Project: Milestone 3

Author: Jordan, Andrew

Date: 07/14/2022

In [123...
```python
import pandas as pd
import numpy as np
import requests

%pwd
```

Out[123]: `'C:\\Users\\Andrew\\Documents\\Grad School\\DSC 540 - Data Preparation\\Project'`

In [124...
```python
#Import data
url = "https://pokemondb.net/pokedex/all"
html = requests.get(url).content
poke_list = pd.read_html(html)
pokesite = poke_list[-1]
pokesite.head(10)
```

Out[124]:

	#	Name	Type	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
0	1	Bulbasaur	Grass Poison	318	45	49	49	65	65	45
1	2	Ivysaur	Grass Poison	405	60	62	63	80	80	60
2	3	Venusaur	Grass Poison	525	80	82	83	100	100	80
3	3	Venusaur Mega Venusaur	Grass Poison	625	80	100	123	122	120	80
4	4	Charmander	Fire	309	39	52	43	60	50	65
5	5	Charmeleon	Fire	405	58	64	58	80	65	80
6	6	Charizard	Fire Flying	534	78	84	78	109	85	100
7	6	Charizard Mega Charizard X	Fire Dragon	634	78	130	111	130	85	100
8	6	Charizard Mega Charizard Y	Fire Flying	634	78	104	78	159	115	100
9	7	Squirtle	Water	314	44	48	65	50	64	43

1. Split dual-type into two columns

In [125…

```python
#Split type into two columns
pokesite[['Type1','Type2']] = pokesite.Type.str.split(' ', expand=True)
#Drop original Type column
pokesite = pokesite.drop('Type', axis=1)
```

2. Change header names

In [126…

```python
#Rename pokedex column number
pokesite = pokesite.rename(columns={'#':'Pokedex','Total':'Total_Stats','Sp. Atk':'SpAtk','Sp. Def':'SpDef'})
pokesite.head()
```

Out[126]:

	Pokedex	Name	Total_Stats	HP	Attack	Defense	SpAtk	SpDef	Speed	Type1	Type2
0	1	Bulbasaur	318	45	49	49	65	65	45	Grass	Poison
1	2	Ivysaur	405	60	62	63	80	80	60	Grass	Poison
2	3	Venusaur	525	80	82	83	100	100	80	Grass	Poison
3	3	Venusaur Mega Venusaur	625	80	100	123	122	120	80	Grass	Poison
4	4	Charmander	309	39	52	43	60	50	65	Fire	None

3. Find and modify duplicates

In [127…
```python
#Change pokedex number of mega evolutions to unique value
pokesite['Pokedex']=np.where(pokesite['Pokedex'].duplicated(keep='first'), pokesite['Pokedex'] + .1, pokesite['Pokede
#Run a second time for those with multiple mega evolution forms
pokesite['Pokedex']=np.where(pokesite['Pokedex'].duplicated(keep='first'), pokesite['Pokedex'] + .1, pokesite['Pokede
pokesite.head(10)
```

Out[127]:

	Pokedex	Name	Total_Stats	HP	Attack	Defense	SpAtk	SpDef	Speed	Type1	Type2
0	1.0	Bulbasaur	318	45	49	49	65	65	45	Grass	Poison
1	2.0	Ivysaur	405	60	62	63	80	80	60	Grass	Poison
2	3.0	Venusaur	525	80	82	83	100	100	80	Grass	Poison
3	3.1	Venusaur Mega Venusaur	625	80	100	123	122	120	80	Grass	Poison
4	4.0	Charmander	309	39	52	43	60	50	65	Fire	None
5	5.0	Charmeleon	405	58	64	58	80	65	80	Fire	None
6	6.0	Charizard	534	78	84	78	109	85	100	Fire	Flying
7	6.1	Charizard Mega Charizard X	634	78	130	111	130	85	100	Fire	Dragon
8	6.2	Charizard Mega Charizard Y	634	78	104	78	159	115	100	Fire	Flying
9	7.0	Squirtle	314	44	48	65	50	64	43	Water	None

4. Clean up name column

```
In [128...    #Remove extra name before "mega" in name column
              pokesite['Name'] = pokesite['Name'].str.split(' ', n=1).str.get(-1)
              pokesite.head(10)
```

Out[128]:

	Pokedex	Name	Total_Stats	HP	Attack	Defense	SpAtk	SpDef	Speed	Type1	Type2
0	1.0	Bulbasaur	318	45	49	49	65	65	45	Grass	Poison
1	2.0	Ivysaur	405	60	62	63	80	80	60	Grass	Poison
2	3.0	Venusaur	525	80	82	83	100	100	80	Grass	Poison
3	3.1	Mega Venusaur	625	80	100	123	122	120	80	Grass	Poison
4	4.0	Charmander	309	39	52	43	60	50	65	Fire	None
5	5.0	Charmeleon	405	58	64	58	80	65	80	Fire	None
6	6.0	Charizard	534	78	84	78	109	85	100	Fire	Flying
7	6.1	Mega Charizard X	634	78	130	111	130	85	100	Fire	Dragon
8	6.2	Mega Charizard Y	634	78	104	78	159	115	100	Fire	Flying
9	7.0	Squirtle	314	44	48	65	50	64	43	Water	None

5. Reorder Columns

Though I could simply type out each column name, I will soon be working with too many columns to do so. Instead I will attempt to order the columns as if I have a large amount and only want to move a few around.

```
In [129...    #Create column orders
              first_cols = ['Pokedex','Name','Type1','Type2']
              last_cols = [col for col in pokesite.columns if col not in first_cols]
              #Re-order columns
              pokesite = pokesite[first_cols+last_cols]
              pokesite.head(10)
```

	Pokedex	Name	Type1	Type2	Total_Stats	HP	Attack	Defense	SpAtk	SpDef	Speed
0	1.0	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45
1	2.0	Ivysaur	Grass	Poison	405	60	62	63	80	80	60
2	3.0	Venusaur	Grass	Poison	525	80	82	83	100	100	80
3	3.1	Mega Venusaur	Grass	Poison	625	80	100	123	122	120	80
4	4.0	Charmander	Fire	None	309	39	52	43	60	50	65
5	5.0	Charmeleon	Fire	None	405	58	64	58	80	65	80
6	6.0	Charizard	Fire	Flying	534	78	84	78	109	85	100
7	6.1	Mega Charizard X	Fire	Dragon	634	78	130	111	130	85	100
8	6.2	Mega Charizard Y	Fire	Flying	634	78	104	78	159	115	100
9	7.0	Squirtle	Water	None	314	44	48	65	50	64	43