

CDS 101 – Final Project Proposal

Samir Dawar, Andrew Lee, Sumera Muhamm

December 08, 2025

Contents

1. Dataset	1
2. Dataset Link	1
3. Research Question	2
4. Modeling Approach / Models Used	2
5. Expected Results	2
6. Potential Challenges (Optional)	3

1. Dataset

Dataset Name: Chernobyl Chemical Radiation / CSV / Country Data

Source / Platform: Kaggle

Short Description:

- This dataset contains radiation measurements after The Chernobyl Incident. Looking at three isotopes, across varying dates in different European countries, and specific cities. - With the main variables of radiation measurements being the three isotopes which are Iodine-131 (I-131), Caesium-134 (Cs-134), and Caesium-137 (Cs-137). - This dataset contains 9 rows, and approximately 2038 columns. - As for it's relevance, The Chernobyl Incident is considered to be the worst nuclear disaster in history, based on cost and casualties caused by the accident. Not only that, Several other European countries beside the Soviet Union were affected by it, due to a massive atmospheric transfer of radioactive material. Highlighting a point of interest on how drastic these effects were, across multiple countries, cities, and communities.

2. Dataset Link

Paste a direct URL to the dataset (or documentation) here:

<https://www.kaggle.com/datasets/brsdincer/chernobyl-chemical-radiation-csv-country-data>

3. Research Question

Main research question:

How did radiation levels of I-131, Cs-134, and Cs-137 vary across different countries after the Chernobyl disaster?

Optional secondary question(s):

How does each radiation isotope vary from each other? what properties of these isotopes could explain any variation?

4. Modeling Approach / Models Used

- Type of task (check all that apply):
 - Exploratory analysis only
 - Regression (predicting a numeric value)
 - Classification (predicting categories/labels)
 - Clustering / unsupervised patterns
 - Other (describe)

Planned models / methods:

- Linear regression:
 - boxplot
 - scatterplot
 - geographic plot

Planned evaluation metrics:

- Regression: RMSE, MAE, R^2
 - Classification: accuracy, precision/recall, F1, confusion matrix
-

5. Expected Results

- We expect distance from Chernobyl to be the strongest determiner of radiation measures. (farther away from Chernobyl = less isotope count).
 - That most of the isotopes will be generally equal in amount, when distance and date are mostly same.
 - Of course, countries closer to Chernobyl will have higher radiation measures (number of isotopes).
-

6. Potential Challenges (Optional)

- What might be difficult? missing data, sample size being potentially too large, determining wind effects to isotopes
 - How might you handle or mitigate these issues? reduce sample used, creating new variables that take outliers to determine potential patterns in isotope movement.
-

Submitted on:

(12/8/2025)