Case Study Data Science Master, Module 2: Analyzing Formula 1 Data

In this case study, you will play with Formula 1 data. The data we will work with is from FastF1 (https://theoehrly.github.io/Fast-F1/), which gives you access to F1 lap timing, car telemetry and position, tire data, weather data and weekend information among others. No Formula 1 account is needed. The module is designed around Pandas, Numpy and Matplotlib. This makes it easy to use while offering lots of possibilities for data analysis and visualization. FastF1 handles big chunks of data (~50-100MB per session) so most of the information is stored locally as cached requests (be aware).

It is recommended to install FastF1 using pip (pip install fastf1) and to familiarize yourself with FastF1 before turning to the case study. Note that Python 3.8 or higher is required. Note that you can change between Python versions using the Anaconda Navigator (detailed instructions: https://docs.anaconda.com/anaconda/navigator/tutorials/use-multiple-python-versions/).

For this case study, we will look at the 2021 Abu Dhabi Grand Prix. Specifically, we will analyze the qualification session, during which Max Verstappen and Lewis Hamilton were going neck-2-neck like they had done all season. Eventually, Verstappen beat Hamilton to pole position with 0,371 seconds. Your goal is to see how you can analyze this final lap and find the difference.

Do the following:

1. Enable the cache
2. Setup plotting
3. Load the data and get the laps
4. Study the data to get an understanding of it
5. Select the specific data you are interested in (e.g., Hamilton, Verstappen, fastest lap, etc.)
6. Construct a single DataFrame called telemetry
7. Break the length of the lap into minisectors of equal size and create a list of all distances at which the next minisector starts
8. Assign a minisector to every row in the dataset and answer the question “in which minisector was the car at the moment the datapoint was recorded?”
9. Calculate the fastest driver per minisector: calculate the average speed per driver per minisector, select the driver with the highest average speed, and join the fastest driver per minisector with the telemetry data.

Obviously, there are still a few roadblocks to circumvent before you get to the result, which should be a plot showing who, Hamilton or Verstappen, was faster in which minisector.

Hints:

1. Before you start, you may want to take a look at the tutorial for beginners.
2. Use the following libraries: fastf1, matplotlib, pandas, numpy.

Please submit a complete package consisting of code, dataset, and output so that your results can easily be verified.