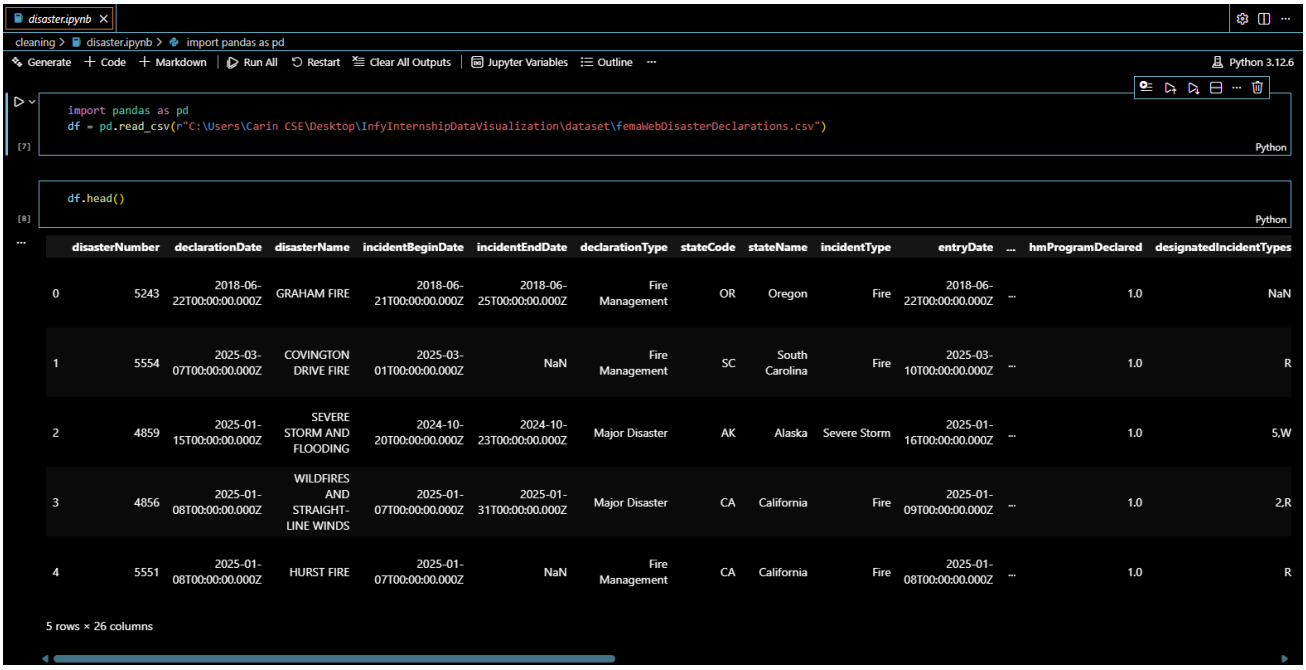
VISUALIZING US NATURAL DISASTER DECLARATION – TRENDS AND PATTERNS

Week 2 Documentation/ Screenshots

Data Cleaning in Python (Pandas)

Data Cleaning in Python in VS Code



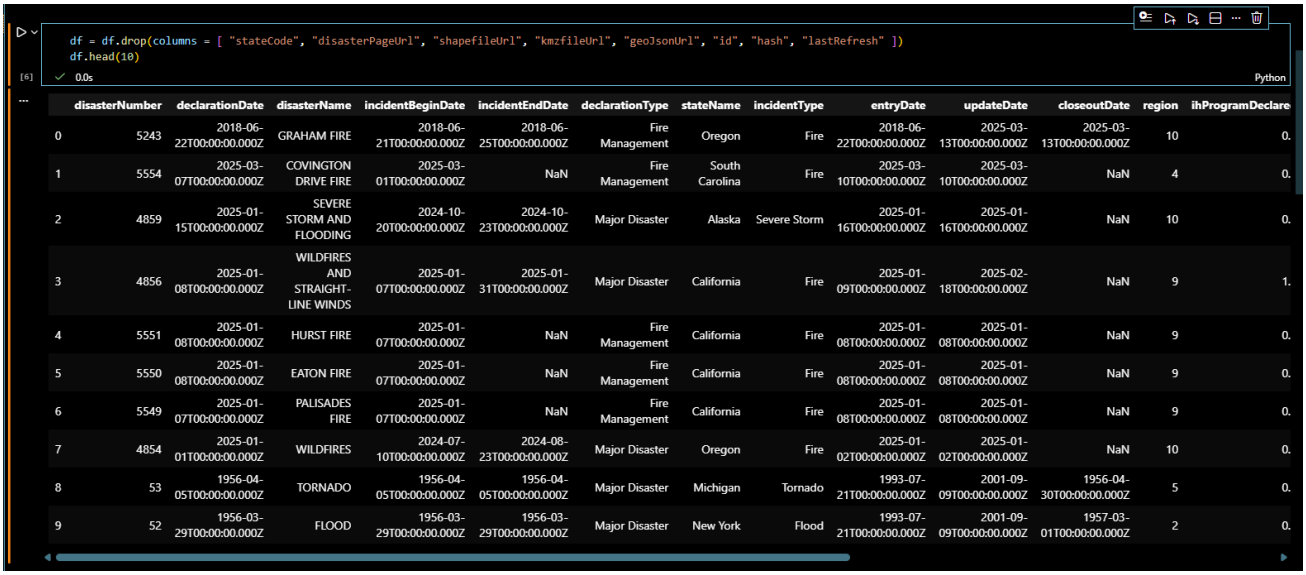
```
import pandas as pd
df = pd.read_csv(r"C:\Users\Carin CSE\Desktop\InfyInternshipDataVisualization\dataset\femaWebDisasterDeclarations.csv")
```

```
df.head()
```

	disasterNumber	declarationDate	disasterName	incidentBeginDate	incidentEndDate	declarationType	stateCode	stateName	incidentType	entryDate	...	hmProgramDeclared	designatedIncidentTypes
0	5243	2018-06-22T00:00:00.000Z	GRAHAM FIRE	2018-06-21T00:00:00.000Z	2018-06-25T00:00:00.000Z	Fire Management	OR	Oregon	Fire	2018-06-22T00:00:00.000Z	...	1.0	NaN
1	5554	2025-03-07T00:00:00.000Z	COVINGTON DRIVE FIRE	2025-03-01T00:00:00.000Z	NaN	Fire Management	SC	South Carolina	Fire	2025-03-10T00:00:00.000Z	...	1.0	R
2	4859	2025-01-15T00:00:00.000Z	SEVERE STORM AND FLOODING	2024-10-20T00:00:00.000Z	2024-10-23T00:00:00.000Z	Major Disaster	AK	Alaska	Severe Storm	2025-01-16T00:00:00.000Z	...	1.0	5,W
3	4856	2025-01-08T00:00:00.000Z	WILDFIRES AND STRAIGHT-LINE WINDS	2025-01-07T00:00:00.000Z	2025-01-31T00:00:00.000Z	Major Disaster	CA	California	Fire	2025-01-09T00:00:00.000Z	...	1.0	2,R
4	5551	2025-01-08T00:00:00.000Z	HURST FIRE	2025-01-07T00:00:00.000Z	NaN	Fire Management	CA	California	Fire	2025-01-08T00:00:00.000Z	...	1.0	R

5 rows × 26 columns

Fig 1.1 Importing .csv file



```
df = df.drop(columns = [ "stateCode", "disasterPageUrl", "shapefileUrl", "kmzfileUrl", "geoJsonUrl", "id", "hash", "lastRefresh" ])
df.head(10)
```

0.0s

	disasterNumber	declarationDate	disasterName	incidentBeginDate	incidentEndDate	declarationType	stateName	incidentType	entryDate	updateDate	closeoutDate	region	ihProgramDeclare
0	5243	2018-06-22T00:00:00.000Z	GRAHAM FIRE	2018-06-21T00:00:00.000Z	2018-06-25T00:00:00.000Z	Fire Management	Oregon	Fire	2018-06-22T00:00:00.000Z	2025-03-13T00:00:00.000Z	2025-03-13T00:00:00.000Z	10	0.
1	5554	2025-03-07T00:00:00.000Z	COVINGTON DRIVE FIRE	2025-03-01T00:00:00.000Z	NaN	Fire Management	South Carolina	Fire	2025-03-10T00:00:00.000Z	2025-03-10T00:00:00.000Z	NaN	4	0.
2	4859	2025-01-15T00:00:00.000Z	SEVERE STORM AND FLOODING	2024-10-20T00:00:00.000Z	2024-10-23T00:00:00.000Z	Major Disaster	Alaska	Severe Storm	2025-01-16T00:00:00.000Z	2025-01-16T00:00:00.000Z	NaN	10	0.
3	4856	2025-01-08T00:00:00.000Z	WILDFIRES AND STRAIGHT-LINE WINDS	2025-01-07T00:00:00.000Z	2025-01-31T00:00:00.000Z	Major Disaster	California	Fire	2025-01-09T00:00:00.000Z	2025-02-18T00:00:00.000Z	NaN	9	1.
4	5551	2025-01-08T00:00:00.000Z	HURST FIRE	2025-01-07T00:00:00.000Z	NaN	Fire Management	California	Fire	2025-01-08T00:00:00.000Z	2025-01-08T00:00:00.000Z	NaN	9	0.
5	5550	2025-01-08T00:00:00.000Z	EATON FIRE	2025-01-07T00:00:00.000Z	NaN	Fire Management	California	Fire	2025-01-08T00:00:00.000Z	2025-01-08T00:00:00.000Z	NaN	9	0.
6	5549	2025-01-07T00:00:00.000Z	PALISADES FIRE	2025-01-07T00:00:00.000Z	NaN	Fire Management	California	Fire	2025-01-08T00:00:00.000Z	2025-01-08T00:00:00.000Z	NaN	9	0.
7	4854	2025-01-01T00:00:00.000Z	WILDFIRES	2024-07-10T00:00:00.000Z	2024-08-23T00:00:00.000Z	Major Disaster	Oregon	Fire	2025-01-02T00:00:00.000Z	2025-01-02T00:00:00.000Z	NaN	10	0.
8	53	1956-04-05T00:00:00.000Z	TORNADO	1956-04-05T00:00:00.000Z	1956-04-05T00:00:00.000Z	Major Disaster	Michigan	Tornado	1993-07-21T00:00:00.000Z	2001-09-09T00:00:00.000Z	1956-04-30T00:00:00.000Z	5	0.
9	52	1956-03-29T00:00:00.000Z	FLOOD	1956-03-29T00:00:00.000Z	1956-03-29T00:00:00.000Z	Major Disaster	New York	Flood	1993-07-21T00:00:00.000Z	2001-09-09T00:00:00.000Z	1957-03-01T00:00:00.000Z	2	0.

Fig 1.2 Drop unnecessary columns

```
# check datatype
df["disasterNumber"] = df["disasterNumber"].astype(int)
date_cols = ["declarationDate","incidentBeginDate","incidentEndDate", "entryDate","closeoutDate","updateDate"]
for col in date_cols:
    df[col] = pd.to_datetime(df[col], errors="coerce")
```

[11] ✓ 0.1s Python

```
flag_cols = ["iaProgramDeclared","ihProgramDeclared","paProgramDeclared","hmProgramDeclared"]
for col in flag_cols:
    df[col] = df[col].astype(bool)
df.head()
```

[12] ✓ 0.0s Python

stateName	incidentType	entryDate	updateDate	closeoutDate	region	ihProgramDeclared	iaProgramDeclared	paProgramDeclared	hmProgramDeclared	designatedIncidentTypes	declarationRequestDate
Oregon	Fire	2018-06-22 00:00:00+00:00	2025-03-13 00:00:00+00:00	2025-03-13 00:00:00+00:00	10	False	False	True	True	NaN	2018-06-21T00:00:00.000Z
South Carolina	Fire	2025-03-10 00:00:00+00:00	2025-03-10 00:00:00+00:00	NaT	4	False	False	True	True	R	2025-03-07T00:00:00.000Z
Alaska	Severe Storm	2025-01-16 00:00:00+00:00	2025-01-16 00:00:00+00:00	NaT	10	False	False	True	True	5,W	2024-12-16T00:00:00.000Z
California	Fire	2025-01-09 00:00:00+00:00	2025-02-18 00:00:00+00:00	NaT	9	True	False	True	True	2,R	2025-01-08T00:00:00.000Z
California	Fire	2025-01-08 00:00:00+00:00	2025-01-08 00:00:00+00:00	NaT	9	False	False	True	True	R	2025-01-08T00:00:00.000Z

Fig 1.3 Check the datatype of every column and convert flag columns to boolean

```
# from closeOutDate
df["status"] = df["closeoutDate"]. apply( lambda x: "Closed" if pd.notnull(x) else "Open" )
df.head()
```

[13] ✓ 0.0s Python

ne	incidentType	entryDate	updateDate	closeoutDate	region	ihProgramDeclared	iaProgramDeclared	paProgramDeclared	hmProgramDeclared	designatedIncidentTypes	declarationRequestDate	status
on	Fire	2018-06-22 00:00:00+00:00	2025-03-13 00:00:00+00:00	2025-03-13 00:00:00+00:00	10	False	False	True	True	NaN	2018-06-21T00:00:00.000Z	Closed
uth ina	Fire	2025-03-10 00:00:00+00:00	2025-03-10 00:00:00+00:00	NaT	4	False	False	True	True	R	2025-03-07T00:00:00.000Z	Open
aka	Severe Storm	2025-01-16 00:00:00+00:00	2025-01-16 00:00:00+00:00	NaT	10	False	False	True	True	5,W	2024-12-16T00:00:00.000Z	Open
nia	Fire	2025-01-09 00:00:00+00:00	2025-02-18 00:00:00+00:00	NaT	9	True	False	True	True	2,R	2025-01-08T00:00:00.000Z	Open
nia	Fire	2025-01-08 00:00:00+00:00	2025-01-08 00:00:00+00:00	NaT	9	False	False	True	True	R	2025-01-08T00:00:00.000Z	Open

Fig 1.4 Create a derived column ‘status’ from closeOutDate column

```
# fiscal year
df["fyDeclared"] = df["declarationDate"]. apply( lambda x: x.year+1 if x.month > 9 else x.year )
df.head()
```

[14] Python

Type	entryDate	updateDate	closeoutDate	region	ihProgramDeclared	iaProgramDeclared	paProgramDeclared	hmProgramDeclared	designatedIncidentTypes	declarationRequestDate	status	fyDeclared
Fire	2018-06-22 00:00:00+00:00	2025-03-13 00:00:00+00:00	2025-03-13 00:00:00+00:00	10	0.0	0.0	1.0	1.0	NaN	2018-06-21T00:00:00.000Z	Closed	2018
Fire	2025-03-10 00:00:00+00:00	2025-03-10 00:00:00+00:00	NaT	4	0.0	0.0	1.0	1.0	R	2025-03-07T00:00:00.000Z	Open	2025
torm	2025-01-16 00:00:00+00:00	2025-01-16 00:00:00+00:00	NaT	10	0.0	0.0	1.0	1.0	5,W	2024-12-16T00:00:00.000Z	Open	2025
Fire	2025-01-09 00:00:00+00:00	2025-02-18 00:00:00+00:00	NaT	9	1.0	0.0	1.0	1.0	2,R	2025-01-08T00:00:00.000Z	Open	2025
Fire	2025-01-08 00:00:00+00:00	2025-01-08 00:00:00+00:00	NaT	9	0.0	0.0	1.0	1.0	R	2025-01-08T00:00:00.000Z	Open	2025

Fig 1.5 Create a derived column ‘fyDeclared’ from declarationDate column

```
df["incidentDuration"] = ( (df["incidentEndDate"] - df["incidentBeginDate"]).dt.days )
df.head()
```

[17] Python

	entryDate	updateDate	...	region	ihProgramDeclared	iaProgramDeclared	paProgramDeclared	hmProgramDeclared	designatedIncidentTypes	declarationRequestDate	status	fyDeclared	incidentDuration
	2018-06-22 00:00:00+00:00	2025-03-13 00:00:00+00:00	...	10	False	False	True	True	NaN	2018-06-21T00:00:00.000Z	Closed	2018	4.0
	2025-03-10 00:00:00+00:00	2025-03-10 00:00:00+00:00	...	4	False	False	True	True	R	2025-03-07T00:00:00.000Z	Open	2025	NaN
	2025-01-16 00:00:00+00:00	2025-01-16 00:00:00+00:00	...	10	False	False	True	True	5,W	2024-12-16T00:00:00.000Z	Open	2025	3.0
	2025-01-09 00:00:00+00:00	2025-02-18 00:00:00+00:00	...	9	True	False	True	True	2,R	2025-01-08T00:00:00.000Z	Open	2025	24.0
	2025-01-08 00:00:00+00:00	2025-01-08 00:00:00+00:00	...	9	False	False	True	True	R	2025-01-08T00:00:00.000Z	Open	2025	NaN

Fig 1.6 Create a derived column ‘incidentDuration’

```
df = df.drop_duplicates()
```

[16] ✓ 0.0s Python

```
df.to_csv("femaDisasterCleaned.csv", index=False)
```

[17] ✓ 0.1s Python

Fig 1.7 Save the cleaned .csv file