

## The main project objective and motivation

The main idea of the project is providing a new implementation of the t-quark pair production analysis from the IRIS-HEP Analysis Grand Challenge (AGC) [1] using ROOT's RDataFrame [2] instead of Coffea [3].

Coffea is the python analysis tool that provides efficient HEP event data manipulation in numpythonic way and additionally to columnar analysis it provides histogramming and plotting utilities [3, 4]. At the same time most analysis codes in high-energy physics are written in ROOT. RDataFrame is ROOT's modern analysis interface providing similar features that Coffea, so it is extremely useful for the community to provide an example implementation of the AGC in RDataFrame to showcase best practices and to investigate possible usability or performance issues, also in comparison with Coffea.

Having a generic description of the analysis in plain English will also make it simpler to provide other alternative implementations, improving knowledge-sharing and offering a common usability benchmark for future analysis APIs.

## Milestones

1. Studying of the existing implementation of Coffea  $t\bar{t}$ -analysis[4] and RDataFrame
2. Writing a description of the  $t\bar{t}$  analysis in plain English, documenting what the analysis logic does (which events are selected, what histograms are produced, etc.)
3. Implementation of another Python version of the analysis using RDataFrame instead of Coffea, demonstrate that the output histograms are the same

## Timelines

The anticipated duration of the project is a two-month period.

- Week 1

Getting familiar with Coffea package

- Week 2

Studying Coffea application to  $t\bar{t}$ -analysis

- Week 3-4

Documenting in plain English what  $t\bar{t}$ -analysis logic does

- Week 5

Getting familiar with RDataFrame. Studying existing examples and writing my own ones.

- Week 6-12

Writing python codes using RDataFrame to do  $t\bar{t}$ -analysis in similar way as Coffea

- Week 13-14

Spillover tasks. Writing documentation. Demonstrate scaling out the RDataFrame version of the analysis on batch resources. Presenting my findings at the ROOT team meeting and IRIS-HEP Fellow meeting at the end of internship.

#### References

1. IRIS-HEP. *Grand Challenge*. URL:  
<https://iris-hep.org/projects/agc.html>
2. ROOT's RDataFrame documentation. URL:  
[https://root.cern/doc/master/classROOT\\_1\\_1RDataFrame.html](https://root.cern/doc/master/classROOT_1_1RDataFrame.html)
3. Coffea documentation. URL:  
<https://coffeateam.github.io/coffea/>
4. IRIS-HEP. *Coffea*. URL:  
<https://github.com/iris-hep/analysis-grand-challenge/tree/main/workshops/agctools2022/coffea>