

```

1  # import the necessary modules
2  import csv # module to read from and write to csv files
3  import pprint as pp
4
5
6  def extract_relevant_cols(csv_file):
7      with open(csv_file, "r") as infile:
8          csv_reader = csv.reader(infile, delimiter=',')
9          lines = [line if len(line) == 5 else line[:5] for line in csv_reader]
10         return lines
11
12
13 def write_to_csv_file(csv_file, data):
14     with open(csv_file, 'w', newline='') as outfile:
15         csv_writer = csv.writer(outfile)
16         csv_writer.writerows(data)
17
18
19 if __name__ == "__main__":
20
21     # open csv file and extract relevant columns
22     data1 = extract_relevant_cols("./gwas_ancestry_01.csv")
23     data2 = extract_relevant_cols("./gwas_ancestry_02.csv")
24
25     # ensure the column headings are the same
26     assert data1[0] == data2[0], 'ensure that the columns match'
27
28     # merge the two collections of data records
29     data = data1 + data2[1:]
30     print("length of cleaned data: {}".format(len(data)))
31
32     # remove rows with any 'NA' values
33     # cleaned_rows = []
34     # for row in data:
35     #     if 'NA' in row:
36     #         continue
37     #     else:
38     #         cleaned_rows.append(row)
39     cleaned_rows = [row for row in data if 'NA' not in row]
40
41     # print out some information about the data
42     print("length of uncleaned data: {}".format(len(cleaned_rows)))
43     pp.pprint(cleaned_rows[:10])
44
45     # write processed data back to a csv file
46     write_to_csv_file("cleaned_gwas_ancestry.csv", cleaned_rows)

```