Missing Values

Setup

```
import numpy as np
import pandas as pd
```

```
df = pd.read_csv('/content/Economy_of_US_na.csv')
```

Detect and report missing values

```
df
```

	Year	GDP_Nominal	GDP_Growth
0	1980.0	2857.3	NaN
1	1981.0	3207.0	0.025
2	1982.0	3343.8	-0.018
3	1983.0	NaN	NaN
4	1984.0	4037.7	0.072
5	1985.0	4339.0	NaN
6	1986.0	NaN	NaN
7	1987.0	4855.3	NaN
8	1988.0	5236.4	0.042
9	1989.0	NaN	NaN
10	1990.0	5963.1	0.019
11	NaN	NaN	NaN
12	1992.0	6520.3	0.035

```
df.isnull()
```

	Year	GDP_Nominal	GDP_Growth
0	False	False	True
1	False	False	False
2	False	False	False
3	False	True	True
4	False	False	False
5	False	False	True
6	False	True	True
7	False	False	True
8	False	False	False
9	False	True	True
10	False	False	False
11	True	True	True
12	False	False	False

```
for c in df.columns:
    miss = df[c].isnull().sum()
    print("{} has {} missing value(s)".format(c,miss))
```

```
Year has 1 missing value(s)
GDP_Nominal has 4 missing value(s)
GDP_Growth has 7 missing value(s)
```

Dropping missing values

```
df2 = df.dropna()
df2
```

	Year	GDP_Nominal	GDP_Growth
1	1981.0	3207.0	0.025
2	1982.0	3343.8	-0.018
4	1984.0	4037.7	0.072
8	1988.0	5236.4	0.042
10	1990.0	5963.1	0.019
12	1992.0	6520.3	0.035

```
df
```

	Year	GDP_Nominal	GDP_Growth
0	1980.0	2857.3	NaN
1	1981.0	3207.0	0.025
2	1982.0	3343.8	-0.018
3	1983.0	NaN	NaN
4	1984.0	4037.7	0.072
5	1985.0	4339.0	NaN
6	1986.0	NaN	NaN
7	1987.0	4855.3	NaN
8	1988.0	5236.4	0.042
9	1989.0	NaN	NaN
10	1990.0	5963.1	0.019
11	NaN	NaN	NaN
12	1992.0	6520.3	0.035

```
df2 = df.dropna(axis=1)
df2
```

0
1
2
3
4
5
6
7
8
9
10
11
12

```
df2 = df.dropna(thresh=3)
df2
```

	Year	GDP_Nominal	GDP_Growth
1	1981.0	3207.0	0.025
2	1982.0	3343.8	-0.018
4	1984.0	4037.7	0.072
8	1988.0	5236.4	0.042
10	1990.0	5963.1	0.019
12	1992.0	6520.3	0.035

```
df2 = df.dropna(thresh=2)
df2
```

	Year	GDP_Nominal	GDP_Growth
0	1980.0	2857.3	NaN
1	1981.0	3207.0	0.025
2	1982.0	3343.8	-0.018
4	1984.0	4037.7	0.072
5	1985.0	4339.0	NaN
7	1987.0	4855.3	NaN
8	1988.0	5236.4	0.042
10	1990.0	5963.1	0.019
12	1992.0	6520.3	0.035

```
df2 = df.dropna(thresh=1)
df2
```

	Year	GDP_Nominal	GDP_Growth
0	1980.0	2857.3	NaN
1	1981.0	3207.0	0.025
2	1982.0	3343.8	-0.018
3	1983.0	NaN	NaN
4	1984.0	4037.7	0.072
5	1985.0	4339.0	NaN
6	1986.0	NaN	NaN
7	1987.0	4855.3	NaN
8	1988.0	5236.4	0.042
9	1989.0	NaN	NaN
10	1990.0	5963.1	0.019
12	1992.0	6520.3	0.035

```
df2 = df.dropna(axis = 1, thresh=7)
df2
```

	Year	GDP_Nominal
0	1980.0	2857.3
1	1981.0	3207.0
2	1982.0	3343.8
3	1983.0	NaN
4	1984.0	4037.7
5	1985.0	4339.0
6	1986.0	NaN
7	1987.0	4855.3
8	1988.0	5236.4
9	1989.0	NaN
10	1990.0	5963.1
11	NaN	NaN
12	1992.0	6520.3

```
df2 = df.dropna(axis = 1, thresh=10)
df2
```

	Year
0	1980.0
1	1981.0
2	1982.0
3	1983.0
4	1984.0
5	1985.0
6	1986.0
7	1987.0
8	1988.0
9	1989.0
10	1990.0
11	NaN
12	1992.0

Filling with constant

```
df2 = df.fillna('NA')
df2
```

	Year	GDP_Nominal	GDP_Growth
0	1980.0	2857.3	NA
1	1981.0	3207.0	0.025
2	1982.0	3343.8	-0.018
3	1983.0	NA	NA
4	1984.0	4037.7	0.072
5	1985.0	4339.0	NA
6	1986.0	NA	NA
7	1987.0	4855.3	NA
8	1988.0	5236.4	0.042
9	1989.0	NA	NA
10	1990.0	5963.1	0.019
11	NA	NA	NA
12	1992.0	6520.3	0.035

```
df['Year_filled'] = df['Year'].fillna('YEAR')
df
```

	Year	GDP_Nominal	GDP_Growth	Year_filled
0	1980.0	2857.3	NaN	1980.0
1	1981.0	3207.0	0.025	1981.0
2	1982.0	3343.8	-0.018	1982.0
3	1983.0	NaN	NaN	1983.0
4	1984.0	4037.7	0.072	1984.0
5	1985.0	4339.0	NaN	1985.0
6	1986.0	NaN	NaN	1986.0
7	1987.0	4855.3	NaN	1987.0
8	1988.0	5236.4	0.042	1988.0
9	1989.0	NaN	NaN	1989.0
10	1990.0	5963.1	0.019	1990.0
11	NaN	NaN	NaN	YEAR
12	1992.0	6520.3	0.035	1992.0

Filling with ffill

```
df['GDP_filled_ffill'] = df['GDP_Nominal'].fillna(method = 'ffill')
df[['GDP_Nominal', 'GDP_filled_ffill']]
```

	GDP_Nominal	GDP_filled_ffill
0	2857.3	2857.3
1	3207.0	3207.0
2	3343.8	3343.8
3	NaN	3343.8
4	4037.7	4037.7
5	4339.0	4339.0
6	NaN	4339.0
7	4855.3	4855.3
8	5236.4	5236.4
9	NaN	5236.4
10	5963.1	5963.1
11	NaN	5963.1
12	6520.3	6520.3

Filling with bfill

```
df['GDP_filled_bfill'] = df['GDP_Nominal'].fillna(method = 'bfill')
df[['GDP_Nominal', 'GDP_filled_bfill']]
```

	GDP_Nominal	GDP_filled_bfill
0	2857.3	2857.3
1	3207.0	3207.0
2	3343.8	3343.8
3	NaN	4037.7
4	4037.7	4037.7
5	4339.0	4339.0
6	NaN	4855.3
7	4855.3	4855.3
8	5236.4	5236.4
9	NaN	5963.1
10	5963.1	5963.1
11	NaN	6520.3
12	6520.3	6520.3

Filling with mean

```
df['GDP_Nominal_filled_mean'] = df['GDP_Nominal'].fillna(df['GDP_Nominal'].mean())
df[['GDP_Nominal', 'GDP_Nominal_filled_mean']]
```

	GDP_Nominal	GDP_Nominal_filled_mean
0	2857.3	2857.300000
1	3207.0	3207.000000
2	3343.8	3343.800000
3	NaN	4484.433333
4	4037.7	4037.700000
5	4339.0	4339.000000
6	NaN	4484.433333
7	4855.3	4855.300000
8	5236.4	5236.400000
9	NaN	4484.433333
10	5963.1	5963.100000
11	NaN	4484.433333
12	6520.3	6520.300000

Filling with mode

```
df['GDP_Nominal_filled_mode'] = df['GDP_Nominal'].fillna(df['GDP_Nominal'].mode()[0
df[['GDP_Nominal', 'GDP_Nominal_filled_mode']]
```

	GDP_Nominal	GDP_Nominal_filled_mode
0	2857.3	2857.3
1	3207.0	3207.0
2	3343.8	3343.8
3	NaN	2857.3
4	4037.7	4037.7
5	4339.0	4339.0
6	NaN	2857.3
7	4855.3	4855.3
8	5236.4	5236.4
9	NaN	2857.3
10	5963.1	5963.1
11	NaN	2857.3
12	6520.3	6520.3

Summary

```
df['GDP_Growth_fill_NA'] = df['GDP_Growth'].fillna('NA')
df['GDP_Growth_fill_0'] = df['GDP_Growth'].fillna(0)
df['GDP_Growth_fill_ffill'] = df['GDP_Growth'].fillna(method = 'ffill')
df['GDP_Growth_fill_bfill'] = df['GDP_Growth'].fillna(method = 'bfill')
df['GDP_Growth_fill_mean'] = df['GDP_Growth'].fillna(df['GDP_Growth'].mean())
df['GDP_Growth_fill_mode'] = df['GDP_Growth'].fillna(df['GDP_Growth'].mode()[0])
df[['GDP_Growth', 'GDP_Growth_fill_NA', 'GDP_Growth_fill_0',
    'GDP_Growth_fill_ffill','GDP_Growth_fill_bfill',
    'GDP_Growth_fill_mean', 'GDP_Growth_fill_mode']]
```

	GDP_Growth	GDP_Growth_fill_NA	GDP_Growth_fill_0	GDP_Growth_fill_ffill	G
0	NaN	NA	0.000	NaN	
1	0.025	0.025	0.025	0.025	
2	-0.018	-0.018	-0.018	-0.018	
3	NaN	NA	0.000	-0.018	
4	0.072	0.072	0.072	0.072	
5	NaN	NA	0.000	0.072	
6	NaN	NA	0.000	0.072	
7	NaN	NA	0.000	0.072	
8	0.042	0.042	0.042	0.042	
9	NaN	NA	0.000	0.042	
10	0.019	0.019	0.019	0.019	
11	NaN	NA	0.000	0.019	
12	0.035	0.035	0.035	0.035	

Colab paid products - Cancel contracts here