

1. Data Validation (validating input)

▾ Validate the data frame

To validate the data frame is empty or not

```
[79] def read_url(url):  
    df = pd.read_csv(url)  
    if(df.empty):  
        print ('url file is empty')  
    else:  
        print ('url file is not empty')  
    return df
```

```
#1st approach get data from our github  
url = 'https://raw.githubusercontent.com/Dabaiee/CS5100group7/main/datasets/TSLA.csv'  
df = pd.read_csv(url)  
df
```

	Date	Open	High	Low	Close	Adj Close	Volume
0	2017-11-17	21.711332	21.778000	20.876667	21.003332	21.003332	206026500
1	2017-11-20	20.919333	21.033333	20.316668	20.582666	20.582666	123715500
2	2017-11-21	20.724001	21.215334	20.580667	21.187332	21.187332	108919500
3	2017-11-22	21.118000	21.161333	20.789333	20.840000	20.840000	73764000
4	2017-11-24	20.919333	21.094000	20.733334	21.036667	21.036667	48661500

✓ 2s completed at 6:41 PM

▾ data types

```
print(df.dtypes)
```

```
Date      object  
Open      float64  
High      float64  
Low       float64  
Close     float64  
Adj Close float64  
Volume    int64  
dtype: object
```

▾ validate data to check missing values

```
[83] df.columns  
Index(['Date', 'Open', 'High', 'Low', 'Close', 'Adj Close', 'Volume'], dtype='object')
```

```
for col in df.columns:  
    miss = df[col].isnull().sum()  
    if miss>0:  
        print("{} has {} missing value(s)".format(col,miss))  
    else:  
        print("{} has NO missing value!".format(col))
```

```
Date has NO missing value!  
Open has NO missing value!  
High has NO missing value!  
Low has NO missing value!  
Close has NO missing value!  
Adj Close has NO missing value!  
Volume has NO missing value!
```

```

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1258 entries, 0 to 1257
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype  
---  -
0   Date        1258 non-null   datetime64[ns]
1   Open        1258 non-null   float64
2   High        1258 non-null   float64
3   Low         1258 non-null   float64
4   Close       1258 non-null   float64
5   Adj Close   1258 non-null   float64
6   Volume      1258 non-null   int64  
dtypes: datetime64[ns](1), float64(5), int64(1)
memory usage: 68.9 KB

[90] df.shape

(1258, 7)

```

2. Testing for Actuators (if any)

3. UI Testing (if applicable)



4. ML Model(s) Testing (validation metrics wrt to various models such as Logistic Regression, Decision tree etc.)