Alien Elements:
Exploring Data Science in Chemistry

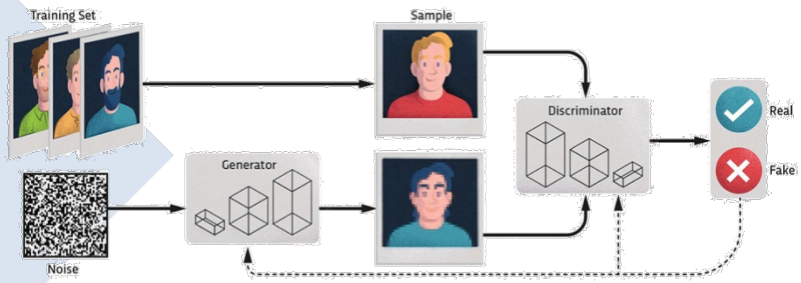
LEARNING UNDEFEATED
Educate | Elevate | Empower

Name: _____

The inception of the periodic table of elements can be traced back to the early 19th century when chemists like Dmitri Mendeleev and Lothar Meyer independently recognized recurring patterns in the properties of elements. In 1869, Mendeleev famously organized the elements into a table based on their atomic weights, grouping them by similarities in chemical properties and leaving gaps for yet-to-be-discovered elements. This groundbreaking arrangement paved the way for the prediction of the properties of elements not yet observed, demonstrating the remarkable power of systematic organization in the field of chemistry.

This analysis of the known elements and their properties would today fall into the field of data science. Data science is an interdisciplinary field that uses statistics, computing, and scientific methods to extract knowledge and insights from noisy data. Data science has ushered in a new era of innovation in chemistry. Chemists are increasingly relying on data-driven approaches to accelerate research and discovery. Data science techniques such as machine learning and deep learning are employed to analyze vast datasets of chemical information, predict molecular properties, and optimize chemical reactions.

One AI tool data scientists are expanding the use of is Generative Adversarial Networks (GANs). GANs are a class of artificial neural networks consisting of two interconnected components: a generator and a discriminator, engaged in a competitive learning process. The generator creates synthetic data, such as images, audio, or text, with the aim of mimicking real data distributions. Simultaneously, the discriminator evaluates the generated data's authenticity by distinguishing between real and synthetic examples.



While GANs have many uses in science when it comes to synthetic data, it's more commonly known for its applications with images and videos, specifically with deepfakes. This AI-generated content manipulates existing content to convincingly depict individuals saying or doing things they never did. For example, replacing a president's face with that of a famous actor. The U.S. Defense Advanced Research Projects Agency (DARPA) has created counter-AI programs to spot deepfakes. But the project continues as a key problem is that machine learning systems can be trained to outmaneuver these tools.

? *Quick start: What are the benefits of synthetic data in science?*

PART 1: DESCRIPTIVE ANALYTICS AND VISUALIZATION

Scientists are studying planet Dod. In doing so they have discovered 36 unique elements in the planet's atmosphere and on its surface. The collected data has been provided to the data scientist (you!) to help communicate the relationships between the elements and their properties.

Look for Unique Properties

1. Drag a column label from the table to the X-axis of the graph. Notice that this plots all 36 points according to this property.
2. Click on the X-axis label to swap it for another category or drag another column label to the X-axis.
3. Look for a graph that is mostly flat (i.e., has as few overlapping or stacked dots as possible).

? *Quick Check: What property is the most unique to each element?*

Find Patterns and Trends

4. Keep the unique property found above on the X-axis.
5. Add different properties to the Y-axis and see if there is a pattern or trend.

? *Quick Check: What properties show patterns in relation to your unique property?*

Visualizing the Data

1. Use the graph of Atomic Radius vs Atomic Weight to sort the cards into 4 rows on your center mat. Ensure each row is ordered left to right by atomic weight.

? *Quick Check: Does this create the same groupings we saw for electronegativity and ionization energy?*

2. Use the graph of RxOy (reactivity with oxygen) vs Atomic Weight to align elements into columns with similar reactivity without changing their position within their row.
3. Label your mat with the trends observed by going left to right on each row and from top to bottom on each column.

? *Quick Check: Periodic means occurring at intervals. Do we see patterns of periodicity with the alien elements?*

PART 2: FROM DESCRIPTIVE TO PREDICTIVE UTILIZING GANs

Scientists have found some compounds, molecules with more than one element, in the atmosphere and surface of Dod. To guide their search for more compounds, they will utilize a Generative Adversarial Network (GAN).

This tool is composed of two neural networks, a generator and a discriminator. The discriminator has been trained to recognize data, in this case, the compounds found on Dod. You will work as the generator to produce new compounds that could possibly exist. But keep in mind that the discriminator is looking for the same patterns as its dataset, so you'll need to learn from your responses to find possible answers.

❓ *Quick Check: Does the generator start with any information about the dataset it's trying to match?*

New Data

Some more properties for each element has been identified. To view them, follow these steps:

1. In CODAP, click on the table of data, then select the eye icon,  , from the menu on the right.
2. Select "Show 1 Hidden Attribute".
3. You should now see valence electrons for each element next to element symbol.

Generating New Compounds

4. On the discriminator input screen, Input one element symbol on the screen and hit enter.
5. Input the number of atoms of this element you'd like present in the subscript. Then hit enter.
6. Input another element symbol and hit enter.
7. Input the number of atoms of this element you'd like present in the subscript. Then hit enter.
8. Click "Send to Discriminator" to gather feedback.
9. Note your results each time in the table to the right to find patterns.

Generated Compound (Ex. I_2T_2)	Number of Valence Electrons	Did it *pass* or *fail* the discriminator?

PART 3: TIES TO CURRENT USES

GANs are a great tool to create synthetic data- in our case, more molecules that we do not have any physical evidence of. It can also be great for masking personal information from healthcare organizations. They can share the synthetic data to show the same trends and information without any information being directly tied to a person, therefore helping to protect their privacy.

Outside of science, GANs can also be used for image, video, and sound generation. These GANs are common in the news because of their use to create deepfakes.

? *Quick Check: What are the dangers of deepfakes?*