data_collection

April 26, 2024

```python
import pandas as pd
import numpy as np

import fastf1
import logging
```

Get weather details from fastf1 python module

```python
logging.getLogger('fastf1').setLevel(logging.CRITICAL)

# Initialize FastF1
fastf1.Cache.enable_cache('in_progress_data')  # Enable caching to speed up
 ↪data retrieval

# Given lists for years and number of rounds
years = [2018, 2019, 2020, 2021, 2022, 2023]
num_races = [22, 22, 18, 22, 23, 23]

weather = pd.DataFrame()

for year, num_race in zip(years, num_races):
    df = pd.DataFrame()

    for x in range(1, num_race):
        race_session = fastf1.get_session(year, x, 'R').event
        df = pd.concat([df, pd.DataFrame([race_session])], ignore_index=True)

    # Get weather data for each race
    for track in df['Location']:
        try:
            race_session = fastf1.get_session(year, track, 'R')
            race_session.load()  # Load the data before accessing it
            weather_data = race_session.weather_data
            round_number = df.loc[df['Location'] == track, 'RoundNumber'].
 ↪values[0]
            weather_data['Round Number'] = round_number
            weather_data['Year'] = year
            weather = pd.concat([weather, weather_data])
```

```
                print(f"{track}, Year: {year}")
            except Exception as e:
                print(f"Error loading weather data for {track}, Year: {year}")
                print(e)

    # Disable caching after the loop
    fastf1.Cache.cache_dir = None
```

[ ]: `weather.drop(columns=['Time'])`

[ ]:
```
     AirTemp  Humidity  Pressure  Rainfall  TrackTemp  WindDirection  \
0       24.1      36.2     997.1     False       38.2            294
1       24.0      36.3     997.1     False       38.6            273
2       24.0      36.3     997.1     False       38.6            273
3       23.9      37.2     997.0     False       38.7            287
4       24.2      35.8     997.1     False       38.7            309
..       ...       ...       ...       ...        ...            ...
151     26.4      54.0    1015.7     False       30.8            279
152     26.4      54.0    1015.7     False       30.8            283
153     26.4      54.0    1015.7     False       30.6            314
154     26.4      54.0    1015.7     False       30.6            273
155     26.3      54.0    1015.7     False       30.5            316

     WindSpeed  Round Number  Year
0          3.0             1  2018
1          1.4             1  2018
2          1.4             1  2018
3          2.3             1  2018
4          3.5             1  2018
..         ...           ...   ...
151        1.7            22  2023
152        1.8            22  2023
153        1.5            22  2023
154        1.5            22  2023
155        1.8            22  2023

[18214 rows x 9 columns]
```

[ ]:
```
weather_file = "./in_progress_data/weather.csv"
weather.to_csv(weather_file, index=False)
```

Get tyre details from fastf1 python module

[ ]:
```
logging.getLogger('fastf1').setLevel(logging.CRITICAL)

# Enable FastF1 cache for faster data retrieval
fastf1.Cache.enable_cache('in_progress_data')
```

```python
# Define the years and rounds for the races
years = [2018, 2019, 2020, 2021, 2022, 2023]
num_races = [22, 22, 18, 22, 23, 23]  # Number of races per year from 2018 to
  ↪2023

# Initialize an empty list to hold data for each row
stint_data_list = []

# Loop through each year and round
for year, num_race in zip(years, num_races):
    for race_num in range(1, num_race):
        # Get the race session information
        session = fastf1.get_session(year, race_num, 'R')
        session.load()

        # Extract stints information
        stints = session.laps[["Driver", "Stint", "Compound", "LapNumber"]]
        stints = stints.groupby(["Driver", "Stint", "Compound"]).
  ↪agg({'LapNumber': ['min', 'max', 'count']})
        stints.columns = ['Stint Start Lap', 'Stint End Lap', 'Stint Length']
        stints = stints.reset_index()

        # Append data to the stint_data list
        for _, row in stints.iterrows():
            stint_data_list.append({
                'Year': year,
                'Round': f'{race_num}',
                'Code': row['Driver'],
                'Stint': row['Stint'],
                'Compound': row['Compound'],
                'Stint Start Lap': row['Stint Start Lap'],
                'Stint End Lap': row['Stint End Lap'],
                'Stint Length': row['Stint Length']
            })

# Convert the list to a DataFrame
stint_data = pd.DataFrame(stint_data_list)

fastf1.Cache.cache_dir = None
```

```python
[ ]: stint_data.head
```

```
[ ]: <bound method NDFrame.head of      Year Round Code  Stint    Compound  Stint
    Start Lap  Stint End Lap  \
    0    2018     1  ALO    1.0  ULTRASOFT            2.0           26.0
    1    2018     1  ALO    2.0       SOFT           27.0           58.0
    2    2018     1  BOT    1.0  ULTRASOFT            2.0           25.0
```

3

```
3      2018     1  BOT    2.0  SUPERSOFT              26.0              58.0
4      2018     1  ERI    1.0  SUPERSOFT               2.0               6.0
...     ...    ..  ...    ...        ...               ...               ...
6653   2023    22  VER    2.0       HARD              17.0              43.0
6654   2023    22  VER    3.0       HARD              44.0              58.0
6655   2023    22  ZHO    1.0     MEDIUM               1.0              13.0
6656   2023    22  ZHO    2.0       HARD              14.0              37.0
6657   2023    22  ZHO    3.0     MEDIUM              38.0              58.0

       Stint Length
0                25
1                32
2                24
3                33
4                 5
...             ...
6653             27
6654             15
6655             13
6656             24
6657             21

[6658 rows x 8 columns]>
```

```python
stint_data.to_csv('./in_progress_data/tire.csv', index=False)
```