Activity: Modifying & Replacing Values

Introduction

In this activity you will practice modifying and replacing values in a DataFrame using the various method that Pandas has to offer. This activity will cover the following, not necessarily in this order:

- Checking for anomalous values
- Using `.isnumeric()`
- Using `min()` and `max()` methods
- Using `.loc[]` to replace values
- Using `isnull()` and `notnull()` methods

Question 1

Create a `DataFrame` called `df` from the given CSV file `employee_data.csv` , and then create a mask called `valid_names` that checks the `Name` column for any non-numeric values.

In [11]:
```
import pandas as pd

# Your code here

df = pd.read_csv('employee_data.csv')

valid_names = ~df.Name.str.isnumeric()

valid_names.head()

df.head()
```

Out[11]:

	Name	Years of Employment	Weeks of Vacation	Position
0	Jennifer Jackson	9	4.0	Engineer
1	Michael Johnson	9	6.0	Analyst
2	Robert Lee	13	3.0	Engineer
3	Linda Jones	3	6.0	Manager
4	Karen Thomas	14	2.0	Intern

In [10]:
```
# Question 1 Grading Checks

assert isinstance(df, pd.DataFrame), 'Have you created a DataFrame named df?'
assert isinstance(valid_names, pd.Series), 'Have you created a Series named valid_names?'
```

Question 2

Using the original `DataFrame` `df`, create a mask called `unknown_position` that checks the `Position` column for any values that are equal to the string `Unknown`. Then, replace all such values with `Engineer`.

In [44]:
```
# Your code here

unknown_position = df['Position'] == 'Unkown'

df.loc[unknown_position, 'Position'] = 'Engineer'

df['Position'].head()
```

Out[44]:
```
0    Engineer
1     Analyst
2    Engineer
3     Manager
4      Intern
Name: Position, dtype: object
```

In [45]:
```
# Question 2 Grading Checks

assert isinstance(unknown_position, pd.Series), 'Have you created a Series named unknown_position?'
```

Question 3

Using the original `DataFrame` `df`, create a mask called `invalid_vacation` that checks the `Weeks of Vacation` column for any values that are null or missing. Then, use that mask to assign the value 0 to them.

In [42]:
```
# Your code here

invalid_vacation = df['Weeks of Vacation'].isnull()

df.loc[invalid_vacation, 'Weeks of Vacation'] = 0

df.head()
```

Out[42]:

	Name	Years of Employment	Weeks of Vacation	Position
0	Jennifer Jackson	9	4.0	Engineer
1	Michael Johnson	9	6.0	Analyst
2	Robert Lee	13	3.0	Engineer
3	Linda Jones	3	6.0	Manager
4	Karen Thomas	14	2.0	Intern

In [43]:
```
# Question 3 Grading Checks

assert isinstance(invalid_vacation, pd.Series), 'Have you created a Series named invalid_vacation?'
```