

# WiDS: Data-Driven Carbon Footprint Analytics for Chemical Processes

Week 1:

Major Things to Learn this week:-

1. Understand carbon footprint, carbon intensity, and emission inventory basics.
2. Learn how chemical processes generate CO<sub>2</sub>, CH<sub>4</sub>, and N<sub>2</sub>O emissions.
3. Familiarize with typical datasets: temperature, pressure, flow rates, energy consumption, utility loads.
4. Learn basic EDA, cleaning, and visualization for process data.

Suggested Resources you can use to understand these concepts :

Reading

- IPCC Guidelines - Industrial Processes & Product Use (free PDF)
- US EPA - Greenhouse Gas Reporting Program (GHGRP)
- Chapter 2 of Sustainability in Chemical Engineering (Process emissions)

Videos

## 1. Carbon Footprint & Carbon Accounting Basics

- Carbon accounting basics:  
<https://www.youtube.com/watch?v=IPgeCdNmZBc>
- "What is Carbon Footprint?" (UN Climate Change / UNFCCC)  
[https://www.youtube.com/watch?v=8q7\\_aV8eLUE](https://www.youtube.com/watch?v=8q7_aV8eLUE)
- Introduction to Greenhouse Gases & Emissions (US EPA)  
<https://www.youtube.com/watch?v=RM49RGn7uQo>

## 2. Process Data, Sensors, and Industrial Analytics

- Introduction to Process Data Analytics  
<https://www.youtube.com/watch?v=cDfTxLtaMHo>
- Industrial time-series data | Sensors | Process parameters  
<https://www.youtube.com/watch?v=A3fowDMo8mM&t=715s>

## 3. EDA + Cleaning + Plotting (Python)

- Data School — Pandas Data Cleaning  
[https://www.youtube.com/watch?v=otCriSKVV\\_8](https://www.youtube.com/watch?v=otCriSKVV_8)
- Corey Schafer — Pandas Basics & Plotting  
<https://www.youtube.com/watch?v=zmdjNSmRXF4>
- StatQuest — Correlation & Regression Basics  
<https://www.youtube.com/watch?v=4EXNedimDMs>

Short Assignments(for practice only, wont count for requirement of completion of WiDS)

- **Task 1:** Identify emission sources for any 3 industrial chemical processes.
- **Task 2:** Do EDA on this dataset:  
<https://www.kaggle.com/datasets/elikplim/eergy-efficiency-dataset>  
Plot temperature, pressure, flow over time if applicable.  
Compute correlation matrix for process variables.
- **Task 3:** Calculate basic CO<sub>2</sub> estimation:  
$$\text{CO}_2 = \text{Fuel Consumption} \times \text{Emission Factor}$$

Week 2:

Major Things to Learn this week:-

1. Build supervised ML models to predict emissions or carbon intensity.
2. Learn feature engineering for chemical process data.
3. Compare baseline regression models.

Suggested Resources you can use to understand these concepts :

Reading

- Hands-On ML  
<https://homl.info/>
- ML in Chemical Engineering — Review Paper  
<https://www.sciencedirect.com/science/article/pii/S2666544121000039>

Videos

- StatQuest - Linear Regression  
[https://www.youtube.com/watch?v=nk2CQITm\\_eo](https://www.youtube.com/watch?v=nk2CQITm_eo)
- StatQuest - Ridge and Lasso Regression  
<https://www.youtube.com/watch?v=Q81RR3yKn30>
- StatQuest - Random Forests  
[https://www.youtube.com/watch?v=J4Wdy0Wc\\_xQ](https://www.youtube.com/watch?v=J4Wdy0Wc_xQ)

- StatQuest - Gradient Boosting  
<https://www.youtube.com/watch?v=3CC4N4z3GJc>
- Train/Test Split, Cross-Validation (Python) – Data School  
<https://www.youtube.com/watch?v=6dbrR-Wymjl>
- Krish Naik — Time-Series Basics  
<https://www.youtube.com/watch?v=zp8clK9yCro>
- ARIMA Model Explanation  
<https://www.youtube.com/watch?v=Y2XcWFkHxHI>

### Practice Tasks for this week:-

- **Task 1:** Train 3 models (LR, RF, XGB) to predict CO<sub>2</sub> emissions on any dataset from net.
- **Task 2:** Create plots:  
predicted vs. actual emissions  
feature importance
- **Task 3:** Identify top 5 parameters contributing to carbon intensity using:  
SHAP values or  
Random Forest feature importance

### Week 3:

### Major Things to Learn this week:-

- Extract keywords from sustainability reports.
- Compare process-level data vs. company sustainability targets.
- Build a lightweight NLP text-mining workflow

## Resources

### Reading

spaCy Documentation

<https://spacy.io/usage/spacy-101>

TF-IDF: Analytics Vidhya Guide

<https://www.analyticsvidhya.com/blog/2015/02/complete-guide-text-classification/>

Access sustainability reports (free PDFs):

- BASF → <https://www.basf.com/global/en.html>
- Dow → <https://corporate.dow.com/en-us/science-and-sustainability>
- Reliance → <https://www.ril.com>

## Videos

- freeCodeCamp — Full NLP With Python (spaCy, NLTK, TF-IDF etc.)  
[https://www.youtube.com/watch?v=I-\\_N7zVOzA0](https://www.youtube.com/watch?v=I-_N7zVOzA0)
- spaCy Crash Course (by spaCy)  
<https://www.youtube.com/watch?v=WnGPv6HnBok>
- TF-IDF explained simply  
<https://www.youtube.com/watch?v=VNWc8m9W9jQ>
- Extract text from PDFs using Python  
[https://www.youtube.com/watch?v=gqd9nDw\\_5Bk](https://www.youtube.com/watch?v=gqd9nDw_5Bk)
- Topic Modelling with Scikit-Learn (LDA)  
<https://www.youtube.com/watch?v=3mHy4OSyRf0>

## Practice Tasks for this week:-

- Task 1: Extract text from 1–2 sustainability report PDFs.
- Task 2: Generate:  
word frequency distribution  
bigram/trigram extraction  
TF-IDF keywords
- Task 3: Extract sustainability metrics, e.g.:  
"Reduce CO<sub>2</sub> by 20% by 2030"
- Task 4: Compare these to actual process data trends:  
Is emission declining?  
Does process efficiency match sustainability claims?