly_avg.rename(columns={'Daily_High': 'weekly_high_avg'})

temps_df.reset_index(inplace=True)
temps_df.head()

merged_df = pd.merge(df, temps_df, on = 'Date', how='inner')
merged_df
```

```
[5]:
```

	Date	ranch_no	ranch_sub	var_cd	variety_desc \
0	2015-07-13	1	NaN	178	ARRA 15 - SWEETIES
1	2015-07-13	14	NaN	16	ARRA 16
2	2015-07-13	14	NaN	16	ARRA 16
3	2015-07-13	29	SW	71	SUGRAONE
4	2015-07-13	29	NW	71	SUGRAONE
...
4262	2018-11-12	32	NE	96	AUTUMN KING
4263	2018-11-12	32	NW	755	ARRA 28 - ARRA PASSION PUNCH
4264	2019-07-01	33	W	90	EARLY SWEET
4265	2019-07-01	33	W	90	EARLY SWEET

4266 2019-07-01 34 E 13 FLAME SEEDLESS

	var_boxes	color	prod_yr	grafted	acres	year_planted_grafted \
0	3905	GREEN	2015	True	36.00	2009
1	3038	GREEN	2015	True	8.00	2017
2	3038	GREEN	2015	True	10.00	2018
3	6389	GREEN	2015	True	18.00	2006
4	23119	GREEN	2015	True	70.00	2006
...
4262	23204	GREEN	2018	False	282.80	2012
4263	667	RED	2018	True	69.35	2017
4264	4155	GREEN	2019	False	63.00	2009
4265	4155	GREEN	2019	True	35.00	2017
4266	632	RED	2019	True	69.20	1977

	yield	age	yield_class	Daily_High	Daily_Low	Precipitation
0	108.472222	6	medium	94	68	0.0
1	379.750000	-2	high	94	68	0.0
2	303.800000	-3	high	94	68	0.0
3	354.944444	9	high	94	68	0.0
4	330.271429	9	high	94	68	0.0
...
4262	82.050919	6	medium	74	39	0.0
4263	9.617880	1	low	74	39	0.0
4264	65.952381	10	medium	93	65	0.0
4265	118.714286	2	medium	93	65	0.0
4266	9.132948	42	low	93	65	0.0

[4267 rows x 17 columns]