Final Project: Data Set Description

Group Members: Adam Vendrasco

What is your data set? Does it contain labels?

- My dataset comprises a collection of root files sourced from CMS at CERN, accessible as open public data. Currently, I possess a text file serving as a pointer to numerous root files (image 1), which I presume are hosted on CERN servers. These root files encapsulate detailed information regarding particle collisions, with a focus on Particle Flow (PF) Candidates. These PF candidates represent the reconstructed objects detected by CMS. My objective is to leverage these PF candidates data to reconstruct the mass of Z-bosons utilizing ML techniques. Success would be gauged by the model's ability to accurately predict the momentum of the Z-boson in the z-direction.

- Yes, my dataset does contain labels. However, these labels do not take the form of binary numbers. Instead, they represent the momentum of the Z-boson in the z-direction, acting as ground truth data for comparison with the predictions generated by my model.

What do you plan to learn from it?

- So this dataset was selected because its a Z -> dimuon dataset. Usually when calculating the invariant mass of this system, using the energy, momentum and other properties of the muon pair you reconstruct the Z. However, since protons are composite particles there is a lot more *“stuff”* happening that is not accounted for. So using these PF candidates, I plan to try and get a more precise measurement of the Z-bosons momentum which can be used to get a better measurement of its mass.

What does it look like? How big is it? Provide samples and/or figures/plots.

- The data is all in the form of root files (See images 1 and 2)
- It is quite large. I have not combined all the files yet because currently, I am trying to see if my code will work just on one root file. It already takes awhile to load as it is so adding an extra 100 files will just slow the process down. But all together the root files in total are probably around 100ish gigs of data. I will not use every aspect of the root file but I am still working out on what aspects of the root file I will need in the analysis.
- For some root generated histograms see image 2.

Where did you obtain it? If not, did you generate it yourself? What methods/software did you use?

- I obtained it from CMS CERN open data so no I did not generate it myself.
- CMS detects hits via sensors/energy deposits and send that information (data) to the CERN Data Centre for digital reconstruction. The digitized summary is recorded as a 'collision event’ in a root file. I am now accessing that file. I plan on using ROOT, uproot, pandas, numpy and awkward array packages to actually evaluate the data.

```
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_1.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_10.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_100.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_101.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_102.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_103.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_104.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_105.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_106.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_107.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_108.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_109.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_11.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_110.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_111.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_112.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_113.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_114.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_115.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_116.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_117.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_118.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_119.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_12.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_120.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_121.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_122.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_123.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_124.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_125.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_126.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_127.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_128.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_129.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_13.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_130.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_131.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_132.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_133.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_134.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_135.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_136.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_137.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_138.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_139.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_14.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_140.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_141.root
root://eospublic.cern.ch//eos/opendata/cms/derived-data/PFNano/29-Feb-24/SingleMuon/Run2016G-UL2016_MiniAODv2_PFNanoAODv1/240207_205649/0000/nano_data2016_142.root
```

Image 1

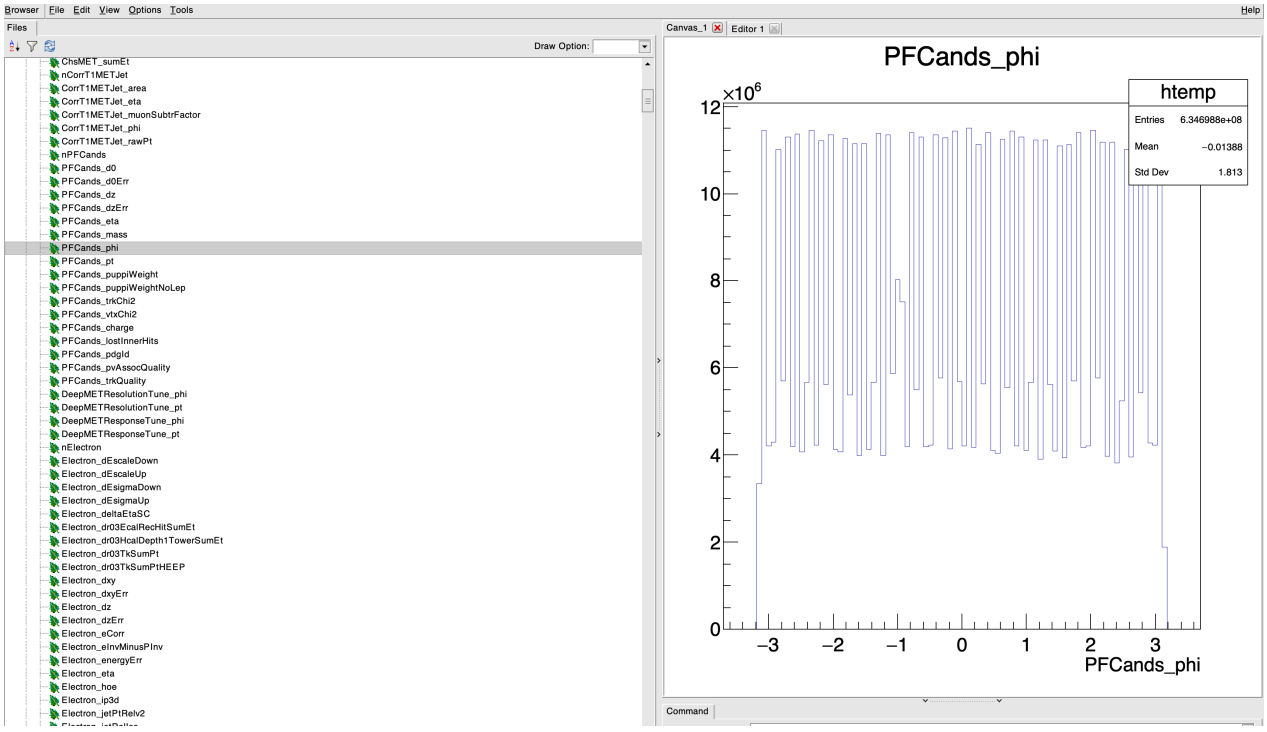

Image 2