**Data Cleaning Functions**

**Checklist**

* Drop any unnecessary columns.
* Check the data types in the dataset. Are they correct or do they need to be changed?
* Check for and drop any duplicate rows.
* Check for missing values. How many are there?
* Determine the best decision for dealing with the missing values and change the dataset. Possible choices are:
* Drop the rows with missing values.
* Drop the columns with missing values.
* Fill in the missing values with a zero.
* Fill in the missing values with an average value.
* Fill in the missing values with the value either before or after the missing value.
* Leave value as missing (missing data can be good data, too!)

**Drop Irrelevant Columns**

df = df.drop(["column1", "column2"], axis=1)

**Drop Missing Values**

df.isnull().sum()

df.dropna()

df.dropna(axis=1)

**Replace Missing Values**

df.fillna(0, inplace=True)

df.fillna(method='bfill')

df.fillna(method='ffill', axis=1)

**Drop Duplicate Rows**

df.duplicated().sum()

df.loc[df.duplicated()])

df.drop\_duplicates(inplace=True)

**Change Data Types**

df.column.astype(data\_type)

pd.to\_numeric(df.column)