**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Programming for DA  Statistics for Data Analytics  Machine Learning for Data Analysis  Data Preparation & Visualisation |
| **Assessment Title:** | MSC\_DA\_CA2 |
| **Lecturer Name:** | Marina Iantorno/John O’Sullivan  Sam Weiss  Muhammad Iqbal  David McQuaid |
| **Student Full Name:** | Jan Andersson |
| **Student Number:** | sba20368 |
| **Assessment Due Date:** | 6th January 2023 |
| **Date of Submission:** |  |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

Contents

[Introduction 3](#_Toc123893908)

[Project aims 4](#_Toc123893909)

[About the datasets 5](#_Toc123893910)

[Irish Cattle Births 5](#_Toc123893911)

[Irish Dairy Production 5](#_Toc123893912)

[EU Dairy cattle supply 5](#_Toc123893913)

[References 6](#_Toc123893914)

# Introduction

Irish dairy farming is based on a system of small, family-owned farms. The majority of Irish dairy farmers are still working the land in this way and producing milk for their own consumption. The main reason why Irish dairy farmers have not diversified into other business areas is because it’s too expensive to do so. There are only very few large dairies that can afford to buy directly from small scale farms; and even then, they don’t always get what they want as some farmers prefer to sell direct rather than through a middleman who would take a percentage off the top – which often means that consumers end up paying more for their milk products than they should be doing. Irish dairy farmers are very proud of their products and they want to make sure that consumers know exactly what’s in the product.

They also want to be able to sell directly to the consumer; and this means that Irish dairy farmers have a great deal of control over their own business. The main problem with direct selling is that it can lead to higher prices for consumers, so some people would prefer if there was a way for small scale farms such as these to sell into larger markets without having all of their costs passed on through middlemen. This is where dairying cooperatives come in – Irish cooperative businesses work by pooling together money from different members who then invest it in order to get more milk production going. The result is that you end up with a much cheaper product than you would otherwise be getting, but you still get quality because each farmer has an equal say in how things should run and they must all abide by the same rules when it comes down to selling direct or selling through middlemen. It’s not uncommon for Irish farmers who have joined an agricultural cooperative such as this one – called Coillte Teoranta (the National Dairy Board) – will receive around €50 per tonne more per year than those who produce milk themselves but don’t use any other means of distribution besides retail outlets like supermarkets or shops; and many families see this kind of extra cash coming into the farmhouse every month as enough reason alone for them not only continue milking cows but also keep doing so through good times and bad! .

It’s said that Ireland is one of the top 10 dairy producing countries in the world, and it’s not hard to see why. The country has a long history of co-operatives; this kind of business structure was originally set up by small farmers in the 19th century who wanted to be able to sell their milk directly to consumers for a fair price. It grew from there, with people joining together over time and forming bigger co-ops which would then go on to provide services such as insurance, banking and even legal advice. However, things took a turn for the worse when supermarkets began taking over more and more aspects of our food supply chain – they started buying milk direct from farms at much lower prices than what farmers were being paid before going through middlemen like dairies or cooperatives.

***[535 Words]***

# Project aims

The use of Data Analytics may be quite beneficial in dairy farming. It is beneficial to farmers for a variety of reasons, including:

1. Estimating milk output. It is simpler to arrange your milking routine if you know how much milk your cows make each day and how much they will provide over time. This enables you to improve cow wellness while avoiding having too few or too many cows available during any particular moment.
2. Recognizing issues with your herd. Identifying what's incorrect with your herd, for instance when an animal isn't giving enough dairy, allows you to address concerns before they become larger ones with long-term ramifications for the overall farm business if left unchecked. You may even be able to avoid them from happening in the future by carefully following their activity
3. Herd management. Understanding what really is going on with your cattle might assist you in determining how to govern your herd the best. For instance, if a cow isn't providing enough dairy, it may be time to give her a dry spell or have her bred so she may produce more milk in the future. This might save you cost while also providing you with higher-quality dairy products.
4. Applying data analytics to boost efficiency and productivity. Data analytics may also allow farmers make better judgements regarding their livestock and their business as a whole, which can enhance overall farm operations. Using this type of technology, farmers can track the condition of their pastures over extended periods of time, enabling them to make informed choices while still on the farm, rather than having those decisions made wirelessly from afar by an individual who may not notice all of the specifics of each circumstance at this time.
5. Making better choices. Data analytics may also be utilised to assist farmers in making better informed decisions regarding their livestock and their whole enterprise. Farmers can far more easily monitor the health of their cattle over lengthy periods of time by employing this sort of equipment.
6. Asset management. This is highly relevant if you have a lot going on with your herd. For instance, if you have an old cow that has begun generating less dairy over time or one that is becoming ill due to poor health, data analytics can help you evaluate if she ought to be put down or managed to sell off so you wouldn't make a loss.

***[406 Words]***

# About the datasets

### Irish Cattle Births

### 

### Irish Dairy Production

### 

### EU Dairy cattle supply

#### Country Key ID's for Numeric Dataframe

This data was acquired from the EU international database of public statistics, I have chosen it as it is the most reliable source and has the most relevant and up to date historical data.

I have supplied the country codes for reference with use with the numeric only dataframe of EU Dairy cattle stocks.

|  |  |
| --- | --- |
| * 0: Austria * 1: Belgium * 2: Bulgaria * 3: Croatia * 4: Cyprus * 5: Czechia * 6: Denmark * 7: Estonia * 8: Finland * 9: France * 10: Germany * 11: Greece * 12: Hungary | * 13: Ireland * 14: Italy * 15: Lithuania * 16: Luxembourg * 17: Malta * 19: Netherlands * 20: Poland * 21: Portugal * 22: Romania * 23: Slovakia * 24: Slovenia * 25: Spain * 26: Sweden |

As we can see the data for the EU countries is formatted where the year is the column featuring rows with the corresponding countries within the EU. I have removed the unnecessary and unformatted data as features such as Unnamed: 0 have no use in any data analysis without the documentation that is not present in this case.

I then created a separated dataframe which contained the EU dairy cattle distribution and formatted it to remove special characters any values that contained strings and converted the whole dataframe to an integer. As the year and country wasn't in the correct format, I transposed the dataframe to make the feature selection the countries within the dataframe rather than the year.

Here I have displayed the averages for the selected countries to get a better understanding of what country might have a similar distribution of the cattle quantity.

|  |  |
| --- | --- |
| 0 557.826087  1 548.695652  2 320.913043  3 103.608696  4 26.260870  5 416.695652  6 585.260870  7 105.652174  8 301.913043  9 3830.695652  10 4267.173913  11 136.956522  12 279.956522  13 1155.869565  14 1927.869565 | 15 175.217391  16 369.086957  17 46.391304  18 5.956522  19 1568.565217  20 2570.086957  21 269.608696  22 1104.304348  23 176.739130  24 118.826087  25 948.565217  26 368.521739 |

***[200 Words]***

# References

Awojide, Margaret. “Statistics for Data Analysts: Inferential Statistics with Python.” *CodeX*, 14 Sept. 2022, medium.com/codex/statistics-for-data-analysts-inferential-statistics-with-python-de8b7f49cfa. Accessed 5 Jan. 2023.

Hirsch, Stefan, et al. “Dataset on the Dairy Processing Industry in France, Italy, Poland and Spain.” *Www.research-Collection.ethz.ch*, 2019, www.research-collection.ethz.ch/handle/20.500.11850/333174, 10.3929/ethz-b-000333174. Accessed 4 Jan. 2023.

“Python Pandas - Descriptive Statistics - Tutorialspoint.” *Www.tutorialspoint.com*, www.tutorialspoint.com/python\_pandas/python\_pandas\_descriptive\_statistics.htm.

“Python Z Test | When to Perform Z Test in Python with Examples?” *EDUCBA*, 13 Feb. 2022, www.educba.com/python-z-test/.

Shane. “Delete Rows & Columns in DataFrames Using Pandas Drop.” *Www.shanelynn.ie*, www.shanelynn.ie/pandas-drop-delete-dataframe-rows-columns/.

Stojiljković, Mirko. “Python Statistics Fundamentals: How to Describe Your Data – Real Python.” *Realpython.com*, realpython.com/python-statistics/.

Thu, et al. “Irish Dairy’s Elevated Position across Range of Global Metrics.” *Irish Examiner*, 7 Oct. 2021, www.irishexaminer.com/farming/arid-40715282.html.

“UK and EU Cow Numbers | AHDB.” *Ahdb.org.uk*, 2019, ahdb.org.uk/dairy/uk-and-eu-cow-numbers.

Zach. “How to Calculate Confidence Intervals in Python.” *Statology*, 16 July 2020, www.statology.org/confidence-intervals-python/.

---. “Pandas: How to Remove Special Characters from Column.” *Statology*, 10 Oct. 2022, www.statology.org/pandas-remove-special-characters/. Accessed 4 Jan. 2023.