ex3

February 18, 2023

[1]:
```
import pandas as pd

tomato = pd.read_csv('tomato-yields.csv')

tomato.head()
```

[1]:

	Entity	Code	Year	Tomatoes | 00000388 || Yield | 005419 || tonnes per hectare
0	Africa	NaN	1961	12.320172
1	Africa	NaN	1962	12.976988
2	Africa	NaN	1963	12.867894
3	Africa	NaN	1964	13.189582
4	Africa	NaN	1965	13.492712

[2]:
```
tomato.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11282 entries, 0 to 11281
Data columns (total 4 columns):
 #   Column                                                  Non-Null Count
Dtype
---  ------                                                  --------------
-----
 0   Entity                                                  11282 non-null
object
 1   Code                                                    8793 non-null
object
 2   Year                                                    11282 non-null
int64
 3   Tomatoes | 00000388 || Yield | 005419 || tonnes per hectare  11282 non-null
float64
dtypes: float64(1), int64(1), object(2)
```

```
memory usage: 352.7+ KB
```

[3]:
```
tomato.isna().sum()
```

[3]:
```
Entity                                                      0
Code                                                     2489
Year                                                        0
Tomatoes | 00000388 || Yield | 005419 || tonnes per hectare  0
dtype: int64
```

[4]:
```
'''
    Obtain a line chart of yeild of United States, Spain, China, India during␣
 ↪1961-2020
'''

tomato = tomato.rename(columns={"Tomatoes | 00000388 || Yield | 005419 ||␣
 ↪tonnes per hectare" : "Yields"})
```

[5]:
```
China = tomato.loc[tomato['Entity'] == "China"]
US = tomato.loc[tomato['Entity'] == "United States"]
Spain = tomato.loc[tomato['Entity'] == "Spain"]
India = tomato.loc[tomato['Entity'] == "India"]
```

[6]:
```
country = ['United States', 'China', 'India', 'Spain']
mask = tomato['Entity'].isin(country)
```

[7]:
```
import plotly.express as px

fig = px.line(tomato[mask], x = "Year", y = "Yields", color = 'Entity')
fig.show()
```