Data Visualisation & Analytics

(CIA1C11)

**EDA ASSIGNMENT**

**AY2023/2024 Apr Semester**

**DECLARATION OF ORIGINALITY**

I am the originator of this work and I have appropriately acknowledged all other original sources used as my reference for this work.

I understand that Plagiarism is the act of taking and using the whole or any part of another person’s work, including work generated by AI, and presenting it as my own.

I understand that Plagiarism is an academic offence and if I am found to have committed or abetted the offence of plagiarism in relation to this submitted work, disciplinary action will be enforced.

☑ I Agree (Please Tick ✓)

|  |  |
| --- | --- |
| **My Information** | |
| Name (as in matriculation card) | Shubham Kaushik |
| Admin Number | 2301418F |
| Class (E.g. P01) | P14 |

**Declaration on the use of Generative AI tools for assignments**

|  |
| --- |
| Describe how you have used Generative AI tools such as ChatGPT or Dall.E-2 in your assignment.  Show snapshots of the conversations with the AI tool (i.e., the prompts you used and the response you get from the AI tool). |
| **Didn’t use chat gpt.** |
| How to indicate the reference?  The content generated by AI tools are not retrievable except by the user who generated them, so they are considered non-recoverable sources. For non-recoverable sources:   * do not include in a reference list * cite within the text as personal communications or correspondence   Based on APA 7th edition referencing format,  (Communicator, personal communication, Month Day, year)  E.g. (Paraphrase from OpenAI's ChatGPT AI language model, personal communication, March 8­, 2023). |

**Important Note:**

* Do not copy answers produced by the AI tool in totality as it is considered as plagiarism.
* Do not rely on any information produced by the AI tool blindly. You should always verify the answer with other sources. Do not assume that these answers provided by the AI tool are correct.

* To achieve quality outputs from the AI tool, you should provide good prompt that is clear and specific. Be precise and provide context. Avoid asking open-ended questions.

# Truth or Myth?

1. The **first** belief I am investigating is: The larger the animal, the higher the GHG emissions per kg.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Conclusion: Based on data, this conventional belief is true

Justification:

A screenshot of a computer

Description automatically generated

A picture containing text, plot, software, line

Description automatically generated

Assumptions made:

Beef(beef herd) food product represents the animal cows. Shrimps(farmed) food product represents the animal shrimps.

Lamb and mutton food product represent the animals Lamb and Sheep. Fish(Farmed) food product represents the animal fish.

Pigs meat food product represents the animal Pigs. Poultry meat food product represents the animal birds.

To start off, I decided to use Bar Chart where the Food Product is the x-axis and the GHG gas emission per kg as the y-axis. However, there was an issue Broad cat contained both plants and animals in the rows. However, using the Food Product, I wanted to only focus on the animal food products. Hence, I decided to use a row filter to filter out the plants from the x-axis and focused on the rows where the Broad cat was animals.

From the Bar chart, we are able to see that Beef (beef herd) has 59 kgCO2eq per kg GHG gas emissions. Beef dairy herd had 21.1 kgCO2eq per kg GHG gas emissions. Lamb and mutton had 24.5 kgCO2eq per kg GHG gas emissions. Pig meat had 7.2 kgCO2eq per kg GHG gas emissions. Poultry meat had 6.1 kgCO2eq per kg GHG cash emissions. Milk had 2.8 kgCO2eq per kg GHG gas emissions. Cheese had 21.1 kgCO2eq per kg GHG gas emissions. Egg had 4.5 kgCO2eq per kg GHG gas emissions. Fish (Farmed) had 5.1 kgCO2eq per kg GHG gas emissions. Shrimp farmed had 11.8 kgCO2eq per kg GHG gas emissions.

Beef(beef herd) food product represents the animal cows. Shrimps(farmed) food product represents the animal shrimps. Cows are much more bigger than shrimps which explains why cow’s GHG gas emissions are much higher. Beef (beef herd) has 59 kgCO2eq per kg GHG gas emissions. Shrimp farmed had 11.8 kgCO2eq per kg GHG gas emissions. This proves the idea that The larger the animal, the higher the GHG emissions per kg.

Lamb and mutton food product represent the animals Lamb and Sheep. Fish(Farmed) food product represents the animal fish. Fishes are much smaller than Lambs and Sheep which explains why Lamb and Sheep was much higher than Fish’s GHG gas emissions. Lamb and sheep had 24.5 kgCO2eq per kg GHG gas emissions. Whereas, the animal Fish had 5.1 kgCO2eq per kg GHG gas emissions.

Pigs meat food product represents the animal Pigs. Poultry meat food product represents the animal birds. Birds are much smaller than pigs which explains why pigs have a higher GHG gas emissions than Birds. Pig meat had 7.2 kgCO2eq per kg GHG gas emissions. Poultry meat had 6.1 kgCO2eq per kg GHG cash emissions

Hence, with the interpretations and comparisons made prior. I am able to conclude that; **The larger the animal, the higher the GHG emissions per kg.**

A picture containing screenshot, text, plot, rectangle

Description automatically generated

Assumptions made:

Beef(beef herd) food product represents the animal cows. Shrimps(farmed) food product represents the animal shrimps. Lamb and mutton food product represent the animals Lamb and Sheep. Fish(Farmed) food product represents the animal fish. Pigs meat food product represents the animal Pigs. Poultry meat food product represents the animal birds.

I view the cows, lambs & sheep, pigs as large animals. Whereas, I view birds, milk, cheese, eggs, fish, and shrimps as small animals. The reason why I am grouping the animals into big and small, is because shrimp GHG gas emissions being relatively higher. This is in order to show the difference and further prove the trend.

Beef (beef herd) had 1.9 GHG gas emissions, Beef(dairy herd) had 2.5 kgCO2eq per kg GHG gas emissions per kg, Lamb& Sheep had 2.4 kgCO2eq per kg GHG gas emissions per kg, Pig had 2.9 kgCO2eq per kg GHG gas emissions per kg, birds had 1.8 kgCO2eq per kg GHG gas emissions per kg, Milk had 0.2 kgCO2eq per kg GHG gas emission per kg, Cheese had 2.3 kgCO2eq per kg GHG gas emission per kg, Eggs had 2.2 kgCO2eq per kg GHG gas emission per kg, Fish had 0.8 kgCO2eq per kg GHG gas emission per kg, Shrimps had 2.5 kgCO2eq per kg GHG gas emission per kg.

The average GHG gas emission for big animals is 2.6 kgCO2eq per kg and for small animals the average 1.63 kgCO2eq per kg. This further proved my point that; The larger the animal, the higher the GHG emissions per kg. This because the food that is processed for the big animals (animal feed) to eat resulted in higher GHG gas emission per kg compared to the type of food that the small animals(animal feed) ate.

File Table:

A screenshot of a computer

Description automatically generated

Bar Chart: (Below)

A screenshot of a graph

Description automatically generated with medium confidence

**This is a Bar Chart where the Y-axis is the total Greenhouse gas emission emitted to produce 1 kg of product. X-axis is the Sub cat. We will be focussing on Dairy & Eggs, Pig, Poultry, Red meat, Sea food. Using the file table, I am able to assume that sub cats; sea food is referring to shrimps, red meat refers to sheeps, lambs and cows(beef herd). Dairy & egg refers to chickens and cows(dairy herd). Poultry refers to birds. Lastly, sub cat pig refers to the animals pigs. Armed with this knowledge, I begin comparing the Dairy & eggs, Pig, Poultry, Red meat and Seafood sub cat. I came to the realisation that the bigger the animal sub cat, the greater the GHG emissions. Animals that are used from red meat and Dairy & Eggs are bigger than animals listed under Poultry and Seafood. Hence, from the histogram, I am able to read that the sub cat red meat and Dairy & Eggs generate 42.05kgCO2eq per kilogram and 12.4kgCO2eq per kilogram of GHG emissions, respectively. Whereas, the sub cat seafood and poultry generated 8.45kgCO2eq per kilogram and 6.1kgCO2eq per kilogram of GHG emissions.**

1. **second** belief I am investigating is: The higher the land use per kg, the higher the GHG emissions per kg.

Conclusion: Based on data, this conventional belief is true.

Justification:

**A screenshot of a computer

Description automatically generated with medium confidence**

**The above diagram is an interactive histogram where the y-axis is greenhouse gas emissions per KG and the x-axis are bins in ranges about land use change per kg (**kgCO2eq per kilogram**)**

**W are able to spit that from the first bin to the fourth bin, as land use change per kg increased, so did the greenhouse gas emissions per KG. However, the from the 4th bin to the 5th bin, greenhouse gas emissions per KG dropped drastically, from 59.6 to 24.5** kgCO2eq per kilogram. However, the 5th bin GHG gas emission is still larger than the ghg gas emissions for the 2th and 3rd bin. **This supports the idea that The higher the land use per kg, the higher the GHG emissions per kg.**

**A screenshot of a computer

Description automatically generated**

**I used Group By node and filtered two columns; Land use change and Greenhouse gas Emissions per KG. I set both the columns to ascending values. We can see on the left column, as the land use change value increased from -2.1** kgCO2eq per kilogram to 16.3 kgCO2eq per kilogram**, the corresponding values of the GHG emissions(on the right column) increased from 0.2** kgCO2eq per kilogram **to 59.6** kgCO2eq per kilogram**. I realised that spotting a trend was not easy using numbers(GroupBy node). Hence, for the sake of Data Visualisation, I decided to use a scatter plot.**

**A screenshot of a computer

Description automatically generated with medium confidence**

**This is a scatter plot, where the x-axis is where the y-axis is Greenhouse Gas Emissions per KG and the x-axis are bins in ranges about land use change(**kgCO2eq per kilogram**)**

**Using the X-axis from -1.101 to around 1.200, most of the points are cluttered in the area. Generally, we are able to see a trend that the greater the land use, the higher the GHG emissions per KG. I was able to spot 2 outliers, they were really far away from rest of the points. One outlier at around 14.200 of land use change(**kgCO2eq per kilogram) had around 18.1 GHG per KG. The outlier at around 16.000**(**kgCO2eq per kilogram) and that corresponded to around 59.6 GHG emission per KG .

**A screenshot of a computer

Description automatically generated with medium confidence**

**I, hence, decided to draw a best fit line on the scatter plot. This helped me to further prove my point that the general trend is that of increasing. The belief “**The higher the land use per kg, the higher the GHG emissions per kg.” is true.

1. The **third** belief I am investigating is: For every 1000 kcal of energy provided to human, plant-based food has lower GHG emissions per 1000kcal than animal-based food.

Conclusion: Based on data, this conventional belief is true

Justification:

**A screenshot of a computer

Description automatically generated with medium confidence**

**I decided to use to conditional box plot, where the. Y axis is configured as greenhouse gas emissions and the X axis is configured as Broad Cat.**

**The category animal under Broad Cat, has UQR of 12.5 , LQR of 5.25, IQR = 7.25, Median of 6.89, Minimum of 3.24, maximum of 12.53.**

**Whereas, the category plant under Broad Cat, has UQR of 1.35 , LQR of 0.52, IQR = 0.83, Median of 0.93 ,Minimum of 0.07, maximum of 1.43**

**From the summary statistics, I am able to learn that animal have a greater GHG emission. This is because, the minimum GHG emissions of animals is 3.24. While the maximum GHG emissions of plants is 1.43. Additionally, the IQR of animals is 7.25 and the IQR of animals is 0.83. This supports the idea that the values of GHG emissions of plants are more closer and consistent than the values of GHG emissions of animals.**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Now that I had found the IQR and Median. I decided to use Interactive histogram to find the mean GHG emissions per 1000kcal of energy. The binning column is set as Broad Cat and the aggregation column is GHG emissions per 1000kcal. Average GHG emissions per 1000kcal is 12 kgCO2eq whereas, the average GHG emission per 1000kcal is 3.37 kgCO2eq. Hence, this proves the point that for every 1000kcal, on average, animals generate higher amounts of GHG emissions.**

A screenshot of a graph

Description automatically generated with medium confidence

Land Use change it is defined How much GHG is emitted to produce 1 kg of product during the land use change stage. This includes above- ground changes in biomass from deforestation, and below ground changes in soil carbon. The graph above is a Bar Chart where the y-axis being the land use change while the x-axis being the Broad Cat. For animals, the land use change value was 2.81000000000000005 while for animals the land use change value was 0.7909090909090909. Since, farming animals result in a higher Land use change compared to plants. This further explains one of the reasons why for every 100kcal of animal-based food is consumed has higher GHG gas emissions than plant-based food. I would like to add on that the land where animals are being used to farm animals could have previously could have been filled with trees and plants which photosynthesis, taking and in carbon dioxide and giving off oxygen. This explains the idea that farming animals and instead of growing crops or letting plants grow, there could have been net negative of GHG gas emitted. Instead of giving off GHG gas emissions, plants/trees would have taken in carbon dioxide which would have converted into oxygen via photosynthesis.

# Screenshot of my overall KNIME workflow is shown below.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| SUBMISSION INSTRUCTIONS   * Save this file as “EDA\_[Your Class]\_[Your Full Name]\_[Your Student ID]”, e.g. “***EDA \_P01\_ALBERT EINSTEIN\_2309999A***”. * Submit your report, together with your KNIME (.knwf) file, to DAVA LMS site > Assessment > EDA Assignment > Look for the appropriate link to submit your work. |

\*\*\*\*\* END OF EDA ASSIGNMENT \*\*\*\*\*