RECIPE INSIGHTS.

RECIPE INSIGHTS.

The advanced stage - using data confidently, critically an strategically.

Key Abilities:

Explain patterns or make predictions, end evaluate their accuracy and limitations.
Communicate insights clearly to different audiences.
Challenge assumptions and spot flawed analyses.
Make data-driven decisions with context and nuance.
Integrate data thinking into strategy, design and problem solving.

Typical Learner Mindset:

""I can think with data, not just about data"

RECIPE INSIGHTS.

In this activity you will develop the ability to interpret and manipulate data by identifying key attributes such as ingredients and cooking techniques within recipes.

AGENDA.

- What Are Recipe Insights?
- Learning Goals
- Understanding And Cleaning The Data
- Analysing Similarities And Clusters
- CoPilot
- Key Takeaways

WHAT ARE RECIPE INSIGHTS?

- In today's data-driven world, **understanding** how to **analyse** and **interpret information** is a crucial skill.

  - When it comes to exploring collections of data - such as a group of recipes - **algorithms** can offer **powerful tools** to uncover hidden **patterns** and **relationships**.
  - By applying computational techniques to recipes, we can **systematically compare** ingredients, cooking methods, and flavour profiles to identify similarities and groupings that might not be immediately obvious.
  - This approach not only deepens our appreciation for culinary creativity but also provides a practical example of how **data literacy** enables us to extract **meaningful insights** from **complex** datasets.
  - In this activity, we will explore how algorithms can be used to analyse recipes, revealing the subtle **connections** that link diverse dishes across cuisines and cultures.

LEARNING GOALS.

- **Interpret** and **manipulate** data to identify key attributes such as ingredients and cooking techniques.
- Apply **algorithmic thinking** to analyse and compare data points, **recognising patterns** and **similarities** within a dataset.
- Use **data exploration** methods to draw meaningful insights from **unstructured** or **semi-structured** data like recipes.
- Communicate findings effectively by translating algorithmic results into **understandable patterns** and **relationships**.

RECIPE DATA.

- Printed out is a list of 29 recipes from around the world.
- Working in groups of **four**, explore the question sets.

UNDERSTANDING AND CLEANING THE DATA.

- Ingredient **Identification** & **Normalisation**.

  - **What** challenges do you notice in the way ingredients are named across different recipes?
  - Can you **identify examples** of the same ingredient being named differently?
    - How would you **standardise** them?

  - Are there any **typographical** errors or **inconsistencies**?
    - How would you **correct** or **handle** them?

- How would you **deal with** plural vs singular forms or ingredients?

UNDERSTANDING AND CLEANING THE DATA.

- Ingredient **Categorisation**.

- Can you **categorize** the ingredients into broader groups (e.g., meat, fish, vegetable, spice, dairy)?
  - Which **categories** do you find most useful?

- **How** could you handle ingredients that are **combinations** or **ambiguous** (e.g., cream or coconut milk)?

- Are there any ingredients unique to certain regions or cuisines?
  - **How** would you treat those in your **analysis**?

UNDERSTANDING AND CLEANING THE DATA.

- **Data Preparation**.

  - How could you deal with **missing** or vague ingredient information (e.g., “spices”)?

ANALYSING SIMILARITIES AND CLUSTERS.

- **Similarity Assessment**.

  - **How** would you **measure** similarity between two recipes based on their ingredients?

ANALYSING SIMILARITIES AND CLUSTERS.

- **Clustering Approaches**.
- **How** would you group recipes based on ingredient **similarity**?
- **What** criteria would you use to decide the number of **clusters** or **groups**?
- After clustering, how would you **name** or **describe** the clusters based on their ingredients?

COPILOT.

- We will use **CoPilot** to investigate our data.

COPILOT.

I have a file containing multiple individual recipes. Our goal is to analyse the similarities between these recipes and group them accordingly.

When doing so, you will need to account for inconsistencies in the data, such as variations in naming, typographical errors, and ingredient categories (e.g., fish-based or vegetable-based). Please create a comprehensive table summarizing the types of inconsistencies you will address.

After cleaning the data, perform Hierarchical Agglomerative Clustering on the recipes based on their ingredients, and show and describe these groups.

Then, generate a table showing the cosine similarity scores between recipes, based on their ingredients. Ultimately, I want the recipes grouped by theme and ingredient similarity.

Finally, provide a cosine similarity table listing the recipes, their cosine similarity scores, the cleansed ingredients, and the clusters they belong to. The table should include at least 15 entries; recipes may appear multiple times. Order the table by cosine similarity score in descending order and display the cluster names.

For the top five entries with the highest cosine similarity scores, investigate if any cultural or historical context—such as trade routes or significant world events—can be added to enrich the data.

KEY TAKEAWAYS.

- **Strengthens Data Cleansing Skills.**
  - **Learning** to identify and correct inconsistencies such as typos, synonyms, and formatting issues **enhances** your ability to prepare raw data for accurate analysis—an **essential** part of data literacy.
- **Improves Understanding of Data Standardization.**
  - **Normalizing** data by **unifying** ingredient names and grouping related items helps build **fluency** in creating **consistent** datasets, which is critical for **reliable comparisons** and insights.
- **Demonstrates the Importance of Data Representation.**
  - **Transforming** messy, **unstructured** information into **structured** formats (like ingredient vectors) teaches how to make real-world data computationally **accessible** and ready for analysis.
- **Highlights the Role of Similarity Measures in Data Grouping.**
  - Applying techniques like **cosine similarity** to clustered data fosters an understanding of how to **quantify relationships** and patterns within datasets, a key step after **data preparation**.