

山东大学计算机科学与技术学院

可视化技术课程实验报告

学号：202200101012	姓名：康海洋	班级：数据 23
实验题目： 数据质量实践		
实验学时：2	实验日期： 2025 10 10	
实验步骤： 1 先导入必要的库： <pre>import pandas as pd df = pd.read_csv('Pokemon.csv', encoding='latin1')</pre> 2 最后两行数据无意义，可直接删去 <pre>df = df.iloc[:-2]</pre> 3 type2 存在异常的数值取值，可清空 <pre>valid_types = ['Normal', 'Fire', 'Water', 'Grass', 'Flying', 'Fighting', 'Poison', 'Electric', 'Ground', 'Rock', 'Psychic', 'Ice', 'Bug', 'Ghost', 'Steel', 'Dragon', 'Dark', 'Fairy'] df.loc[~df['Type 2'].isin(valid_types), 'Type 2'] = None</pre> 4 数据集中存在重复值 <pre># 7 删除重复值 # ===== df = df.drop_duplicates(subset=['Name', 'Type 1', 'Type 2'], keep='first')</pre> 5 Attack 属性存在过高的异常值		

```
# 将 Attack 转为数值（字符串会变为 NaN）
df['Attack'] = pd.to_numeric(df['Attack'], errors='coerce')

# 处理 Attack 异常高值 (>200)
df.loc[df['Attack'] > 200, 'Attack'] = None
```

6 有两条数据的 generation 与 Legendary 属性被置换

```
mask = df['Generation'].isin([True, False]) & df['Legendary'].astype(str).str.isnumeric()
```

```
# 将错误的两列交换回来
df.loc[mask, ['Generation', 'Legendary']] = \
df.loc[mask, ['Legendary', 'Generation']].values
```

7 将清洗出来的数据保存

```
# 8 保存清洗后数据
# =====
df.to_csv('Pokemon_cleaned.csv', index=False)
print("清洗完成！已保存为 Pokemon_cleaned.csv")
```

清洗完成！已保存为 Pokemon_cleaned.csv

结果图片：

Delimiter:

	#	Name	Type 1	Type 2	Total	HP	Attack	Def
1	1	Bulbasaur	Grass	Poison	318	45	49.0	
2	2	Ivysaur	Grass	Poison	405	60	62.0	
3	3	Venusaur	Grass	Poison	525	80	82.0	
4	3	VenusaurMega Venusaur	Grass	Poison	625	80	100.0	
5	4	Charmander	Fire		309	39	52.0	
6	5	Charmeleon	Fire		405	58	64.0	
7	6	Charizard	Fire	Flying	534	78	84.0	
8	6	CharizardMega Charizard X	Fire	Dragon	634	78	130.0	
9	6	CharizardMega Charizard Y	Fire	Flying	634	78	104.0	
10	7	Squirtle	Water		314	44		
11	8	Wartortle	Water		405	59	63.0	
12	9	Blastoise	Water		530	79	83.0	
13	9	BlastoiseMega Blastoise	Water		630	79	103.0	
14	10	Caterpie	Bug		195	45	30.0	
15	11	Metapod	Bug		205	50	20.0	
16	12	Butterfree	Bug	Flying	395	60	45.0	
17	13	Weedle	Bug	Poison	195		35.0	
18	14	Kakuna	Bug	Poison	205	45	25.0	
19	15	Beedrill	Bug	Poison	395	65	90.0	
20	15	BeedrillMega Beedrill	Bug	Poison	495	65	150.0	
21	17	Pidgeotto	Normal	Flying	349	63	60.0	
22	16	Pidgey	Normal	Flying	251	40	45.0	
23	18	Pidgeot	Normal	Flying	479	83	80.0	