1. Gathering data

February 22, 2024

Gathering Data from Kisckstarter

https://webrobots.io/kickstarter-datasets/

[1]:
```
import pandas as pd
import csv

inputs = ["2022-01.csv", "2022-02.csv","2022-03.csv","2022-04.csv","2022-05.
 ↪csv","2022-06.csv",
        "2022-07.csv","2022-08.csv","2022-09.csv","2022-10.csv","2022-11.
 ↪csv","2022-12.csv",
        "2021-01.csv", "2021-02.csv","2021-03.csv","2021-04.csv","2021-05.
 ↪csv","2021-06.csv",
        "2021-07.csv","2021-08.csv","2021-09.csv","2021-10.csv","2021-11.
 ↪csv","2021-12.csv",
        "2020-01.csv", "2020-02.csv","2020-03.csv","2020-04.csv","2020-05.
 ↪csv","2020-06.csv",
        "2020-07.csv","2020-08.csv","2020-09.csv","2020-10.csv","2020-11.
 ↪csv","2020-12.csv",
        "2023-01.csv", "2023-02.csv","2023-03.csv","2023-04.csv","2023-05.
 ↪csv","2023-06.csv",
        "2023-07.csv","2023-08.csv","2023-09.csv","2023-10.csv","2023-11.
 ↪csv","2023-12.csv"]
fieldnames = set()  # Use a set to avoid duplicates

# Collect unique field names from all input files
for filename in inputs:
    with open(filename, "r", newline="") as f_in:
        reader = csv.reader(f_in)
        headers = next(reader)
        fieldnames.update(headers)

# Then copy the data
with open("uncleaned_4years.csv", "w", newline="") as f_out:    # Comment 2 below
    writer = csv.DictWriter(f_out, fieldnames=fieldnames)
    writer.writeheader()  # Write header with field names
    for filename in inputs:
        with open(filename, "r", newline="") as f_in:
```

```
            reader = csv.DictReader(f_in)  # Uses the field names in this file
            for line in reader:
                # Comment 3 below
                writer.writerow(line)
```