Observing our data (Types of Attributes)

We will use 5 dataset in the assignment, assignment will be developed for each data set. For each one of the data sets, we describe each one of the fields. We show a screenshot of the data. We show information related to the dataset such as number of records, number if entries. We can see what fields are null.

1. **Covid_19_india dataset**

These are the columns of the dataset and its description:
Sno - Serial Number
Date - Date
Time - Time
State/UnionTerritory - State
ConfirmedIndianNational - Indian National or not
ConfirmedForeignNational - Foreign National or not
Cured - Number of Patients Cured
Deaths - Number of patients expired
Confirmed - Number of confirmed Covid case

Screenshot of the data set:

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
0	1	2020-01-30	6:00 PM	Kerala	1	0	0	0	1
1	2	2020-01-31	6:00 PM	Kerala	1	0	0	0	1
2	3	2020-02-01	6:00 PM	Kerala	2	0	0	0	2
3	4	2020-02-02	6:00 PM	Kerala	3	0	0	0	3
4	5	2020-02-03	6:00 PM	Kerala	3	0	0	0	3
5	6	2020-02-04	6:00 PM	Kerala	3	0	0	0	3
6	7	2020-02-05	6:00 PM	Kerala	3	0	0	0	3
7	8	2020-02-06	6:00 PM	Kerala	3	0	0	0	3
8	9	2020-02-07	6:00 PM	Kerala	3	0	0	0	3
9	10	2020-02-08	6:00 PM	Kerala	3	0	0	0	3
10	11	2020-02-09	6:00 PM	Kerala	3	0	0	0	3

```
df_data = pd.read_csv('covid_19_india.csv', encoding='latin1')
df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15554 entries, 0 to 15553
Data columns (total 9 columns):
 #    Column                      Non-Null Count   Dtype
---   ------                      --------------   -----
 0    ï»¿Sno                      15554 non-null   int64
 1    Date                        15554 non-null   object
 2    Time                        15554 non-null   object
 3    State/UnionTerritory        15554 non-null   object
 4    ConfirmedIndianNational     15554 non-null   object
 5    ConfirmedForeignNational    15554 non-null   object
 6    Cured                       15554 non-null   int64
 7    Deaths                      15554 non-null   int64
 8    Confirmed                   15554 non-null   int64
dtypes: int64(4), object(5)
memory usage: 1.1+ MB
```

2. Covid_vaccine_statewise dataset

These are the columns and its description:

Updated On – report update date

State – State in India

Total Individuals Vaccinated - Total Individuals Vaccinated patients

Total Sessions Conducted - Total Sessions Conducted

Total Sites – Total Sites in India

First Dose Administered - First Dose Administered or not

Second Dose Administered - Second Dose Administered or not

Male(Individuals Vaccinated) - Male count Vaccinated

Female(Individuals Vaccinated) - Female count Vaccinated

Transgender(Individuals Vaccinated) - Transgender Vaccinated

Total Covaxin Administered - Total Covaxin Administered

Total CoviShield Administered - Total CoviShield Administered

Total Sputnik V Administered - Total Sputnik V Administered

AEFI - Adverse Events Following Immunization

18-45 years (Age) - Age group 18-45

45-60 years (Age) - Age group 45-60

60+ years (Age) - Age group 60+

Total Doses Administered - Total Doses Administered

Screenshot of the data set:

	Updated On	State	Total Individuals Vaccinated	Total Sessions Conducted	Total Sites	First Dose Administered	Second Dose Administered	Male(Individuals Vaccinated)	Female(Individuals Vaccinated)	Transgender(Individuals Vaccinated)	Total Covaxin Administered	A
0	16/01/2021	India	48276.0	3455.0	2957.0	48276.0	0.0	23757.0	24517.0	2.0	579.0	
1	17/01/2021	India	58604.0	8532.0	4954.0	58604.0	0.0	27348.0	31252.0	4.0	635.0	
2	18/01/2021	India	99449.0	13611.0	6583.0	99449.0	0.0	41361.0	58083.0	5.0	1299.0	
3	19/01/2021	India	195525.0	17855.0	7951.0	195525.0	0.0	81901.0	113613.0	11.0	3017.0	
4	20/01/2021	India	251280.0	25472.0	10504.0	251280.0	0.0	98111.0	153145.0	24.0	3946.0	
5	21/01/2021	India	365965.0	32226.0	12600.0	365965.0	0.0	132784.0	233143.0	38.0	5367.0	
6	22/01/2021	India	549381.0	36988.0	14115.0	549381.0	0.0	193899.0	355402.0	80.0	8128.0	
7	23/01/2021	India	759008.0	43076.0	15605.0	759008.0	0.0	267856.0	491049.0	103.0	11192.0	
8	24/01/2021	India	835058.0	49851.0	18111.0	835058.0	0.0	296283.0	538647.0	128.0	13156.0	
9	25/01/2021	India	1277104.0	55151.0	19682.0	1277104.0	0.0	444137.0	832766.0	201.0	18858.0	
10	26/01/2021	India	1293784.0	60821.0	21457.0	1293784.0	0.0	449119.0	844448.0	217.0	19504.0	

Total CoviShield Administered	Total Sputnik V Administered	AEFI	18-45 years (Age)	45-60 years (Age)	60+ years (Age)	Total Doses Administered
47697.0	NaN	NaN	NaN	NaN	NaN	48276.0
57959.0	NaN	NaN	NaN	NaN	NaN	58604.0
98150.0	NaN	NaN	NaN	NaN	NaN	99449.0
192508.0	NaN	NaN	NaN	NaN	NaN	195525.0
247334.0	NaN	NaN	NaN	NaN	NaN	251280.0
360598.0	NaN	NaN	NaN	NaN	NaN	365965.0
541253.0	NaN	NaN	NaN	NaN	NaN	549381.0
747816.0	NaN	NaN	NaN	NaN	NaN	759008.0
821902.0	NaN	NaN	NaN	NaN	NaN	835058.0
1258246.0	NaN	NaN	NaN	NaN	NaN	1277104.0

```
df_data = pd.read_csv('covid_vaccine_statewise.csv', encoding='latin1')
df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5144 entries, 0 to 5143
Data columns (total 18 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Updated On                            5144 non-null   object
 1   State                                 5144 non-null   object
 2   Total Individuals Vaccinated          5024 non-null   float64
 3   Total Sessions Conducted              5024 non-null   float64
 4   Total Sites                           5024 non-null   float64
 5   First Dose Administered               5024 non-null   float64
 6   Second Dose Administered              5024 non-null   float64
 7   Male(Individuals Vaccinated)          5024 non-null   float64
 8   Female(Individuals Vaccinated)        5024 non-null   float64
 9   Transgender(Individuals Vaccinated)   5024 non-null   float64
 10  Total Covaxin Administered            5024 non-null   float64
 11  Total CoviShield Administered         5024 non-null   float64
 12  Total Sputnik V Administered          400 non-null    float64
 13  AEFI                                  2841 non-null   float64
 14  18-45 years (Age)                     2840 non-null   float64
 15  45-60 years (Age)                     2841 non-null   float64
 16  60+ years (Age)                       2841 non-null   float64
 17  Total Doses Administered              5030 non-null   float64
dtypes: float64(16), object(2)
memory usage: 723.5+ KB
```

3. Database dataset (Fema database)

These are the columns and its description:
Declaration Number - One of "DR" (= major disaster), "EM" (= emergency management), or "FM" (= "fire management")
Declaration Type - Type of incident such as "Fire", "Flood", or "Hurricane". The incident type will affect the types of assistance available.
Declaration Date - Date the disaster was declared.
State - US state, district, or territory.
County
Disaster Type – Type of Disaster
Disaster Title – Disaster title
Start Date – Disaster Start Date
End Date – Disaster End Date
Close Date – Disaster Close Date
Individual Assistance Program - Individual Assistance Program declared or not

Individuals & Households Program - Individuals & Households Program declared or not
Public Assistance Program - Public Assistance Program declared or not
Hazard Mitigation Program - Hazard Mitigation Program declared or not

Screenshot of the data set:

	Declaration Number	Declaration Type	Declaration Date	State	County	Disaster Type	Disaster Title	Start Date	End Date	Close Date	Individual Assistance Program	Individuals & Households Program	Public Assistance Program	Ha: Mitiga Prog
0	DR-1	Disaster	05/02/1953	GA	NaN	Tornado	Tornado	05/02/1953	05/02/1953	06/01/1954	Yes	No	Yes	
1	DR-2	Disaster	05/15/1953	TX	NaN	Tornado	Tornado and Heavy Rainfall	05/15/1953	05/15/1953	01/01/1958	Yes	No	Yes	
2	DR-3	Disaster	05/29/1953	LA	NaN	Flood	Flood	05/29/1953	05/29/1953	02/01/1960	Yes	No	Yes	
3	DR-4	Disaster	06/02/1953	MI	NaN	Tornado	Tornado	06/02/1953	06/02/1953	02/01/1956	Yes	No	Yes	
4	DR-5	Disaster	06/06/1953	MT	NaN	Flood	Floods	06/06/1953	06/06/1953	12/01/1955	Yes	No	Yes	
5	DR-6	Disaster	06/09/1953	MI	NaN	Tornado	Tornado	06/09/1953	06/09/1953	03/30/1956	Yes	No	Yes	
6	DR-7	Disaster	06/11/1953	MA	NaN	Tornado	Tornado	06/11/1953	06/11/1953	06/01/1956	Yes	No	Yes	
7	DR-8	Disaster	06/11/1953	IA	NaN	Flood	Flood	06/11/1953	06/11/1953	11/01/1955	Yes	No	Yes	
8	DR-9	Disaster	06/19/1953	TX	NaN	Flood	Flood	06/19/1953	06/19/1953	01/01/1958	Yes	No	Yes	
9	DR-11	Disaster	07/02/1953	NH	NaN	Fire	Forest Fire	07/02/1953	07/02/1953	02/01/1956	Yes	No	Yes	
10	DR-12	Disaster	10/22/1953	FL	NaN	Flood	Flood	10/22/1953	10/22/1953	05/01/1956	Yes	No	Yes	

```
df_data = pd.read_csv('database.csv', encoding='latin1')
```

```
df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 46185 entries, 0 to 46184
Data columns (total 14 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Declaration Number                    46185 non-null  object
 1   Declaration Type                      46185 non-null  object
 2   Declaration Date                      46185 non-null  object
 3   State                                 46185 non-null  object
 4   County                                45988 non-null  object
 5   Disaster Type                         46185 non-null  object
 6   Disaster Title                        46185 non-null  object
 7   Start Date                            46185 non-null  object
 8   End Date                              45843 non-null  object
 9   Close Date                            35210 non-null  object
 10  Individual Assistance Program         46185 non-null  object
 11  Individuals & Households Program      46185 non-null  object
 12  Public Assistance Program             46185 non-null  object
 13  Hazard Mitigation Program             46185 non-null  object
dtypes: object(14)
memory usage: 4.9+ MB
```

These are the columns and its description:
Unnamed: - Serial Number

Name – Name of Movie
Year – Year of Release
Runtime – Total Duration of the movie
Genre – Genre of movie
Rating – Rating of the Movie out of 10
Metascore – Metascore of Movie
Timeline – Description of the Movie
Votes – Number of Votes received
Gross – Gross Income

Screenshot of the data set:

	Unnamed: 0	name	year	runtime	genre	rating	metascore	timeline	votes	gross
0	0	The Shawshank Redemption	1994	142	Drama	9.3	80.0	Two imprisoned men bond over a number of years...	2,394,059	$28.34M
1	1	The Godfather	1972	175	Crime, Drama	9.2	100.0	An organized crime dynasty's aging patriarch t...	1,658,439	$134.97M
2	2	Soorarai Pottru	2020	153	Drama	9.1	NaN	Nedumaaran Rajangam "Maara" sets out to make t...	78,266	NaN
3	3	The Dark Knight	2008	152	Action, Crime, Drama	9.0	84.0	When the menace known as the Joker wreaks havo...	2,355,907	$534.86M
4	4	The Godfather: Part II	1974	202	Crime, Drama	9.0	90.0	The early life and career of Vito Corleone in ...	1,152,912	$57.30M
5	5	12 Angry Men	1957	96	Crime, Drama	9.0	96.0	A jury holdout attempts to prevent a miscarria...	706,079	$4.36M
6	6	The Lord of the Rings: The Return of the King	2003	201	Action, Adventure, Drama	8.9	94.0	Gandalf and Aragorn lead the World of Men agai...	1,672,460	$377.85M
7	7	Pulp Fiction	1994	154	Crime, Drama	8.9	94.0	The lives of two mob hitmen, a boxer, a gangst...	1,862,472	$107.93M
8	8	Schindler's List	1993	195	Biography, Drama, History	8.9	94.0	In German-occupied Poland during World War II,...	1,236,213	$96.90M
9	9	Inception	2010	148	Action, Adventure, Sci-Fi	8.8	74.0	A thief who steals corporate secrets through t...	2,113,984	$292.58M
10	10	Fight Club	1999	139	Drama	8.8	66.0	An insomniac office worker and a devil-may-car...	1,892,181	$37.03M

```
df_data = pd.read_csv('IMDB_movie_reviews_details.csv', encoding='latin1')
```

df_data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Unnamed: 0   1000 non-null   int64
 1   name         1000 non-null   object
 2   year         1000 non-null   object
 3   runtime      1000 non-null   int64
 4   genre        1000 non-null   object
 5   rating       1000 non-null   float64
 6   metascore    841 non-null    float64
 7   timeline     1000 non-null   object
 8   votes        1000 non-null   object
 9   gross        829 non-null    object
dtypes: float64(2), int64(2), object(6)
memory usage: 78.2+ KB
```

5. StatewiseTestingDetails

Columns and its description:

Date - Date of entry
State – State for total update
Total Samples – Total samples collected on particular a date
Negative – Number of Covid patients that are tested Negative
Positive - Number of Covid patients that are tested Positive

Screenshot of the data set:

	Date	State	TotalSamples	Negative	Positive
0	2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0
1	2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2	2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
3	2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
4	2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0
5	2020-05-19	Andaman and Nicobar Islands	6965.0	NaN	33.0
6	2020-05-20	Andaman and Nicobar Islands	7082.0	NaN	33.0
7	2020-05-21	Andaman and Nicobar Islands	7167.0	NaN	33.0
8	2020-05-22	Andaman and Nicobar Islands	7263.0	NaN	33.0
9	2020-05-23	Andaman and Nicobar Islands	7327.0	NaN	33.0
10	2020-05-24	Andaman and Nicobar Islands	7327.0	NaN	33.0

```
df_data = pd.read_csv('StatewiseTestingDetails.csv', encoding='latin1')
df_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13852 entries, 0 to 13851
Data columns (total 5 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Date           13852 non-null  object
 1   State          13852 non-null  object
 2   TotalSamples   13852 non-null  float64
 3   Negative       6284 non-null   object
 4   Positive       5483 non-null   float64
dtypes: float64(2), object(3)
memory usage: 541.2+ KB
```