

Loading data into Pandas

```
In [10]: import pandas as pd

# df = pd.read_csv('pokemon_data.csv')

# print(df.head(5))

# df_xlsx = pd.read_excel('pokemon_data.xlsx')
# print(df_xlsx.head(3))

# df = pd.read_csv('pokemon_data.txt', delimiter='\t')

# print(df.head(5))

df['HP']

Out[10]: 0      45
1      60
2      80
3      80
4      39
..
795    50
796    50
797    80
798    80
799    80
Name: HP, Length: 800, dtype: int64
```

Reading Data in Pandas

```
In [19]: ##### Read Headers
df.columns

## Read each Column
# print(df[['Name', 'Type 1', 'HP']])

# print(df.Name)

## Read Each Row
# print(df.iloc[0:4])
# for index, row in df.iterrows():
#     print(index, row['Name'])
# df.loc[df['Type 1'] == "Grass"]

## Read a specific location (R,C)
# print(df.iloc[2,1])

Out[19]: Index(['#', 'Name', 'Type 1', 'Type 2', 'HP', 'Attack', 'Defense', 'Sp. Atk',
      'Sp. Def', 'Speed', 'Generation', 'Legendary'],
      dtype='object')
```

Sorting/Describing Data

```
In [20]: df.sort_values(['Type 1', 'HP'], ascending=[1,0])

Out[20]:
```

	#	Name	Type 1	Type 2	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary	
	520	469	Yanmega	Bug	Flying	86	76	86	116	56	95	4	False
	698	637	Volcarona	Bug	Fire	85	60	65	135	105	100	5	False
	231	214	Heracross	Bug	Fighting	80	125	75	40	95	85	2	False
	232	214	HeracrossMega Heracross	Bug	Fighting	80	185	115	40	105	75	2	False
	678	617	Accelgor	Bug	NaN	80	70	40	100	60	145	5	False

	106	98	Krabby	Water	NaN	30	105	90	25	25	50	1	False
	125	116	Horsea	Water	NaN	30	40	70	70	25	60	1	False
	129	120	Staryu	Water	NaN	30	45	55	70	55	85	1	False
	139	129	Magikarp	Water	NaN	20	10	55	15	20	80	1	False
	381	349	Feebas	Water	NaN	20	15	20	10	55	80	3	False

800 rows × 12 columns

Making changes to the data

```
In [27]: # df['Total'] = df['HP'] + df['Attack'] + df['Defense'] + df['Sp. Atk'] + df['Sp. Def'] + df['Speed']

# df = df.drop(columns=['Total'])

# df['Total'] = df.iloc[:, 4:10].sum(axis=1)

cols = list(df.columns)
df = df[cols[0:4] + [cols[-1]]+cols[4:12]]

df.head(5)
```

```
Out[27]:
```

	#	Name	Type 1	Type 2	Legendary	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation
0	1	Bulbasaur	Grass	Poison	False	318	45	49	49	65	65	45	1
1	2	Ivysaur	Grass	Poison	False	405	60	62	63	80	80	60	1
2	3	Venusaur	Grass	Poison	False	525	80	82	83	100	100	80	1
3	3	VenusaurMega Venusaur	Grass	Poison	False	625	80	100	123	122	120	80	1
4	4	Charmander	Fire	NaN	False	309	39	52	43	60	50	65	1

Saving our Data (Exporting into Desired Format)

```
In [31]: # df.to_csv('modified.csv', index=False)

# df.to_excel('modified.xlsx', index=False)

df.to_csv('modified.txt', index=False, sep='\t')
```

Filtering Data

```
In [37]: new_df = df.loc[(df['Type 1'] == 'Grass') & (df['Type 2'] == 'Poison') & (df['HP'] > 70)]

new_df.reset_index(drop=True, inplace=True)

new_df

new_df.to_csv('filtered.csv')
```

```
Out[37]:
```

	#	Name	Type 1	Type 2	Legendary	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation
0	3	Venusaur	Grass	Poison	False	525	80	82	83	100	100	80	1
1	3	VenusaurMega Venusaur	Grass	Poison	False	625	80	100	123	122	120	80	1
2	45	Vileplume	Grass	Poison	False	490	75	80	85	110	90	50	1
3	71	Victreebel	Grass	Poison	False	490	80	105	65	100	70	70	1
4	591	Amoonguss	Grass	Poison	False	464	114	85	70	85	80	30	5

Conditional Changes

```
In [41]: # df.loc[df['Total'] > 500, ['Generation','Legendary']] = ['Test 1', 'Test 2']
# df.loc[df['Type 1'] == 'Fire', 'Legendary'] = True

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

df
```

```
Out[41]:
```

	#	Name	Type 1	Type 2	Legendary	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation
0	1	Bulbasaur	Grass	Poison	False	318	45	49	49	65	65	45	1
1	2	Ivysaur	Grass	Poison	False	405	60	62	63	80	80	60	1
2	3	Venusaur	Grass	Poison	False	525	80	82	83	100	100	80	1
3	3	VenusaurMega Venusaur	Grass	Poison	False	625	80	100	123	122	120	80	1
4	4	Charmander	Fire	NaN	False	309	39	52	43	60	50	65	1
...
795	719	Diancie	Rock	Fairy	True	600	50	100	150	100	150	50	6
796	719	DiancieMega Diancie	Rock	Fairy	True	700	50	160	110	160	110	110	6
797	720	HoopaHoopa Confined	Psychic	Ghost	True	600	80	110	60	150	130	70	6
798	720	HoopaHoopa Unbound	Psychic	Dark	True	680	80	160	60	170	130	80	6
799	721	Volcanion	Fire	Water	True	600	80	110	120	130	90	70	6

800 rows × 13 columns

Aggregate Statistics (Groupby)

```
In [45]: df = pd.read_csv('modified.csv')

df.groupby(['Type 1']).count()
df['count'] = 1

df.groupby(['Type 1', 'Type 2']).count()['count']

Out[45]:
```

Type 1	Type 2	
Bug	Electric	2
	Fighting	2
	Fire	2
	Flying	14
	Ghost	1

Water	Ice	3
	Poison	3
	Psychic	5
	Rock	4
	Steel	1

Name: count, Length: 136, dtype: int64

Working with large amounts of data

```
In [67]: new_df = pd.DataFrame(columns=df.columns)

for df in pd.read_csv('modified.csv', chunksize=5):
    results = df.groupby(['Type 1']).count()
new_df = pd.concat([new_df, results])
```

```
In [68]: new_df
```

```
Out[68]:
```

	#	Name	Type 1	Type 2	Legendary	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation
Fire	1	1	NaN	1	1	1	1	1	1	1	1	1	1
Psychic	2	2	NaN	2	2	2	2	2	2	2	2	2	2
Rock	2	2	NaN	2	2	2	2	2	2	2	2	2	2

```
In [ ]:
```