

PANDAS

WHAT DO THE FIRST FEW ROWS OF MY DATAFRAME LOOK LIKE?

```
food_df = pd.read_csv("food_usa.csv")  
food_df.head()
```

ITEM	VALUE	UNIT
Almonds, in shell	2000340.0	t
Apples	4922840.0	t
Apricots	31120.0	t
Artichokes	30481.0	t
Asparagus	30481.0	t

WHAT DO THE LAST FEW ROWS OF MY DATAFRAME LOOK LIKE?

```
food_df = pd.read_csv("food_usa.csv")  
food_df.tail()
```

ITEM	VALUE	UNIT
Tung nuts	NaN	t
Unmanufactured tobacco	147520.0	t
Walnuts, in shell	547030.0	t
Watermelons	1677612.0	t
Wheat	53650020.0	t

HOW MANY ROWS (DATA POINTS) DO I HAVE IN MY DATAFRAME?

```
food_df = pd.read_csv("food_usa.csv")  
len(food_df)
```

```
126
```

WHAT COLUMNS DO I HAVE IN MY DATAFRAME?

```
food_df = pd.read_csv("food_usa.csv")  
food_df.columns
```

```
Index(['Domain Code', 'Domain', 'Area Code (M49)',  
      'Area', 'Element Code', 'Element',  
      'Item Code (CPC)', 'Item', 'Year Code',  
      'Year', 'Unit', 'Value', 'Flag',  
      'Flag Description', 'Note'],  
      dtype='object')
```

HOW MANY COLUMNS DO I HAVE IN MY DATAFRAME?

```
food_df = pd.read_csv("food_usa.csv")  
len(food_df.columns)
```

15

HOW MANY ROWS & COLUMNS DO I HAVE IN MY DATAFRAME?

```
food_df = pd.read_csv("food_usa.csv")  
food_df.shape
```

```
(126, 15)
```

WHAT ARE THE DIFFERENT VALUES PRESENT IN A SINGLE COLUMN?

```
food_df = pd.read_csv("food_usa.csv")  
food_df["Flag Description"].value_counts()
```

```
Flag Description  
Official figure          86  
Value imputed by a receiving agency  23  
Estimated value          9  
Missing value; data cannot exist      8  
Name: count, dtype: int64
```

WHERE ARE THERE VALUES MISSING FROM MY DATAFRAME?

"Give me a view of my DataFrame that replaces the value in each cell with **True** if the value is missing; **False** otherwise."

```
food_df = pd.read_csv("food_usa.csv")
food_df.isnull()
```

ITEM	...	VALUE	NOTE
Almonds, in shell	...	False	True
Apples	...	False	True
...	...	...	...
Tung nuts	...	True	True
Wheat	...	False	True

I WANT TO DELETE ROWS WITH MISSING INFORMATION IN MY DATAFRAME

"If a row is missing a number in the **Value** column, I want to exclude it from my analysis."

```
food_df_no_missing = food_df.dropna(subset=["Value"])  
food_df_no_missing["Flag Description"].value_counts()
```

```
Flag Description  
Official figure          86  
Value imputed by a receiving agency  23  
Estimated value          9  
Name: count, dtype: int64
```

HOW CAN I COMPUTE THE SUM OF A COLUMN?

"How many total tons of foodstuff did the US produce in the year 2024?"

```
food_df_no_missing["Value"].sum()
```

```
1001865585.81
```

WHICH ROWS HAVE THE HIGHEST VALUES FOR A CERTAIN COLUMN?

"Which items did the USA produce in the highest quantities?"

```
food_df_no_missing.nlargest(5, "Value")
```

ITEM	VALUE	UNIT
Maize (corn)	377632690.0	t
Soya beans	118836440.0	t
Hen eggs in shell, fresh	108514900.0	1000 No
Raw milk of cattle	102452123.0	t
Wheat	53650020.0	t

HOW CAN I LOOK AT JUST THE ROWS THAT MEET A CERTAIN CONDITION?

"What are the products that are measured by weight in tons?"

```
food_df_tons = food_df_no_missing[  
    food_df_no_missing["Unit"] == "t"  
]  
food_df_tons.shape
```

```
(117, 15)
```

FILTERING WITH MULTIPLE CONDITIONS

"What are some of the products of which we produce less than 5000 tons?"

```
food_df_tons_small = food_df_tons[
    food_df_tons["Value"] < 5000
]
food_df_tons_small.head()
```

ITEM	VALUE
Bananas	3560.59
Beeswax	1586.43
Coffee, green	3817.00
Currants	0.00
Edible offal of goat...	1943.65

STRING SEARCHING IN COLUMNS

"What does US production of meat look like?"

```
just_meat_df = food_df_no_missing[
    food_df_no_missing["Item"].str.contains(
        "meat", case=False
    )
]
just_meat_df
```

ITEM	VALUE
Game meat, fresh...	271430.20
Horse meat, fresh...	0.00
Meat of cattle with bone...	12289768.00
Meat of chickens...	20176637.00
...	...