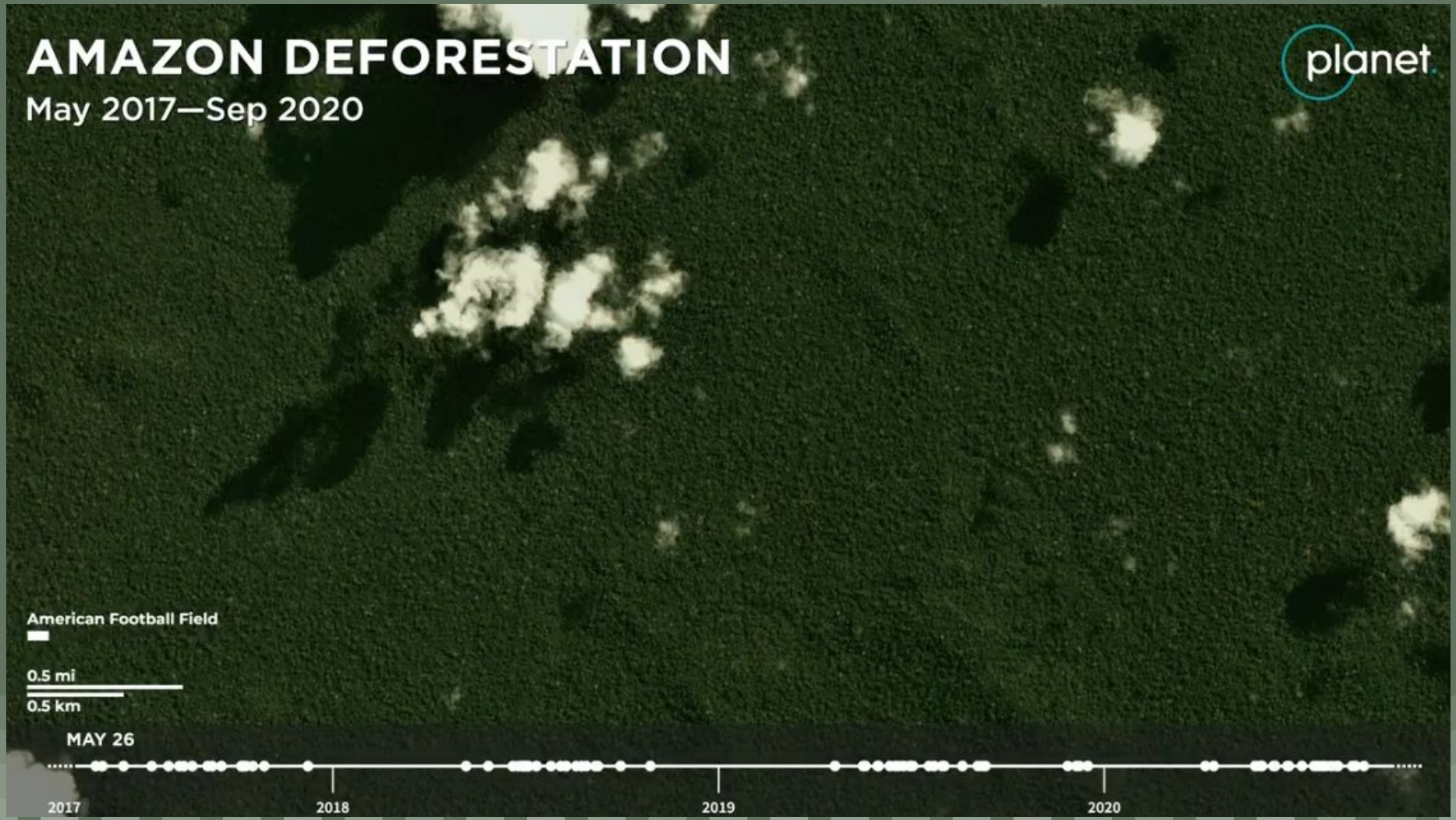
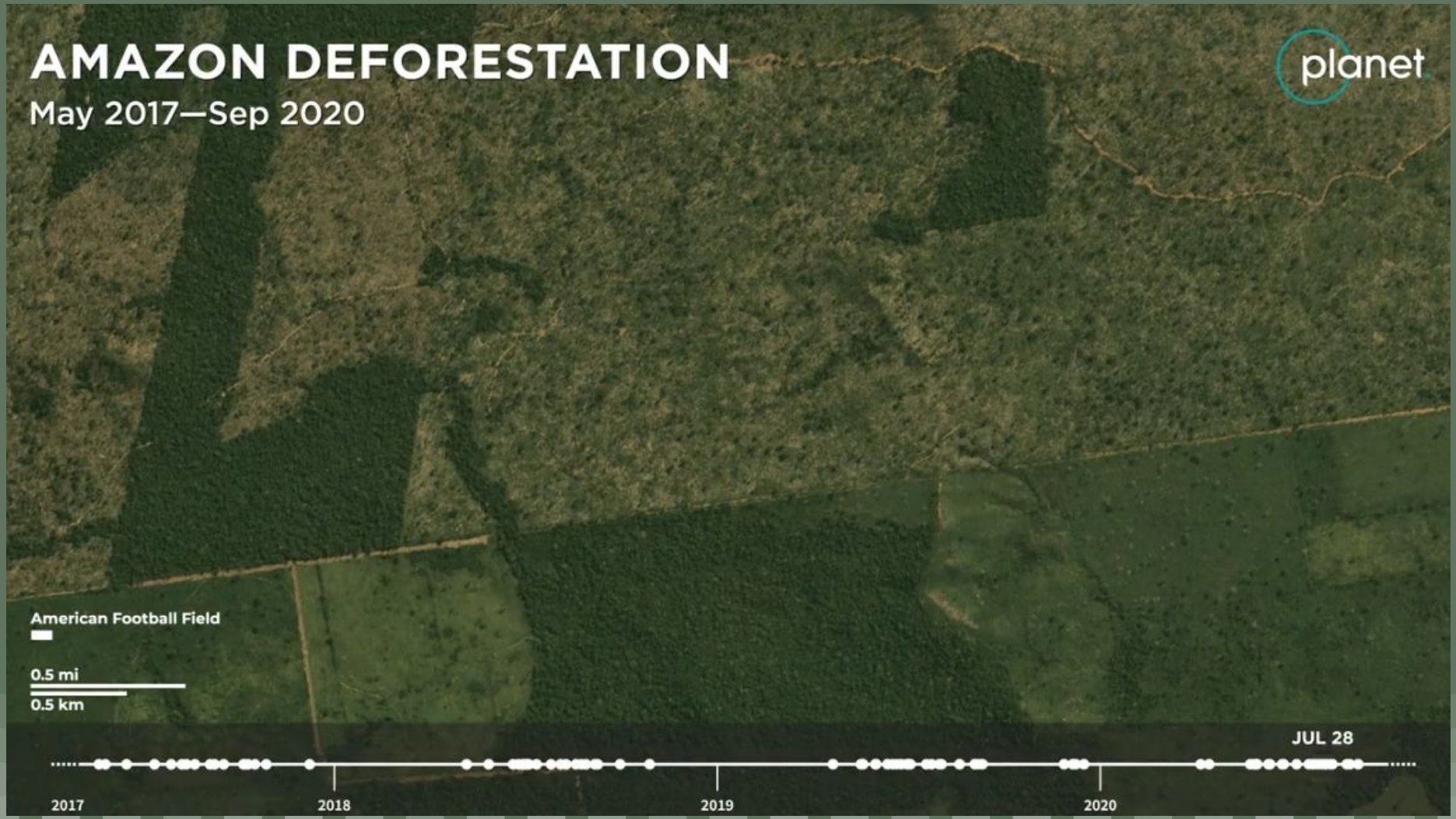
# Satellite Data Monitoring Global Deforestation

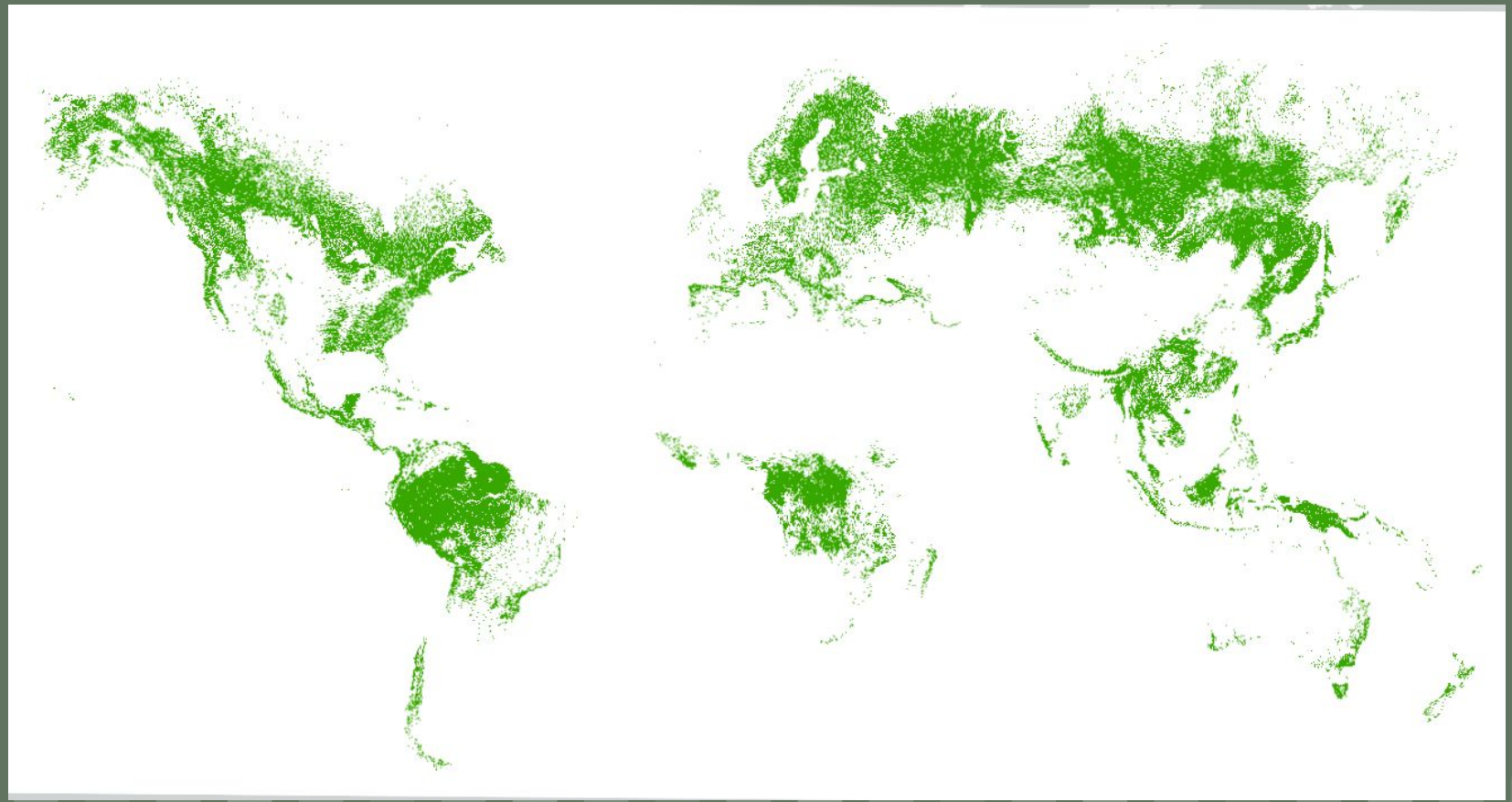
By James Hoang


AMAZON DEFORESTATION
May 2017—Sep 2020
planet.
American Football Field
0.5 mi
0.5 km
MAY 26
2017
2018
2019
2020

AMAZON DEFORESTATION
May 2017—Sep 2020
planet.
American Football Field
0.5 mi
0.5 km
JUN 20
2017
2018
2019
2020


AMAZON DEFORESTATION
May 2017—Sep 2020
planet
American Football Field
0.5 mi
0.5 km
JUL 28
2017
2018
2019
2020

# Global Forest Coverage 2018



https://hub.arcgis.com/datasets/f0cc32be502b49bc87711249ff5dcdfb/explore?location=6.772634%2C7.042178%2C1.48

# Overview of The Subject Area

**Problem Statement:** Using machine learning, how might we predict areas at risk for deforestation such that we can protect and sustain the health of our planet?

**Granularity**: Year over year change as CO2 levels as carbon cycle much better tracked this way as well as numerical canopy density data.

- Also focuses on spatial granularity like regions such as country.

**Model Output**: Ideal output of a model is to predict potential areas at risk for deforestation based on carbon emission and tree coverage. (potential to include climate date)

# Vision for Tackling the Problem using Data Science

Next Point Forecast ML model would help predict carbon emissions and or tree cover density changes for high risk areas

- Main reach for this kind of model is for environmental companies or government agencies to help with sustainability goals or controlled logging
- The metric to add value or success is the models ability and accuracy to predict CO2 and canopy density such that a reduction in annual deforestation occurs
- Model quality can be measure through scikit-learn's train_test_split

# An Estimate of the potential impact of such a solution

Big Picture: While we aim to improve the health and sustainability of our planet through forest conservation efforts, these models can also help companies decide appropriate logging areas that have less of an impact on the ecosystem relative to other areas

Other Benefits:

- Early startup companies can benefit for agriculture or infrastructure growth such that they take a green approach or adhere to environmental regulations
- Detection of illegal logging for quicker responses from authorities

# An introduction to the Dataset

Dataset:

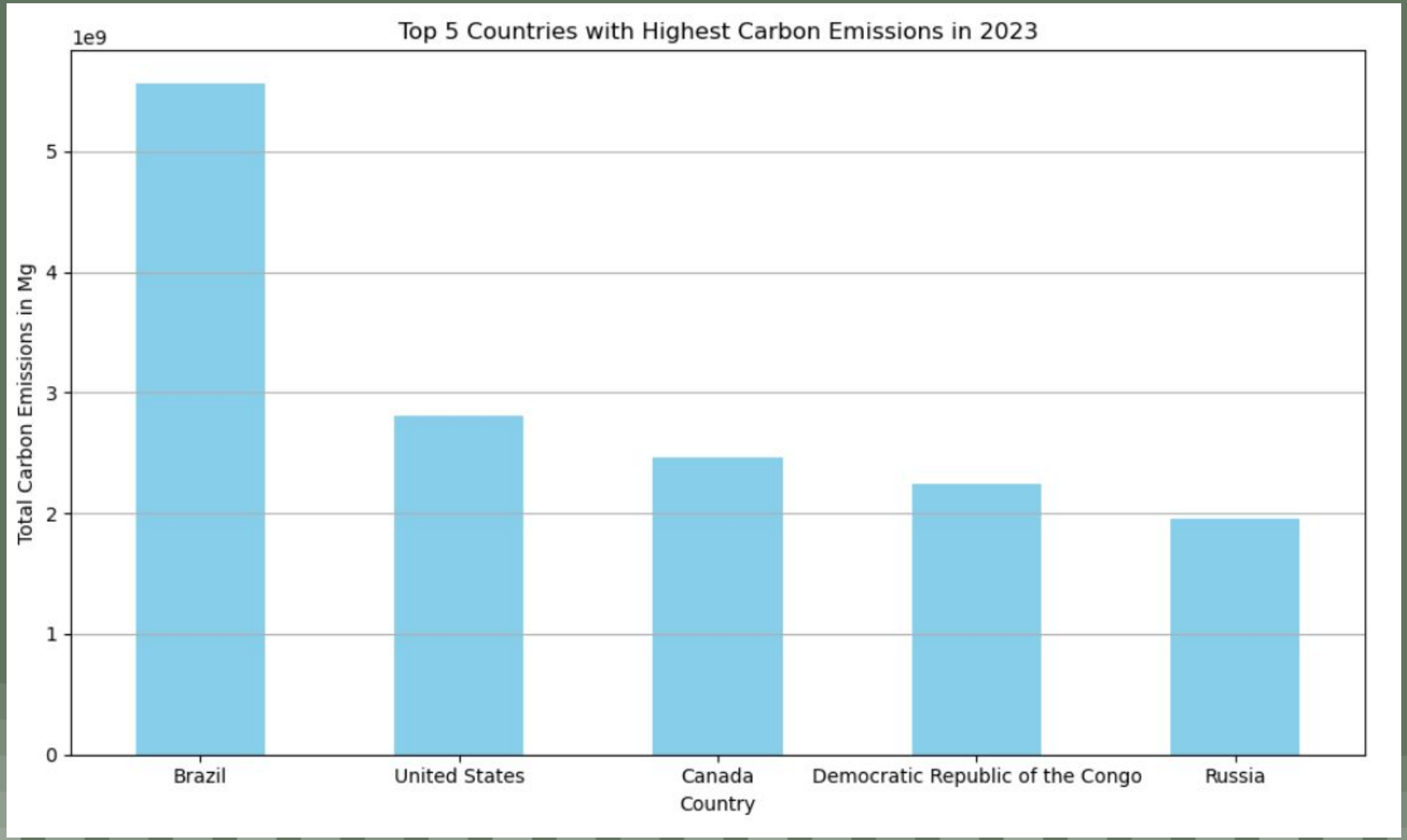
- Global Forest Watch uses dashboards and interactive maps that take on real-time data on ($CO_2$, canopy density, etc)
- Smaller environmental datasets due to its geographical nature

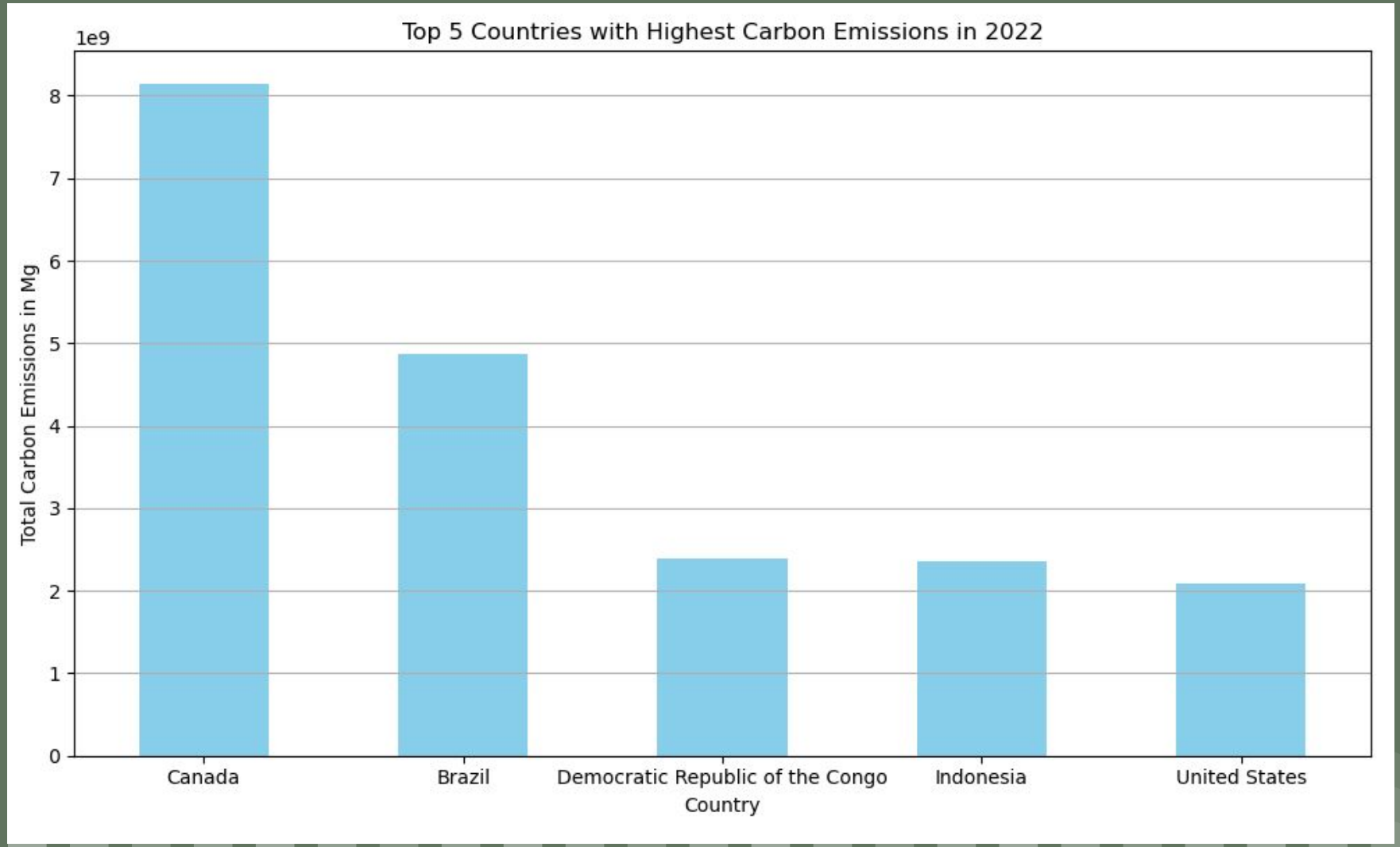
Quality Concerns:

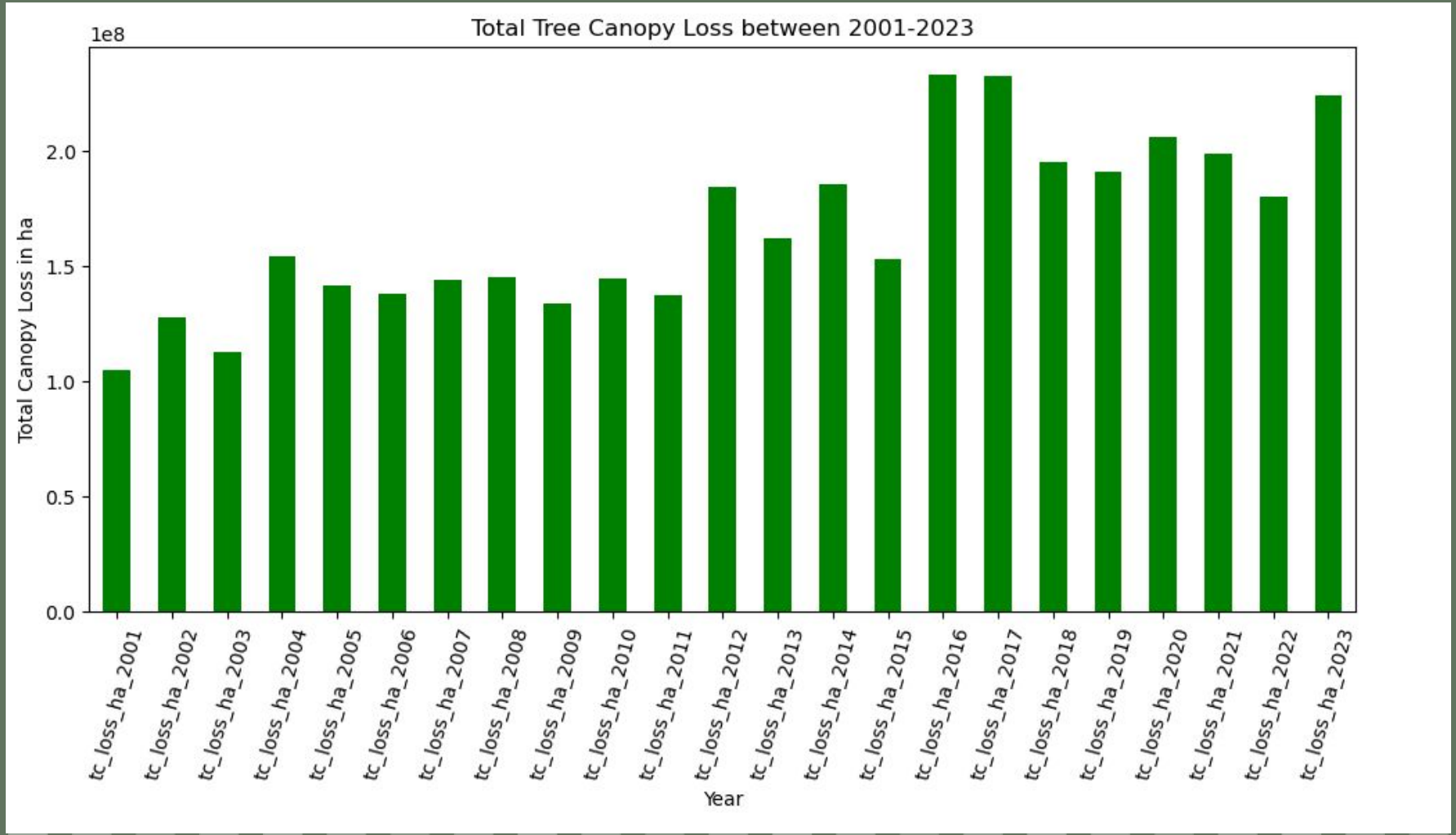
- Had a lot of nulls because it excluded 30 threshold and less
- More robust model might require additional datasets like climate (temperature)

Preliminary EDA Findings:

- Gradual canopy density loss over the years (spike in 2015-2016)
- Brazil followed by Canada are huge carbon emitters


Top 5 Countries with Highest Carbon Emissions in 2023
1e9
Total Carbon Emissions in Mg
0
1
2
3
4
5
Brazil
United States
Canada
Democratic Republic of the Congo
Russia
Country


Top 5 Countries with Highest Carbon Emissions in 2022
1e9
Total Carbon Emissions in Mg
0
1
2
3
4
5
6
7
8
Canada
Brazil
Democratic Republic of the Congo
Indonesia
United States
Country


Total Tree Canopy Loss between 2001-2023
1e8
Total Canopy Loss in ha
0.0
0.5
1.0
1.5
2.0
tc_loss_ha_2001
tc_loss_ha_2002
tc_loss_ha_2003
tc_loss_ha_2004
tc_loss_ha_2005
tc_loss_ha_2006
tc_loss_ha_2007
tc_loss_ha_2008
tc_loss_ha_2009
tc_loss_ha_2010
tc_loss_ha_2011
tc_loss_ha_2012
tc_loss_ha_2013
tc_loss_ha_2014
tc_loss_ha_2015
tc_loss_ha_2016
tc_loss_ha_2017
tc_loss_ha_2018
tc_loss_ha_2019
tc_loss_ha_2020
tc_loss_ha_2021
tc_loss_ha_2022
tc_loss_ha_2023
Year

# Next steps in Feature Engineering & Modeling

Evaluation criteria: is to ensure high-level documentation and reproducibility including comprehensive assumption and context of the analysis

Feature Engineering:

- Year-over-year change
- Rolling averages,
- Additional datasets (climate/temperature)

Baseline modeling:

- Time series Analysis
- Target variables: net carbon emissions/canopy density
- Output: Regions in risk of deforestation

# Supplementary References

https://www.noaa.gov/climate

https://earthdata.nasa.gov/data/catalog?granule_data_format_h[]=CSV

https://research.wri.org/gfr/data-methods#data-sets