Intro to HPC Bootcamp Final Project Presentations

Session 2

August 15, 2025

Lightning Talk Session 2	Group
10:15 - 10:25	Group 3
10:25 - 10:35	
10:35 - 10:45	Group 2a
10:45 - 10:55	Group 1b
10:55 - 11:05	Group 4a
11:05 - 11:15	Group 6b

Groups 3a and 3b

Project 3: The Anatomy of a Power Outage

Teams 3A & 3B
August 15, 2025

Group 3A: The Anatomy of a Power Outage

Kayla Bentley
DePaul University
Computer Science
+ Economics

Bianca Udeogu
University of Illinois at Urbana-Champaign
Math + Computer Science

Justin Dunbar
Howard University
Computer Science
Business Administration

Gleniarys N. Rivera-Torres
University of Puerto Rico at Mayaguez
Chemical Engineering

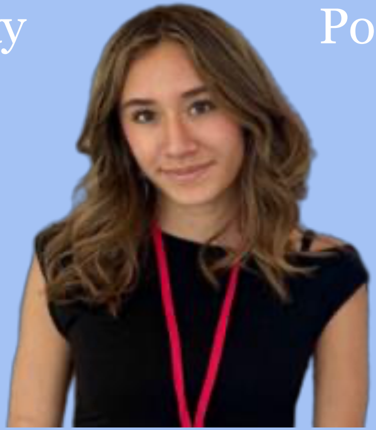
Elijah Potter
Harold Washington College
Biochemistry

Group 3B: The Anatomy of a Power Outage

Leen Abdel Karim
Computer Science - Graduate
Governors State University

Taiwo Quadri
MS. Geography Grad.
Portland State University

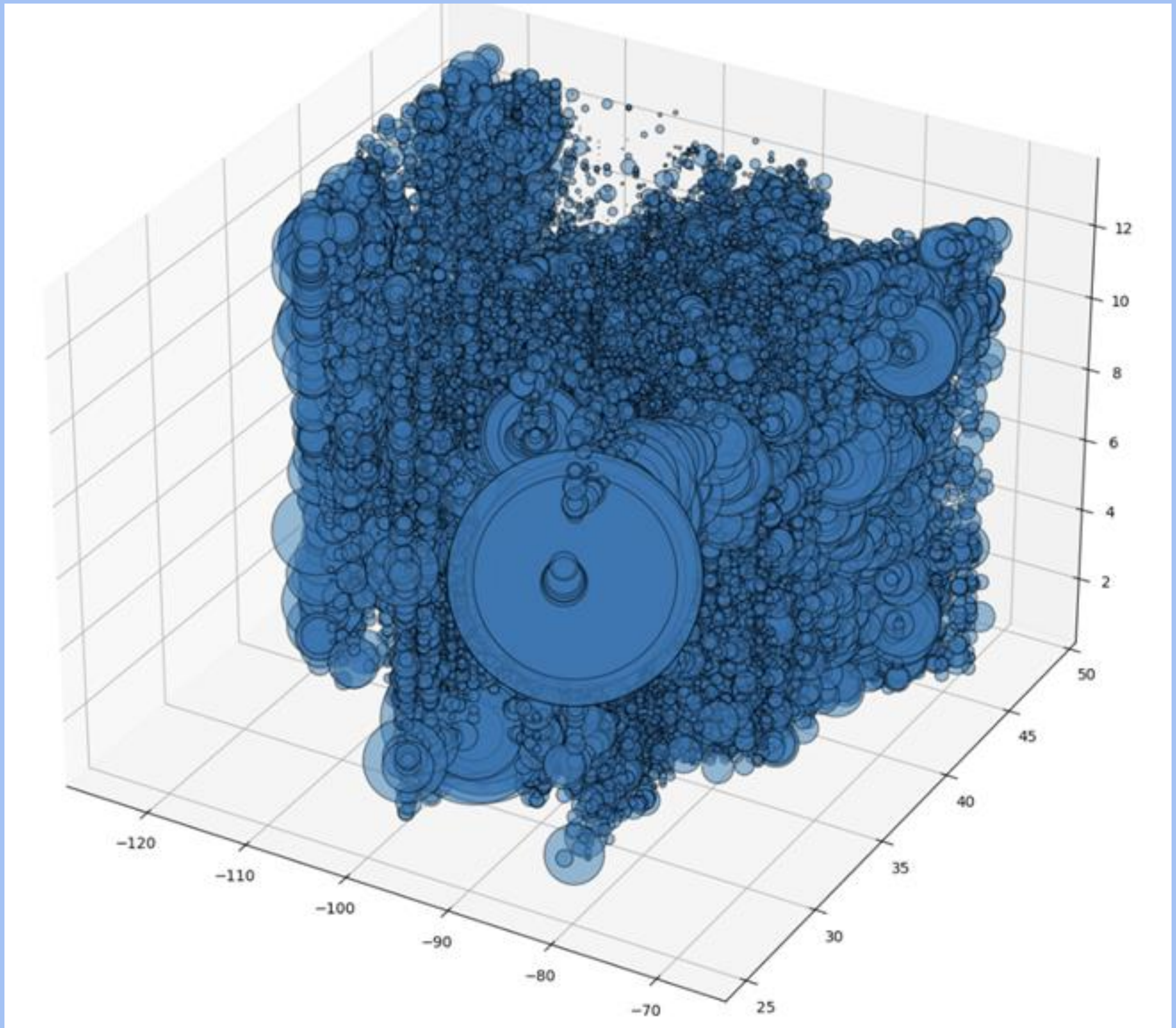
Nick Chaiyachakorn
MS Biostatistics – Candidate
Oregon Health and Science University-Portland State University School of Public Health



Valeria Gonzalez
Chemical Engineering - Freshman
College of DuPage

Joe Udani
Computer Science - Senior
Governors State University

Anatomy of a Power Outage



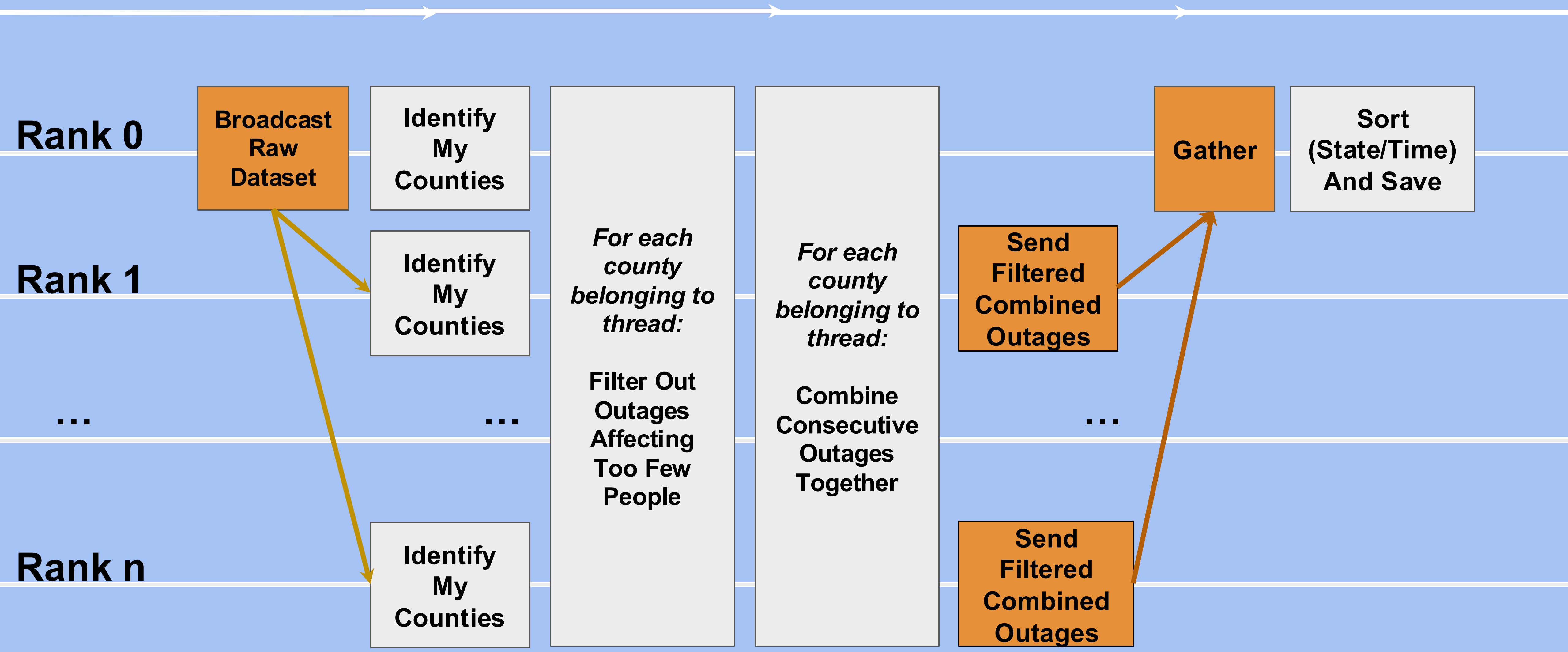
How was the data cleaned?

State, FIPS code, State Number, Region, Latitude, Longitude, Month, Month_Sin, Month_Cos, Outage Start, Outage End, Outage Length, Sum

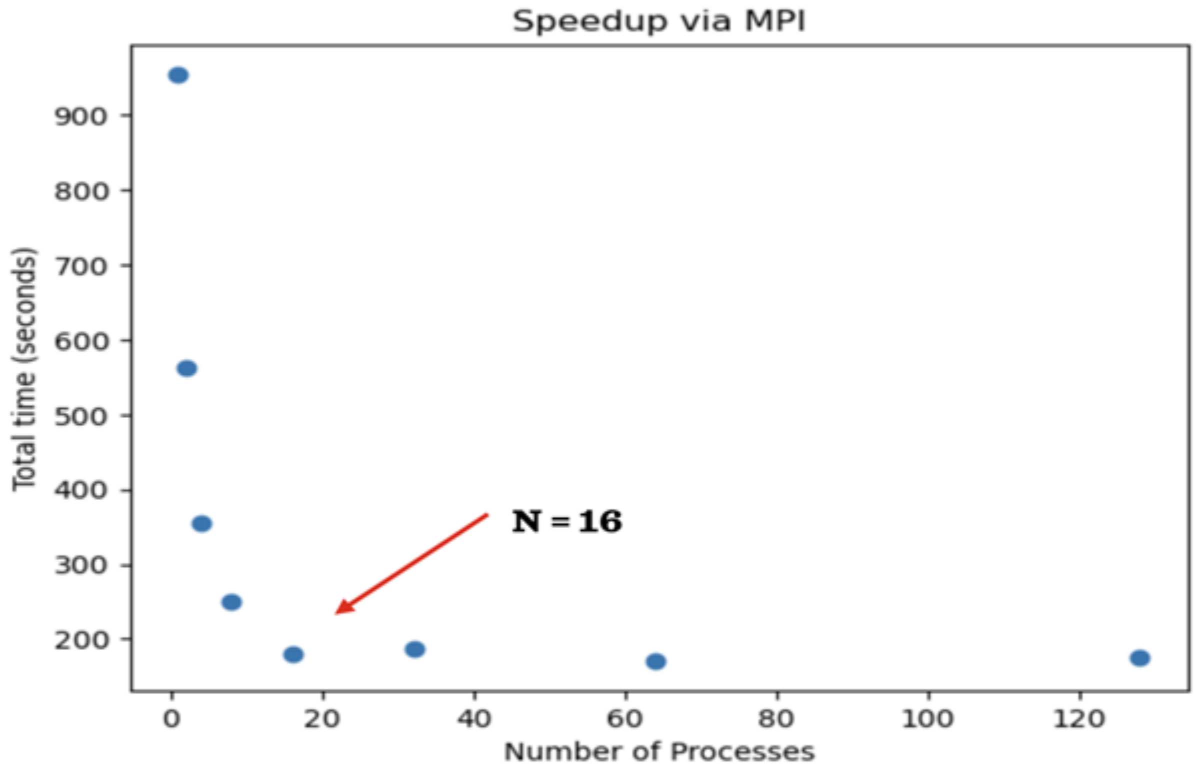
Alabama	1051	01	6	32.597	-86.149	11.15	-0.824	0.567	2014-11-05 13:30:00	2014-11-06 15:45:00	26.25	1096

- The Message Passing Interface (**MPI**) was used to process 6 GB of power outage data recorded at 15-minute intervals for events in every county in the US from 2014 to 2022 to a 16 MB file that contained more insight about how the outages were grouped and located.
- MPI enabled multiple processors of Perlmutter's computing nodes to communicate and collaborate.
- This allowed the dataset to be **sorted and analysed in parallel** through efficient data sharing and coordination.

Parallelizing Cleaning with MPI



Speed Up with MPI



Tasks

- Use MPI, assigned multiple CPU cores to our program that parses data and finds values of interest.
- Members of our team tested values from N = 1, 2, 4, 8, 16, 32, 64, and 128.
- Execution time dropped sharply from 1 to 16, then asymptotes with minimal benefits.

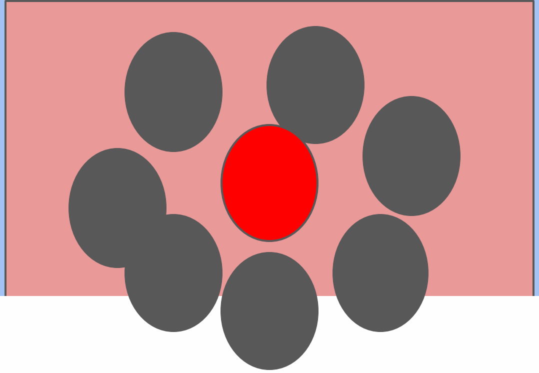
Findings

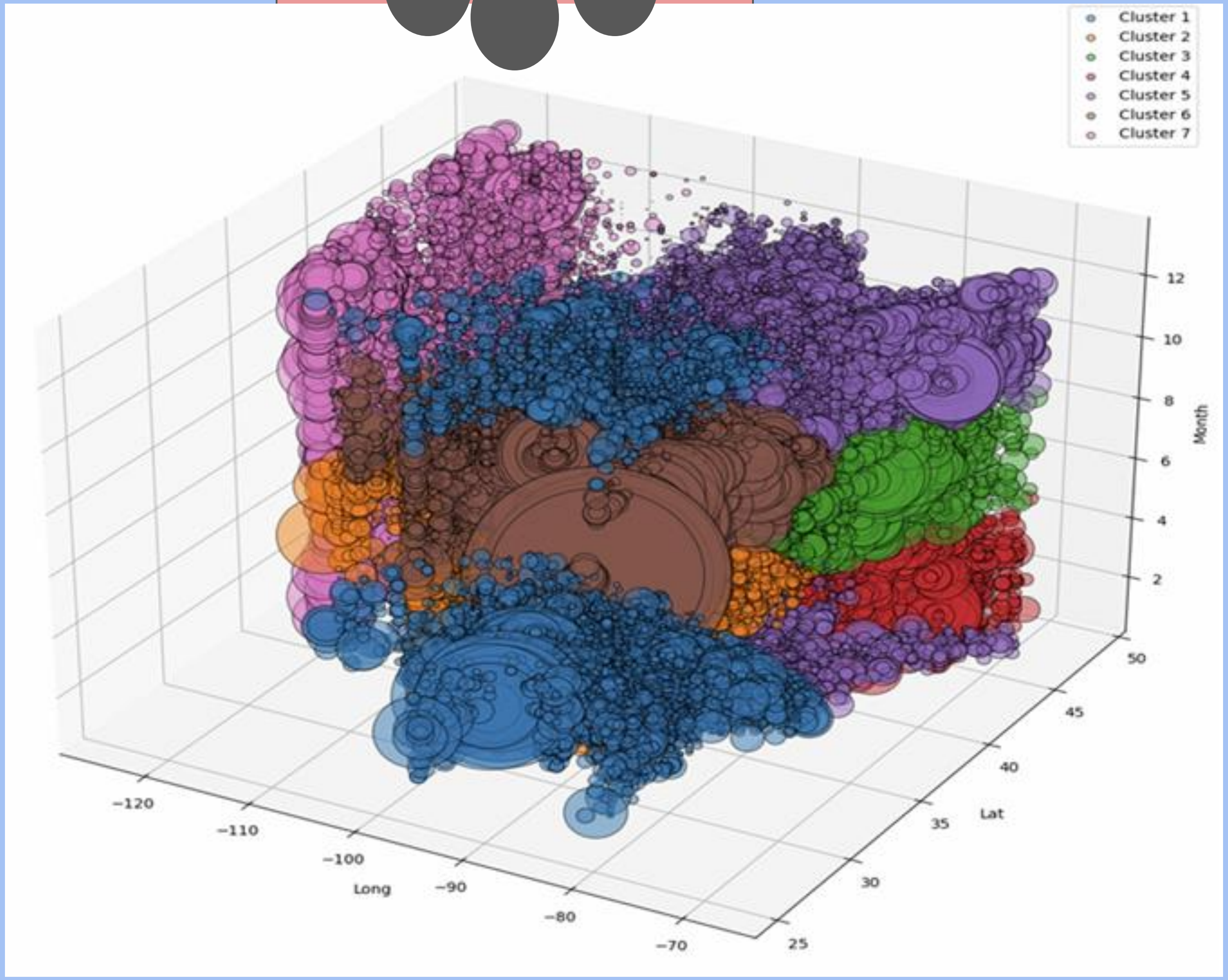
- 16 cores was optimal improvement. Beyond 16 cores, communication overhead between processes limited further speedup.
- Going back to the puzzle example at Argonne Lab, if a large number of us are around the table trying to reach and grab for pieces, at some point it will not help as much as a smaller number with more communication.

Anatomy of a K: *K-Means*

K- Means = *Unsupervised ML Algorithm*

K = ***Number of Clusters***





Centroid for Cluster 1 occurs during time of year: 1.711128004738062

Centroid for Cluster 2 occurs during time of year: 5.735178050639295

Centroid for Cluster 3 occurs during time of year: 7.405455893723389

Centroid for Cluster 4 occurs during time of year: 3.6990201263266322

Centroid for Cluster 5 occurs during time of year: 11.496448998781847

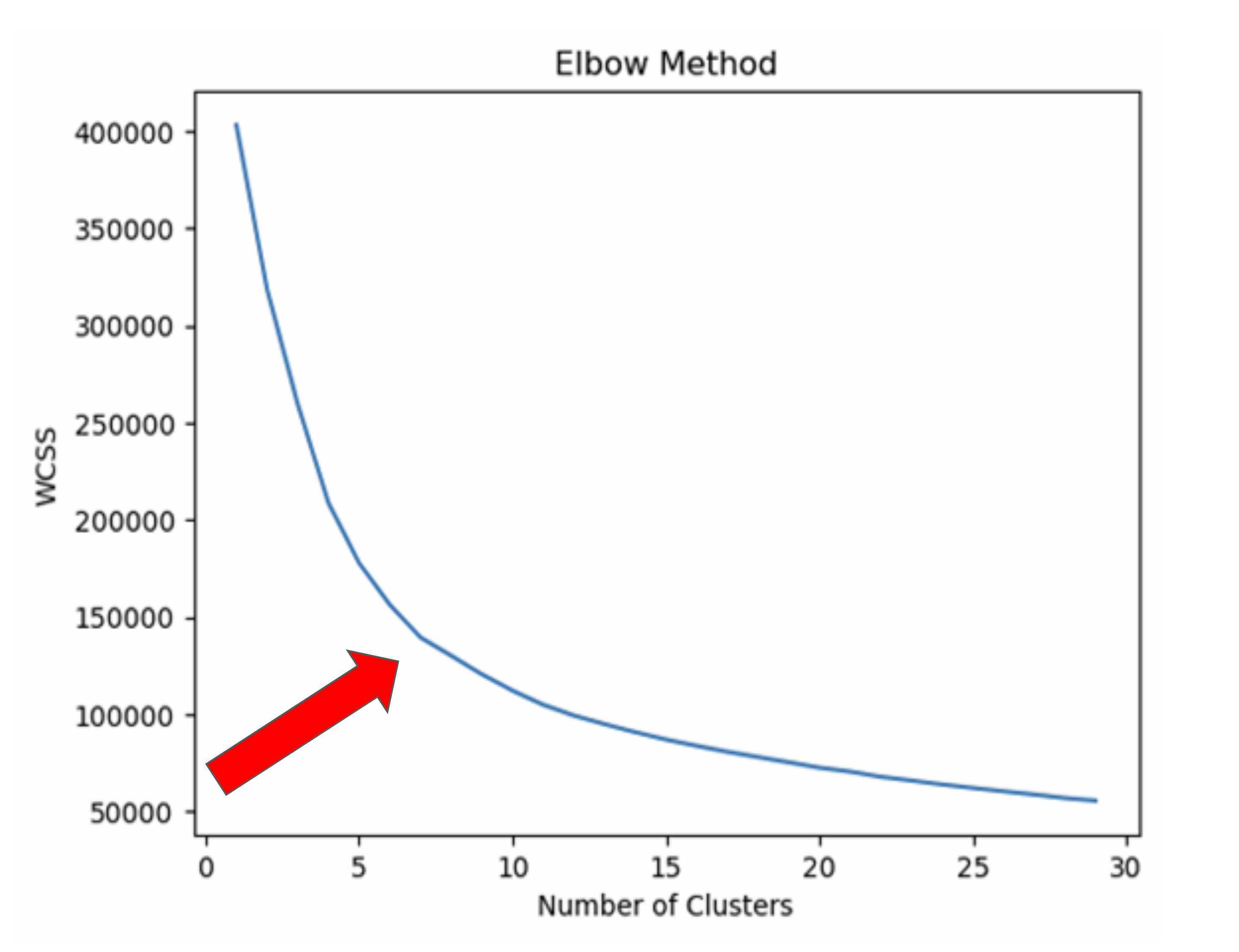
Centroid for Cluster 6 occurs during time of year: 9.457025596951343

Centroid for Cluster 7 occurs during time of year: 10.334768536531733

Anatomy of a K: *Optimization*

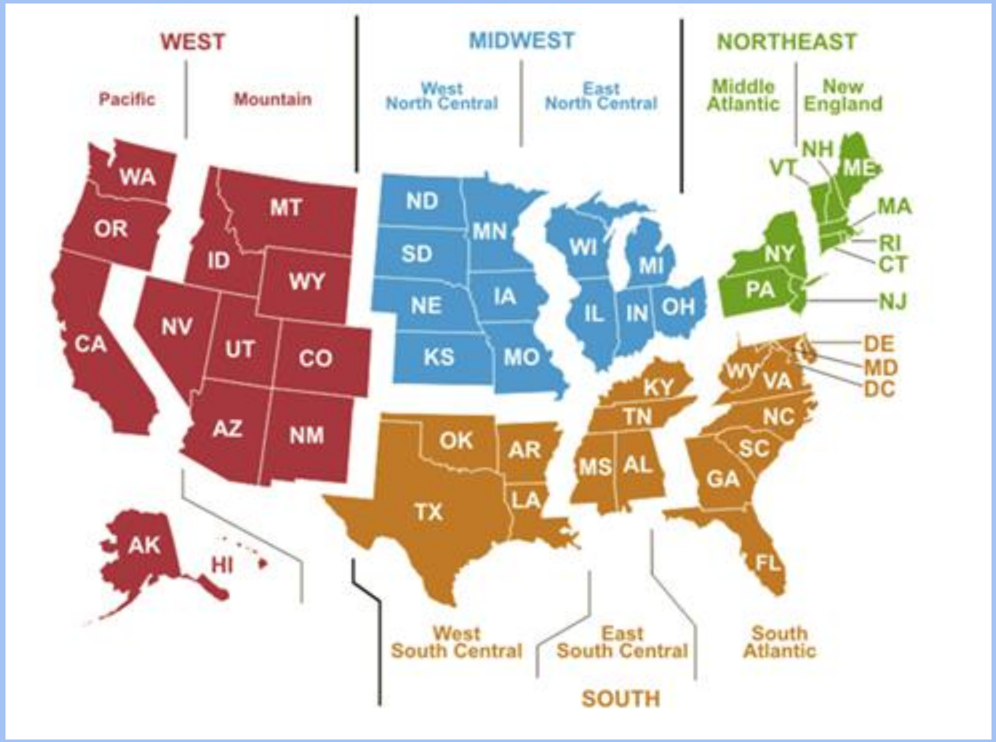
Methods / Procedure:

- ***Elbow***
- ***(WSSC) -*** *within-cluster sum of squares cost*
- ***Davies-Bouldin***
- ***Silhouette***
- ***Calinski-Harabasz***



Analysis Results (2016-2022)

Data shows that power outages have regional seasonality

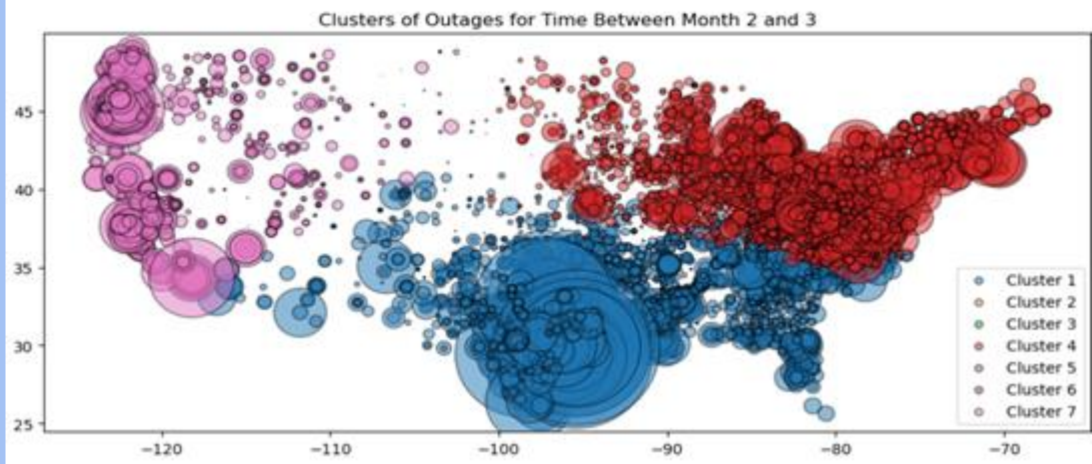


Larger power outage

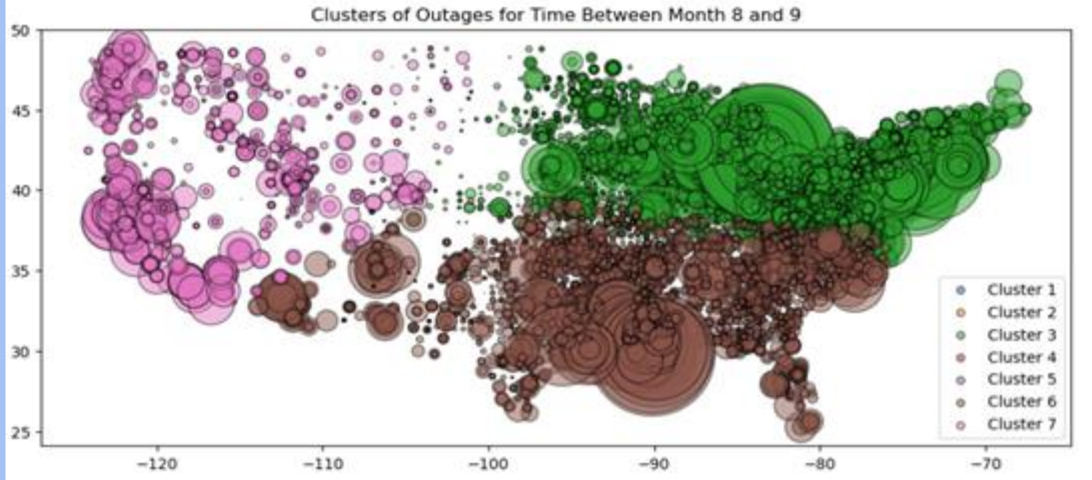
Smaller power outage

Cluster 1: Jan.
Cluster 2: May
Cluster 3: Jul.
CLuster 4: Mar.
Cluster 5: Nov.
Cluster 6: Sept.
Cluster 7: Oct.

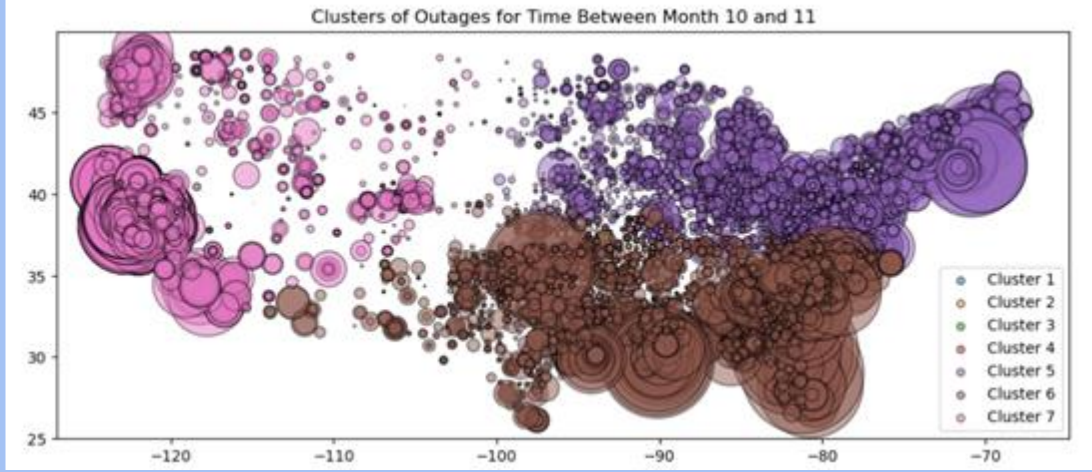
Winter



Summer

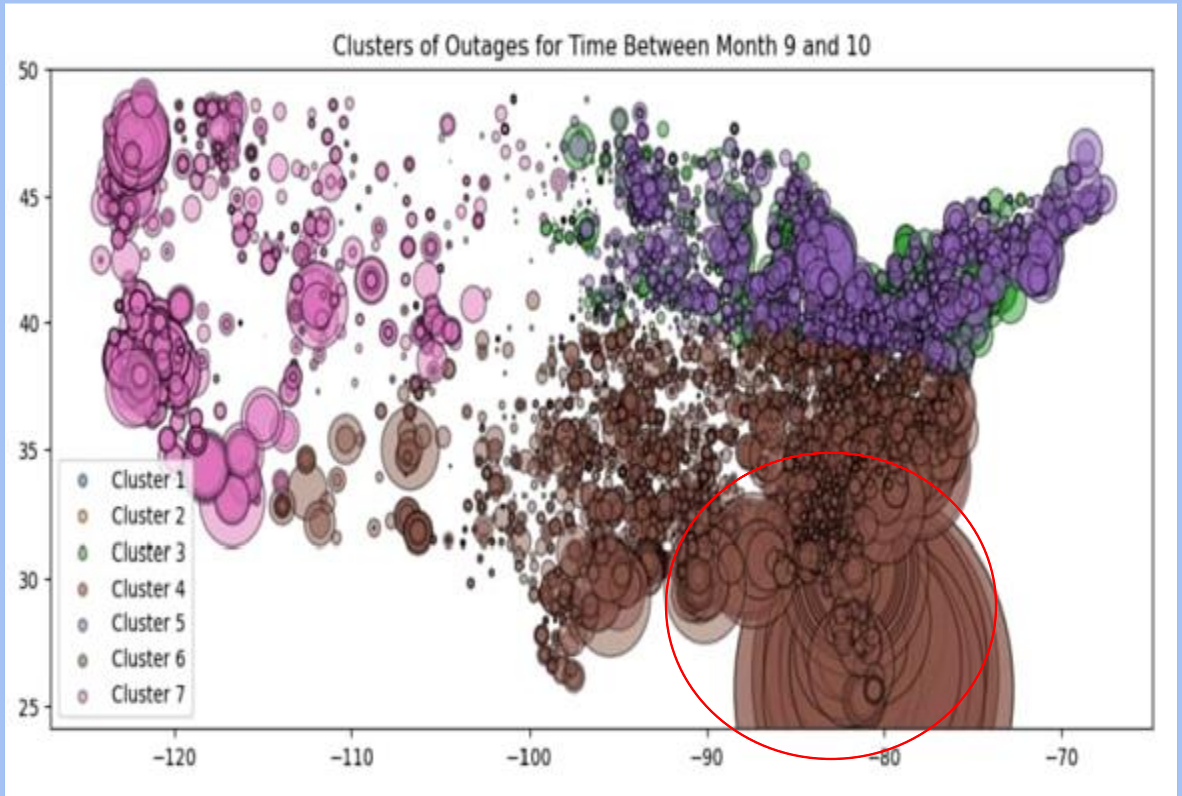


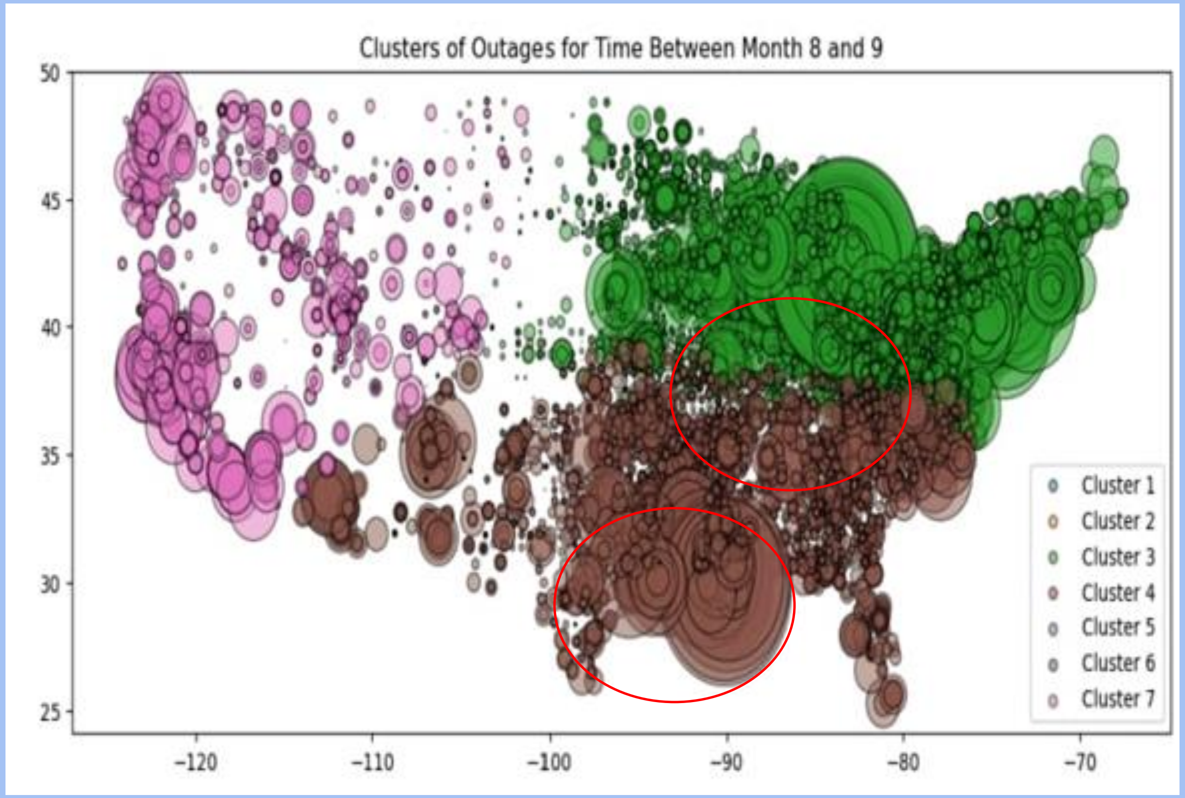
Fall



Are there types of weather that impact certain regions?

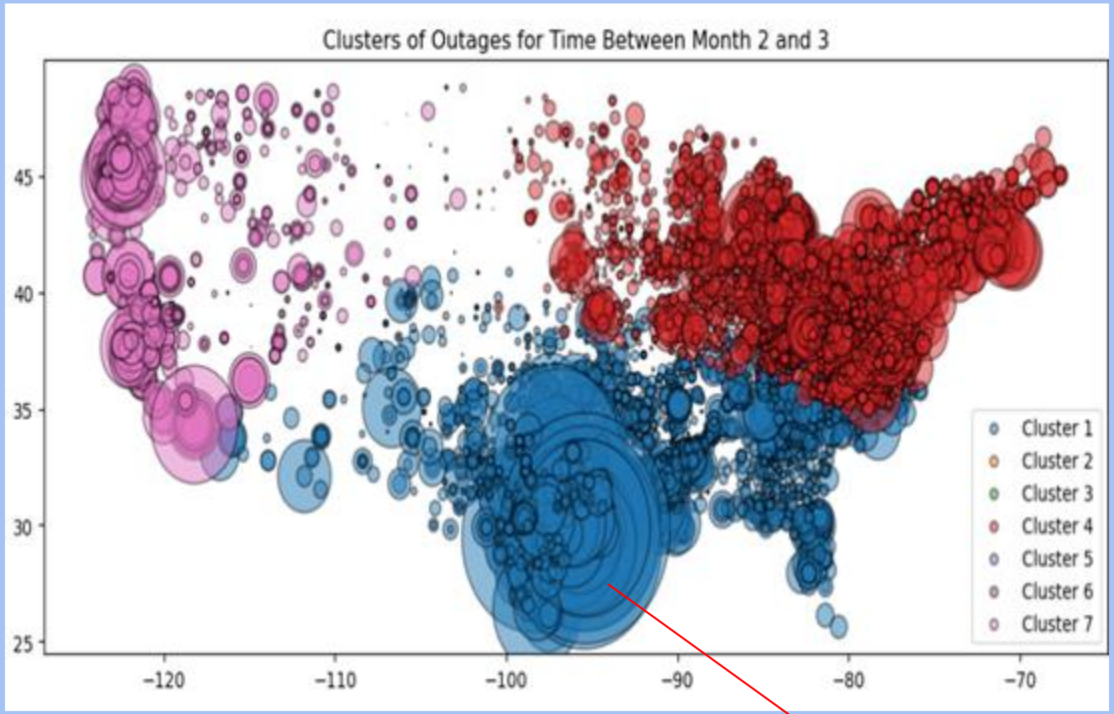
- **South Atlantic** - 2022 Hurricane Ian, 2018 Hurricane Michael
- **Midwest and South** - 2020 Derecho Events (Severe Thunderstorms & High Winds)

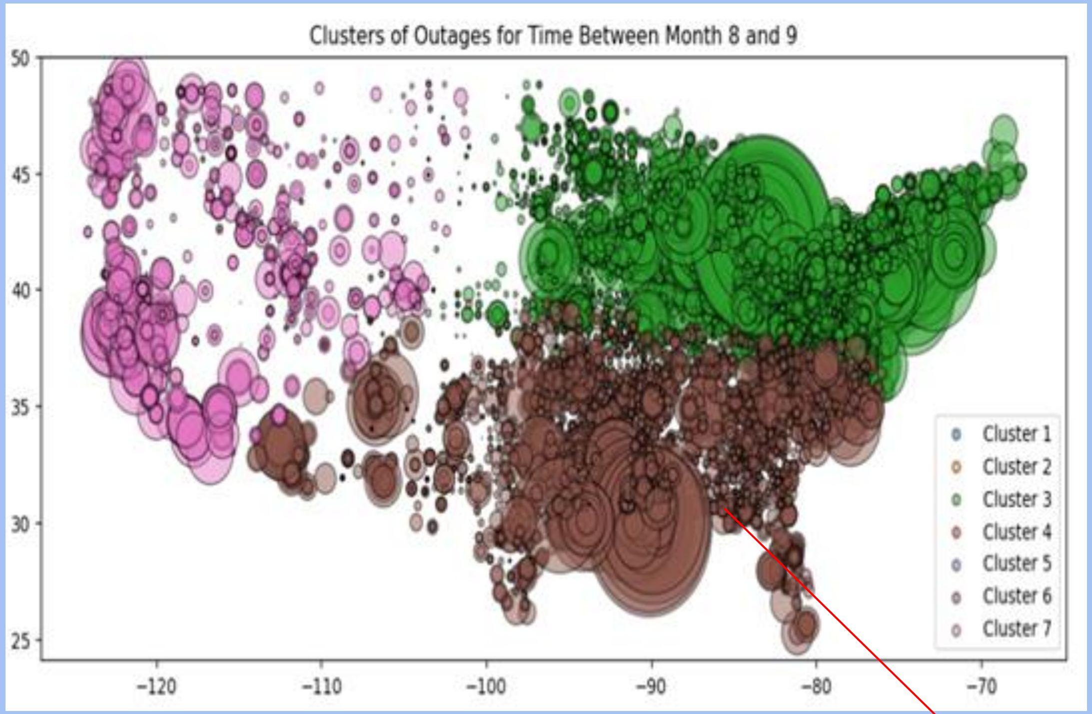




Are there types of weather that impact certain regions?

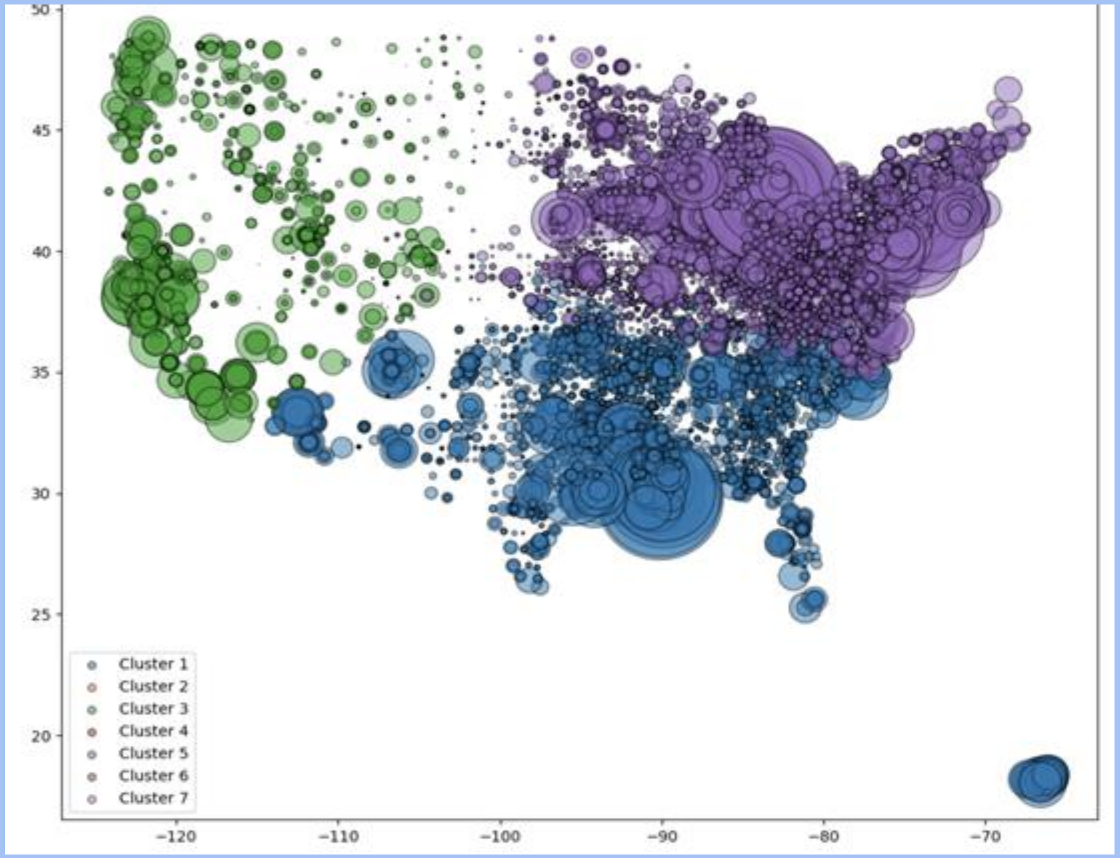
- **West South Central** - The 2021 Greater Texas Freeze
- **Hurricane Ida** - August 2021 | Louisiana and Mississippi
- **Hurricane Harvey** - August 2017 | Texas, Louisiana, Mississippi



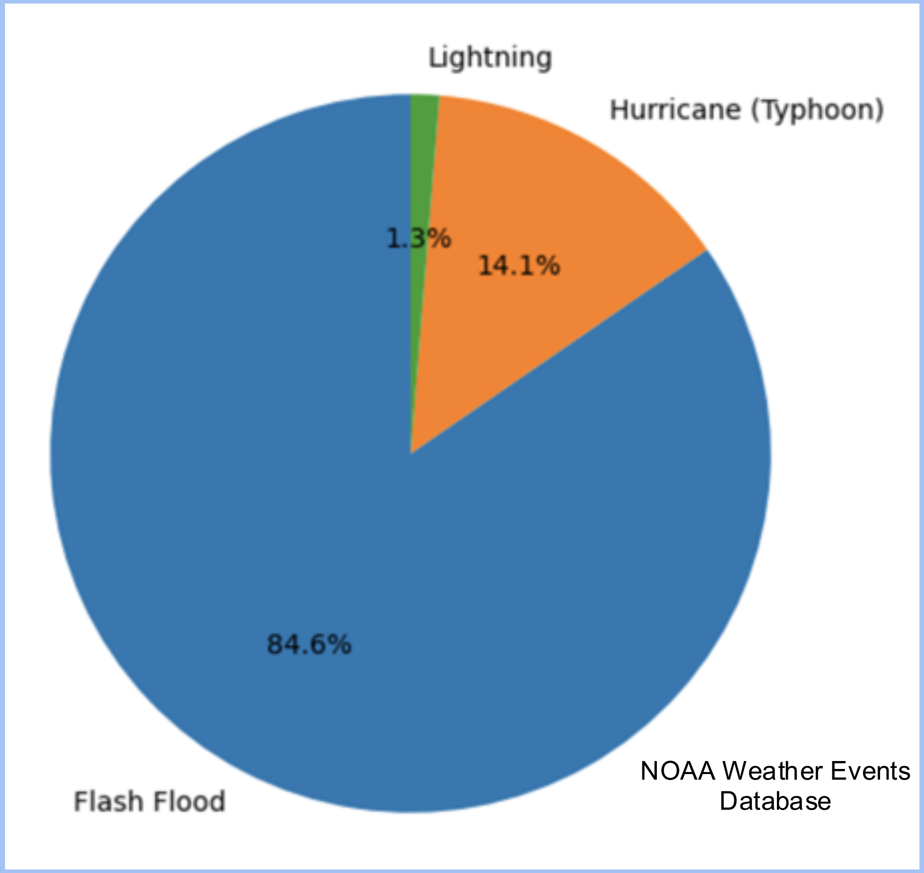


Are there types of weather that impact certain regions?

Clusters of Outages for Time Month 8 and 9

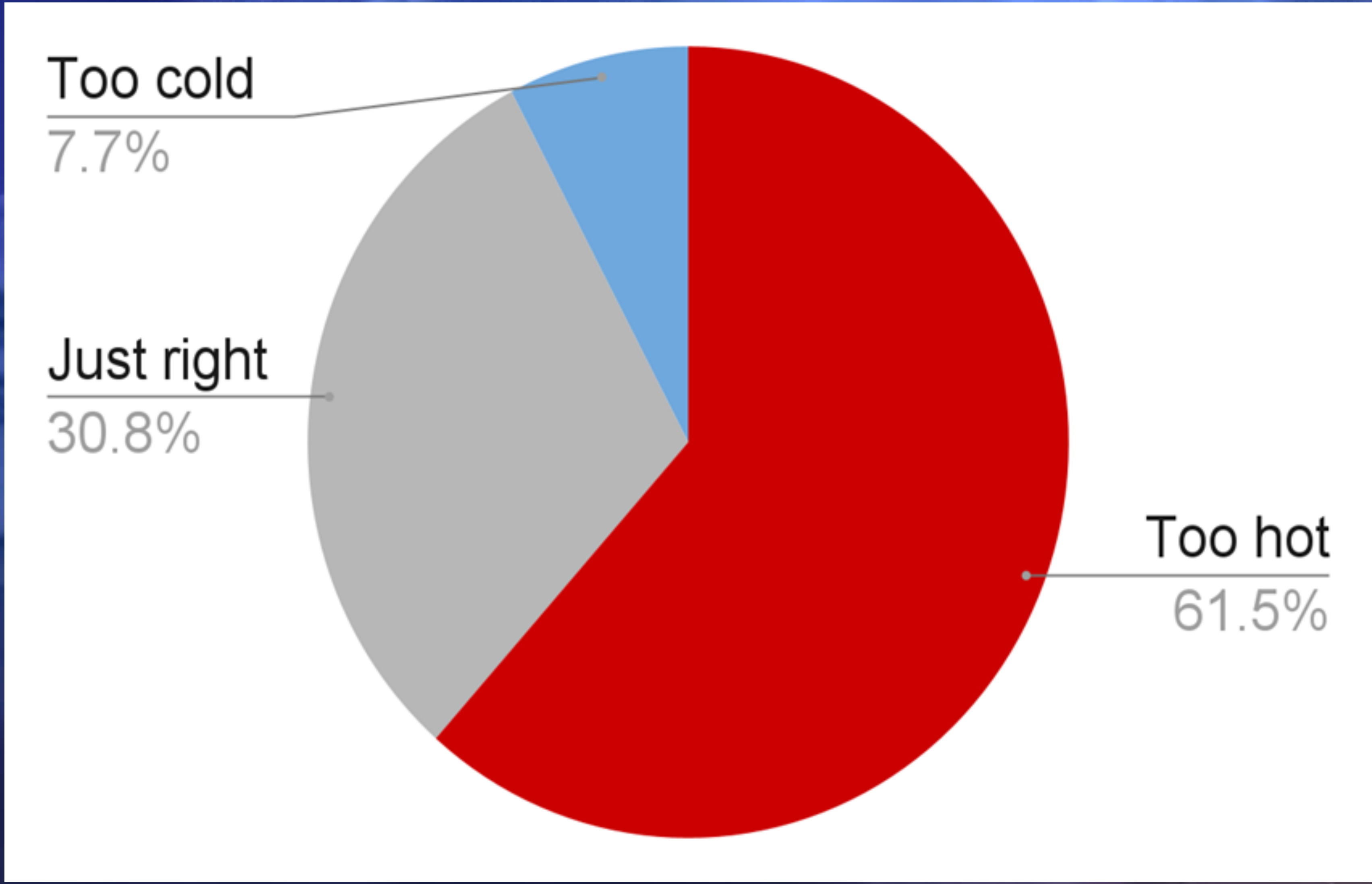


Fractions of Events for Month 9 in Puerto Rico



NOAA Weather Events Database

Challenge: Temperature

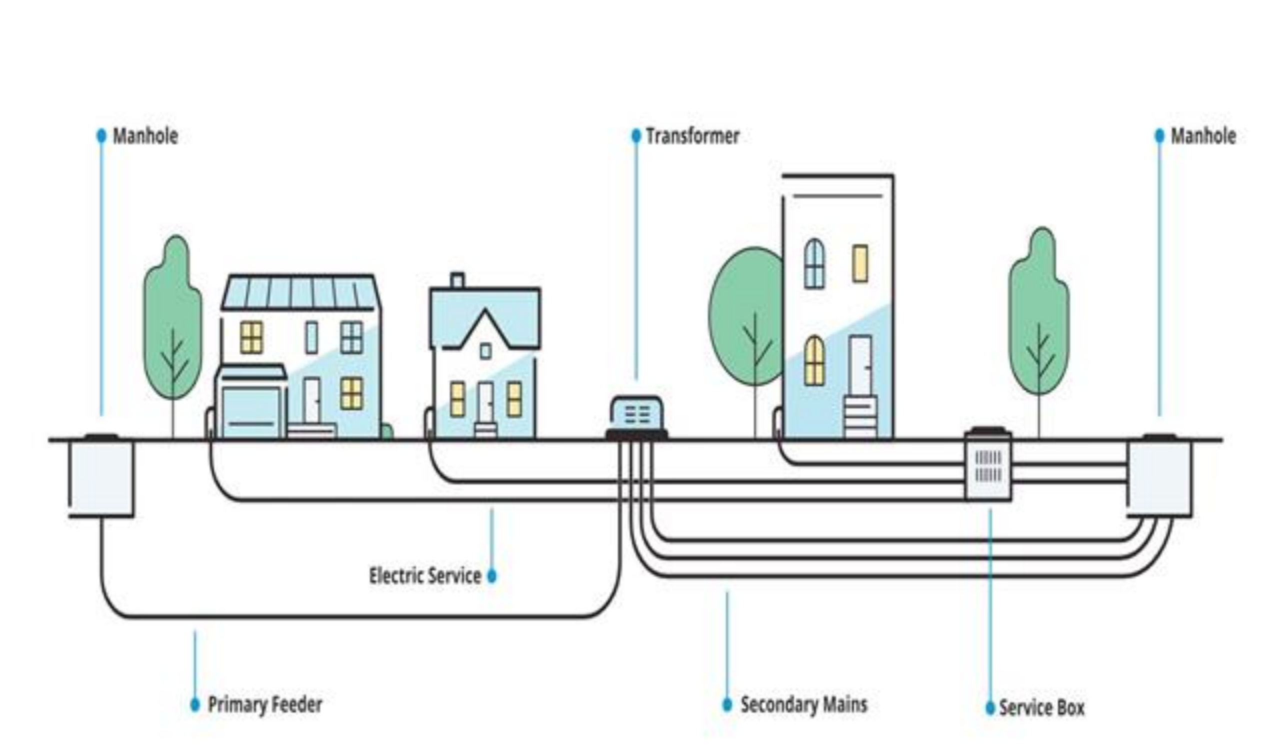


Supercomputer reservations

NERSC

Challenge: Varied skill sets

Conclusions: *Energy Security*



Underground power grid connection

Source: ***ABC7 NY 2018 News***

Clean energy alternatives

Source: ***UN Report 2025***

Acknowledgements

Thomas Fillers

Suzanne Parete-Koon

Pierre Harbin

Wiktoria Zielinska

OLCF, NOAA, EAGLE-I, ALCF

Perlmutter (NERSC)

Group 2a

Energy-Efficient HPC Design for TechVille

- The Best Team

Learn About Us

Samantha Dertouzos
The College of New Jersey

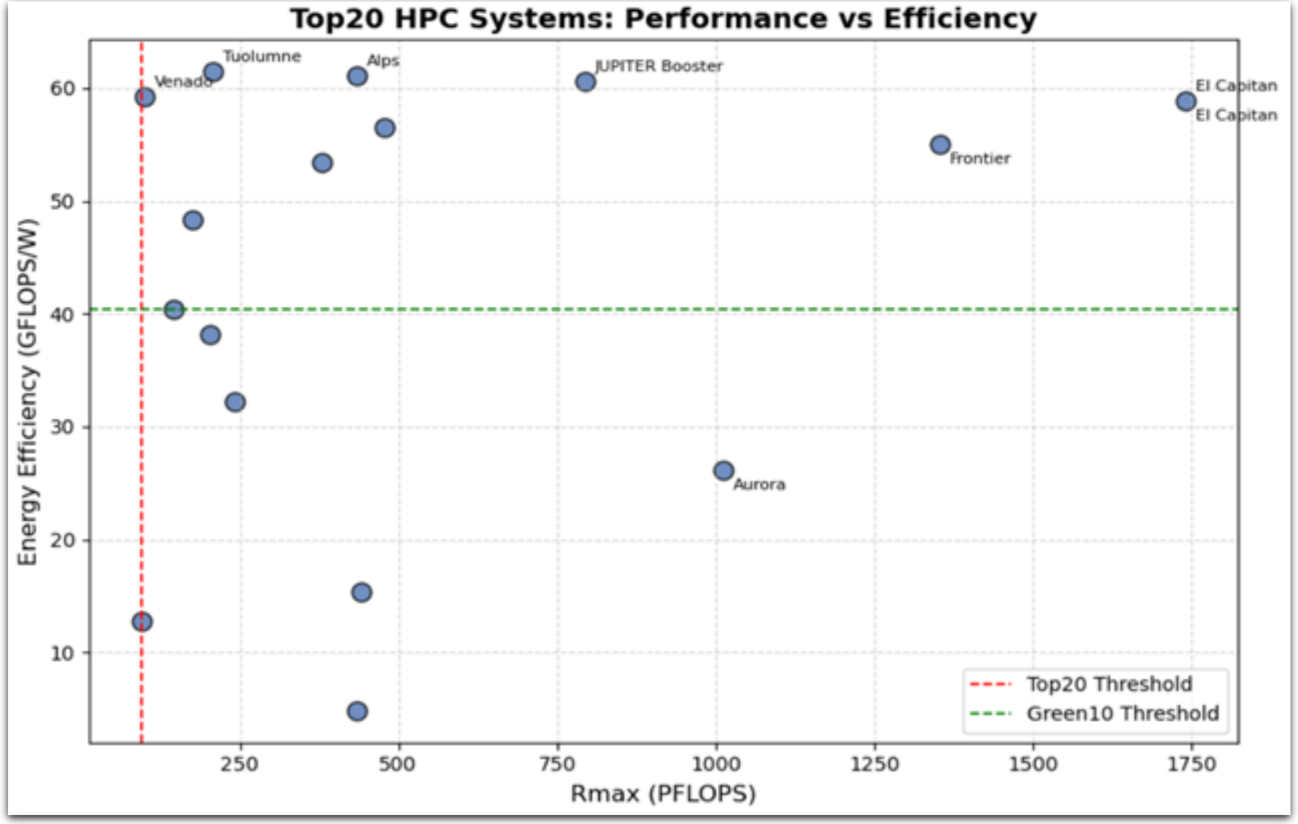
Kirtan Patel
University of Illinois Chicago

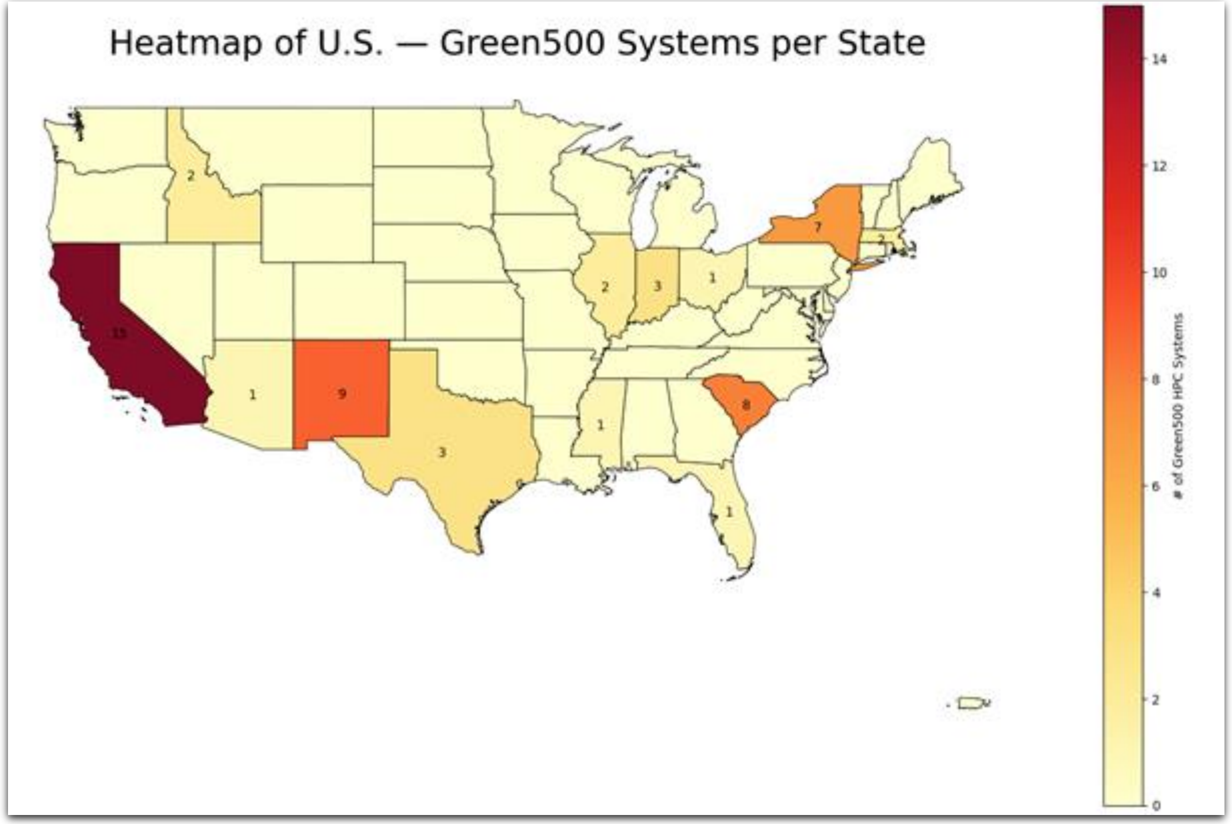
Mereum Fernando
University of Illinois Chicago

Laurence Boadilla
Harold Washington College

Juve Barajas
San Francisco State University

HPC Center Location





HPC Center Location

	Washington (Columbia Basin, TechVille HPC)	Tuolumne (LLNL, California)
Efficiency	~65 GFLOPS/W (Green10-level)	~62 GFLOPS/W (Green10-level)
Grid	Clean hydropower, very low carbon	higher carbon (natural gas heavy)
Cooling Climate	Cool & dry (efficient dry cooling)	Hot, humid summers (higher cooling load)
Land & Cost	Large, affordable	Limited, expensive Bay Area land

Machine Sizing & Specifications

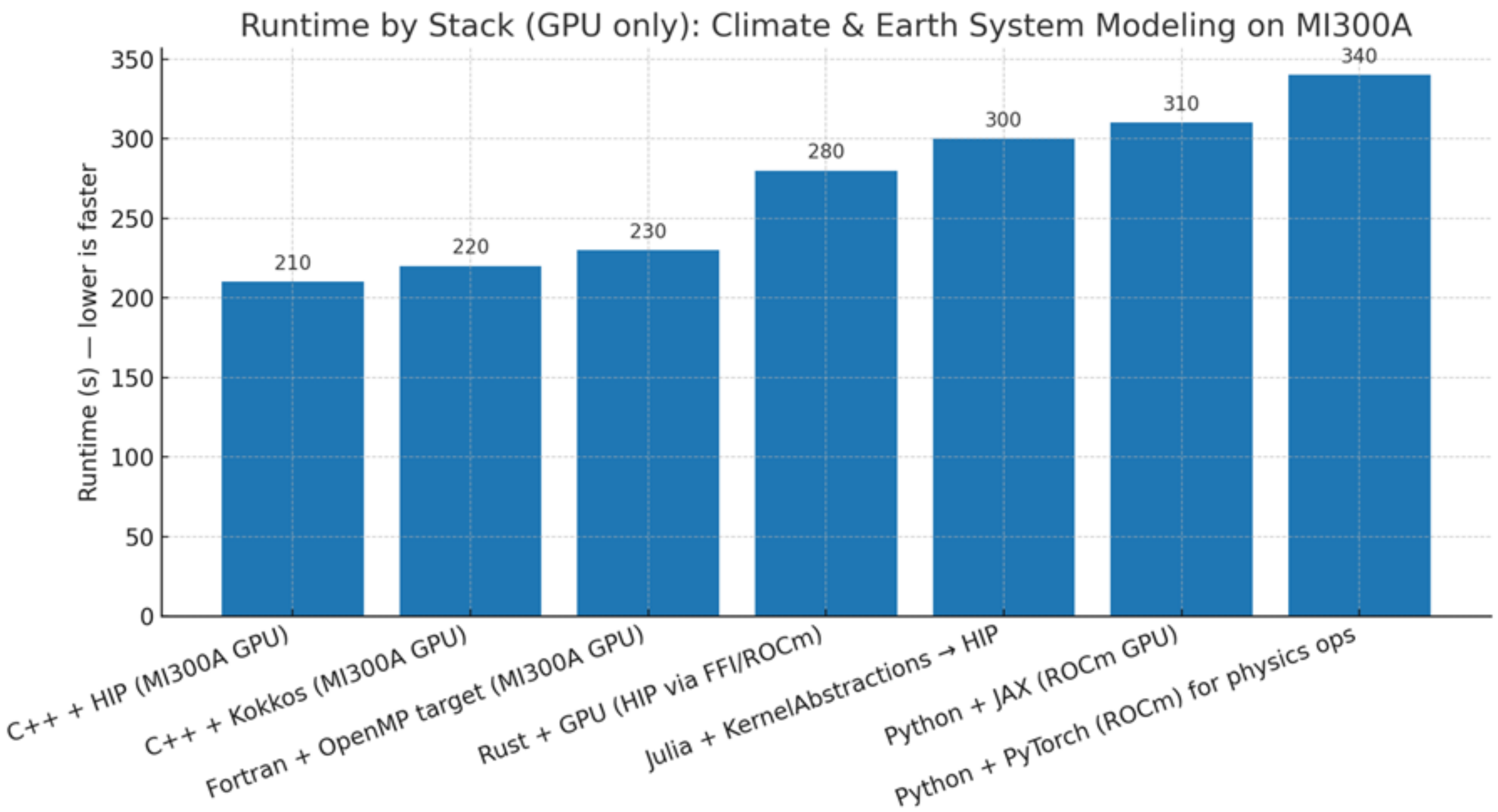
	Archetype	Nodes	Per-node PFLOPS	Total PFLOPS	IT Power (MW)	Facility Power (MW) @ PUE=1.2	Racks	Floor Area (m^2)	Cooling Flow (L/s)	Total Memory (TB)	Local NVMe (TB)	Shared Scratch FS (TB)	Interconnect	Switches	Meets Top20 Perf?	Meets Green10 Eff?
1	MI300A-like	190	0.50	95.00	1.45	1.74	6	6.7	34.6	114.0	608.0	95.0	Slingshot-11	6	True	True
2	GH200-like	211	0.45	94.95	1.45	1.74	7	7.8	34.6	106.0	675.0	106.0	Slingshot-11	7	True	True
3	MI350-like	158	0.60	94.80	1.45	1.74	5	5.6	34.6	101.0	506.0	79.0	Slingshot-11	5	True	True

TechVille HPC – Recommended Machaine Size & Specs

- Target ≥ 94.64 PFLOPS | Efficiency ≥ 65.40 GFLOPS/W
- Node archetype: MI300A-like | Interconnect: Slingshot-11
- Nodes: 190 | Racks: 6 | Switches (est): 6
- Memory:114 TB | Local NVMe: 608 TB | Scratch: 95 TB
- Power: IT 1.45 MW → Facility 1.74 MW (PUE=1.2)
- Cooling water flow (ΔT=10.0°C): 34.6 L/s
- Estimated floor area (incl. aisles/utilities): 6.7 m^2

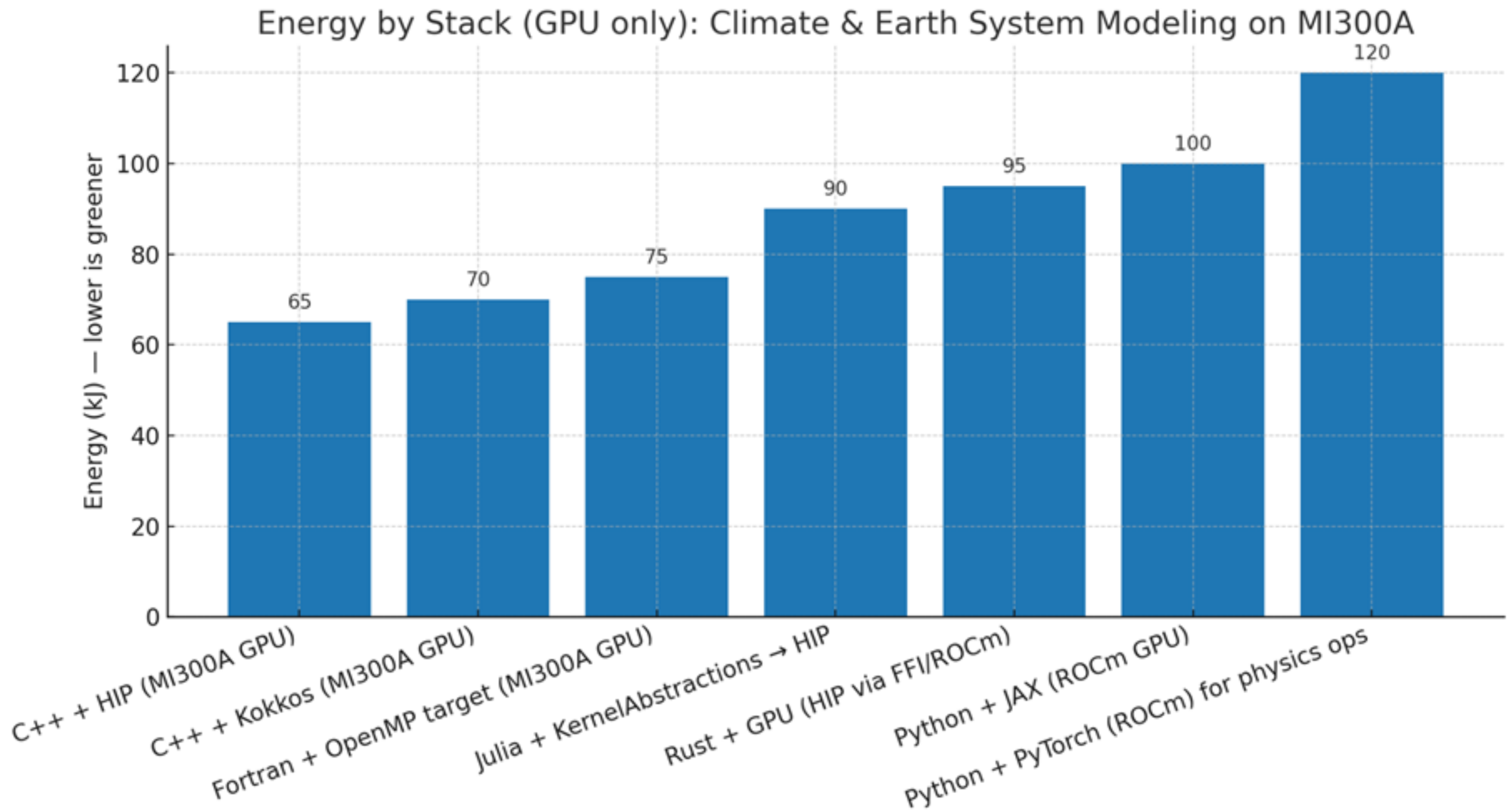
Optimal Language-Hardware Pairings For Green HPC

Fast & Clean Way To Communicate



Optimal Language-Hardware Pairings For Green HPC

Fast & Clean Way To Communicate



Vendor & Technology Analysis

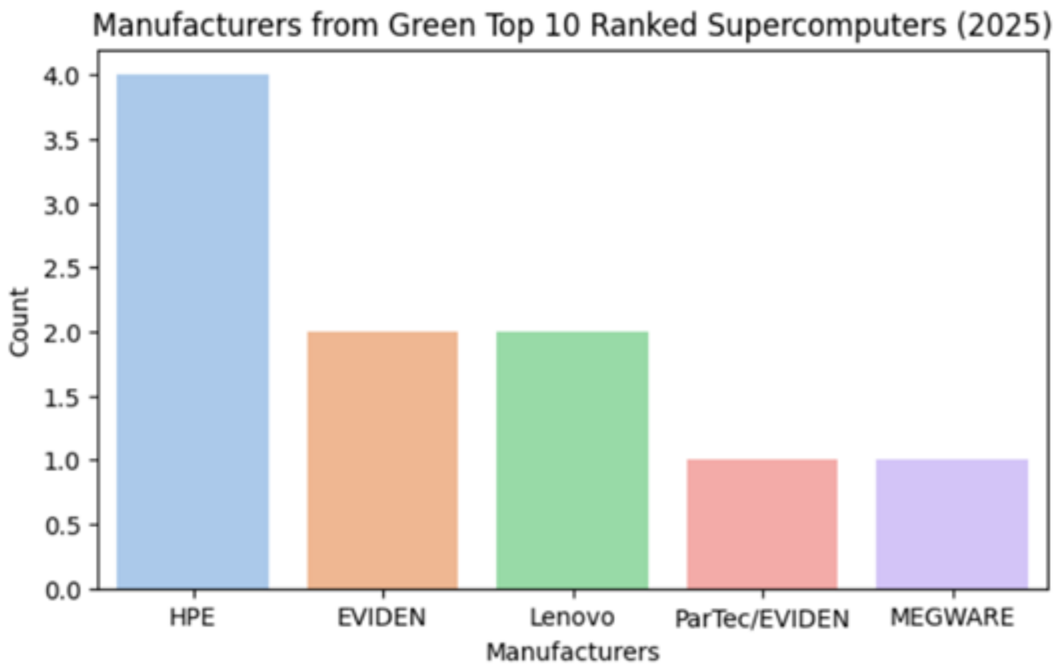
TechVille HPC Technologies:

Hardware

- AMD Instinct MI300A: Integrated CPU/GPU accelerated processing unit
- NVIDIA DGX™ B200

Facility Energy

- Liquid Cooling → lower Power Usage Effectiveness (PUE)
- Solar Panels, Windmills, Hydropower
- Lower CPU frequency for computationally intensive jobs
- Schedule computationally intensive jobs during off-peak hours



Vendors:

Manufacturer: HPE

Carbon-Emissions Tracking: Eviden (EcoDesignCloud)

Our Conclusions and Challenges

- Top 10 greenest and top 20 fastest
- Washington was chosen because of its many benefits

- Learning curve
- Data visualization

Acknowledgments




Argonne NATIONAL LABORATORY

BERKELEY LAB

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

NERSC Perlmutter

Group 1b

FUSION ENERGY: PERFORMANCE & ENERGY TRADE-OFF ANALYSIS

CREATIVE

2025

2025 HPC BOOTCAMP

PROJECT 1B: MASIA WISDOM, KEIOKO HAYNES, MARIA JIMINEZ GUILLERMO, SUNSHINE PANG, KHUAJA SEDIQI

MEMBER INTRODUCTION

KEIOKO HAYNES

UNIVERSITY OF ILLINOIS CHICAGO

MARIA JIMENEZ GUILLERMO

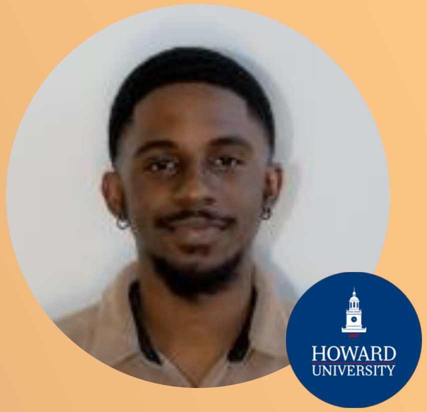
ELMHURST UNIVERSITY

SUNSHINE PANG

UNIVERSITY OF ILLINOIS CHICAGO

KHUAJA SEDIQI

UNIVERSITY OF ILLINOIS - URBANA CHAMPAIGN



MASIA WISDOM

HOWARD UNIVERSITY

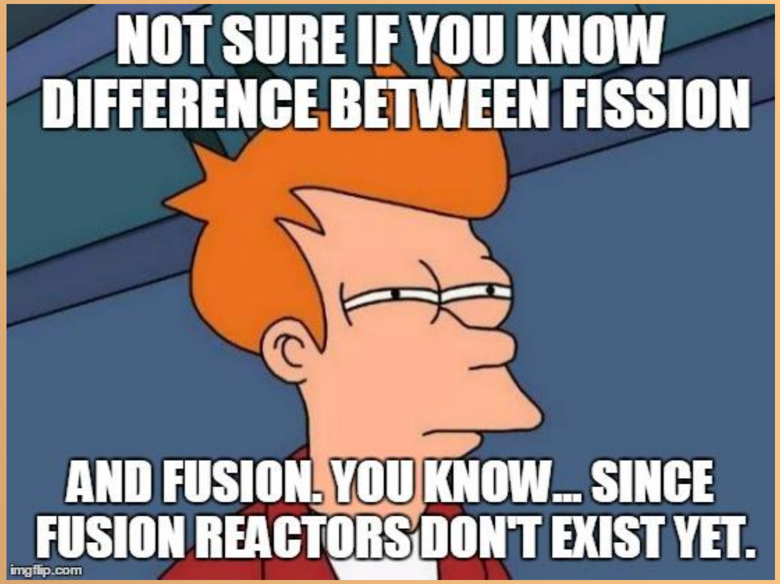
PROJECT INTRODUCTION

1. **Background**:
    a. Fusion energy offers a clean and limitless alternative to fossil fuels, but current supercomputers are not optimized for fusion-specific workloads.
    b. Using simulations rather than experiments allows researchers to study plasma behavior and reactor conditions more safely, cost-effectively, and at a scale that would be difficult to achieve in live testing.
2. **Goal**:
    a. Design a next-generation supercomputer for efficient fusion energy research.

FUSION WORKLOAD TYPE ANALYSIS

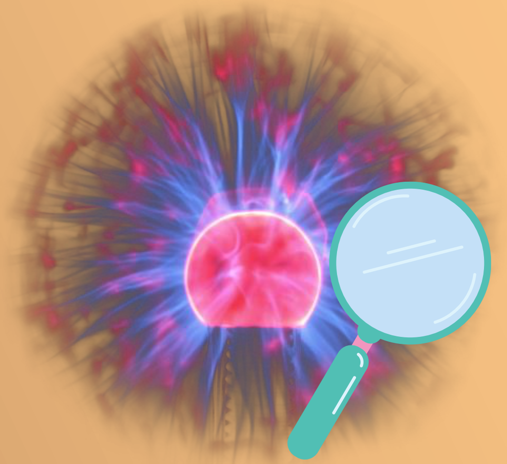
3 Types of workloads were analyzed…

1. Large Scale Plasma Simulations
2. Diagnostics Data Analysis
3. ML-Based Prediction Model

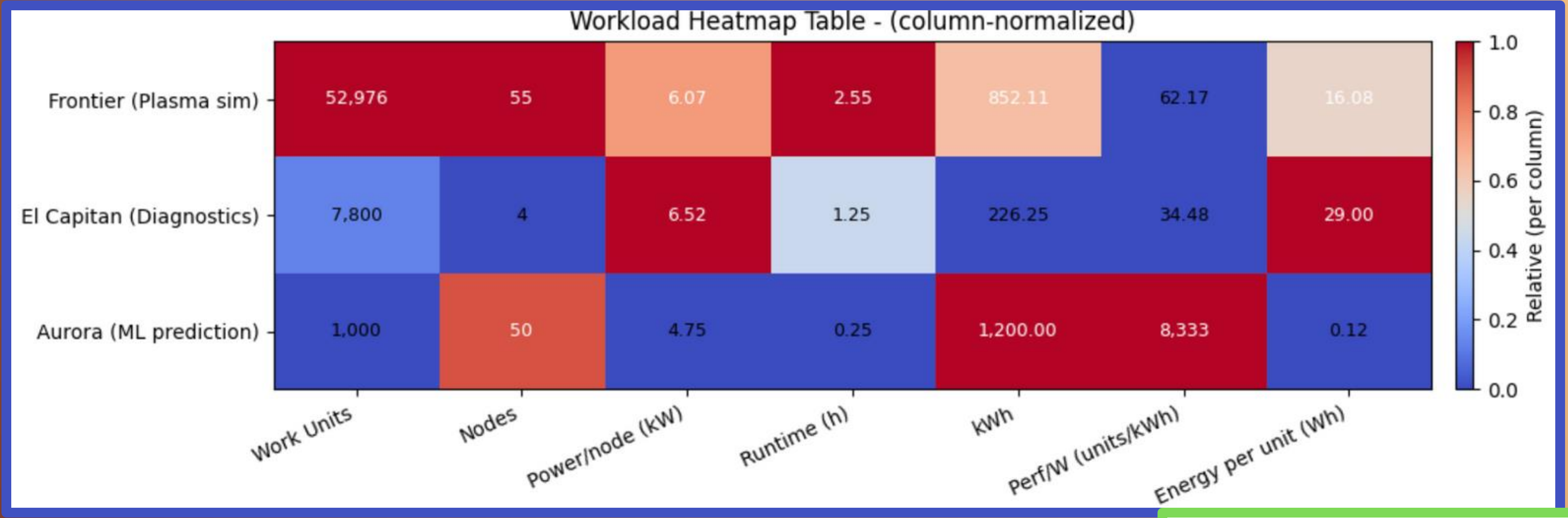


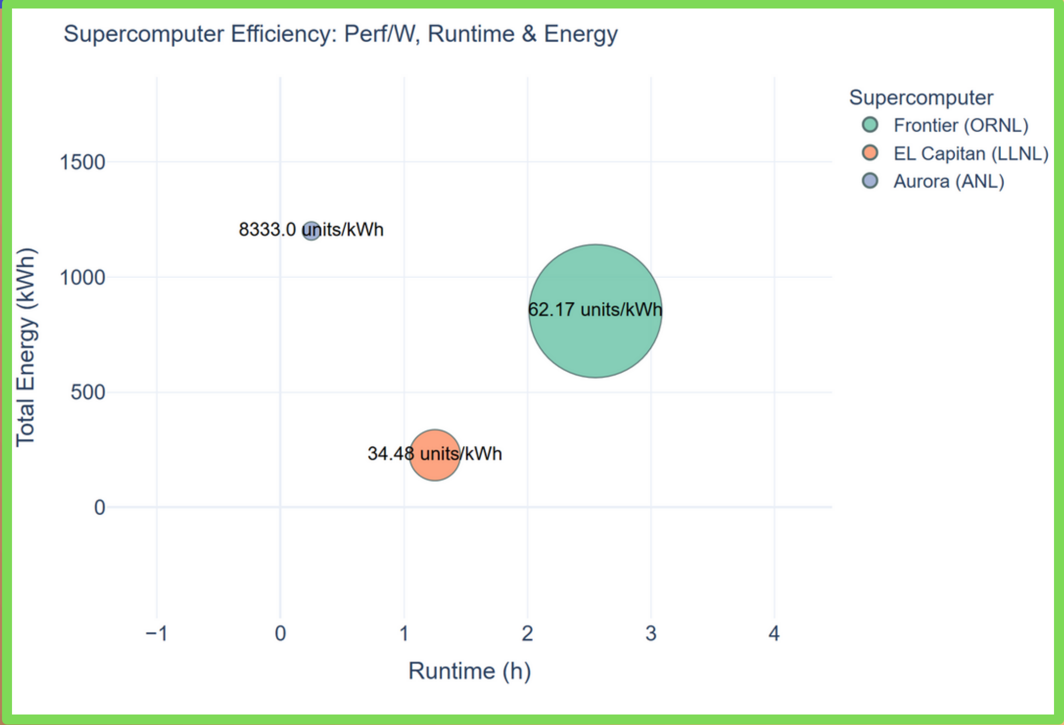
WE COMPARED…

- ARCHITECTURE - **CPU, GPU, Quantum, etc.**
- Power Distribution - **Peak vs. Idle Usage**
- Runtime - **Total Execution Time**
- Efficiency - **Energy Per Job, Scaling**
- Node Footprint - **Physical Space**



WORKLOAD TYPE ANALYSIS

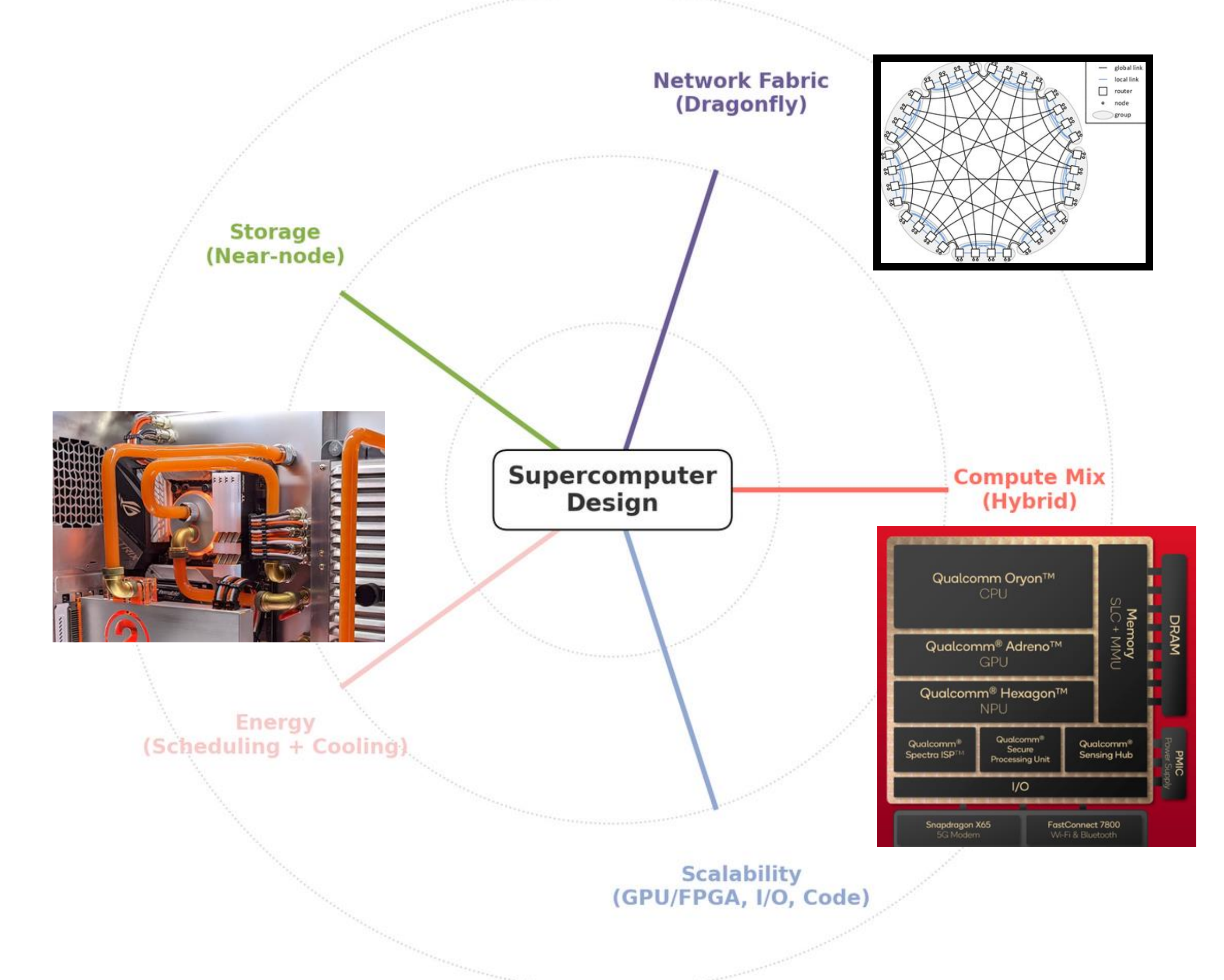




KEY TAKEAWAYS

- El Capitan: Short runtime, most efficient
- Frontier: Longest sustained runtime >> workload
- Aurora: Powerful but energy-hungry

INFLUENCE ON SUPERCOMPUTER DESIGN



Compute Mix (Hybrid)

- Includes CPUs/GPUs for orchestration & control inspired by Qualcomm.

Storage (Near-Node)

- Near-node storage boosts speed by reducing latency

Energy Optimization

- Active Water Cooling
- Smart Scheduling Algorithms reduce energy usage during idle

Network Fabric (Dragonfly)

- Chosen for its low latency, high bandwidth, and scalability

WHAT NEXT?

Location and Cooling

- Perlmutter is easier to cool compared to Aurora due to the stable temperature of its location
- Thousands of gallons of water is required to cool supercomputers

Server Flexibility

- Adapting and scaling to larger or smaller projects

Optimal GPU/CPU Ratios

- GPU/CPU Hybrid builds are more energy efficient for fusion
- Fusion code is expensive to run

THANK YOU

Group 4a

Project 4a: Short-Term Load Forecasting Using Machine Learning (ML)

Alex Gorczowski, Amy Kodama, Claudia Jimenez, and Fatima Rasheed

Team Members

Alex Gorczowski

Amy Kodama

Claudia Jimenez

Fatima Rasheed

Introduction

- Short Term Load Forecasting (STLF) predicts near term energy demand for buildings from hours to a month ahead
- Enables utilities and grid operators to balance supply and demand efficiently

Goals:

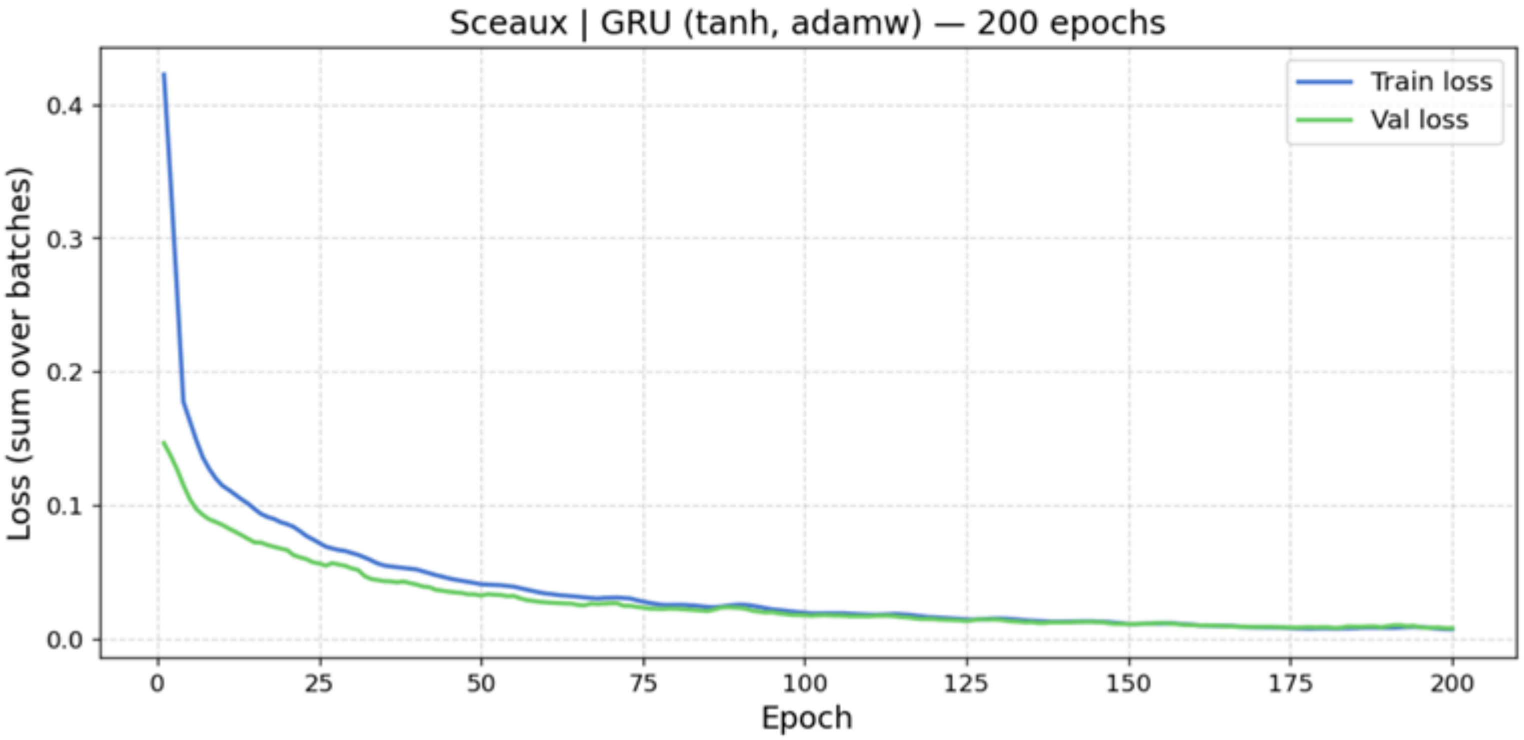
- Develop hands-on skills in STLF with deep learning
- Practice object-oriented programming, PyTorch, and Matplotlib in an applied setting
- Learn to train and tune models on GPUs

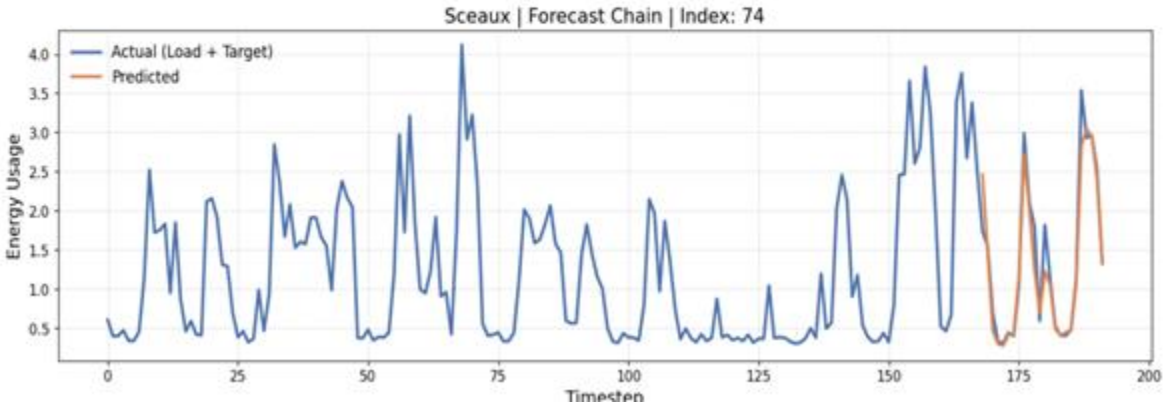
Sceaux Dataset

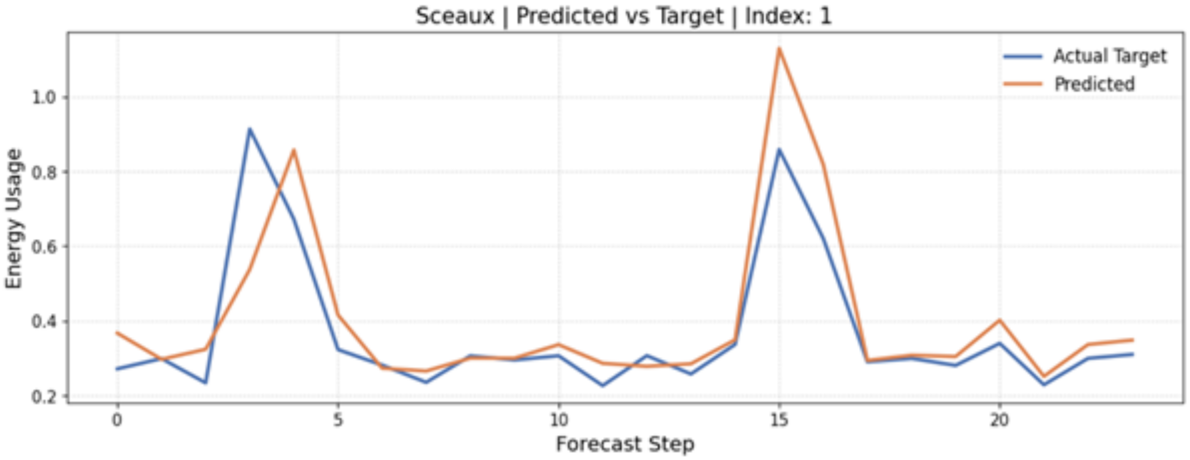
Electricity meter data from 1 home in Sceaux, Paris

Best Model: GRU-tanh-adamw-200 epochs

R^2 **= 0.934**





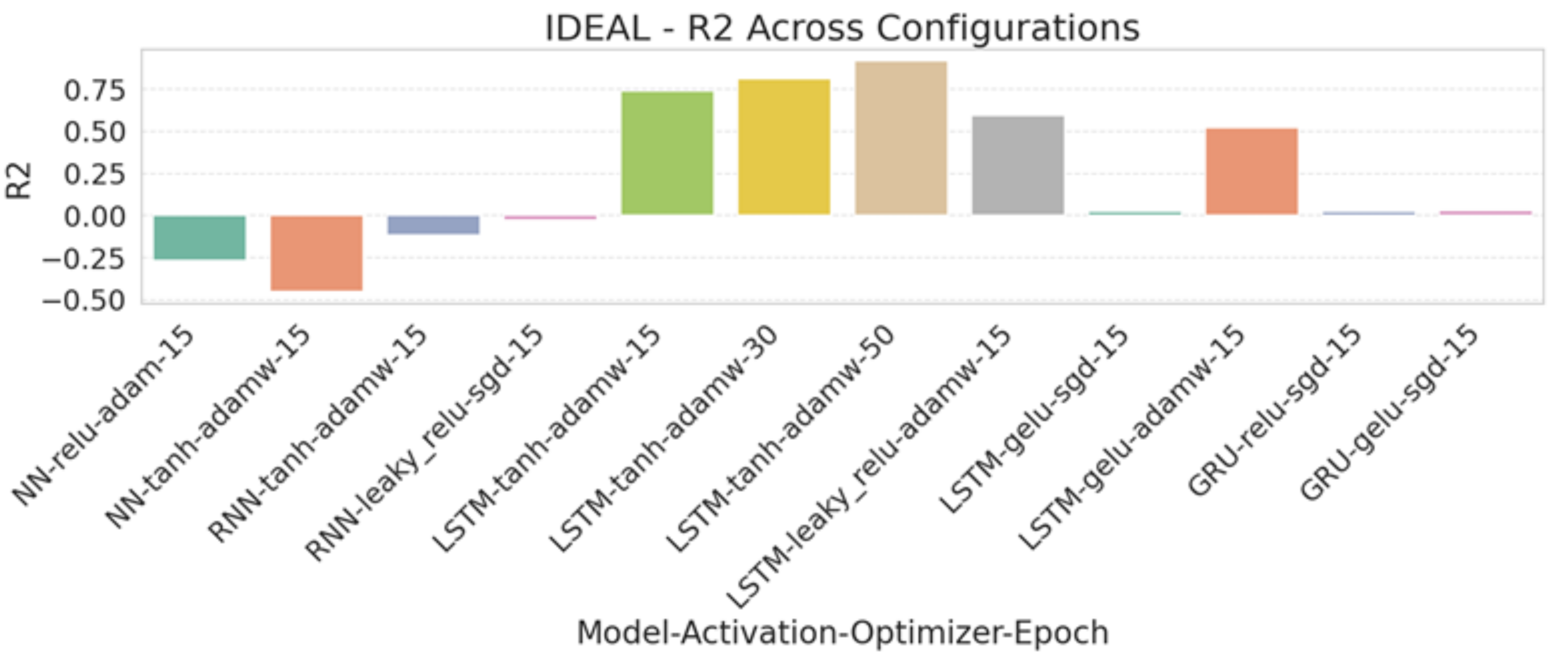


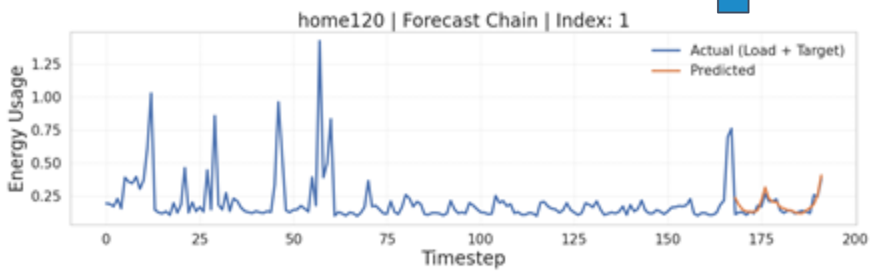
IDEAL Dataset

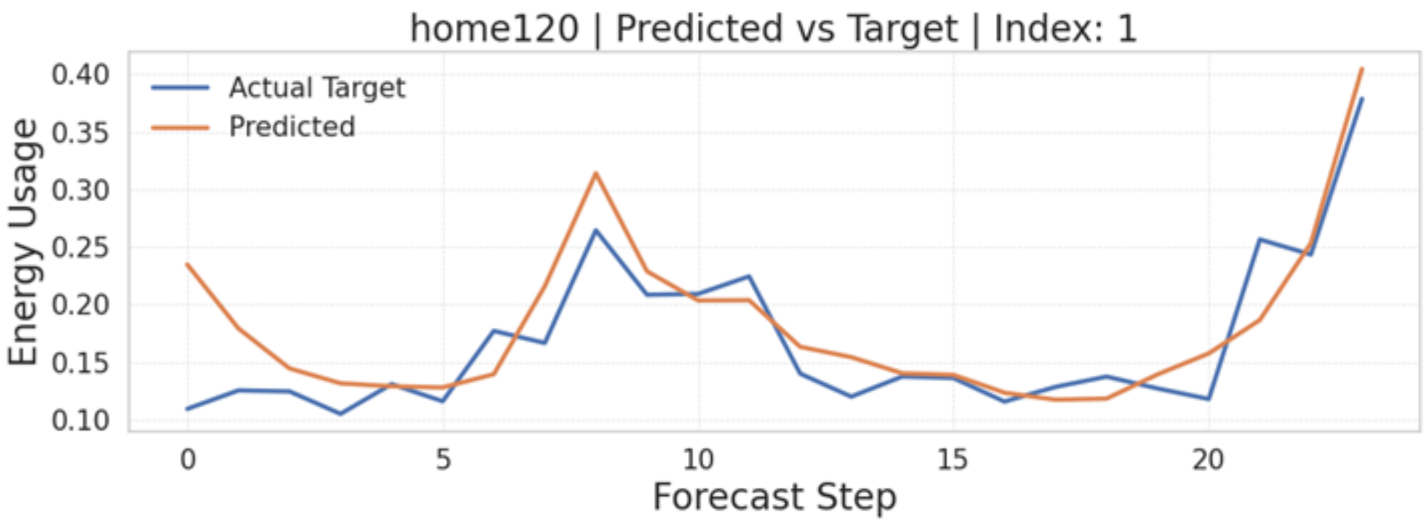
Electricity meter data from 255 homes in Edinburgh, UK

Best Model: LSTM-tanh-adamw

Epochs	R^2
15	0.7391
30	0.8109
50	0.9184





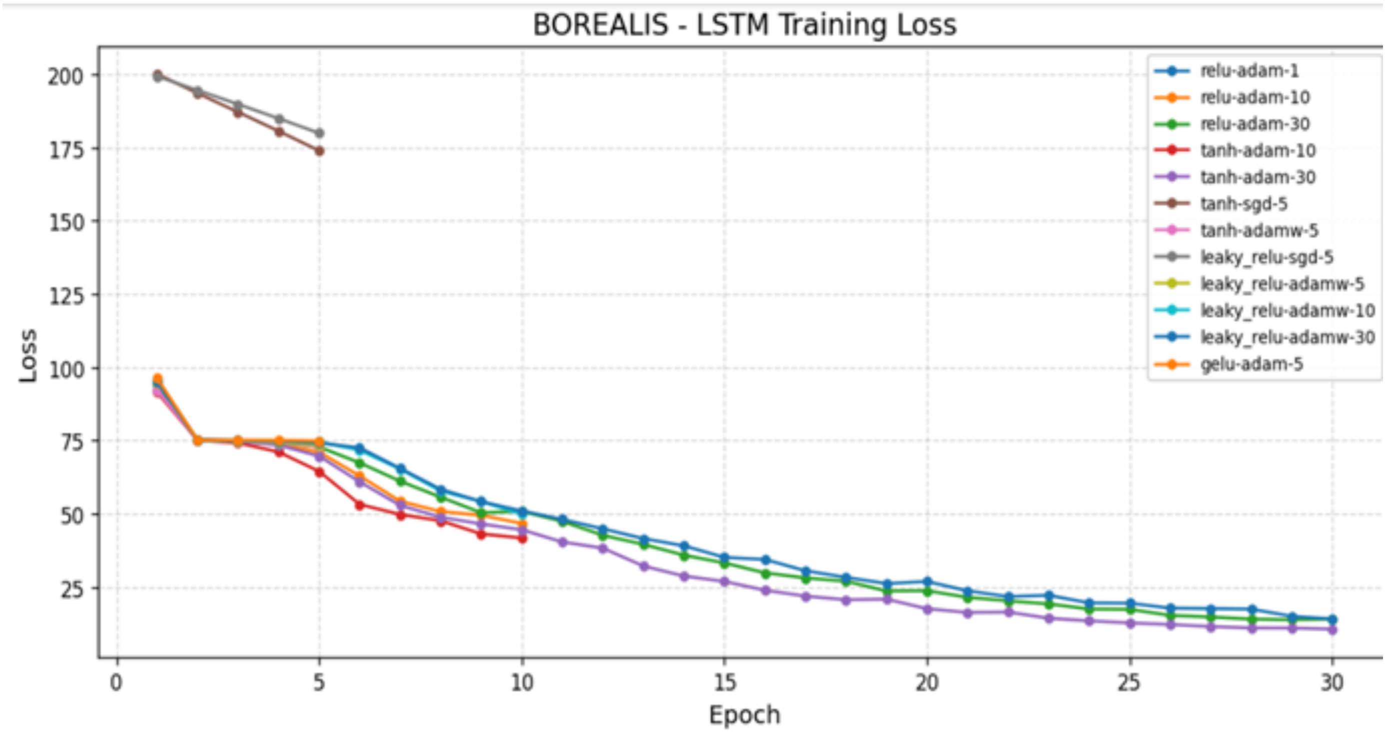


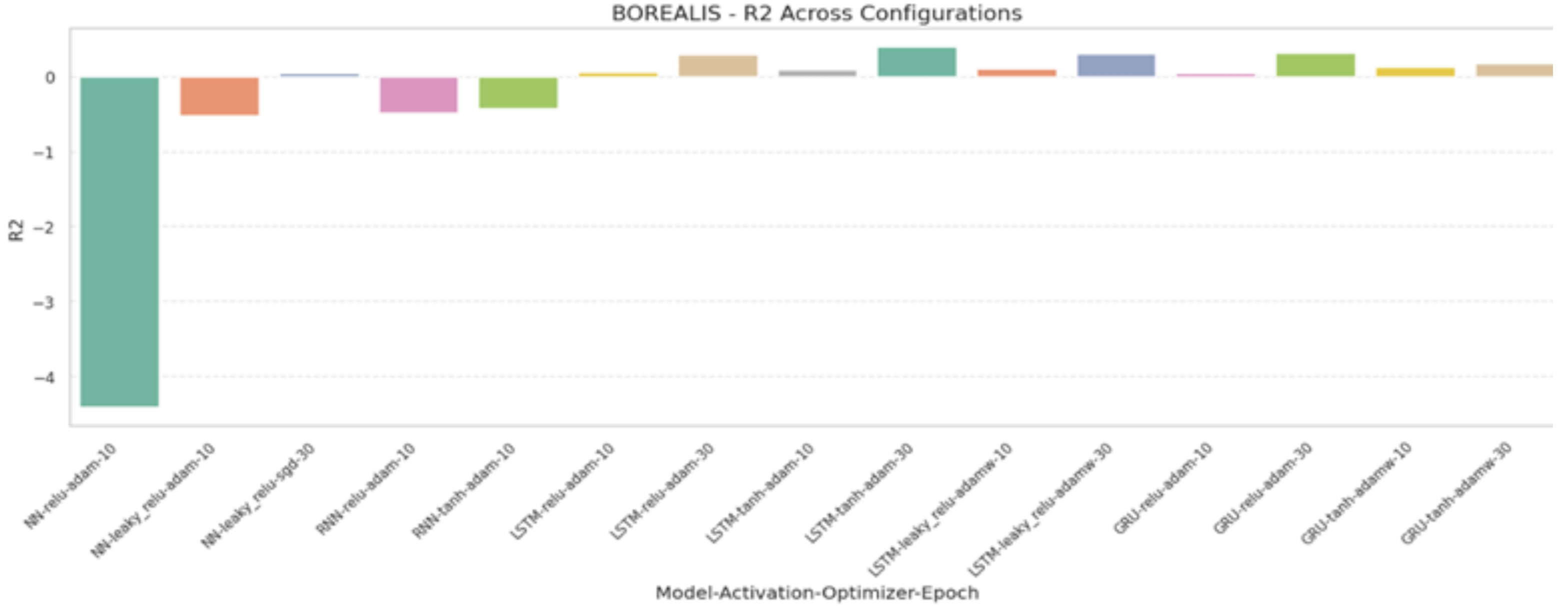
Borealis Dataset

6-second load measurements recorded for 30 homes in Waterloo, ON

Best Model: LSTM-tanh-adam

Epochs	R^2
10	0.0948
30	0.4004



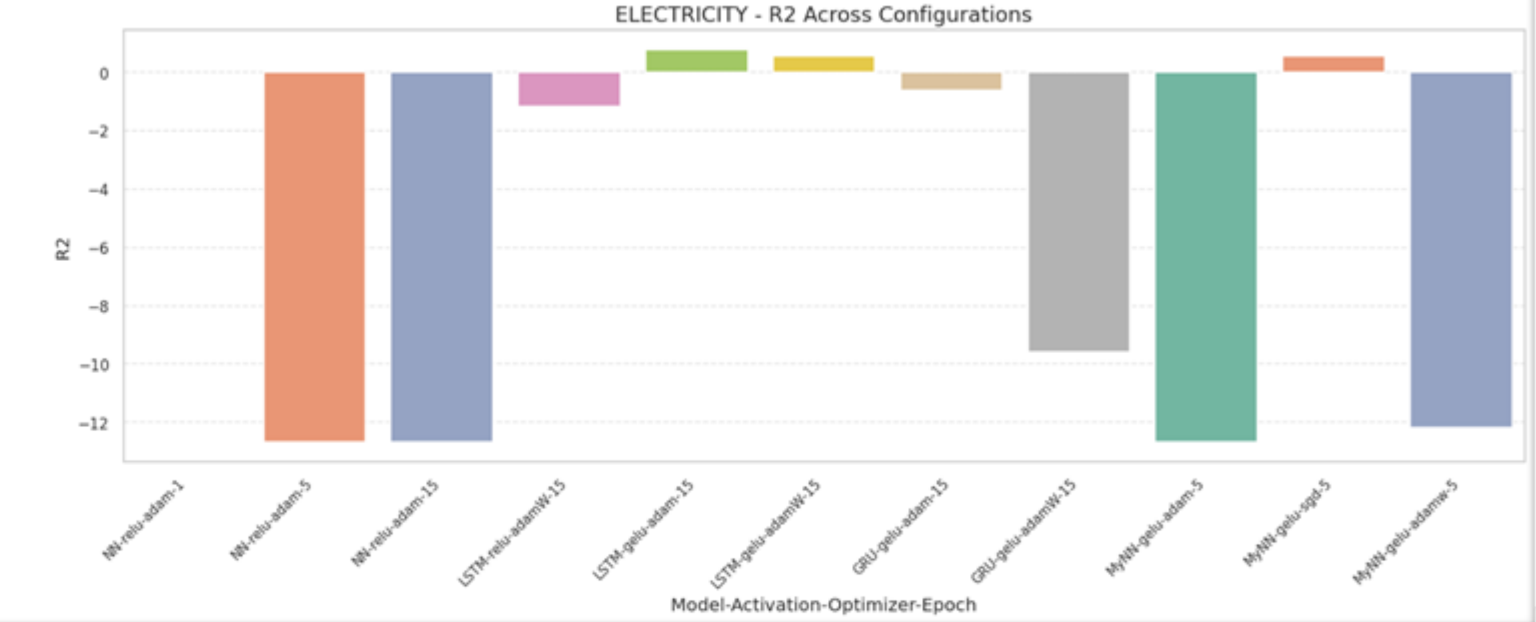


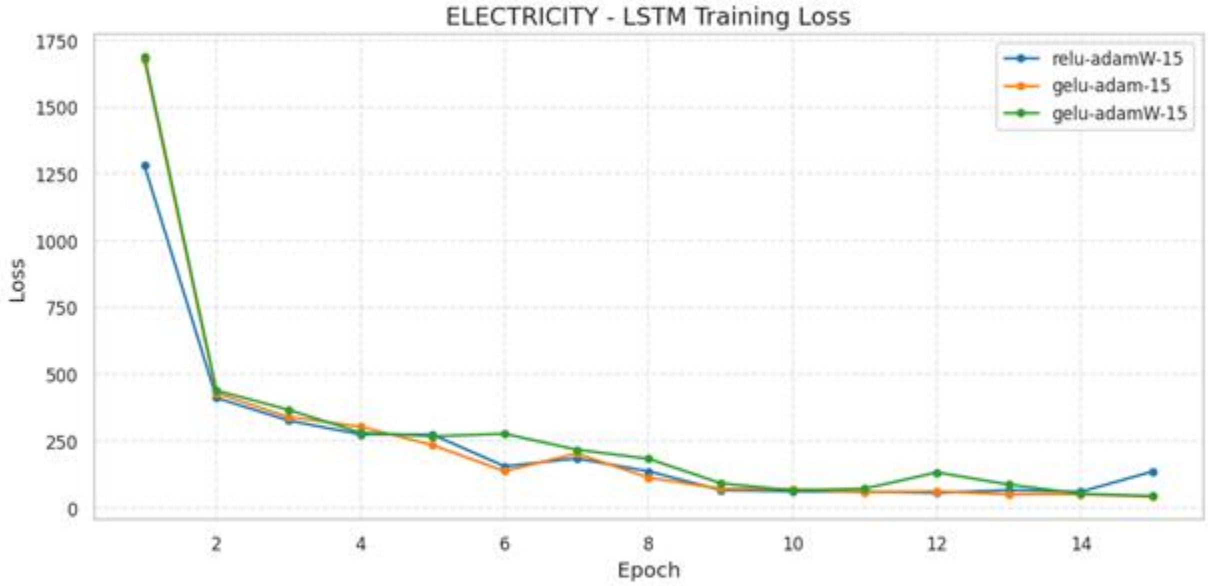
Electricity Dataset

Electricity meter data from 370 commercial buildings in Portugal

Best Model: LSTM-gelu-adam-15 epochs

Epochs	R^2
3	-1.4392
15	0.7610





Argonne NATIONAL LABORATORY

Challenges And Next Step

- **Lack of Parallelism**
  — Unable to train model on LCL dataset
- **Model Architecture Limitations**
  — Poor performance on LCL and Borealis datasets compared to Ideal results
- **Bug in Sceaux**
  — Issue in train–test dataset split

- **Next Step**
  — Address challenges
  — Simplify instructions

Conclusion

- LSTM > GRU > RNN | NN
  — Aligns with literature

- **Influential Factors**
  — Number of Epochs
  — Batch Size
  — Learning Rate
  — Layers | Activation | Optimization

Acknowledgments








Argonne NATIONAL LABORATORY

NERSC

OAK RIDGE National Laboratory

U.S. DEPARTMENT of ENERGY

Perlmutter

Argonne Intro to HPC Bootcamp

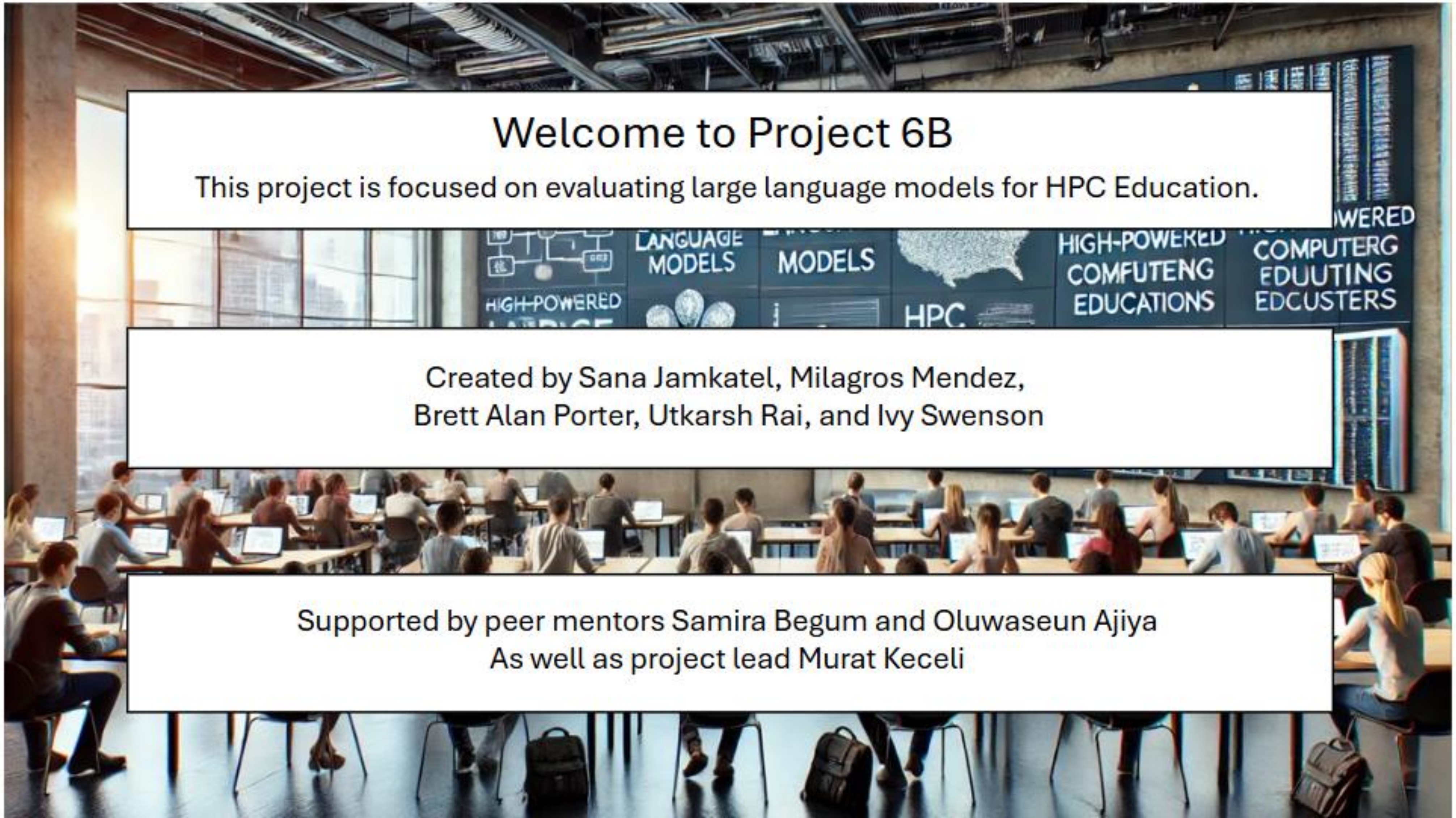
Questions?

Argonne Intro to HPC Bootcamp

Group 6b

High-Performance Computing Bootcamp 2025

INTRODUCTION



Welcome to Project 6B

This project is focused on evaluating large language models for HPC Education.

Created by Sana Jamkatel, Milagros Mendez,
Brett Alan Porter, Utkarsh Rai, and Ivy Swenson

Supported by peer mentors Samira Begum and Oluwaseun Ajiya
As well as project lead Murat Keceli

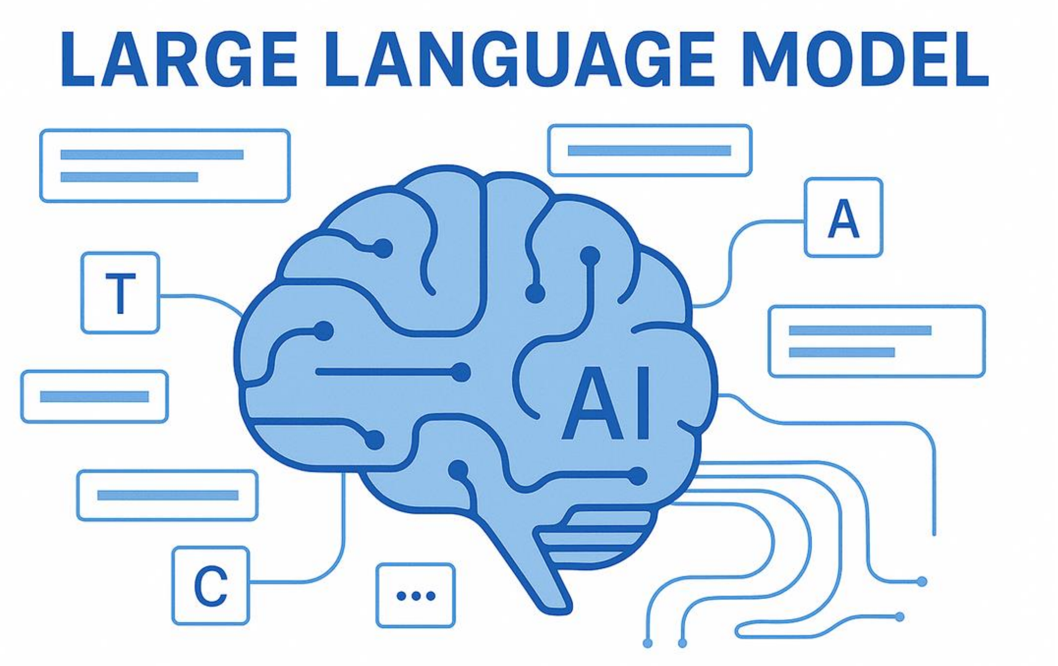
Argonne NATIONAL LABORATORY

PROJECT GOAL

Can LLMs assist in teaching HPC concepts?

Can LLMs answer basic HPC Questions?

Evaluate accuracy of LLM responses



Argonne NATIONAL LABORATORY

CHATBOT

Gemini

python™

LangChain

Argonne NATIONAL LABORATORY

DISCOVERIES

Gemini Model

Retrieval Augmented Generation (RAG)

Evaluation

Evaluating Model

Argonne NATIONAL LABORATORY

CHALLENGES

Too many Python libraries !!

Trouble with creating GUI with Tkinter - layout and interaction issues

- Pre-trained Model Limitations – Scores can drop when model isn't domain-tuned.
- Data Quality Issues – Single PDF with pre-set Q&A limits coverage.
- Limited Knowledge Base – More documents would improve retrieval & answers.
- BERTScores have limitations - effectiveness is tied to the quality of the pre-trained model it uses.

Running on NERSC Jupyter vs on Google Colab - required extra debugging

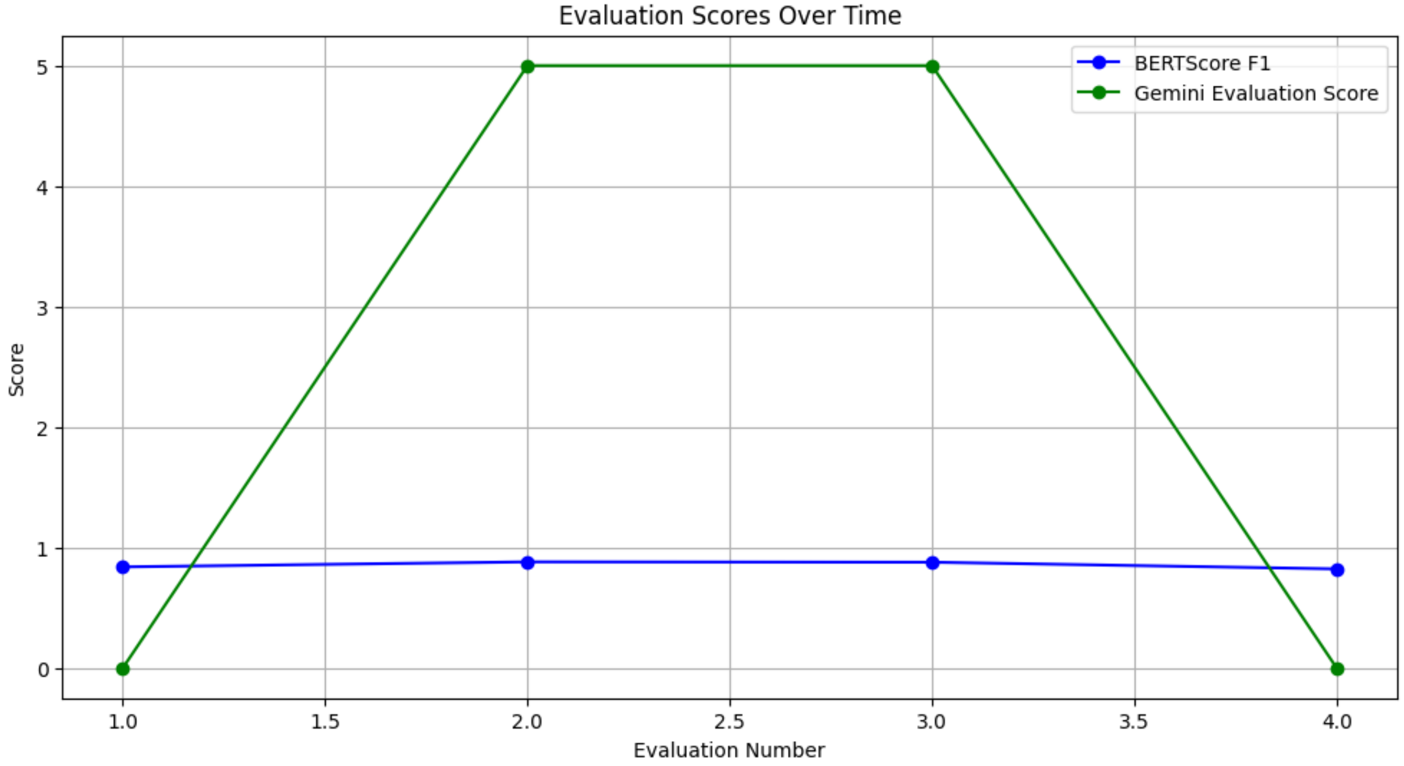
CONCLUSION

Developing our own model would take more time

Using Gemini model:

BertScore F1 stays consistent

Gemini Evaluation Score increases, maintains, and then decreases



Evaluation:

Score: 40/100 Explanation: The answer correctly identifies HPC's role in solving complex problems faster than traditional computers. However, it lacks the full definition (resource aggregation, performance gain over single machines), key benefits (scientific/commercial advancements), and explicit mention of massive data processing. It also includes specific technologies not present in the ground truth's general definition.

NEXT STEPS

We would pre-train the model on our data

We would define our own accuracy points

We would design a better UI/UX

Argonne NATIONAL LABORATORY