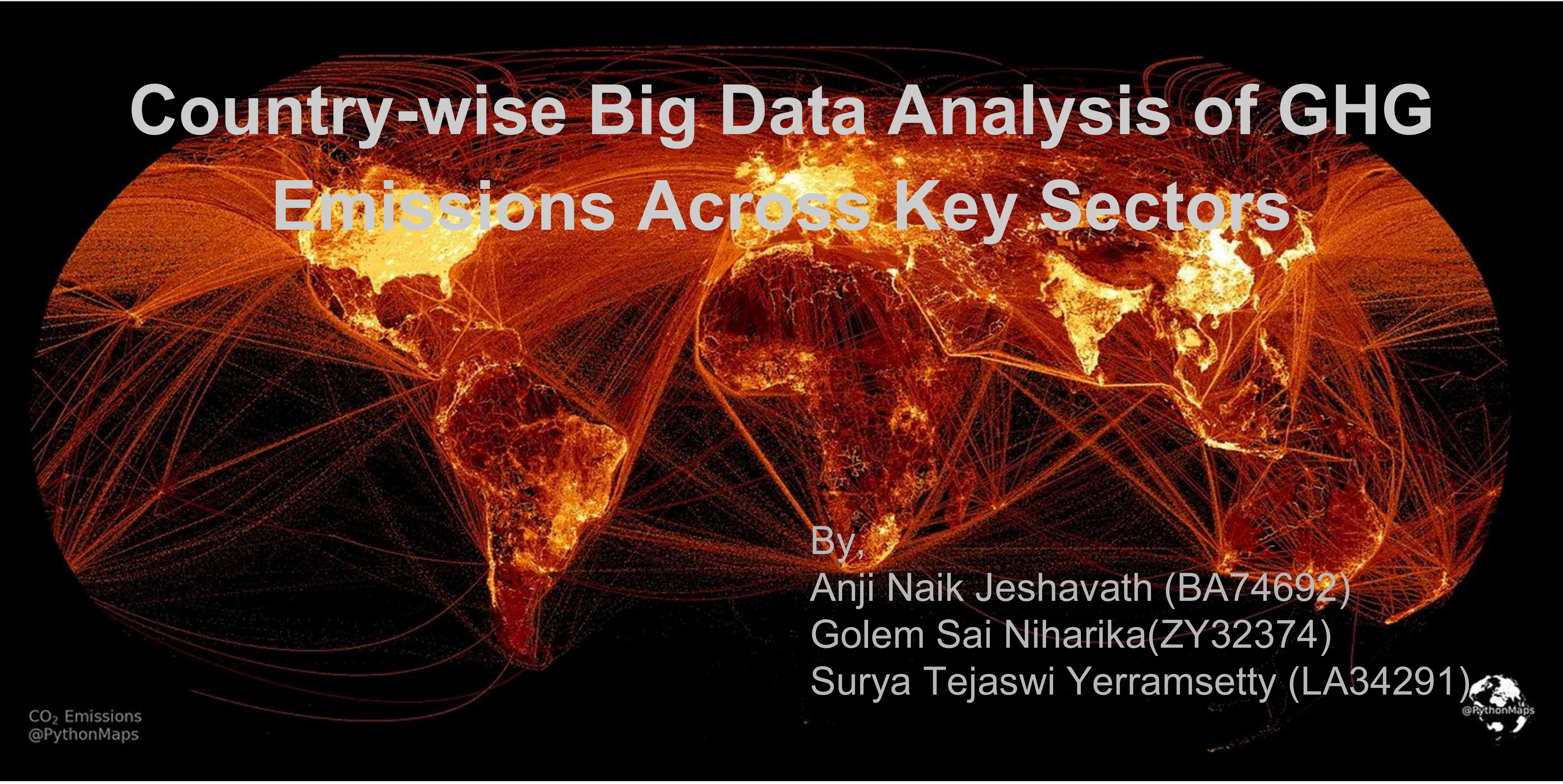
UMBC

Country-wise Big Data Analysis of GHG Emissions Across Key Sectors



CO$_2$ Emissions
@PythonMaps

By,
Anji Naik Jeshavath (BA74692)
Golem Sai Niharika(ZY32374)
Surya Tejaswi Yerramsetty (LA34291)

Introduction

- The project "Country-wise Big Data Analysis of GHG Emissions Across Key Sectors" represents a critical initiative in understanding and mitigating global greenhouse gas emissions. Through a detailed analysis of emissions data from 2015 to 2022 across eight sectors, the project seeks to identify the key contributors and trends in greenhouse gas production.
- We are contributing this project to US EPA - Environmental Protection Agency
- This analysis is not only significant for environmental policy-making but also demonstrates the vital role of big data in addressing global challenges. The project's methodology, which employs advanced data processing and visualization tools, underscores its relevance to the data science course, showcasing practical applications of these skills in a real-world context.

Objective

- Objective 1: Country-Specific Analysis.
- Objective 2: Emissions Assessment.
- Objective 3: Understanding emissions sources and trends

This project aims to address the specific problem of understanding the global landscape of greenhouse gas emissions. By using big data platforms, it seeks to answer questions such as which countries are the largest contributors in each sector, how these emissions have evolved over time, and what this implies for global environmental strategies.

Data Source

1.We have collected element-wise emissions for each sector in various countries.

Example: For each Agriculture sector – CH4, C02, N20 emissions due to various factors/ elements like cropland fires, rice cultivation, synthetic fertilizers, etc.,

2.All this data is collected for all the countries over the period [2015-2022]

3.Size: 17.8 Gb, 48 CSV Files

https://climatetrace.org/

Tools and Technologies

The project utilizes a specific set of big data tools and technologies, each chosen for its unique strengths:

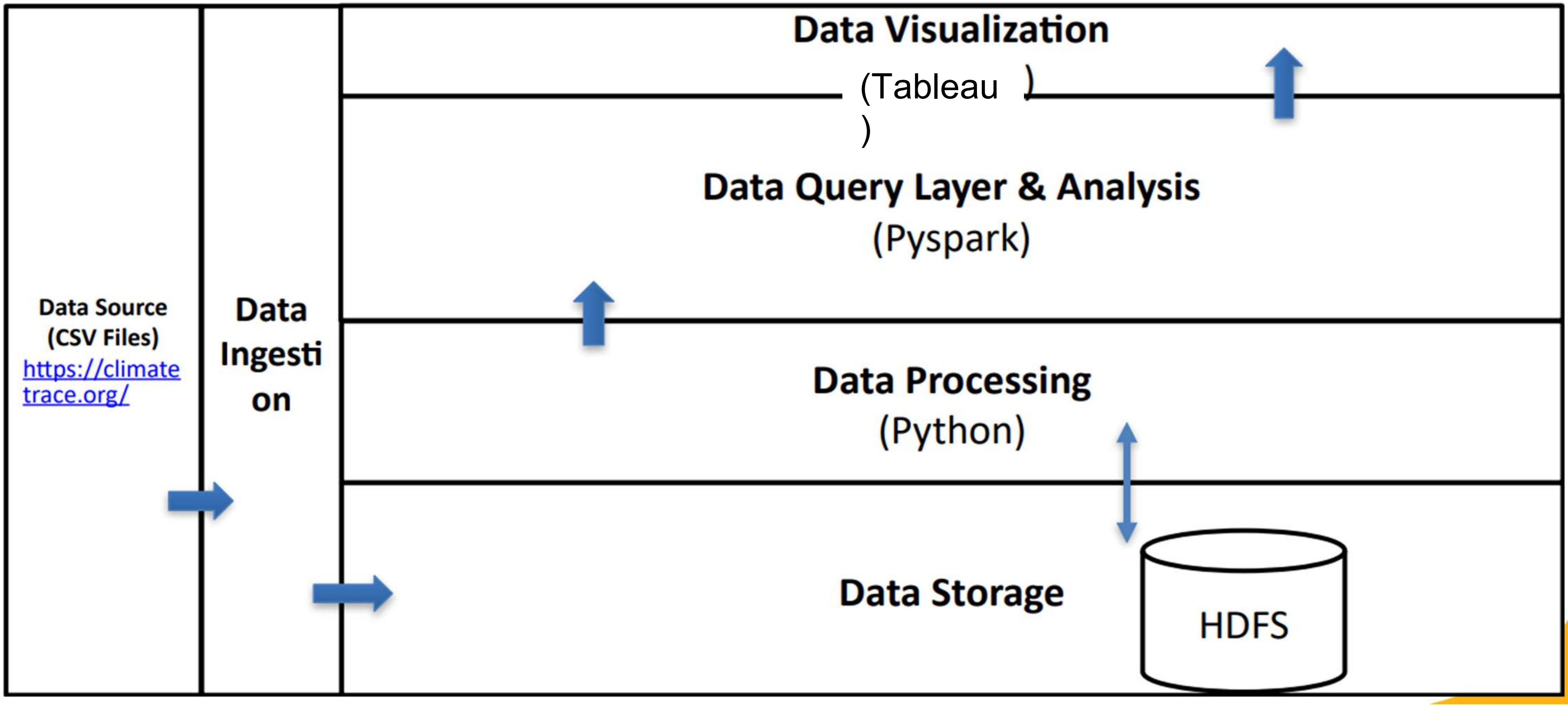
HDFS (Hadoop Distributed File System): A key component for storing large volumes of data. Its reliability and scalability make it an ideal choice for big data projects.

Python: Known for its extensive libraries and versatility, Python is used for data cleaning, preprocessing, and initial analysis. Its wide range of data manipulation capabilities makes it indispensable.

PySpark: Chosen for its ability to handle large-scale data transformation efficiently. PySpark, as part of the Apache Spark ecosystem, is particularly effective in processing big data quickly.

Tableau: Used for its powerful data visualization capabilities. It enables the creation of interactive and informative visual reports, crucial for understanding complex data patterns.

Architecture



Data Visualization

(Tableau)

)

Data Query Layer & Analysis

(Pyspark)

Data Processing

(Python)

Data Storage

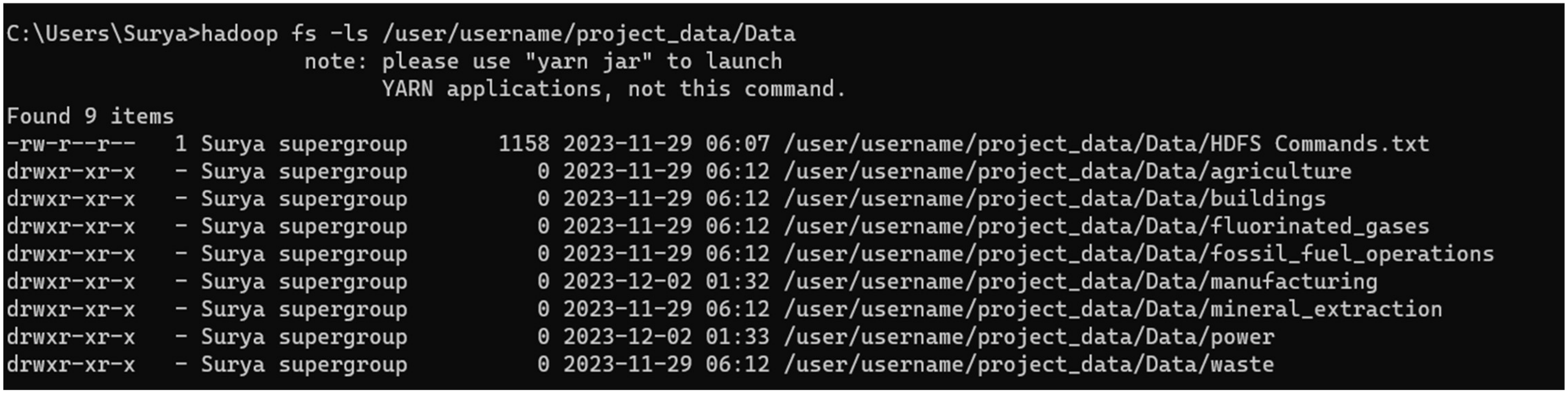
HDFS

Data Source

(CSV Files)

https://climatetrace.org/

Data Ingestion



```
C:\Users\Surya>hadoop fs -ls /user/username/project_data/Data
                       note: please use "yarn jar" to launch
                             YARN applications, not this command.
Found 9 items
-rw-r--r--   1 Surya supergroup       1158 2023-11-29 06:07 /user/username/project_data/Data/HDFS Commands.txt
drwxr-xr-x   - Surya supergroup          0 2023-11-29 06:12 /user/username/project_data/Data/agriculture
drwxr-xr-x   - Surya supergroup          0 2023-11-29 06:12 /user/username/project_data/Data/buildings
drwxr-xr-x   - Surya supergroup          0 2023-11-29 06:12 /user/username/project_data/Data/fluorinated_gases
drwxr-xr-x   - Surya supergroup          0 2023-11-29 06:12 /user/username/project_data/Data/fossil_fuel_operations
drwxr-xr-x   - Surya supergroup          0 2023-12-02 01:32 /user/username/project_data/Data/manufacturing
drwxr-xr-x   - Surya supergroup          0 2023-11-29 06:12 /user/username/project_data/Data/mineral_extraction
drwxr-xr-x   - Surya supergroup          0 2023-12-02 01:33 /user/username/project_data/Data/power
drwxr-xr-x   - Surya supergroup          0 2023-11-29 06:12 /user/username/project_data/Data/waste
```

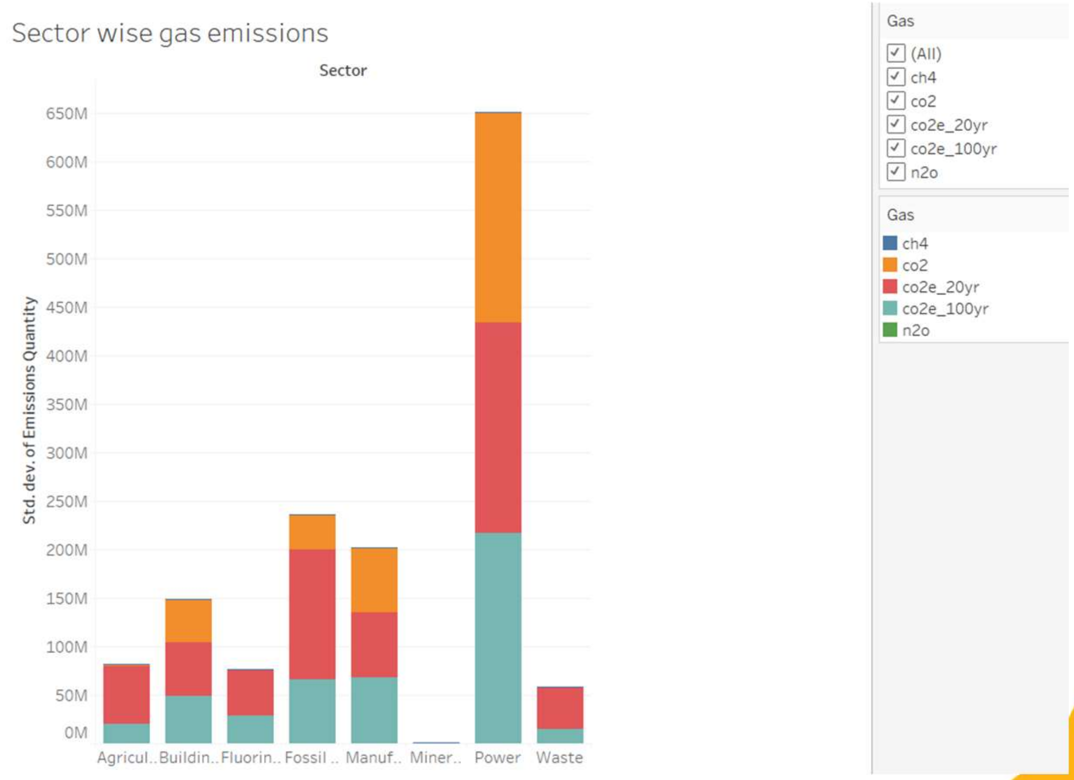
Input files are stored in HDFS

Challenges faced:

- Data Accuracy and Consistency: Ensuring the accuracy and standardization of data across different countries remains a challenge.
- Handling Increasing Data Volumes: As data grows in volume and complexity, finding efficient ways to process and analyze it will be crucial.
- Hardware Limitations: Processing large volumes of emissions data requires substantial computational resources. System constraints, such as limited memory, CPU capacity, or disk I/O, might impact the performance and scalability of data processing.
- Software Compatibility: Different versions of software tools or frameworks used for data processing might not be compatible. Upgrading or transitioning between versions while maintaining compatibility with existing systems can be complex.

Key Insights

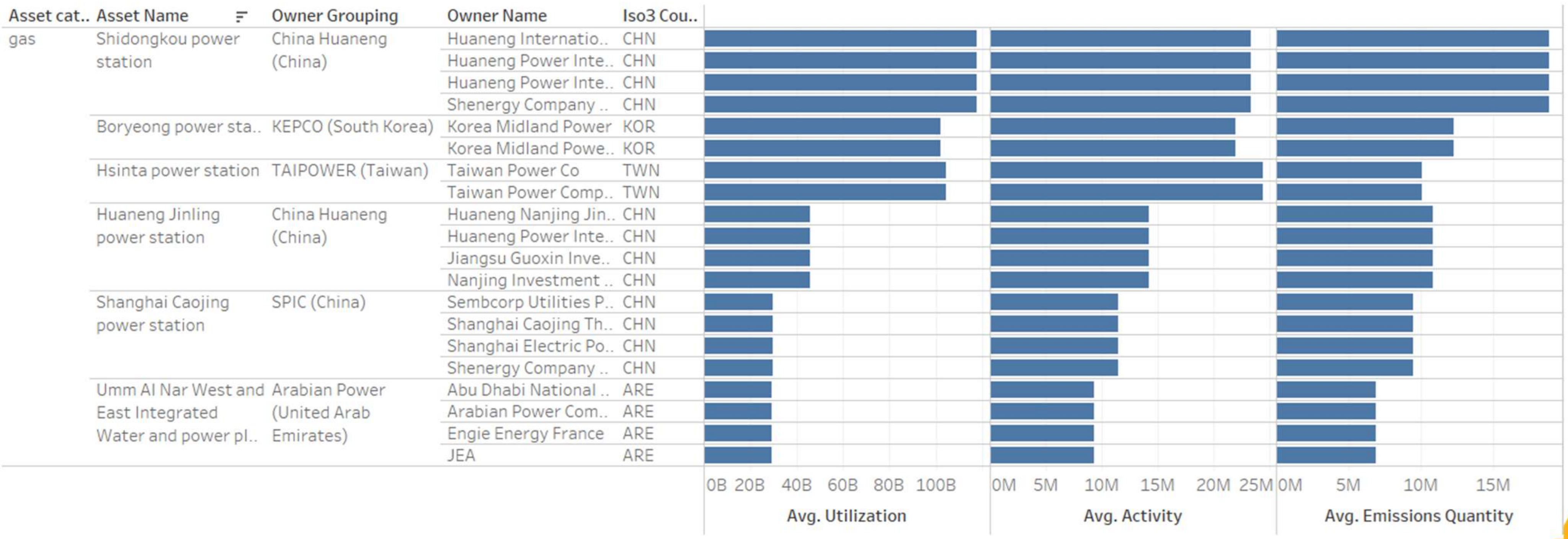
- China is the Top 1 emitter.
- Co2 emissions are majorly due to Power, and manufacturing sectors.
- The Power Sector is major contributor



Gas

☑ (All)
☑ ch4
☑ co2
☑ co2e_20yr
☑ co2e_100yr
☑ n2o

Emissions due to Electricity generating Stations gas



Asset cat..	Asset Name	Owner Grouping	Owner Name	Iso3 Cou..
gas	Shidongkou power station	China Huaneng (China)	Huaneng Internatio..	CHN
			Huaneng Power Inte..	CHN
			Huaneng Power Inte..	CHN
			Shenergy Company ..	CHN
	Boryeong power sta..	KEPCO (South Korea)	Korea Midland Power	KOR
			Korea Midland Powe..	KOR
	Hsinta power station	TAIPOWER (Taiwan)	Taiwan Power Co	TWN
			Taiwan Power Comp..	TWN
	Huaneng Jinling power station	China Huaneng (China)	Huaneng Nanjing Jin..	CHN
			Huaneng Power Inte..	CHN
			Jiangsu Guoxin Inve..	CHN
			Nanjing Investment ..	CHN
	Shanghai Caojing power station	SPIC (China)	Sembcorp Utilities P..	CHN
			Shanghai Caojing Th..	CHN
			Shanghai Electric Po..	CHN
			Shenergy Company ..	CHN
	Umm Al Nar West and East Integrated Water and power pl..	Arabian Power (United Arab Emirates)	Abu Dhabi National ..	ARE
			Arabian Power Com..	ARE
			Engie Energy France	ARE
			JEA	ARE

Future Opportunities

Opportunities:

- Real-Time Insights: Access to more granular and real-time emissions data can offer a comprehensive understanding of major contributors within sectors, aiding proactive responses and policy implementations.
- Holistic Environmental Strategies: Country-wise big data analysis enables the development of targeted and efficient strategies to reduce emissions, fostering a more sustainable and environmentally conscious approach.
- Predictive Analytics: Utilizing big data for predictive modeling can forecast potential emissions trends, assisting in the creation of preemptive measures and adaptive policies.
- Global Collaboration: Sharing standardized emissions data across borders encourages collaborative efforts, fostering a unified global response to combat climate change.

THANK
YOU

References

J. Ravi et al.,2020, "Identification of Greenhouse Gases Emission Thorough Exploration of The Emission from Different Sectors," 2020 *6th International Conference on Computing Engineering and Design (ICCED)*.

A. C. Riekstin, R. Hedjam, T. Dandrest and M. Cheriet, "Statistical-based method to determine the best hour of the day regarding GHG emissions for a smart home appliance," 2017 IEEE SmartWorld, *Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, (SMARTWORLD/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*.

A. Sahebalam, C. Ding, G. Michor, G. Doucette and M. Ahadi, 2018, "Hourly Live GHG Calculation and Forecasting," 2018 *IEEE Canadian Conference on Electrical & Computer Engineering (CCECE).*