**A Data based Analysis of the Irish Dairy sector and how it compares to its European Counterparts**

Author: Teresa Quain

Student ID:  [sba22235](mailto:sba22235@student.cct.ie)

Student email: [sba22235@student.cct.ie](mailto:sba22235@student.cct.ie)

CA2 50% Integrated Assessment, MSc in Data Analytics

Github address*:* [*https://github.com/teresaq1/Agricultural-Dairy-Milk-Data-Analysis*](https://github.com/teresaq1/Agricultural-Dairy-Milk-Data-Analysis)

Wordcount: 3841



Contents

[Table of Figures 2](#_Toc123844547)

[Tables 2](#_Toc123844548)

[Abstract 2](#_Toc123844549)

[Introduction 3](#_Toc123844550)

[Methodology 4](#_Toc123844551)

[Software 4](#_Toc123844552)

[Dataset 1, Trends in Bovine data, 2011 to 2021 in In[38] 4](#_Toc123844553)

[Dataset 2, Milk Sales, 1980 to 2021 in In[47] 5](#_Toc123844554)

[Dataset 3, Milk Prices, 2007 to 2021 in In[54] 6](#_Toc123844555)

[Dataset 4, Milk financials, 2007 to 2021 in In[56] 7](#_Toc123844556)

[Dataset 5, Dairy Production, 2010 to 2022 in In[69] 8](#_Toc123844557)

[Dataset 6, EU Dairy Production, 1990 to 2022 9](#_Toc123844558)

[Plotly, Dash Visualisations 11](#_Toc123844559)

[Times-Series Forecasting in In [89] 13](#_Toc123844560)

[Decomposition 13](#_Toc123844561)

[Autocorrelation and Stationarity testing 14](#_Toc123844562)

[Autofitting and Forecasting of the model 15](#_Toc123844563)

[Conclusion 20](#_Toc123844564)

[Acknowledgements 20](#_Toc123844565)

[References 20](#_Toc123844566)

[Appendices 22](#_Toc123844567)

[Appendix 1 Seasonal production of Butter and Skimmed milk powder in Ireland 22](#_Toc123844568)

[Appendix 2 European Butter and Skimmed Milk produce 24](#_Toc123844569)

[Appendix 3 Time series calculations 26](#_Toc123844570)

[Appendix 4 Additional Visualisations 29](#_Toc123844571)

# Table of Figures

[Figure 1 Trends in Bovine Data 2011-21 4](#_Toc123845648)

[Figure 2 All Milk sales, dataset 2 5](file:///C:\Users\teresa\Documents\cct\data%20viz\ca2\data%20source\New%20folder\Data%20based%20Analysis%20of%20the%20Irish%20Dairy%20sector%20and%20how%20it%20compares%20to%20its%20European%20Counterparts.docx#_Toc123845649)

[Figure 3 Skimmed milk sales 1997-20 6](file:///C:\Users\teresa\Documents\cct\data%20viz\ca2\data%20source\New%20folder\Data%20based%20Analysis%20of%20the%20Irish%20Dairy%20sector%20and%20how%20it%20compares%20to%20its%20European%20Counterparts.docx#_Toc123845650)

[Figure 4 Manufacturing milk prices 2007-22 7](#_Toc123845651)

[Figure 5 Milk sales vs. price scatterplot 7](#_Toc123845652)

[Figure 6 Irish cheese production 2010-22 ln [47] 8](#_Toc123845653)

[Figure 7 Milk production by country in In [17] 9](file:///C:\Users\teresa\Documents\cct\data%20viz\ca2\data%20source\New%20folder\Data%20based%20Analysis%20of%20the%20Irish%20Dairy%20sector%20and%20how%20it%20compares%20to%20its%20European%20Counterparts.docx#_Toc123845654)

[Figure 8 Milk production chloropleth map 2020 in In [24] 10](#_Toc123845655)

[Figure 9 Cheese production by country in In [27] 10](#_Toc123845656)

[Figure 10 Cheese production chloropleth map 2020 in In [35] 11](#_Toc123845657)

[Figure 11 Sire breed In [115] 11](#_Toc123845658)

[Figure 12 Milk sales in Ireland In [159] 12](#_Toc123845659)

[Figure 13 EU Milk produce In [144] 12](#_Toc123845660)

[Figure 14 Decomposition of time series in In [89] 13](#_Toc123845661)

[Figure 15 ADF test on milk sales 1980-22 in In [90] 14](#_Toc123845662)

[Figure 16 ACF testing of time series data in In [92] 15](#_Toc123845663)

[Figure 17 Visualisation of training + test dataset in In [98] 16](#_Toc123845664)

[Figure 18 Forecast of time series Milk data in In [99] 17](#_Toc123845665)

[Figure 19 Diagnostics of fitted model In [112] 19](#_Toc123845666)

[Figure 20 Irish butter production 2010-20 in In [70] 22](#_Toc123845667)

[Figure 21 Irish skimmed milk production 2010-21 in In [71] 23](#_Toc123845668)

[Figure 22 Butter production by country in In [30] 24](#_Toc123845669)

[Figure 23 Butter production chloropleth map 2020 24](#_Toc123845670)

[Figure 24 Skimmed Milk produce by country 25](#_Toc123845671)

[Figure 25 Skimmed milk production chloropleth map 2020 25](#_Toc123845672)

[Figure 26 Stationarity testing with differencing in In [93] 26](#_Toc123845673)

[Figure 27 Stationarity testing of 1st difference of seasonal plot in In [95] 27](#_Toc123845674)

[Figure 28 Stationarity visualisation 27](#_Toc123845675)

[Figure 29 Decomposition after one hot encoding 28](#_Toc123845676)

[Figure 30 Overview Dashboard of Irish + European Dairy produce 29](#_Toc123845677)

[Figure 31 Github desktop screenprint 30](#_Toc123845678)

# Tables

[Table 1 Shapiro Wilks test, dataset 1 In [41] 5](#_Toc123844602)

[Table 2 Cheese production in Ireland ln[50] 9](#_Toc123844603)

[Table 3 Auto Arima model results in In [97] 16](#_Toc123844604)

[Table 4 SARIMAX Results In [111] 18](#_Toc123844605)

# Abstract

*Ireland’s entry to the EU in 1973 transformed the Irish agricultural sector. The implications of EU membership for the Irish dairy included evolution of milk prices, dairy production and new opportunities for dairy product exports. A combination of these circumstances led to the introduction of the milk quota system at the EU level in 1984 and eventually being abolished on 1 April 2005, nothing to do with April fool’s day. In general dairy produce demand is rising annually 3-4%, and Ireland (like New Zealand) is unique for our climate. This climate allows good quality grass to be grown easily keeping the price of milk comparatively low compared to say France or Germany. The risk of no quota at all is of course that the market may become flooded, causing the prices to drop and profit margins to fall, pushing smaller farmers out of the sector. This project focuses on Irish data sourced from the central statistics website and comparing it to European dairy production, sales and prices of milk, butter, cheese and skimmed milk powder. A deep analysis using python and powerful visualisation tools such as seaborn and matplotlib through the attached Jupyter notebook have supported the investigation and comparison of agricultural findings. Version control has been completed with Github address: https://github.com/teresaq1/Agricultural-Dairy-Milk-Data-Analysis*

*This type of analysis can also be a valuable informative tool for farmers and service providers to prevent unnecessary expenditure and support modern methods and products aligned with public dietary requirements.*

# Introduction

This report is focused on the subject of dairy produce in Ireland and abroad, how much is being produced, how much is being sold and for how much money? Datasets were sourced from the central statistics office website, the national Irish data site (data.gov.ie) and the Eurostat website, (ec.europa.eu). The data tends from 1980 to 2022, the present year. Ireland is one of the few EU member states which is rearing the majority of EU livestock. Spain, Germany, France, Italy and Poland rear the majority of the EU’s pigs, bovines, sheep and goats, while Denmark, Netherland and Greece are specialised in pigs, sheep and goats. In 2015 according to the Eurostat livestock population, Ireland accounted for 9% of the EU’s bovine population, impressive since we counted for just under 1% of the human population!

In the attached jupyter file ‘CA2\_jupyter\_analysis ‘, confidence intervals were calculated for price/production demand. Anova analysis allowed the comparison of Irish bovine populations on a statistical level. Choropleth mapping allow the data visualisation to be viewed on a geographical level. Sales of milk forecasting was performed using a time series statistician approach with an ARIMA model. Plotly and Dash were used to visualize the quantitve and categorical information over the time period. One hot encoding was used to convert the categorical variables and Lowes method of nonlinear regression was employed for the purposes of forecasting. The demand of bovine milk will naturally have an effect on the production other dairy products.

The colour scheme for the document was focused on a simple idea of monochrome for dairy cow produce, green colours for the Irish landscape, with some exceptions for the colours chosen for the categorical visualisations to ensure clarity for the reader. Reference to code lines are included throughout this report.

# Methodology

## Software

The data analysis was performed largely through python in the attached jupyter notebook (CA2\_jupyter\_analysis.ipynb). Libraries included pandas, numpy, datetime, sklearn and math. From sklearn libraries normalize, train\_test\_split, StandardScaler, skmetrics. Visualisation tools such as seaborn,plotly, dash and matplotlib aided the plots produced throughout this report.

Five datasets were loaded in In[13], four Irish and one European in order to cover as many aspects of the lifecycle of dairy products from the farmers field to the supermarket between 1980 and 2022. The Irish datasets are based around bovine breeding trends, milk sales, (whole/skimmed and semi-skimmed), milk price trends and the production of dairy products such as cheese, butter and skimmed milk powder. The European dataset contains data about European produce for cheese, butter and skimmed milk powder which was used to compare against the Irish data.

### Dataset 1, Trends in Bovine data, 2011 to 2021 in In[38]

(*df\_trends\_bovine\_data*). This table was published by Department of Agriculture, Food and Marine and contains data from 2011-2021 on: calf births categorised as beef and dairy. They are classified as such depending on whether the sire breed type is a Beef or Dairy Breed. Calf births broken down by the breed of the sire. It contains data on the top 6 breeds only as these account for approx. 90% of all births. The data was melted using a panda’s function to allow a seaborn bar plot to be fitted to the table. Organising the data by sire breed product fig. x below. In Ireland about 90% of all dairy animals are Friesian breed. It’s consistent milk yields and high butterfat content ensure this black and white animal remains the most popular with farmers. Colours for this graph are by sire breed.

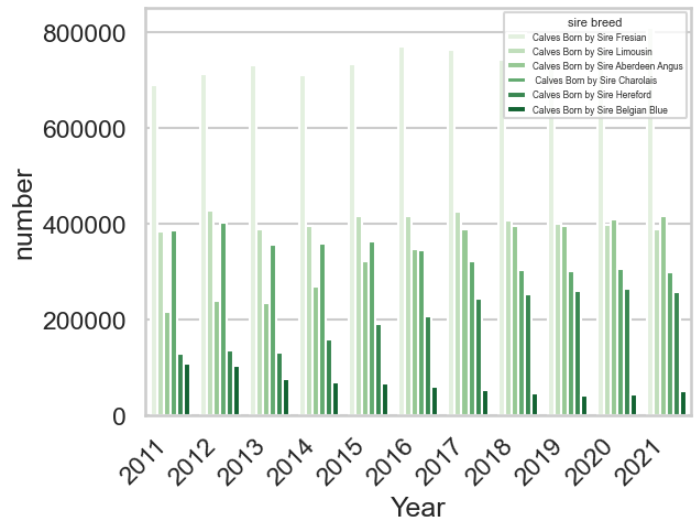


Figure 1 Trends in Bovine Data 2011-21

Shapiro Wilks test performed on dataset 1 showed a normal distribution, and a p value of 0.003, well below 0.05 for a 95% confidence interval, while Levene’s test produced a p value of 0.05 indicating the variances are not significantly different from each other. This, combined with the knowledge that our sire breeds are independent from each other fulfil the criteria for Anova execution. Performing a one-way Anova analysis on the melted dataset 1 produced the results in table 1 below. The p value of 1, is greater than 0.05 hence we reject the null hypothesis, H0 that there is no difference between the means and accept the alternative hypothesis H1.

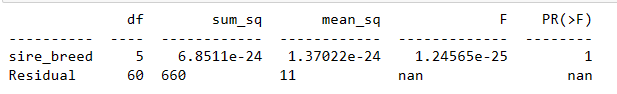
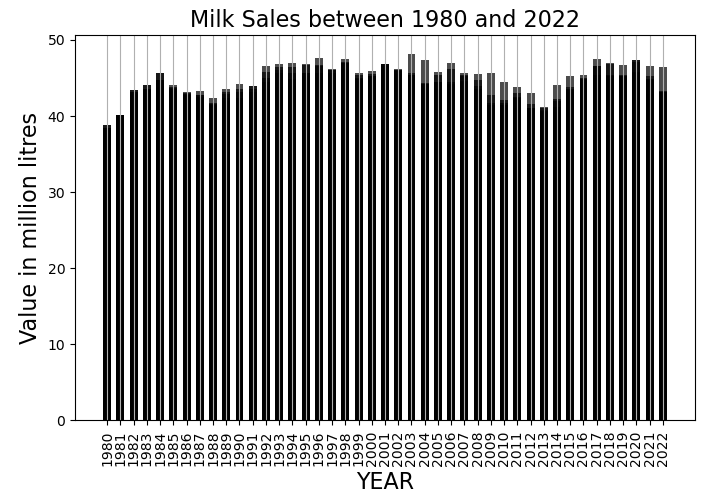


Table 1 Shapiro Wilks test, dataset 1 In [41]

### Dataset 2, Milk Sales, 1980 to 2021 in In[47]

(*df\_milk\_sales8021*). Dataset 2 was sourced from the central statistics office website. It details the quantity of 3 types of milk sold (Whole, skimmed and semi-skimmed) between 1980 and 2021 by month. Quantity sold unit is in million litres. This data was filtered and plotted in matplotlib via bar charts to aid the visualisation of these quantities. As the population of Ireland is naturally increasing, it was 3.4 million in 1980 and 5 million in 2021 according to census data, the demand for milk has slightly increased from less than 40 million litres in 1980 to between 40 to 50 million litres in 2022, but remained relatively static.



Lower fat varieties of milk such as skimmed and semi-skimmed have become more popular over time in Ireland as people are becoming more diet conscious. Figure 3 shows the whole milk sales decreasing from a peak of over 40 million litres in the early 1980’s to a low of approx. 25 million litres in 2013-2014. Skimmed milk and semi-skimmed milk have enjoyed an increased popularity year on year.

Figure 2 All Milk sales, dataset 2

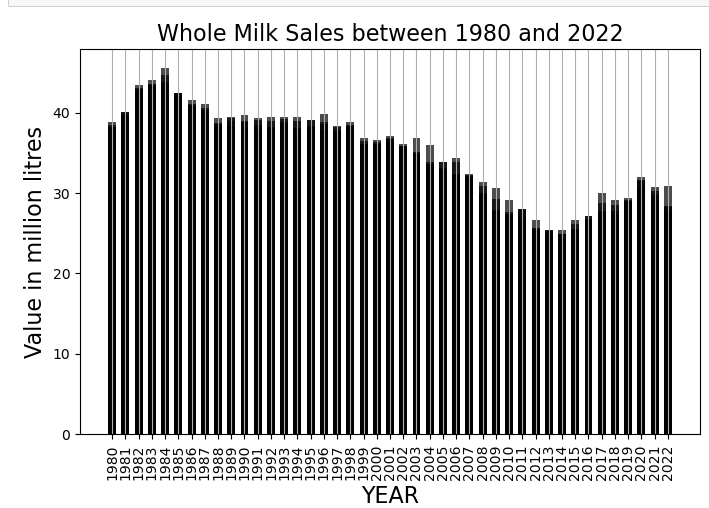
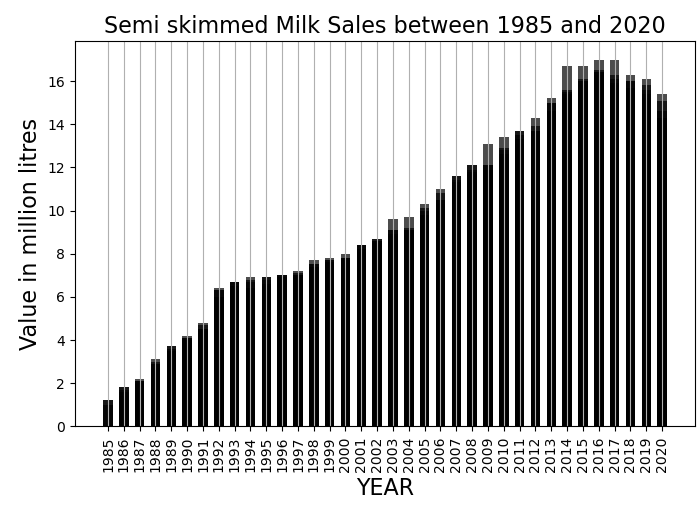


Figure 4 Semi skimmed milk 1985-20

Figure 3 Whole milk sales 1980-22

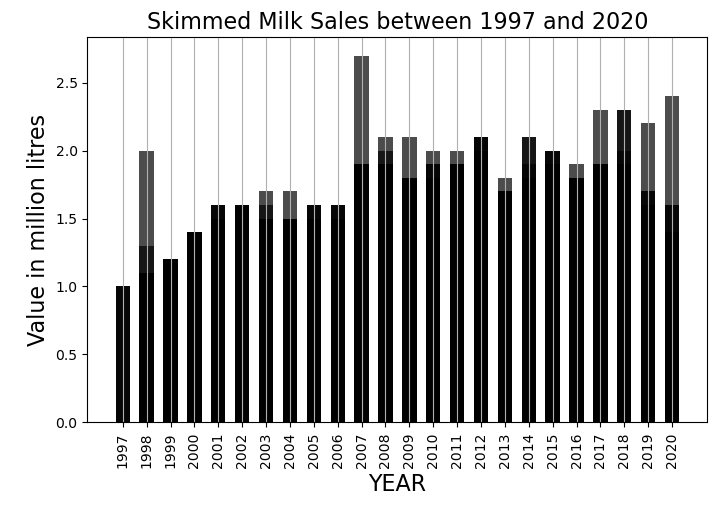


Figure 3 Skimmed milk sales 1997-20

Skimmed milk has increased from 1 million litres sold in 1997 to 2 million in 2020. Semi skimmed milk shows a steady linear upward trend from 2 million in 1985 to 16 million litres sold in 2014-15.

### Dataset 3, Milk Prices, 2007 to 2021 in In[54]

(*df\_milk\_prices0721*). Dataset 3 contains information about the prices of manufacturing milk between 2007 and 2022 in euros per litre, also sourced from the central statistics office website. The two milk products included are 1. Milk (per litre), actual fat and protein and 2. Milk (per litre), 3.7% fat and 3.3% protein. This price increased and reduced over the years, the lowest was in 2009 with a value of between 20 to 25c and the highest was in 2021 between 35 to 40 cents.

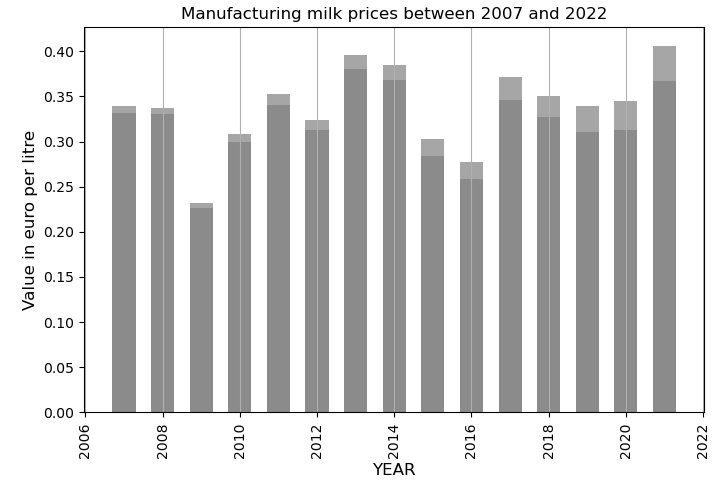


Figure 4 Manufacturing milk prices 2007-22

### Dataset 4, Milk financials, 2007 to 2021 in In[56]

Merging dataset 2, sales and dataset 3, manufacturing prices on the ‘year’ column allowed us to check for correlation in dataset 4, ‘*df\_milkfinancials072’*. The milk sales data was filtered on ‘whole milk’ only to allow for comparison between 2007 and 2021 on the manufacturing price dataset for ‘Milk (per litre), actual fat and protein’. Seaborn join plot in figure 7 below shows the data is normally distributed with a negative slope, indicating that as manufacturing price increases, milk price potentially decreases and sales decrease. Correlation coefficient, Pearson’s R value was calculated at 0.28, and a p value of 0.129.

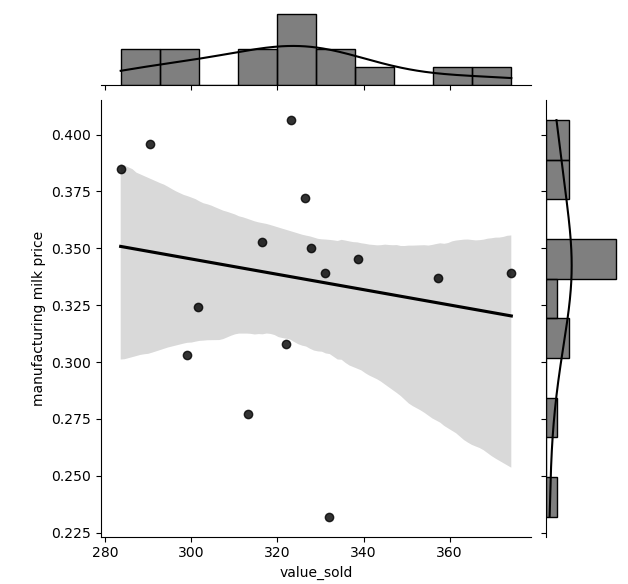


Figure 5 Milk sales vs. price scatterplot

### Dataset 5, Dairy Production, 2010 to 2022 in In[69]

(*df\_irish\_cheese1022, df\_irish\_butter1022, df\_irish\_skimmedmilkpowder1022*)

Dataset 5 contains 3 sets of csv files, with information about Irish dairy production between 2010 and 2022 for cheese, butter and skimmed milk powder respectively. 26 null values were identified for skimmed milk powder production in the ‘Value’ column for quantity of 000’s tonnes produced per month. Since the aim of processing these datasets was for comparative reasons, the null values were supplemented with a mean value assuming there is a certain amount produced every year.

Dataset 5, was subset per product using a panda’s loc function and melted using the panda’s melt function. This allows the user to see production throughout the year for each product between 2010 and 2022. Four columns of year were selected: 2010,2014,2018 and 2022 to allow the graph to be more easily read. From the 3 figures below, we can see an increase in production in the summer period of May, June, July and August. This makes sense as when temperatures drop, milk production generally drops also and dairy cows divert their energy into maintain body temperature rather than producing milk.

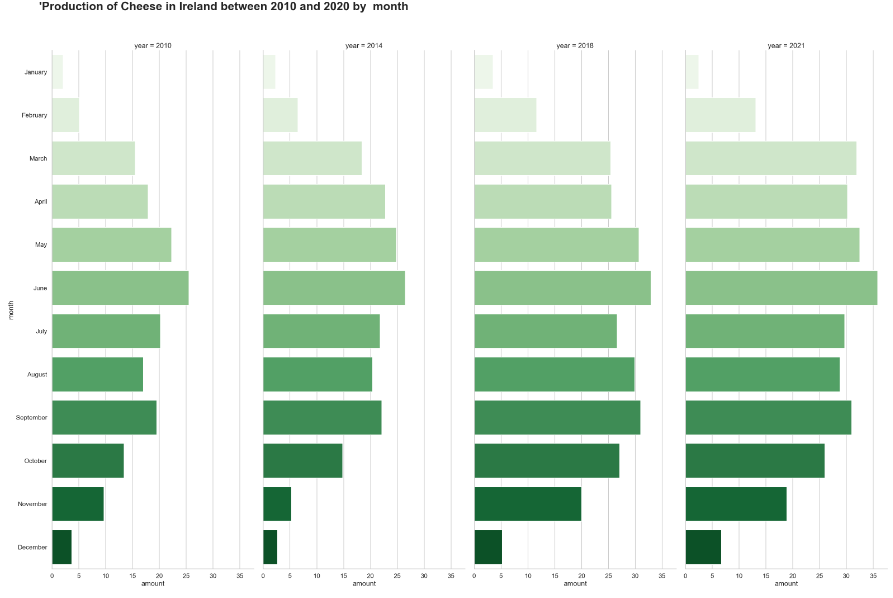


Figure 6 Irish cheese production 2010-22 ln [47]

In June in figure 6 above, production appears to peak annually around June with 25,000 tonnes of cheese being produced in 2010, increasing to 35,000 tonnes in 2021. January and December are the slowest months with approximately 5,000 tonnes being produced each month. There is a 40% increase in June between 2010 and 2021.

Table 3 below shows the tabulated version of figure 6. Again the increase year on year is visible (171,000 tonnes in 2010 and 287,00 tonnes in 2021).

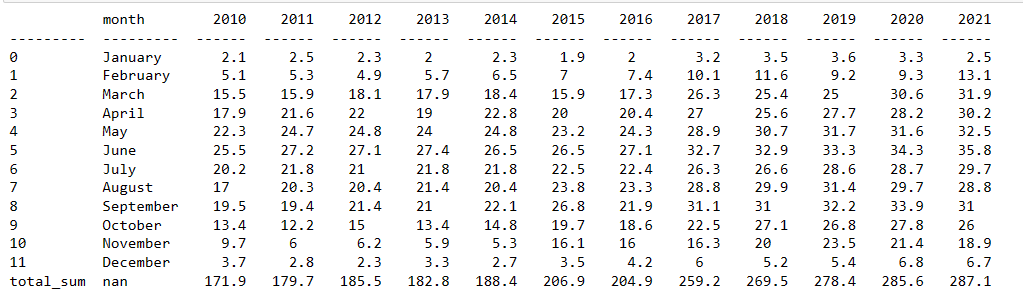


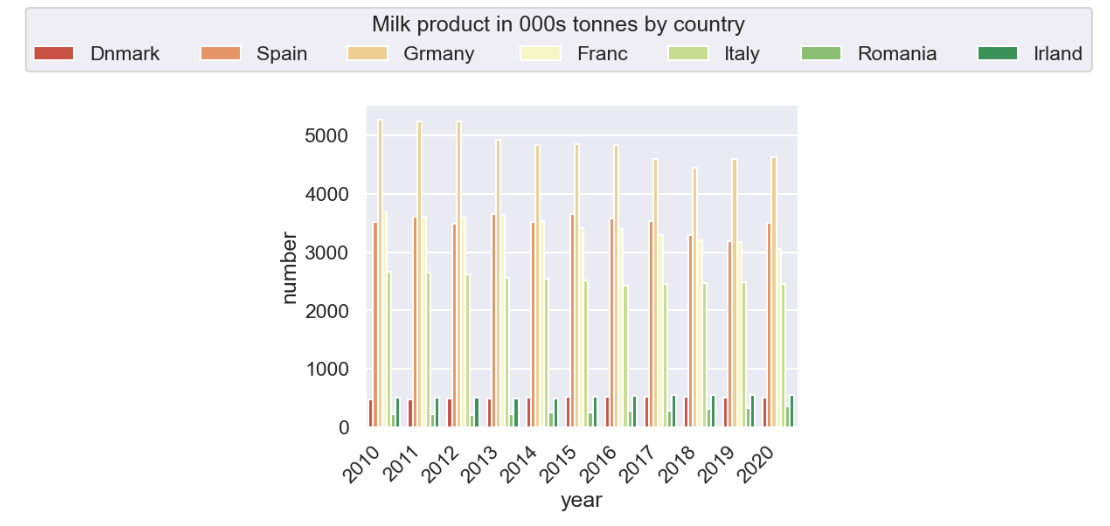
Table 2 Cheese production in Ireland ln[50]

### Dataset 6, EU Dairy Production, 1990 to 2022

*(df\_eumilk\_produce9021, df\_eucheese\_produce9021, df\_eubutter\_produce9021,*

*df\_euskimmedmilkpowder\_produce9021).* Dataset 6 details information about fresh dairy produce in the European union for 27 countries between 1990 and 2022. The data is based on the DG Agriculture and Rural development Eurostat annual milk production statistics. The four dairy products chosen for consumption were Milk, Butter, Cheese and Skimmed milk production produced in thousand tonnes, per year. Irelands was compared with six of Europe’s main producers; Denmark, Spain, Germany, France, Italy and Romania between 2010 and 2020 with seaborn bar charts. The categorical colour palettes were chosen from types of pairplot palettes to ensure sufficient contrast between countries, as opposed to diverging colour palettes.

These visualisations were further supplemented by European choropleth maps generated with the plotly package, showing a snapshot of the situation in 2020 as a graphical visualisation. Country codes were converted from Alpha 2 code to alpha 3 type code for ISO 3166 country codes. Europe was used for the geo scope and the ‘Country’ column was chosen to be displayed in the hover information.



.

.

Figure 7 Milk production by country in In [17]

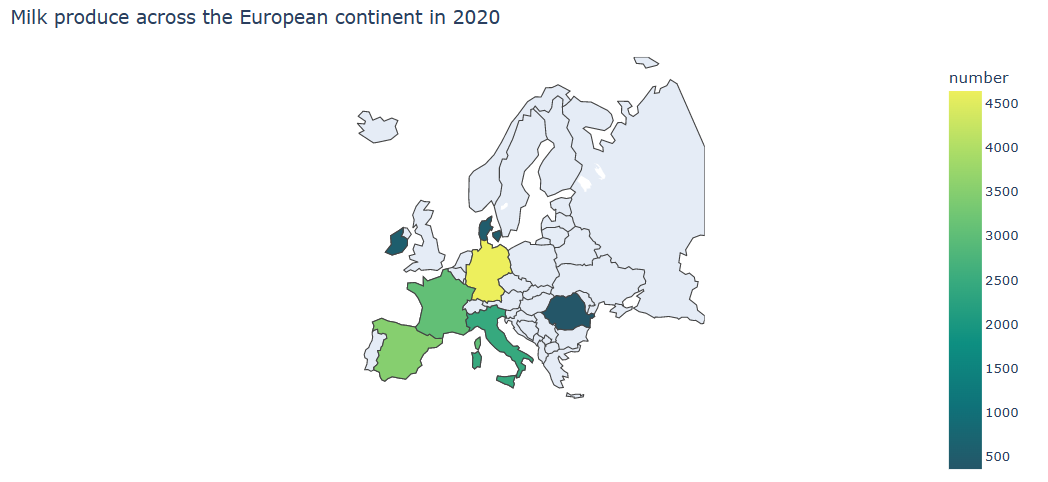


Figure 8 Milk production chloropleth map 2020 in In [24]

Ireland produces a steady 500,00 tonnes per year of fresh milk between 2010 and 2022, similar to Denmark. Germany is the highest producing over 5 million tonnes between 2010 and 2012 and between 4 -5 million thereafter.

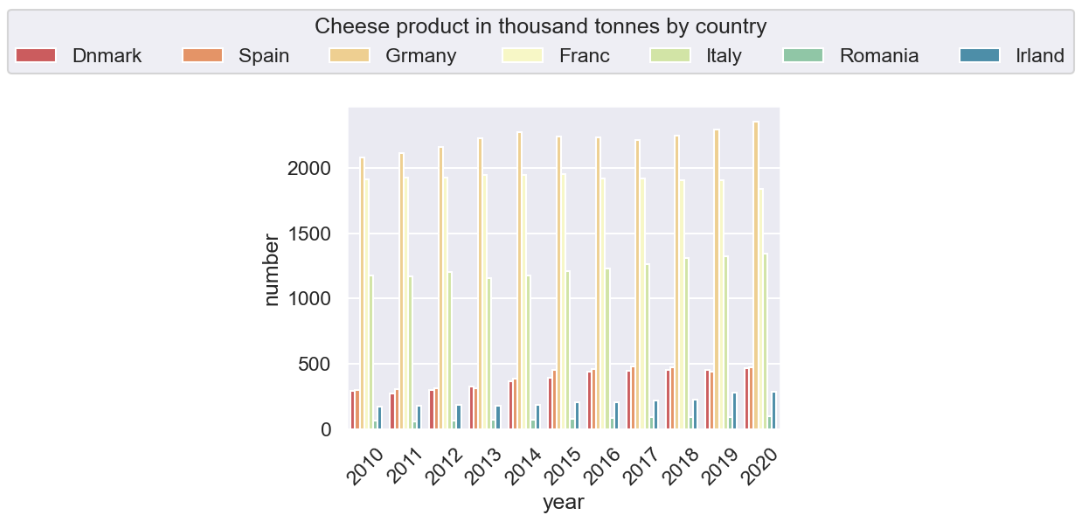


Figure Cheese production by country in In [27]

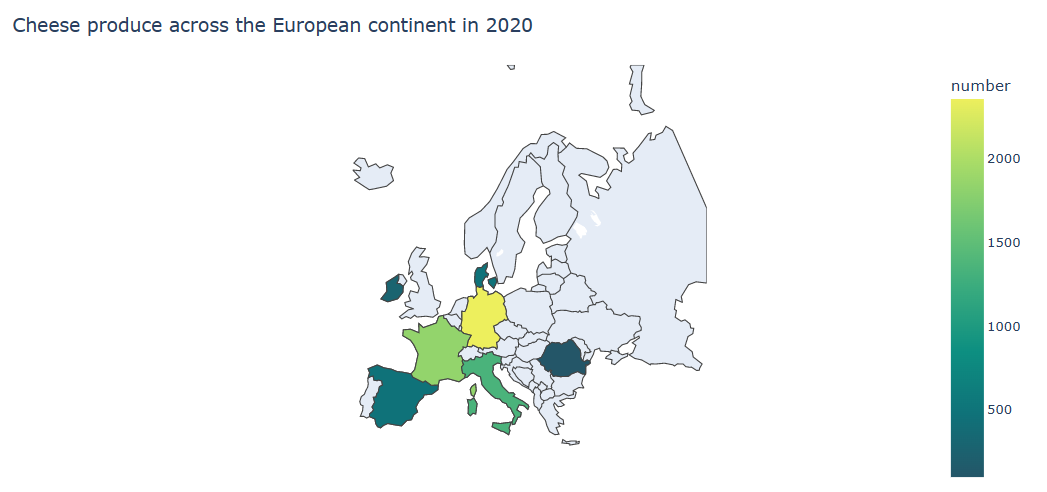


Figure 10 Cheese production chloropleth map 2020 in In [35]

For Cheese production, Germany tops the list with over 2 million tonnes per year and increasing to nearly 5 million tonnes a year from 2015. Ireland is just above Romania with around 200,000 tonnes/year produced between 2010 and 2020, compared to 100,000 tonnes/year.

increasing from 1 to 2 million tonnes between 2010 and 2020.

## Plotly, Dash Visualisations

Colour palettes were chosen for the following visualisations based on their vibrancy and to give the reader clear differentiation between features. Libraries used included plotly.graph, plotly.express, and dash. Explanation of the graphs are included in the Methodology section under each dataset description. These graphs could be collected to form a complete dashboard with plotly, which wasn’t included in this report due to time constraints.

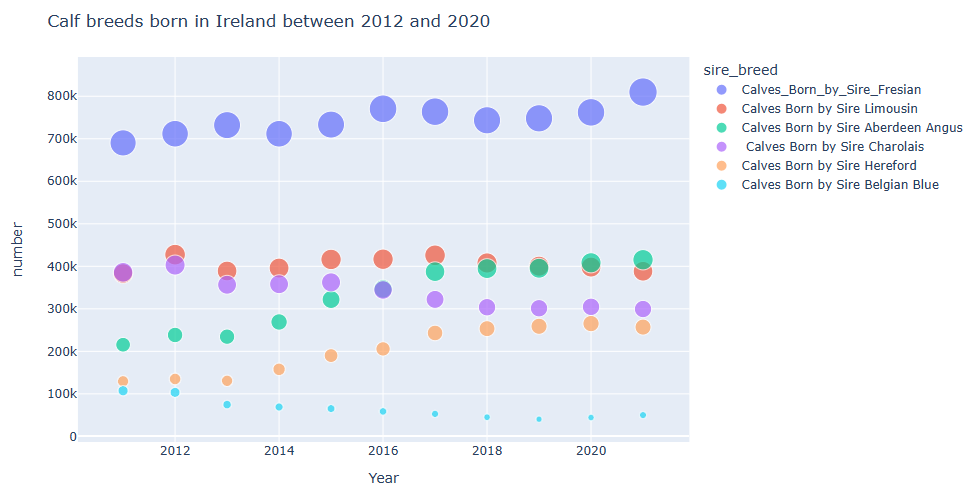


Figure 11 Sire breed In [115]



Figure 12 Milk sales in Ireland In [159]

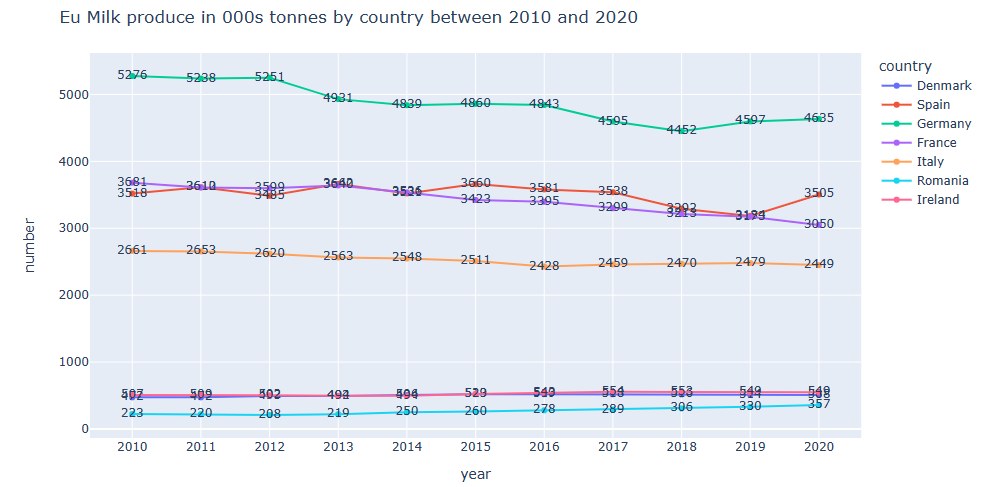


Figure EU Milk produce In [144]

# Times-Series Forecasting in In [89]

## Decomposition

Dataset 6 used for time series analysis was sourced from the cso website shows monthly milk sales between January 1980 to October 2020 in Ireland for human consumption in million litres.

By forecasting the future sales of milk, the production of milk and other dairy products can be planned with farmers to reduce waste and ensure adequate food supply. (S)ARIMA is a powerful statistical modelling technique for time series analysis. The milk sales dataset has a clear pattern with a seasonal element as seen in fig 19. Both manual and auto differencing was completed on the model to ensure stationary and produce accurate sales forecasting figures.

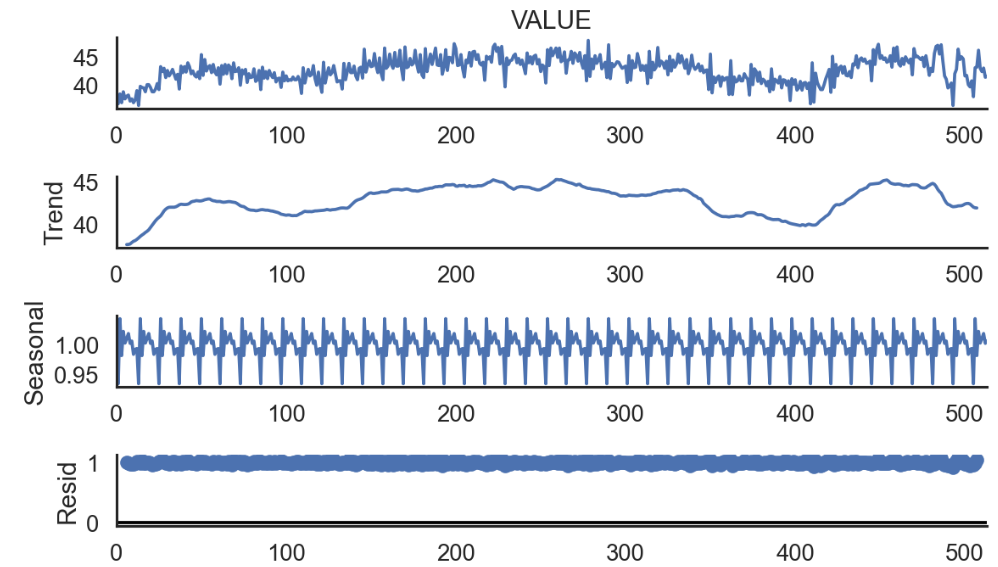


Figure 14 Decomposition of time series in In [89]

The model was initially decomposed in fig 19 above. Since the magnitude of the seasonal component is not changing with time, it was assumed that this was a multiplicative model, with period 12 as these are monthly figures.

Augmented Dicky Fuller (ADF) test was performed to test stationarity in fig 20. The series is increasing with time from less than 38 million litres in 1980 to over 40 million in 2022 and showing annual seasonality as is natural with bovine milk production.

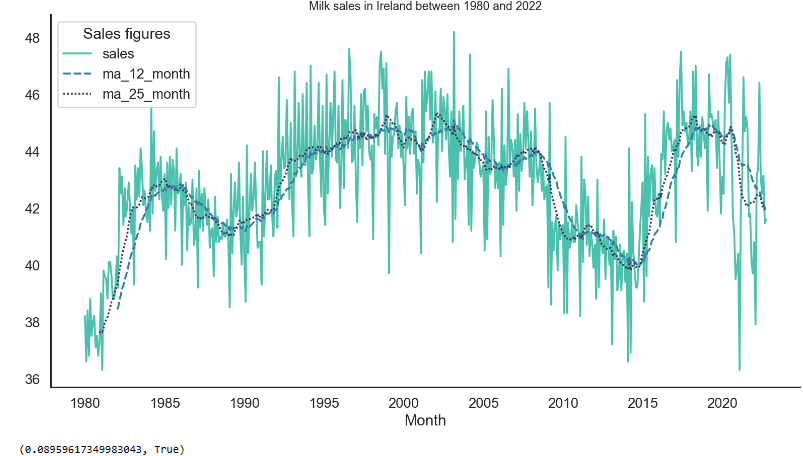


Figure 15 ADF test on milk sales 1980-22 in In [90]

## Autocorrelation and Stationarity testing

Autocorrelation function (ACF) was used to identify how correlated the values in the time series are with each other by plotting the correlation coefficient against the lag. Manual differencing was performed three times to remove trend from the series. The shaded area represents the default 95% confidence interval produced by white noise. The p value of 0.0896 is greater than 0.05, hence we fail to reject the null hypothesis and therefore the mean of the data is not stationary. Ma 12 months and 25 months were performed for 12- and 25-month rolling windows respectively for the observations included in that time period.

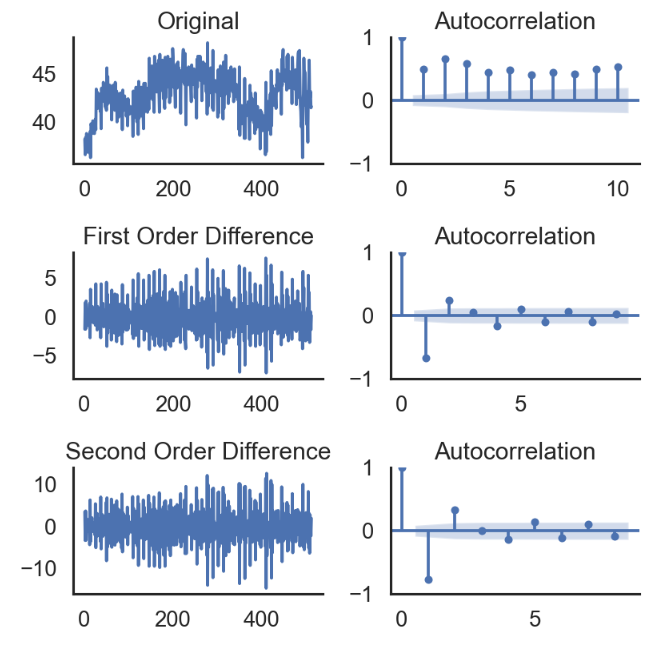


Figure 16 ACF testing of time series data in In [92]

The correlation coefficient tends to decrease from the first differencing from a strong positive 1 to zero and minus 1. The trend is relatively flat and continuous. The seasonality is visible in each plot. After the first differencing all the values are inside the confidence interval and are therefore statistically insignificant. The time series becomes stationary after the first order differencing.

A linear regression was applied to the milk sales data using the months as variables. The categorical column ‘Month’ was converted to numerical with the one hot encoder initially. The trend was removed and a temporary data frame was used to perform the regression In [86]. The model coefficients and seasonality were predicted month by month In [115].

## Autofitting and Forecasting of the model

After the first order differencing, the seasonal part of the time series appears stationary. Second-order differencing does not seem to improve these values. Therefore, it’s concluded that first-order differencing is a good choice for the D parameter. An optimal model was auto-fitted to the time series after splitting the data into training and test sets. Max d was set to 1 for first order differencing only, max p and q were set to 3. Seasonal parameter was set to ‘True’. The best fit model p, d,q was (3,1,1)(2,0,1) with a total fit time of 157 seconds. Ljung Box test result is 0.16 (greater than p value of 0.05) hence we fail to reject the null hypothesis, that the residuals in our time series are independent. The jarque-bera test results of 0 and skew result which is close to zero (0.18) indicate a normal distribution.

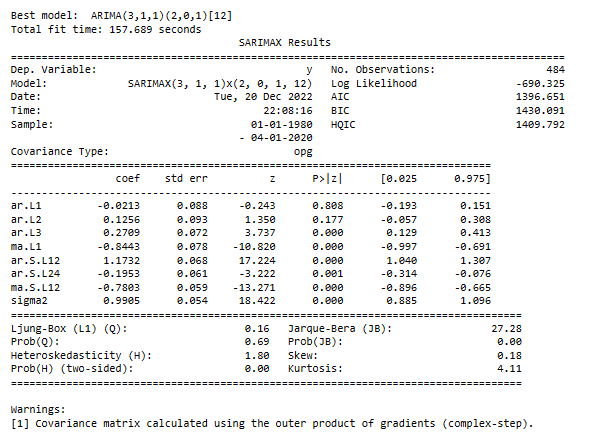


Table 3 Auto Arima model results in In [97]

After training the simulated model, it was used to forecast the time series by calling the predict function in the sample function. The prediction corresponds to the original time series of 12, on which the model was trained.

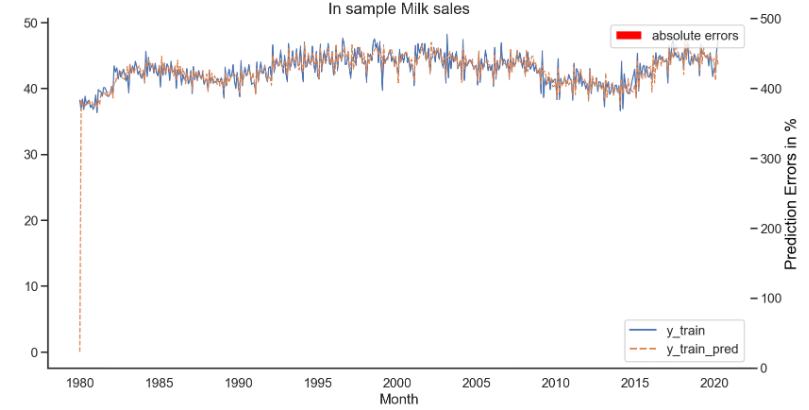


Figure 17 Visualisation of training + test dataset in In [98]

The dataset ‘df\_union’ was concatenated from the training dataset ‘df\_train’ and the testing dataset, ‘df\_test’. After training an optimal model, the dataset ‘df\_union’ was used to generate a sales forecast for the following 2.5 years. An index was generated from the number of predictions adjacent to the original time series and then continued (prediction index).

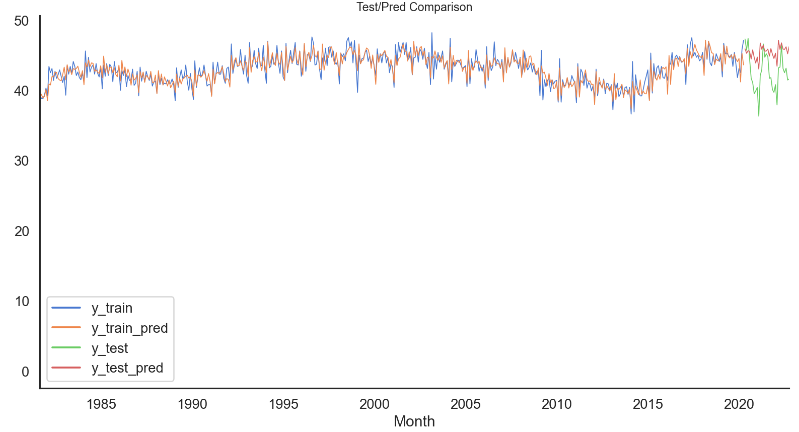


Figure 18 Forecast of time series Milk data in In [99]

It can be seen from fig. 26 that the model’s forecast continues the seasonal pattern of the milk sales time series, which indicates that the model is working and the milk sales will continue to rise. Mean and median absolute percentage error (MAPE, MDAPE) show respective errors of 7.61% and 7.7% showing the model achieves a decent forecasting performance.



The statsmodel library was used to model the seasonal arima with parameters AR=1 and MA=0, with difference set to 1 to remove trend and seasonality. Best ARIMA model results from table 4 (3,1,1)(2,0,1)[12] were used. None of the p weight P>[z] are greater than 0.05, hence not significant.

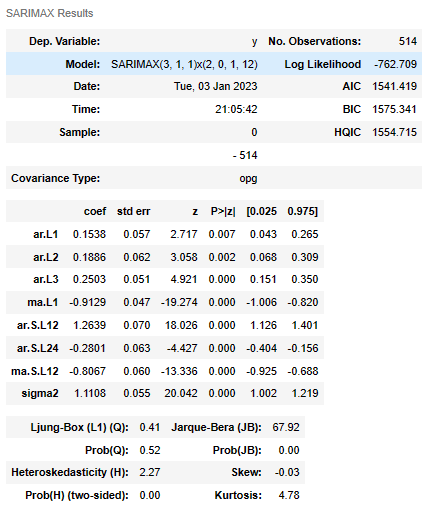


Table 4 SARIMAX Results In [111]

Plotting the residual plot shows white noise, with no seasonality or trend. A normal distribution visible in the histogram, confirmed with qq plot and correlogram with few lagged residuals. Generally, model is a good fit, could be improved by increasing the parameter space for grid search.

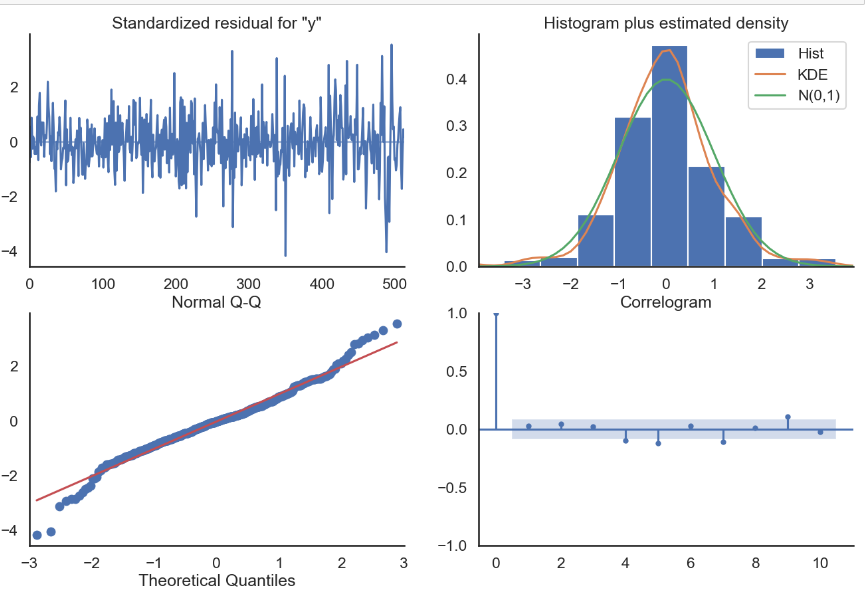


Figure 19 Diagnostics of fitted model In [112]

# Conclusion

The subject of this report is intended to provide the reader with a broad overview of the subject of bovine agricultural data between 1980 and 2022 from bovine breeding trends to the profit associated with selling of milk products such as butter, cheese and milk powder, both in Ireland and Europe during the years when a milk quota existed in Ireland (1984) and after its existence in 2005. The visualisations such as scatter plots, bar charts and tables generated the public datasets are intended to convey this information in an accessible manner.

Figure 4 and 5 show a steady increasing trend for sales of semi skimmed milk 2 million in 1985 to 16 million litres sold in 2014-15, which indicates a change of public taste preference. Figure 7 shows a correlation between price of a litre of milk and sales between 2007 and 2021, indicating the public are aware of the price and purchase accordingly. Figures 8,9 and 10 show the production of dairy products (butter, cheese and milk powder) is very seasonal dependent but production levels are increasing over time suggesting that demand is also increasing. May and June are the highest producing months 15,000 tonnes in 2010 and 30,000 tonnes of butter in 2021, an increase of 66%. Germany is consistently the highest producer of dairy food in western Europe from figures 12,14,16 and 18 producing consistently more that 20% more than other countries such as France and Italy indicating that they have efficient farming methods which other countries may benefit from adapting.

Time series was used to forecast milk sales 2.5 years into the future with data from 1985 to 2020. The model shows less than 8% mean absolute error and could be a valuable forecasting tool. The sale of milk indicates similar demand for other dairy products such as butter and cheese in the future.

The issue of planning the farming in Europe requires smart, responsive solutions which can adapt and change to the public and commercial habits to ensure food security in the future.

# Acknowledgements

The idea and parts of the formatting for this template came from CCT College, Dublin with thanks to my lecturers Sam Weiss, Muhammad Iqbal, David McQuaid, Marina Iantorno

# References

* agriculture.ec.europa.eu. (n.d.). Documents. [online] Available at: <https://agriculture.ec.europa.eu/documents_en?f%5B0%5D=document_title%3Abeef> [
* Agriland.ie. (n.d.). Agriland Farmgate Milk Price – Explanatory Notes. [online] Available at: <https://www.agriland.ie/agriland-farmgate-milk-price-explanatory-notes/>
* ec.europa.eu. (n.d.). Livestock population in numbers. [online] Available at: <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200923-1>.
* Irish Farmers Association. (n.d.). Dairy Factsheet. [online] Available at: <https://www.ifa.ie/dairy-factsheet/>.
* Lazzeri, F. (2021). Machine learning for time series forecasting with Python. Indianapolis Wiley.
* Mueller, J. (2018). Beginning programming with Python. Hoboken, Nj: For Dummies.
* www.farmersjournal.ie. (n.d.). A bluffer’s guide to the end of milk quotas. [online] Available at: <https://www.farmersjournal.ie/a-bluffers-guide-to-the-end-of-milk-quotas-178026>
* Albani, M. (2021). *Grassroots Machine Learning*. [online] GitHub. Available at: https://github.com/Mashimo/datascience/blob/master/01-Regression/dairy-time-forecast.ipynb
* CCT college notes from subjects Data preparation and Visualization, Programming for Data Analytics, Statistics for Data Analytics and Machine learning for Data Analytics

# Appendices

## Appendix 1 Seasonal production of Butter and Skimmed milk powder in Ireland

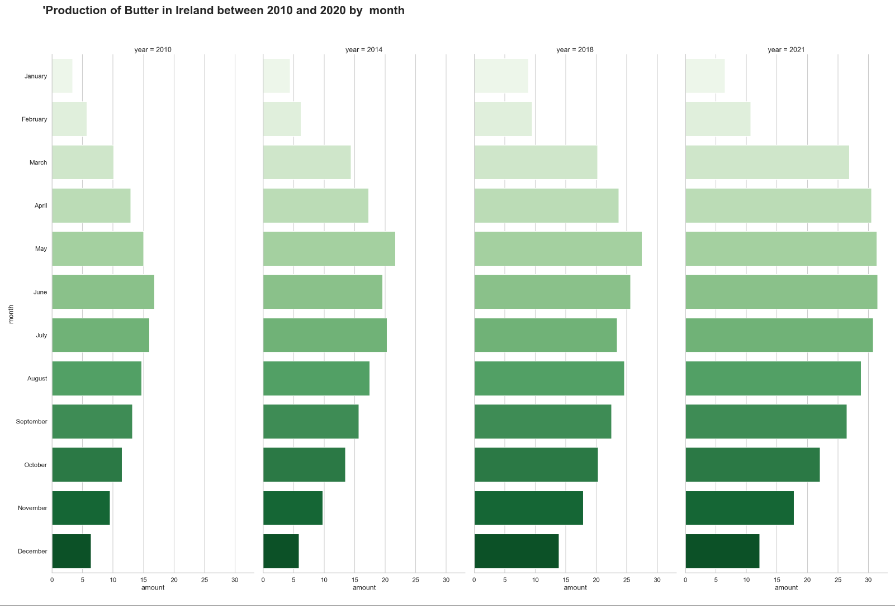


Figure 20 Irish butter production 2010-20 in In [70]

Similar to cheese, May and June are the highest producing months 15,000 tonnes in 2010 and 30,000 tonnes of butter in 2021, an increase of 66%. Less than 7,000 tonnes are being produced in November and December in 2010, this increases to 12,000-15,000 in December 2021.

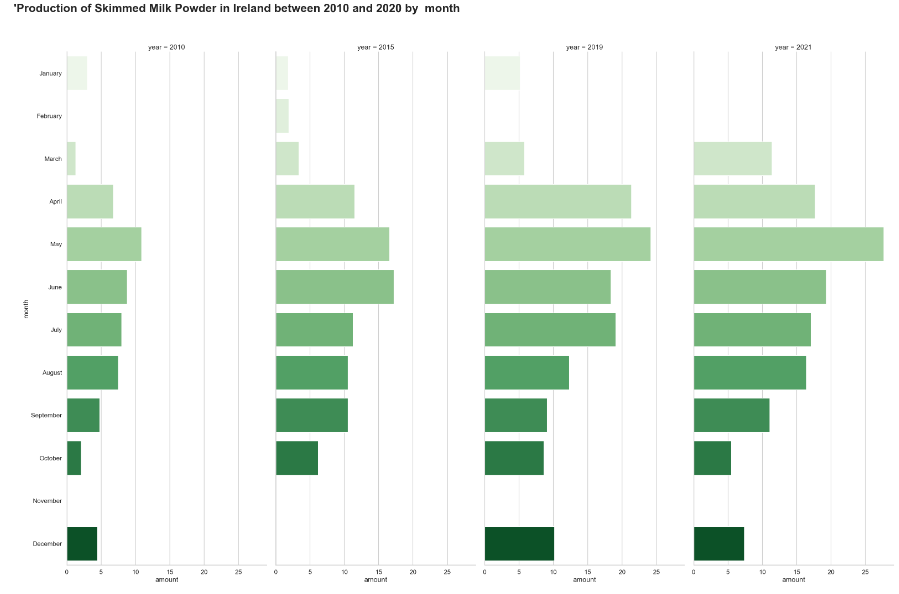


Figure 21 Irish skimmed milk production 2010-21 in In [71]

Skimmed milk powder has some missing values for the end of 2015 and the beginning of 2021. But we can see a similar trend of higher production in May, and an increase year on year of tonnes produced between 2010 and 2021

## Appendix 2 European Butter and Skimmed Milk produce

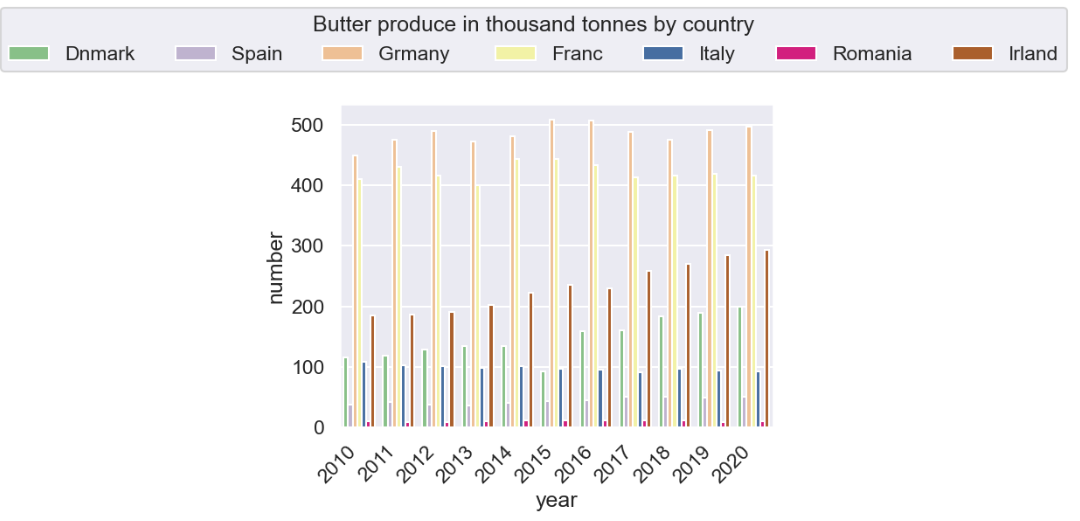


Figure 22 Butter production by country in In [30]

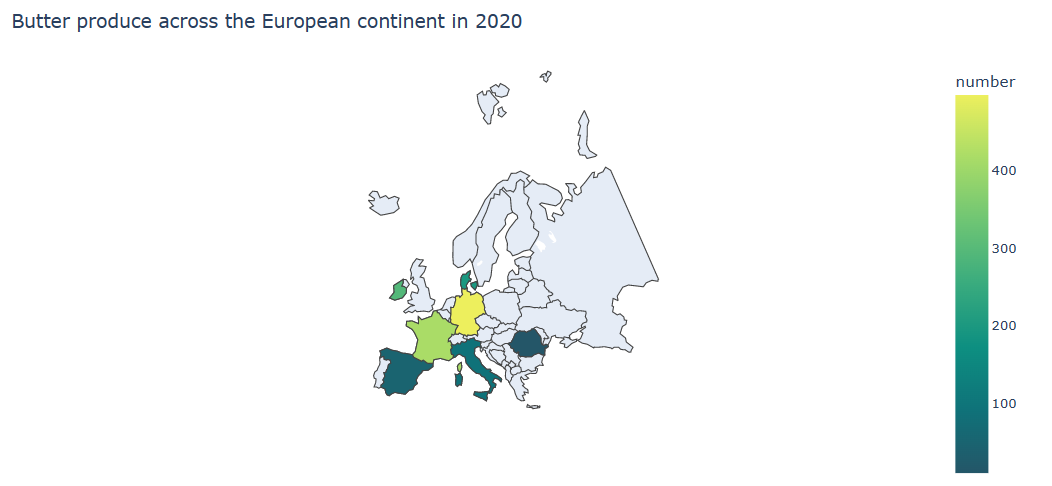


Figure 23 Butter production chloropleth map 2020

Regarding Butter production, Germany unsurprisingly tops the list with between 4-5 million tonnes per year. Ireland is increasing year on year growing from less than 2 million tonnes in 2010 to nearly 3 million in 2020.

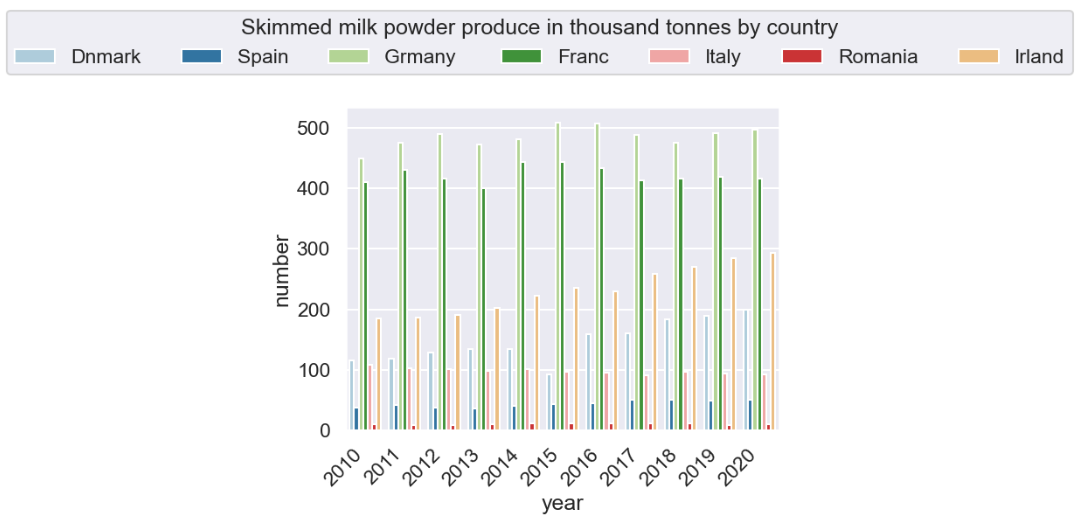


Figure 24 Skimmed Milk produce by country

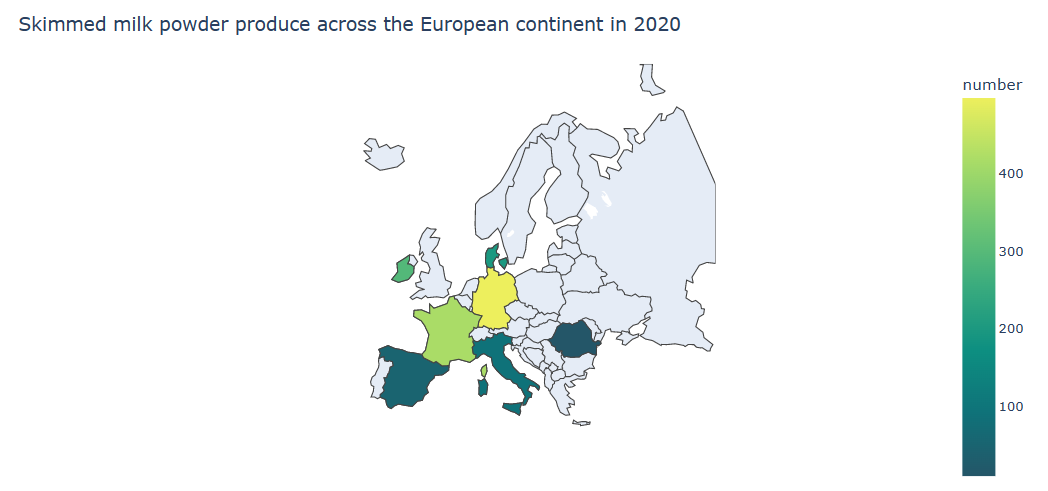


Figure 25 Skimmed milk production chloropleth map 2020

For skimmed milk powder production, Germany tops the list with between 4 to 5 million tonnes/year. Ireland is the 3rd highest, increasing steadily from less from less than 2 million tonnes in 2010 to nearly 3 million in 2020. France’s production remains relatively constant at around 4 million tonnes per year. Denmark’s produce trend is similar to Ireland

## Appendix 3 Time series calculations

The differencing method was repeated for the seasonal part of the time series, with the time frame reduced to a single seasonal period and compared with an autocorrelation function.

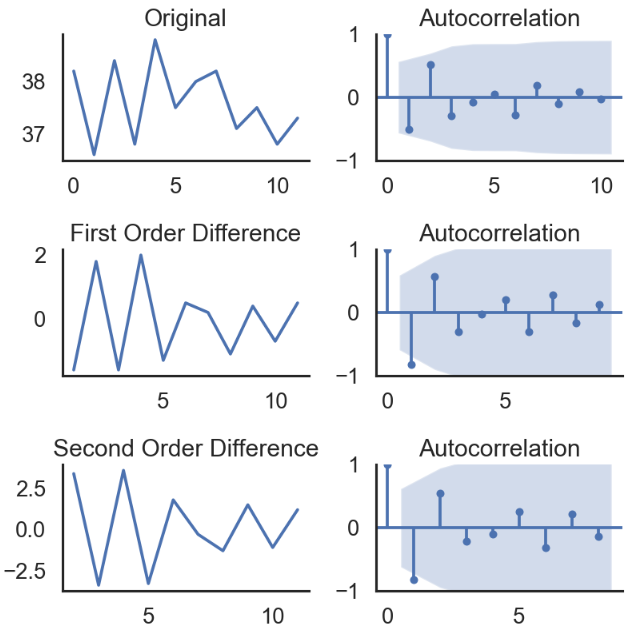


Figure 26 Stationarity testing with differencing in In [93]

Hyper parameter tuning such as a Gridsearch could also be used to find the best parameters for the ARIMA model by simply looping through p,d,q values and checking which model has the lowest AIC values. In [104]

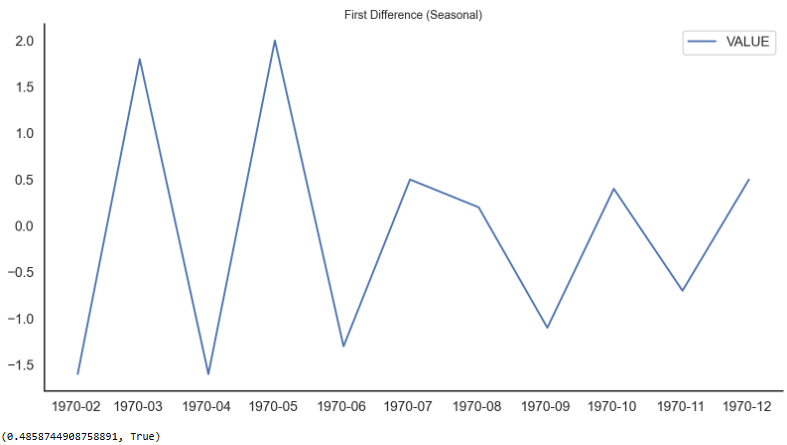


Figure 27 Stationarity testing of 1st difference of seasonal plot in In [95]

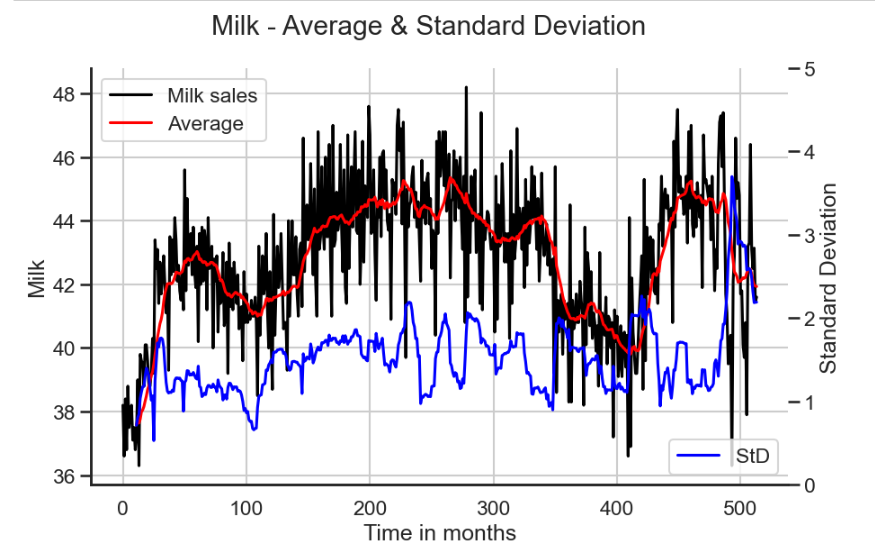


Figure 28 Stationarity visualisation

Graphing the time series (milk sales) against the moving average in red, and standard deviation in blue, it can be seen that the series is not stationary, trend and seasonality are visible. The standard deviation is not constant over the time.

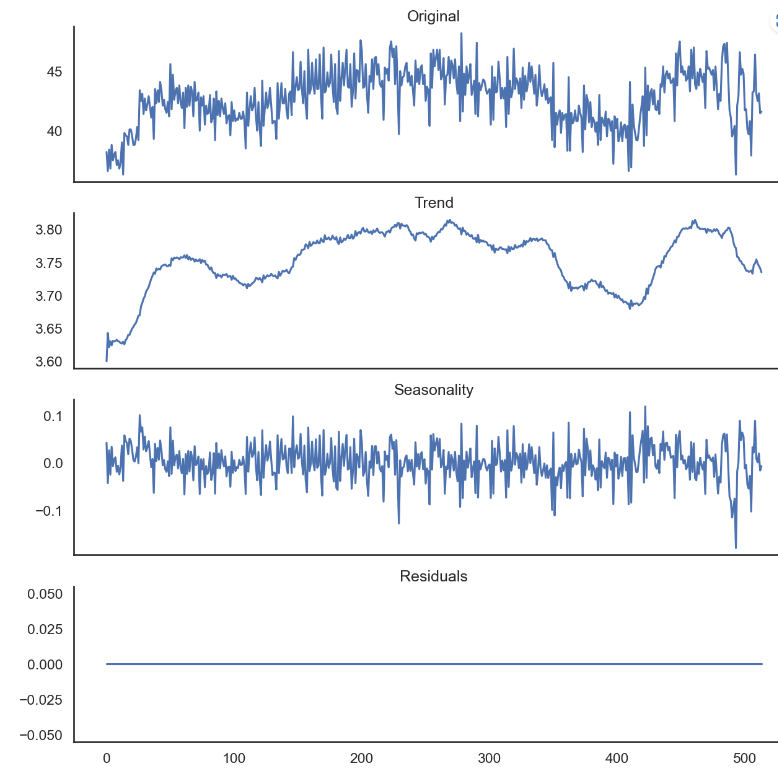


Figure 29 Decomposition after one hot encoding

The trend in figure 27 is of increasing milk sales, beside the last months when it is stable; while there is a clear yearly seasonality. Decomposition could also be compared using statsmodel package.

## Appendix 4 Additional Visualisations

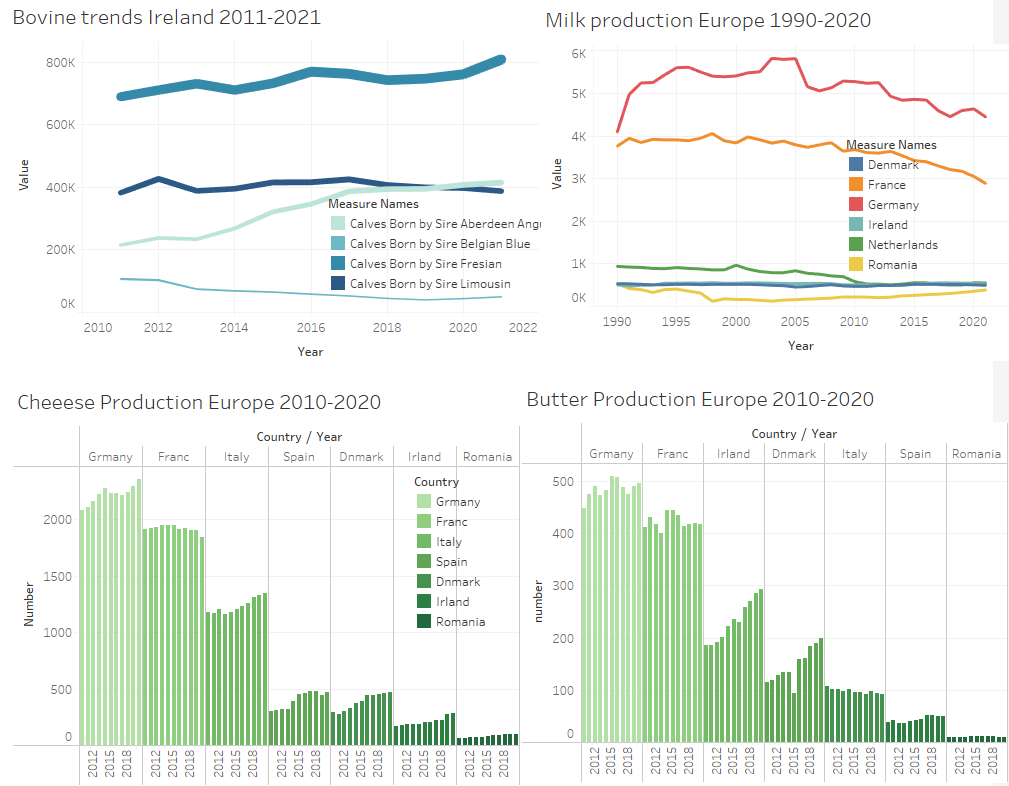


Figure 30 Overview Dashboard of Irish + European Dairy produce

The dashboard in figure 17 above was generated with Tableau to show an overview of some of the visualisations. Friesian cows are the strongest growing breed in Ireland due to their high milk quality and longevity reaching over 800k calves born in 2022. Germany is dominating the cheese, milk and butter production in western Europe, with France a close second.

Appendix 5 Version Control

Version control was completed with Github Desktop

https://github.com/teresaq1/Agricultural-Dairy-Milk-Data-Analysis

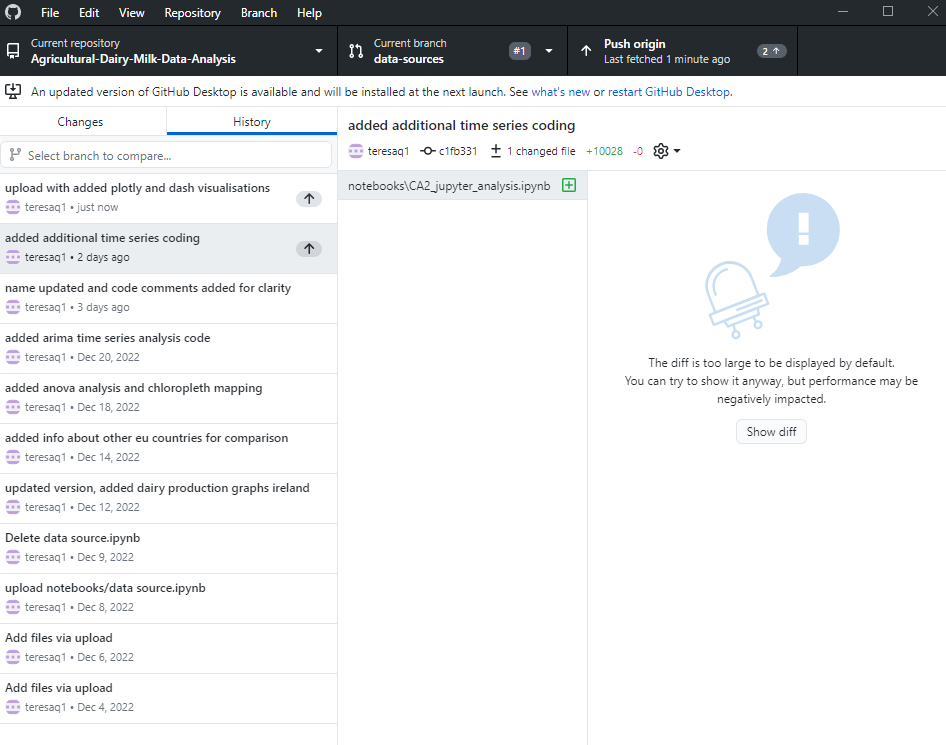


Figure 31 Github desktop screenprint