**GHG Emission Prediction Project**

# **Day 3 -** Understanding Step 2 Code for Week 01

# **Date -** 19 June 2025 (Week 1)

# **Internship Name** - Edunet-Shell Skills4Future AICTE Internship

# **Intern Name:** Abhinay Singh

**✅ Tasks Completed:**

1. **Executed Code to Combine Datasets from All Years (2010–2016):**

**Standardized column names and added two new columns:**

* **Source:** To differentiate between Industry vs Commodity.
* **Year:** The year of the data (added for tracking trends).
* Combined both datasets using pd.concat and appended to a list named all\_data.

2. **Created Final DataFrame (df)**

* Used df = pd.concat(all\_data, ignore\_index=True) to merge all yearly data into one master dataframe.

**This final df now contains:**

* **All 7 years' data**
* Unified format and structure
* Both Industry & Commodity data in a single clean table.

**3. Checked for Null/Missing Values**

**#Step 3: Data Preprocessing**

**df.columns # Checking columns**

**df.isnull().sum()**

* Found that all columns except Unnamed: 7 are clean (i.e., 0 missing values).
* Column Unnamed: 7 has **2100+ missing values**, indicating it's **likely an empty or unnecessary column**, probably coming from Excel formatting.
* Plan to drop it in preprocessing ( commented in code).

🧩 **Key Concepts Learned Today:**

* **DataFrame (df)**: A tabular structure (rows × columns) in pandas, similar to an Excel table.
* **Concatenation (pd.concat):** Used to stack multiple DataFrames together (vertically).
* **Missing Values (NaN):** Gaps in data that need to be handled before modeling.
* **Data Cleaning:** Essential step before feeding data into ML models.

**❓ Doubts Clarified:**

* Understood the concept and structure of a pandas DataFrame.
* Understood the purpose of combining multiple Excel sheets into a single consistent dataset.
* Understood what df.isnull().sum() does and how to interpret the output.

**Differentiated between:**

* all\_data: List containing one combined dataframe per year (7 total).
* df: Single final dataframe combining all years into one.

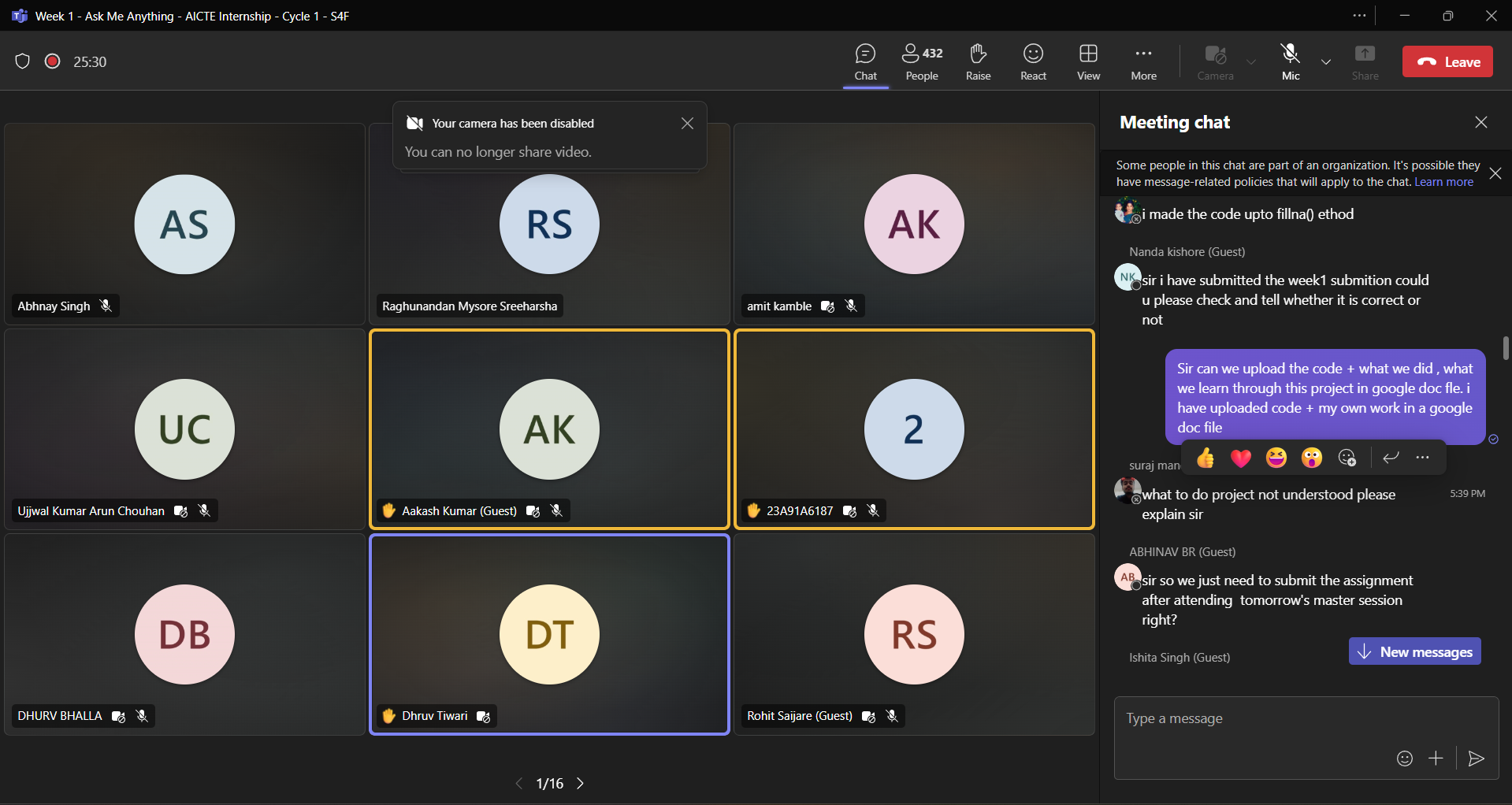
💭 **Self-Reflection:**

* “Today’s session was highly conceptual and boosted my confidence. The data merging part was very interesting. I understood how multiple datasets are brought together in real-world ML projects. I also understood how to detect and handle unnecessary columns. I feel more comfortable using pandas now and will revise this section from my handwritten notes.”
* I also attended ask me anything session taken by our mentor

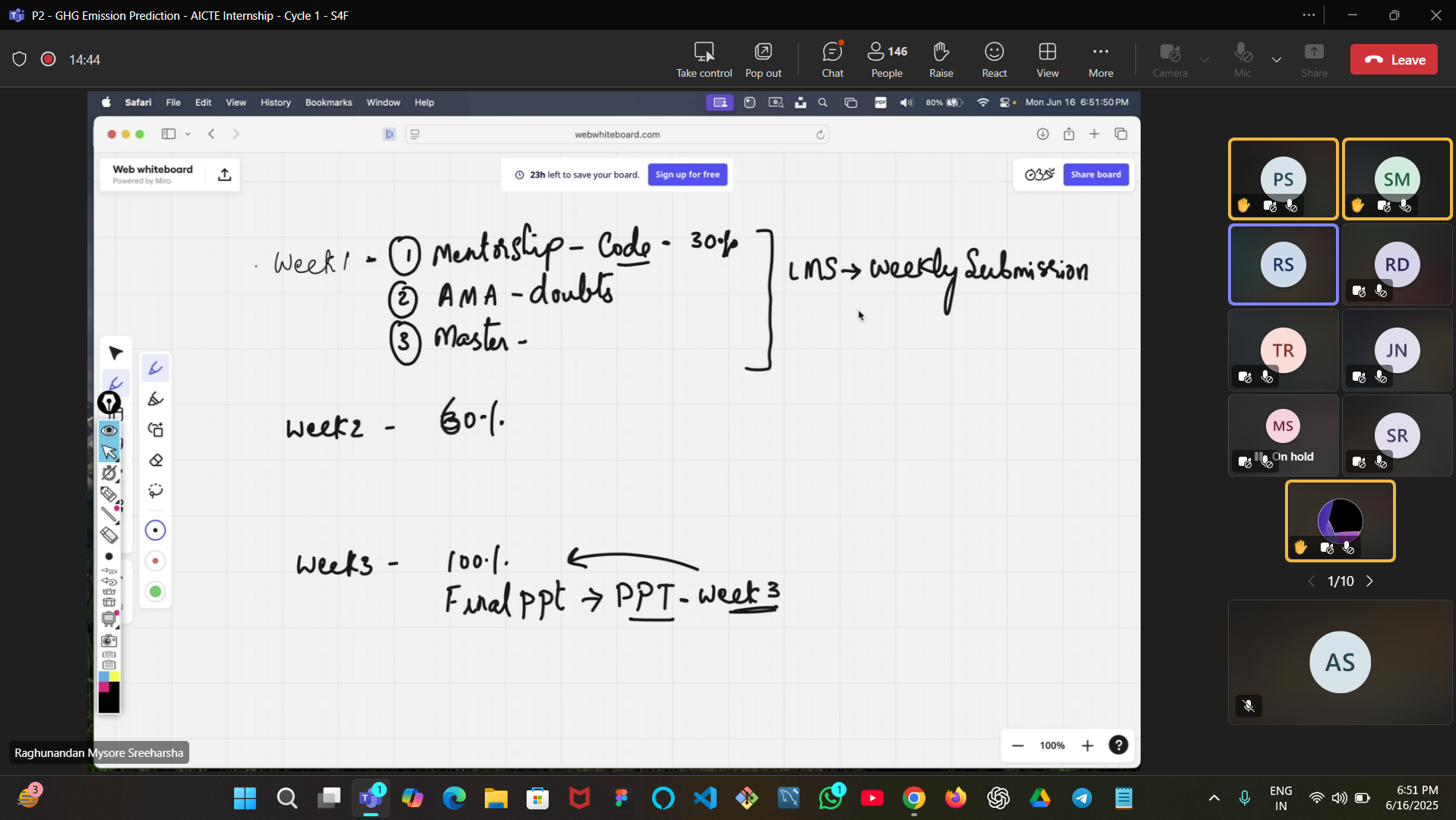
**Raghunandan Mysore Sreeharsha Sir,** Where I asked about how to submit project code alongside daily logsmaintained in the github and LMS portal.

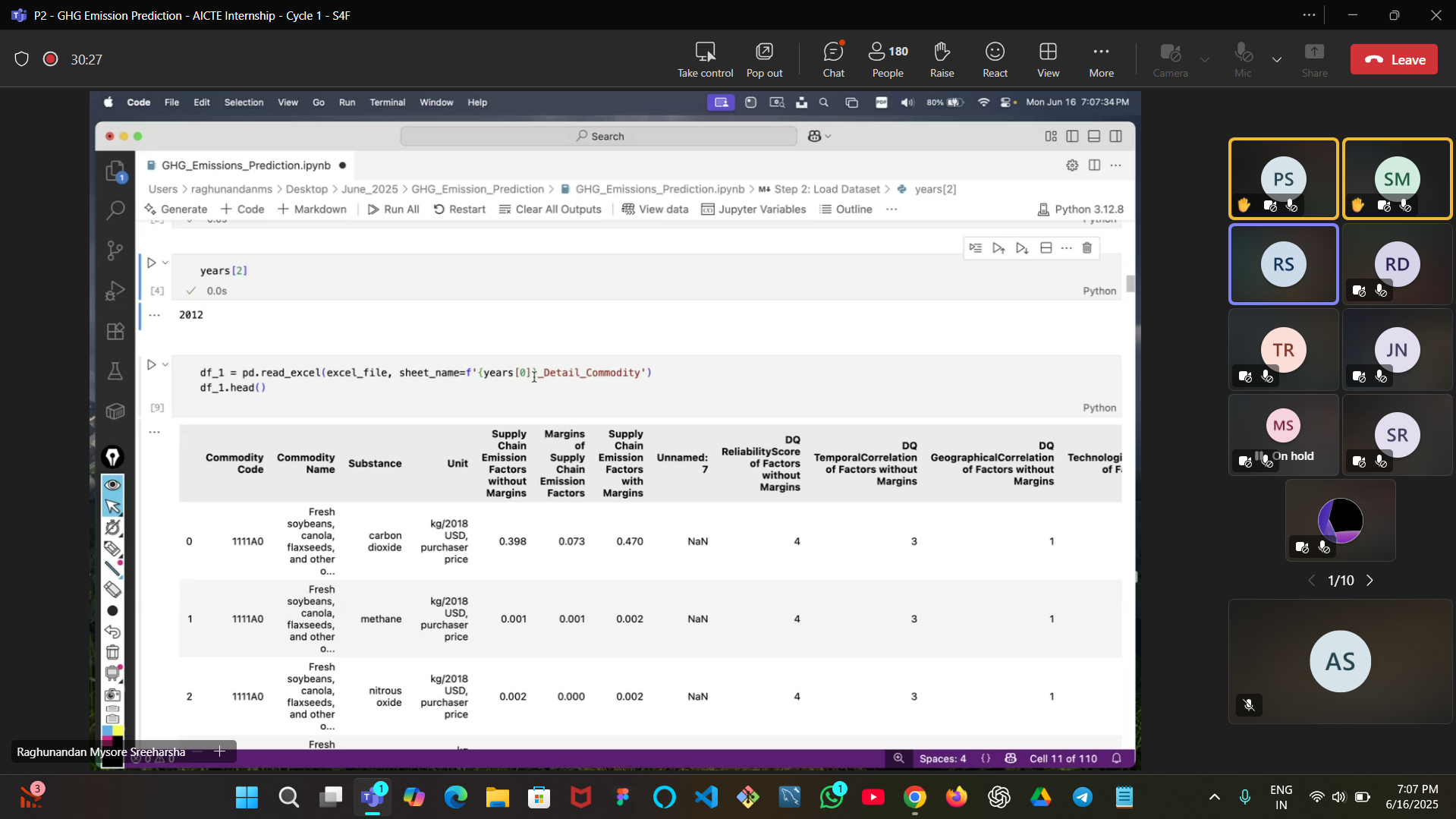
* I also asked sir about the way we can use ChatGPT, Sir told us you can use it to understand the code, concept not just to copy.

**Ask Me Anything Session:**



**Live Session:**

****

****