Project: Milestone 2

Author: Jordan, Andrew

Date: 06/28/2022

In [178...

```
import pandas as pd
import numpy as np

%pwd
```

Out[178...

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

In [179...

```
pokemon = pd.read_csv("data/Pokemon_data.csv")
pokemon.head(10)
```

Out[179...

	pokedex_number	name	generation	classfication	abilities	height_m	weight_kg	type1	type2	base_total	...	against_psychic
0	1	Bulbasaur	1	Seed Pokemon	['Overgrow', 'Chlorophyll']	0.7	6.9	Grass	Poison	318	...	2.0
1	2	Ivysaur	1	Seed Pokemon	['Overgrow', 'Chlorophyll']	1.0	13.0	Grass	Poison	405	...	2.0
2	3	Venusaur	1	Seed Pokemon	['Overgrow', 'Chlorophyll']	2.0	100.0	Grass	Poison	525	...	2.0
3	3	Mega Venusaur	1	Seed Pokemon	['Thick Fat']	2.4	155.5	Grass	Poison	625	...	2.0
4	4	Charmander	1	Lizard Pokemon	['Blaze', 'Solar Power']	0.6	8.5	Fire	None	309	...	1.0
5	5	Charmeleon	1	Flame Pokemon	['Blaze', 'Solar Power']	1.1	19.0	Fire	None	405	...	1.0
6	6	Charizard	1	Flame Pokemon	['Blaze', 'Solar	1.7	90.5	Fire	Flying	534	...	1.0

	pokedex_number	name	generation	classfication	abilities	height_m	weight_kg	type1	type2	base_total	...	against_psychic
					Power']							
7	6	Mega Charizard X	1	Flame Pokemon	['Tough Claws']	1.7	110.5	Fire	Dragon	634	...	1.0
8	6	Mega Charizard Y	1	Flame Pokemon	['Drought']	1.7	100.5	Fire	Flying	634	...	1.0
9	7	Squirtle	1	Tiny Turtle Pokemon	['Torrent', 'Rain Dish']	0.5	9.0	Water	None	314	...	1.0

In [180…

```
#Create copy for later practice transformations
pokemon_copy = pokemon.copy(deep=True)
```

1.Remove non-relevant variables

In [181…

```
#Back-up the classification column data
classifications = pokemon.classfication
print(classifications)
```

```
0              Seed Pokemon
1              Seed Pokemon
2              Seed Pokemon
3              Seed Pokemon
4            Lizard Pokemon
               ...
1028       Big Fish Pokemon
1029     Free Climb Pokemon
1030    Pin Cluster Pokemon
1031      Love-Hate Pokemon
1032      Love-Hate Pokemon
Name: classfication, Length: 1033, dtype: object
```

In [182…

```
#Drop the classification, heaight, and weight columns
pokemon = pokemon.drop(['classfication','height_m','weight_kg'], axis=1)
pokemon.head()
```

Out[182…

	pokedex_number	name	generation	abilities	type1	type2	base_total	hp	attack	defense	...	against_psychic	against_rock

	pokedex_number	name	generation	abilities	type1	type2	base_total	hp	attack	defense	...	against_psychic	against_rock
0	1	Bulbasaur	1	['Overgrow', 'Chlorophyll']	Grass	Poison	318	45	49	49	...	2.0	1.0
1	2	Ivysaur	1	['Overgrow', 'Chlorophyll']	Grass	Poison	405	60	62	63	...	2.0	1.0
2	3	Venusaur	1	['Overgrow', 'Chlorophyll']	Grass	Poison	525	80	82	83	...	2.0	1.0
3	3	Mega Venusaur	1	['Thick Fat']	Grass	Poison	625	80	100	123	...	2.0	1.0
4	4	Charmander	1	['Blaze', 'Solar Power']	Fire	None	309	39	52	43	...	1.0	2.0

The classification of a Pokemon has no affect on performance in game, thus despite the flavor it adds to the game it is not necessary for the sake of this project. In case of a change of heart regarding the scope of the project, I have saved the information in that column under its own variable. The height and weight attributes have also been dropped for relevance to in-game performance, but no back-up has been created in this case as other datasets I will be combining with this one contain the information in case of a change in project scope.

2.Split singular ability column into one column for each pokemon's ability

In [183...

```
#Split each ability into a separate column
pokemon = pd.concat([pokemon, pokemon['abilities'].str.split(', ', expand=True)], axis=1)
pokemon = pokemon.drop('abilities', axis=1)
pokemon.head()
```

Out[183...

	pokedex_number	name	generation	type1	type2	base_total	hp	attack	defense	sp_attack	...	against_water	capture_rate	base
0	1	Bulbasaur	1	Grass	Poison	318	45	49	49	65	...	0.5	45	
1	2	Ivysaur	1	Grass	Poison	405	60	62	63	80	...	0.5	45	
2	3	Venusaur	1	Grass	Poison	525	80	82	83	100	...	0.5	45	
3	3	Mega Venusaur	1	Grass	Poison	625	80	100	123	122	...	0.5	45	

	pokedex_number	name	generation	type1	type2	base_total	hp	attack	defense	sp_attack	...	against_water	capture_rate	base
4	4	Charmander	1	Fire	None	309	39	52	43	60	...	2.0	45	

In [184...

```
#Rename new ability columns
pokemon = pokemon.rename(columns={0:'ability_1', 1:'ability_2',2:'ability_3'})
pokemon.head(25)
```

Out[184...

	pokedex_number	name	generation	type1	type2	base_total	hp	attack	defense	sp_attack	...	against_water	capture_rate	b
0	1	Bulbasaur	1	Grass	Poison	318	45	49	49	65	...	0.5	45	
1	2	Ivysaur	1	Grass	Poison	405	60	62	63	80	...	0.5	45	
2	3	Venusaur	1	Grass	Poison	525	80	82	83	100	...	0.5	45	
3	3	Mega Venusaur	1	Grass	Poison	625	80	100	123	122	...	0.5	45	
4	4	Charmander	1	Fire	None	309	39	52	43	60	...	2.0	45	
5	5	Charmeleon	1	Fire	None	405	58	64	58	80	...	2.0	45	
6	6	Charizard	1	Fire	Flying	534	78	84	78	109	...	2.0	45	
7	6	Mega Charizard X	1	Fire	Dragon	634	78	130	111	130	...	1.0	45	
8	6	Mega Charizard Y	1	Fire	Flying	634	78	104	78	159	...	2.0	45	
9	7	Squirtle	1	Water	None	314	44	48	65	50	...	0.5	45	
10	8	Wartortle	1	Water	None	405	59	63	80	65	...	0.5	45	
11	9	Blastoise	1	Water	None	530	79	83	100	85	...	0.5	45	
12	9	Mega Blastoise	1	Water	None	630	79	103	120	135	...	0.5	45	
13	10	Caterpie	1	Bug	None	195	45	30	35	20	...	1.0	255	

	pokedex_number	name	generation	type1	type2	base_total	hp	attack	defense	sp_attack	...	against_water	capture_rate	l
14	11	Metapod	1	Bug	None	205	50	20	55	25	...	1.0	120	
15	12	Butterfree	1	Bug	Flying	395	60	45	50	90	...	1.0	45	
16	13	Weedle	1	Bug	Poison	195	40	35	30	20	...	1.0	255	
17	14	Kakuna	1	Bug	Poison	205	45	25	50	25	...	1.0	120	
18	15	Beedrill	1	Bug	Poison	395	65	90	40	45	...	1.0	45	
19	15	Mega Beedrill	1	Bug	Poison	495	65	150	40	15	...	1.0	45	
20	16	Pidgey	1	Normal	Flying	251	40	45	40	35	...	1.0	255	
21	17	Pidgeotto	1	Normal	Flying	349	63	60	55	50	...	1.0	120	
22	18	Pidgeot	1	Normal	Flying	479	83	80	75	70	...	1.0	45	
23	18	Mega Pidgeot	1	Normal	Flying	579	83	80	80	135	...	1.0	45	
24	19	Rattata	1	Normal	None	253	30	56	35	25	...	1.0	255	

Each pokemon has one special ability, and for many there are multiple possibilities. As a pokemon's ability greatly affects its in game performance, it is necessary to clearly distinguish each possible ability.

3. Reorder, remove NaN, and clean-up punctuation of new columns

In [185…

```
#Get column list
pokemon.columns.tolist()
```

Out[185…

```
['pokedex_number',
 'name',
 'generation',
 'type1',
 'type2',
 'base_total',
 'hp',
```

```
'attack',
'defense',
'sp_attack',
'sp_defense',
'speed',
'against_bug',
'against_dark',
'against_dragon',
'against_electric',
'against_fairy',
'against_fighting',
'against_fire',
'against_flying',
'against_ghost',
'against_grass',
'against_ground',
'against_ice',
'against_normal',
'against_poison',
'against_psychic',
'against_rock',
'against_steel',
'against_water',
'capture_rate',
'base_egg_steps',
'base_happiness',
'is_legendary',
'is_mythical',
'is_mega',
'ability_1',
'ability_2',
```

In [186…

```
#Re-order columns
pokemon = pokemon[['pokedex_number',
 'name',
 'generation',
 'type1',
 'type2',
 'ability_1',
 'ability_2',
 'ability_3',
 'base_total',
 'hp',
 'attack',
 'defense',
 'sp_attack',
 'sp_defense',
 'speed',
 'against_bug',
 'against_dark',
 'against_dragon',
 'against_electric',
 'against_fairy',
 'against_fighting',
 'against_fire',
 'against_flying',
 'against_ghost',
 'against_grass',
 'against_ground',
 'against_ice',
 'against_normal',
 'against_poison',
 'against_psychic',
 'against_rock',
 'against_steel',
 'against_water',
 'capture_rate',
 'base_egg_steps',
 'base_happiness',
 'is_legendary',
 'is_mythical',
 'is_mega']]
```

In [187…

```
#Remove None values
pokemon['ability_2'].fillna("", inplace=True)
pokemon['ability_3'].fillna("", inplace=True)
pokemon.head()
```

Out[187…

	pokedex_number	name	generation	type1	type2	ability_1	ability_2	ability_3	base_total	hp	...	against_psychic	against_
0	1	Bulbasaur	1	Grass	Poison	['Overgrow'	'Chlorophyll']		318	45	...	2.0	
1	2	Ivysaur	1	Grass	Poison	['Overgrow'	'Chlorophyll']		405	60	...	2.0	
2	3	Venusaur	1	Grass	Poison	['Overgrow'	'Chlorophyll']		525	80	...	2.0	
3	3	Mega Venusaur	1	Grass	Poison	['Thick Fat']			625	80	...	2.0	
4	4	Charmander	1	Fire	None	['Blaze'	'Solar Power']		309	39	...	1.0	

5 rows × 39 columns

In [188…

```
#Clean up ability variables
pokemon['ability_1'] = pokemon['ability_1'].apply(lambda x: x.lstrip("['").rstrip("']"))
pokemon['ability_2'] = pokemon['ability_2'].apply(lambda x: x.lstrip("'").rstrip("']"))
pokemon['ability_3'] = pokemon['ability_3'].apply(lambda x: x.lstrip("'").rstrip("']"))
pokemon.head(25)
```

Out[188…

	pokedex_number	name	generation	type1	type2	ability_1	ability_2	ability_3	base_total	hp	...	against_psychic	agains
0	1	Bulbasaur	1	Grass	Poison	Overgrow	Chlorophyll		318	45	...	2.0	
1	2	Ivysaur	1	Grass	Poison	Overgrow	Chlorophyll		405	60	...	2.0	
2	3	Venusaur	1	Grass	Poison	Overgrow	Chlorophyll		525	80	...	2.0	
3	3	Mega Venusaur	1	Grass	Poison	Thick Fat			625	80	...	2.0	
4	4	Charmander	1	Fire	None	Blaze	Solar Power		309	39	...	1.0	
5	5	Charmeleon	1	Fire	None	Blaze	Solar Power		405	58	...	1.0	

	pokedex_number	name	generation	type1	type2	ability_1	ability_2	ability_3	base_total	hp	...	against_psychic	agains
6	6	Charizard	1	Fire	Flying	Blaze	Solar Power		534	78	...	1.0	
7	6	Mega Charizard X	1	Fire	Dragon	Tough Claws			634	78	...	1.0	
8	6	Mega Charizard Y	1	Fire	Flying	Drought			634	78	...	1.0	
9	7	Squirtle	1	Water	None	Torrent	Rain Dish		314	44	...	1.0	
10	8	Wartortle	1	Water	None	Torrent	Rain Dish		405	59	...	1.0	
11	9	Blastoise	1	Water	None	Torrent	Rain Dish		530	79	...	1.0	
12	9	Mega Blastoise	1	Water	None	Mega Launcher			630	79	...	1.0	
13	10	Caterpie	1	Bug	None	Shield Dust	Run Away		195	45	...	1.0	
14	11	Metapod	1	Bug	None	Shed Skin			205	50	...	1.0	
15	12	Butterfree	1	Bug	Flying	Compound Eyes	Tinted Lens		395	60	...	1.0	
16	13	Weedle	1	Bug	Poison	Shield Dust	Run Away		195	40	...	2.0	
17	14	Kakuna	1	Bug	Poison	Shed Skin			205	45	...	2.0	
18	15	Beedrill	1	Bug	Poison	Swarm	Sniper		395	65	...	2.0	
19	15	Mega Beedrill	1	Bug	Poison	Adaptability			495	65	...	2.0	
20	16	Pidgey	1	Normal	Flying	Keen Eye	Tangled Feet	Big Pecks	251	40	...	1.0	
21	17	Pidgeotto	1	Normal	Flying	Keen Eye	Tangled Feet	Big Pecks	349	63	...	1.0	
22	18	Pidgeot	1	Normal	Flying	Keen Eye	Tangled Feet	Big Pecks	479	83	...	1.0	
23	18	Mega Pidgeot	1	Normal	Flying	No Guard			579	83	...	1.0	
24	19	Rattata	1	Normal	None	Run Away	Guts	Hustle	253	30	...	1.0	

25 rows × 39 columns

Practice Transformations

4.Change specific values based on condition

In [189…

```
#Change pokedex_number to a string
pokemon_copy['pokedex_number'] = pokemon_copy['pokedex_number'].astype(str)
#Change pokedex_number of any pokemon with a mega evolution
pokemon_copy['pokedex_number'] = pokemon_copy['pokedex_number'].mask(pokemon_copy['is_mega'].eq(1), pokemon_copy['poke
pokemon_copy.head()
```

Out[189…

	pokedex_number	name	generation	classfication	abilities	height_m	weight_kg	type1	type2	base_total	...	against_psychic
0	1	Bulbasaur	1	Seed Pokemon	['Overgrow', 'Chlorophyll']	0.7	6.9	Grass	Poison	318	...	2.0
1	2	Ivysaur	1	Seed Pokemon	['Overgrow', 'Chlorophyll']	1.0	13.0	Grass	Poison	405	...	2.0
2	3	Venusaur	1	Seed Pokemon	['Overgrow', 'Chlorophyll']	2.0	100.0	Grass	Poison	525	...	2.0
3	3.5	Mega Venusaur	1	Seed Pokemon	['Thick Fat']	2.4	155.5	Grass	Poison	625	...	2.0
4	4	Charmander	1	Lizard Pokemon	['Blaze', 'Solar Power']	0.6	8.5	Fire	None	309	...	1.0

5 rows × 40 columns

5.Fix typo and remove duplicates

In [191…

```
#Fix classification column typo
pokemon_copy = pokemon_copy.rename(columns={'classfication': 'classification'})
#Remove duplicate classifications
pokemon_copy.drop_duplicates(subset=['classification'], keep='first')
```

Out[191…

	pokedex_number	name	generation	classification	abilities	height_m	weight_kg	type1	type2	base_total	...	against_
0	1	Bulbasaur	1	Seed Pokemon	['Overgrow', 'Chlorophyll']	0.7	6.9	Grass	Poison	318	...	
4	4	Charmander	1	Lizard Pokemon	['Blaze', 'Solar Power']	0.6	8.5	Fire	None	309	...	
5	5	Charmeleon	1	Flame Pokemon	['Blaze', 'Solar Power']	1.1	19.0	Fire	None	405	...	
9	7	Squirtle	1	Tiny Turtle Pokemon	['Torrent', 'Rain Dish']	0.5	9.0	Water	None	314	...	
10	8	Wartortle	1	Turtle Pokemon	['Torrent', 'Rain Dish']	1.0	22.5	Water	None	405	...	
...	...	...	...	...	...	...	...	...	...	...	...	
1026	901	Ursaluna	8	Peat Pokemon	['Guts', 'Bulletproof', 'Unnerve']	2.4	290.0	Normal	Ground	550	...	
1027	902	Basculegion male	8	Big Fish Pokemon	['Rattled', 'Adaptability', 'Mold Breaker']	3.0	110.0	Water	Ghost	530	...	
1029	903	Sneasler	8	Free Climb Pokemon	['Pressure', 'Poison Touch']	1.3	43.0	Poison	Fighting	510	...	
1030	904	Overqwil	8	Pin Cluster Pokemon	['Poison Point', 'Swift Swim', 'Intimidate']	2.5	60.5	Dark	Poison	510	...	
1031	905	Enamorus Incarnate Forme	8	Love-Hate Pokemon	['Healer', 'Contrary']	1.6	48.0	Fairy	Flying	580	...	