CCT College Dublin

Assessment Cover Page

Module Title:	Statistics for Data Analytics Programming for Data Analytics Data Preparation and Visualisation Machine Learning for Data Analytics
Assessment Title:	CA2 50% Integrated Assessment
Lecturer Name:	John O'Sullivan, Sam Weiss, David McQuaid, Muhammad Iqbal
Student Full Name:	Zhongjie Fei
Student Number:	2022173
Assessment Due Date:	06/01/2023(11/01/2023 after PMC)
Date of Submission:	11/01/2023

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.

Author: Zhongjie Fei
Student ID: 2022173

Report for Beef Price Analysis in Ireland

and

a quick comparison to Spain

MSc in Data Analytics

e-mail: 2022173@student.cct.ie

Abstract

Agriculture has played a significant role in human progress, providing food and income for the majority of the global population. Among the various forms of animal production, beef production has remained popular throughout history. The complexity of the beef production industry has resulted in fluctuations in beef prices over time. In this study, the author aimed to analyse and visualise the changes in beef prices from 2017 to 2021 in Ireland and conduct a brief comparison with Spain. Two machine learning models were built to train and predict, and a sentiment analysis was created lastly to gain insight into people's opinions on recent beef prices worldwide. By examining these factors and their possible impacts, the study aims to provide a better understanding of the beef production industry and its potential future trends.

Keywords: agriculture, beef prices, data analysis, statistics, visualisation, machine learning, sentimental analysis

0. Introduction

Agriculture has long been an essential part of human history and with the advent of data science, modern agriculture has seen significant improvements in production efficiency. As people become increasingly concerned with environmental sustainability and health, there is a growing focus on organic and environmentally friendly agricultural practices. The beef production industry, particularly in Ireland, remains popular, leading the author to investigate the fluctuations in beef prices over time and their relation to other factors such as beef production, beef feeding prices and more. Corresponding data was collected and processed for analysing statistical aspects and creating interactive visualisations to provide a more direct understanding. Machine learning models, including linear regression and random forest, were applied to train models and make predictions, and a sentiment analysis was conducted to understand people's perspectives on recent beef prices at last.

1. Agriculture and beef production

The emergence of agriculture, dating back thousands of years ago, marked a significant milestone in the historical development of human civilisation. It is the major source of food and income for majority of people around the world (Branco, 2020).

According to Branco (2020), food production in agriculture is a complex and multi-faceted process that involves a range of factors including farming size, soil

management, water and electricity usage, pest and disease control, food security monitoring etc. Data science, for example artificial intelligence, now plays an important role in modern agriculture management, which have led to significant increases in food production, allowing for the feeding of growing global populations. However, the fast development of modern agriculture consequently raises concerns about the environmental and the potential loss of biodiversity (ibid.). Additionally, the use of chemicals for fertilising and pest control can have negative effects on soil and water quality but also on food safety. Therefore, organic farming and organic food production are being promoted in order to maintain sustainability.

Animal production is one of the two major branches in agriculture, alongside crop production. Within animal production, there are various sub-sectors, including beef production, dairy production, pig production, and poultry production, to name a few. Despite the increasing demand for chicken and other meats, beef production continues to play a crucial role in meeting the protein needs of the global population, providing essential nutrients and energy for human health and development (Kahn and Cottle, 2014).

The price of beef is determined by a variety of factors, including supply and demand, production costs, and government policies. To elaborate, the demand for beef is affected by consumer preferences, population growth, and income levels, and the supply of beef is influenced by factors such as weather conditions, disease outbreaks, and production costs as well as government policies (Croxton, 1905). An increase in demand will lead to a higher price, however, an increase in supply will lead to a lower price. The production costs of beef encompass a wide range of expenses, e.g. equipment and facilities, veterinary care, feeding and labour.

1.1 Common Agriculture Policy (CAP)

The European Union's Common Agriculture Policy (CAP), first established in 1962 and regularly amended, plays a significant role in shaping the agricultural industry in Europe. It was proposed to, according to official website of European Commission:

1. Ensure the continuity of food production and distribution;
2. Ensure a supply of safe and affordable food;
3. Encourage younger generations to take up farming and promote environmentally friendly farming;

4. Practice LEADER method to thrive remote and mountainous areas with disadvantages;
5. Reduce power imbalance and help small farms etc.

The latest publication of the CAP places a stronger emphasis on environmental protection and organic farming, while also promoting fairness and innovation within the industry (European Commission, 2022).

As reported by the European Commission (2021), in Ireland in 2020, the allocation of CAP expenditure was 76.4% for direct payments, 19.9% for rural development, and 3.8% for market measures. Additionally, the beef (cattle) sector accounted for 29% of the output component, making it the second largest sector behind dairy. Feedingstuffs represented the highest percentage of intermediate consumption, at 47.2%.

1.2 EU Beef Carcass Classification Scheme

There are various beef carcass classification schemes used around the world, with a common scheme adopted by European countries. According to the Department of Agriculture, Food, and the Marine (2022) in Ireland, the European scheme categorises beef carcass from three perspectives:

- Sex: denoted by the letters A (young bull), B (bull), C (steer), D (cow), E (heifer);
- Conformation (Quality): E, U, R, O, P with E being the best and P the poorest;
- Fatness: the degree of fat is denoted by the numbers 1, 2, 3, 4, 5 in order of increasing fatness.

2. Study objective

The main purpose of this study was to investigate and visualise the dynamic changes in beef prices in Ireland and to analyse the relationship with other potential influencing factors such as feedingstuffs and beef production. Moreover, the study aimed to compare the beef prices and production in Ireland to that of Spain, another major beef-producing country in Europe. Given the assumption that beef prices in Ireland may differ significantly from those in Spain due to its high consumption level, several hypothesis tests were used to test this assumption. To further understand the data, two machine learning models were applied and compared. Finally, a sentiment

analysis was implemented using Twitter API to observe people's recent (past 7 days) comments on beef prices worldwide.

Jupyter notebook version 6.4.8 and Python version 3.9.12 were used for exploratory data analysis (EDA), visualisation and machine learning in our report.

3. Data collection and preparation

Due to the complexity of beef prices and the limitations of this study, we chose to focus on the R3 level of beef carcass, including all sexes, from 2017 to 2021 for our analysis. Furthermore, we selected three potential factors that could affect beef prices for examination: beef production, beef feedingstuff prices, and pigmeat prices. However, during our research, we were unable to obtain reliable data on beef feedingstuff prices in Spain. As a result, this factor had to be omitted when comparing the two countries. We later also discovered that Spain was missing data for the bull and steer on R3 level. As a result, we had to drop these elements when comparing.

During the data collection period, datasets related to beef prices, beef production, and pigmeat prices were collected from the European Commission's agri-food data portal (https://agridata.ec.europa.eu/extensions/DataPortal/home.html). The data explorer function on the website allowed us to download the desired data for beef prices and beef production directly through filtering. However, for the pigmeat prices, we had to use the open API function to collect the data as the filtering function was not working properly within the data explorer mode (see Jupyter notebook). This thus avoided complication for preparation.

Data related to beef feedingstuff prices in Ireland was downloaded from Central Statistics Office website (https://data.cso.ie/#). While there are several types of beef feedingstuffs, we only selected two types for our analysis. At this stage, four datasets were ready for processing. Through examination the structure of the original datasets (Table 1-4), few insights were gained to process the four datasets for later merging:

- For datasets that contains weekly information, groupby function based on date (and country) was used to transform the weekly price into average monthly price, i.e. beef prices and pigmeat prices, since the other two data only contains monthly information;
- Due to the inconsistency of time/date format, a new column was added in each dataset with the same format of year and month (%Y-%m).

3.1 Processing of beef price dataset

Table 1: Details of features in beef price dataset

FN	Attribute name	Description
1	Year	year
2	Week	week number
3	Begin Date	begin date of the week
4	End Date	end date of the week
5	Member State	country name
6	Category	beef sex
7	Product	classification code
8	Price	price in euro per 100 kg

The main modification applied to this dataset was to calculate the mean beef price based on the date and country. This was done in order to have a monthly beef prices representation rather than weekly. Subsequently, duplicate rows and some unnecessary columns were removed, and columns name and their order were reorganised to simplify the dataframe and prepare it for further processing (Figure 1).

	date	country	category	product	beef price/100kg
24	2021-12	Ireland	Bulls	B R3	321.9675
25	2021-12	Spain	Cows	D R3	304.8400
26	2021-12	Ireland	Cows	D R3	378.2125
27	2021-12	Spain	Heifers	E R3	426.5175
28	2021-12	Ireland	Heifers	E R3	431.5850

Figure 1: Beef price dataset after preparation

3.2 Processing of beef production dataset

Table 2: Details of features in beef production dataset

FN	Attribute name	Description
1	Member State	country name
2	Member State Code	country code
3	Category	beef sex
4	Year	year
5	Month	month
6	1000 Heads	measure unit
7	kg/head	measure unit

The main modification applied in this dataset was to combine the Year and Month columns and format the date to match the previously used date format. The unique values in the Category feature were also modified to match the names used in the beef price dataset. Finally, the columns were reorganised to make the dataframe more succinct (Figure2).

	date	country	category	1000 heads	kg/head
0	2021-01	Ireland	Bulls	15.67	376.515635
1	2021-02	Ireland	Bulls	14.85	383.164983
2	2021-03	Ireland	Bulls	14.19	394.644116
3	2021-04	Ireland	Bulls	11.95	403.347280
4	2021-05	Ireland	Bulls	15.61	393.978219

Figure 2: beef production dataset after preparation

3.3 Processing of beef feeding price dataset

Table 3: Details of features in beef feeding price dataset

FN	Attribute name	Description
1	STATISTIC Label	label
2	Month	date (year and month)
3	Type of Feedstuff	feedstuff type
4	UNIT	measure unit
5	VALUE	price

In this dataset, the main change applied was the Month column to match the date format above (Figure 2). Also, the dataset was split into two parts according to its feedstuff type (see attached).

	date	feedstuff	beef feeding price/tonne
0	2017-01	Cattle fattening nuts and cubes (13-15% protein)	262.44
1	2017-01	Cattle fattening meal (13-15% protein)	250.44
2	2017-02	Cattle fattening nuts and cubes (13-15% protein)	264.56
3	2017-02	Cattle fattening meal (13-15% protein)	248.93
4	2017-03	Cattle fattening nuts and cubes (13-15% protein)	265.00

Figure 3: beef feeding price dataset after preparation

3.4 Processing of pigmeat price dataset

Table 4: Details of features in pigmeat price dataset

FN	Attribute name	Description
1	memberStateCode	country code
2	memberStateName	country name
3	beginDate	begin date of the week
4	endDate	end date of the week
5	weekNumber	week number
6	price	price in euro
7	unit	measure unit
8	pigClass	pig classification (E)

Similar to the processing methods used for the beef price dataset, the pigmeat prices were also converted to average monthly prices (Figure 4). It is worth noting that the values of pigmeat prices were in string format with euro signs after requesting from API. Thus, a conversion function was applied to change it to the appropriate numerical format (see attached).

	date	country	statecode	pig price/100kg
6	2021-12	Ireland	IE	143.550
7	2021-12	Spain	ES	125.880
16	2021-11	Ireland	IE	143.442
17	2021-11	Spain	ES	126.170
24	2021-10	Ireland	IE	147.770

Figure 4: pigmeat price dataset after preparation

3.5 Data merging

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 60 entries, 295 to 0
Data columns (total 20 columns):
 #   Column                                      Non-Null Count  Dtype
---  ------                                      --------------  -----
 0   date                                        60 non-null     object
 1   country                                     60 non-null     object
 2   bulls price(€/100kg)                        60 non-null     float64
 3   bulls(1000 heads)                           60 non-null     float64
 4   bulls(kg/head)                              60 non-null     float64
 5   cows price(€/100kg)                         60 non-null     float64
 6   cows(1000 heads)                            60 non-null     float64
 7   cows(kg/head)                               60 non-null     float64
 8   heifers price(€/100kg)                      60 non-null     float64
 9   heifers(1000 heads)                         60 non-null     float64
 10  heifers(kg/head)                            60 non-null     float64
 11  steers price(€/100kg)                       60 non-null     float64
 12  steers(1000 heads)                          60 non-null     float64
 13  steers(kg/head)                             60 non-null     float64
 14  young bulls price(€/100kg)                  60 non-null     float64
 15  young bulls(1000 heads)                     60 non-null     float64
 16  young bulls(kg/head)                        60 non-null     float64
 17  pigmeat price(€/100kg)                      60 non-null     float64
 18  cattle fattening nuts and cubes price(€/tonne)  60 non-null  float64
 19  cattle fattening meal(€/tonne)              60 non-null     float64
dtypes: float64(18), object(2)
memory usage: 9.8+ KB
```

Figure 5: Information of ireland_beef

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 120 entries, 472 to 5
Data columns (total 13 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   date                        120 non-null    datetime64[ns]
 1   country                     120 non-null    object
 2   statecode                   120 non-null    object
 3   cows price(€/100kg)         120 non-null    float64
 4   cows(1000 heads)            120 non-null    float64
 5   cows(kg/head)               120 non-null    float64
 6   heifers price(€/100kg)      120 non-null    float64
 7   heifers(1000 heads)         120 non-null    float64
 8   heifers(kg/head)            120 non-null    float64
 9   young bulls price(€/100kg)  120 non-null    float64
 10  young bulls(1000 heads)     120 non-null    float64
 11  young bulls(kg/head)        120 non-null    float64
 12  pigmeat price(€/100kg)      120 non-null    float64
dtypes: datetime64[ns](1), float64(10), object(2)
memory usage: 13.1+ KB
```

Figure 6: information of ie_es_beef

After the preparation of the original datasets, we applied the merge and pivot function in pandas to convert the original datasets into two final datasets for analysis, in addition to some essential reorganising. The first dataset, “ireland_beef” contains the data for analysing beef prices in Ireland alone, while the second dataset, “ie_es_beef” includes the data for comparing the beef prices between Ireland and Spain (Figure 5-6, see attached also).

Finally, the datasets were checked for missing values and duplicates before proceeding to the analysis. Missing values were not addressed in this study.

4. Statistics

In this section, we first provided a brief overview of the basic statistics in our two final datasets. We then used the "ie_es_beef" dataset to test our hypothesis. The hypothesis tests were used to determine if there is a statistically significant difference in beef prices between Ireland and Spain.

4.1 Basic statistics

	bulls price(€/100kg)	bulls(1000 heads)	bulls(kg/head)	cows price(€/100kg)	cows(1000 heads)	cows(kg/head)	heifers price(€/100kg)	heifers(1000 heads)	heifers(kg/head)
count	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000
mean	294.580325	17.327333	384.297193	325.900542	31.001833	311.474277	388.191783	48.098333	312.970262
std	25.763776	5.422438	10.354916	26.991156	4.927993	6.408008	21.982764	5.061223	4.759718
min	243.172500	7.210000	365.921788	271.657500	14.880000	297.777778	347.292500	33.300000	302.729529
25%	274.060000	13.580000	375.801697	301.582375	27.785000	308.249192	372.559875	44.702500	310.312813
50%	294.393750	16.180000	382.995730	322.200000	30.610000	312.547406	383.966500	47.060000	312.751249
75%	314.607500	21.775000	391.458790	346.495000	34.372500	315.855635	403.563125	51.070000	316.202140
max	347.186000	29.450000	408.235294	378.212500	39.590000	323.833386	431.585000	59.200000	322.634332

Figure 7: First half of the basic statistics related to Ireland

steers price(€/100kg)	steers(1000 heads)	steers(kg/head)	young bulls price(€/100kg)	young bulls(1000 heads)	young bulls(kg/head)	pigmeat price(€/100kg)	cattle fattening nuts and cubes price(€/tonne)	cattle fattening meal(€/tonne)
60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.00000	60.000000
379.568917	56.101000	354.316466	366.843308	0.799333	180.187593	157.527942	285.49450	278.645333
22.312636	11.017707	6.688084	25.048188	1.729346	43.036141	15.201829	19.12555	21.011908
342.272500	41.330000	342.585404	320.637500	0.060000	100.543478	136.727500	262.44000	248.930000
365.921250	46.805000	349.133236	346.525625	0.117500	142.857143	142.754000	267.39250	264.250000
373.610000	51.760000	353.159400	365.441750	0.170000	181.818182	157.773000	282.35000	275.940000
392.879750	67.142500	359.626800	383.860750	0.267500	222.222222	166.271250	293.21250	287.680000
427.835000	77.110000	370.541612	413.590000	9.090000	250.000000	190.757500	345.13000	339.300000

Figure 8: Second half of the basic statistics related to Ireland

Overview of beef statistics in Ireland from 2017 to 2021:

- The average bull price is 294.58€/100kg, with the highest being 347.19€/100kg and the lowest being 243.17€/100kg
- The average cow price is 325.90€/100kg, with the highest being 378.21€/100kg and the lowest being 271.66€/100kg
- The average heifer price is 388.19€/100kg, with the highest being 431.59€/100kg and the lowest being 347.29€/100kg
- The average steer price is 379.57€/100kg, with the highest being 427.84€/100kg and the lowest being 342.27€/100kg
- The average young bull price is 366.84€/100kg, with the highest being 413.59€/100kg and the lowest being 320.64€/100kg
- The average pigmeat price is 157.53€/100kg, with the highest being 190.76€/100kg and the lowest being 136.73€/100kg

- The average cattle fattening nuts and cubes price is 285.49€/tonne, with the highest being 345.13€/tonne and the lowest being 262.44€/tonne
- The average cattle fattening meal price is 278.65€/tonne, with the highest being 339.30€/tonne and the lowest being 248.93€/tonne, etc

	cows price(€/100kg)	cows(1000 heads)	cows(kg/head)	heifers price(€/100kg)	heifers(1000 heads)	heifers(kg/head)	young bulls price(€/100kg)	young bulls(1000 heads)	young bulls(kg/head)	pigmeat price(€/100kg)
count	60.000000	60.00000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000
mean	263.161305	29.18300	292.688712	380.474257	34.942833	261.656069	370.949797	76.352333	244.823747	155.511700
std	14.272293	5.10234	5.534573	17.485301	4.415289	3.488231	19.772318	6.901685	3.714054	19.803813
min	233.507500	15.89000	280.660377	343.895000	25.180000	251.758087	333.142500	60.810000	236.713768	123.386000
25%	252.244250	25.16500	288.347952	370.683875	32.177500	259.396981	356.801250	70.895000	241.680233	140.052250
50%	265.811000	30.15000	292.937349	379.817500	35.450000	261.938723	372.418900	76.510000	245.209189	156.268500
75%	271.872500	32.97750	297.623890	391.625000	37.792500	263.419675	382.818750	83.012500	247.696544	175.363375
max	304.840000	37.96000	301.875558	426.517500	43.780000	269.726973	436.437500	87.730000	252.231500	189.092000

Figure 9: Basic statistics related to Spain

Overview of beef statistics in Spain from 2017 to 2021:

- The average cow price is 263.16€/100kg, with the highest being 304.84€/100kg and the lowest being 233.51€/100kg
- The average heifer price is 380.47€/100kg, with the highest being 426.52€/100kg and the lowest being 343.90€/100kg
- The average young bull price is 370.95€/100kg, with the highest being 436.44€/100kg and the lowest being 333.14€/100kg
- The average pigmeat price is 155.51€/100kg, with the highest being 189.09€/100kg and the lowest being 123.39€/100kg, etc

4.2 Hypothesis test

A hypothesis test is a procedure for making a decision about the value of a population parameter on the basis of sample data (Reif, 2008). The goal is to determine whether there is enough evidence to support the alternative hypothesis, or whether the data is more likely to have occurred if the null hypothesis is true.

There are two type of hypothesis tests. Parametric statistical tests are based on assumptions about the distribution of data and the parameters of the population from which the sample is drawn, usually assuming normality in the distribution. These tests are more powerful when the assumptions are met but can be unreliable if the assumptions are not met. Non-parametric statistical tests, on the other hand, make fewer or no assumptions about the population distribution, and instead rely on the order or

rank of the data. These tests are considered robust and can be used when the assumptions of parametric tests are not met but are generally less powerful than parametric tests.

Since there are three types of beef prices, i.e. cow, heifer, and young bull, in Ireland and Spain, we therefore separated the null hypothesis into three distinct parts. Normality tests using the Shapiro-Wilk test with a significance level of 0.05 were performed to check whether the data was normally distributed. The results of the normality test indicated that, with the exception of cow prices in Spain, the prices for all other categories of beef were normally distributed.

To test the null hypotheses, we used both parametric and non-parametric hypothesis testing methods. After analysing the data, we found that:

- Reject the null hypothesis that cow prices in Ireland were significantly different from those in Spain;
- Reject the null hypothesis that heifer prices in Ireland were significantly different from those in Spain (since the sample for heifer prices were more normally distributed, we had more trust in the results of parametric testing);
- Failed to reject the null hypothesis that young bull prices in Ireland were significantly different from those in Spain.

5. Visualisation

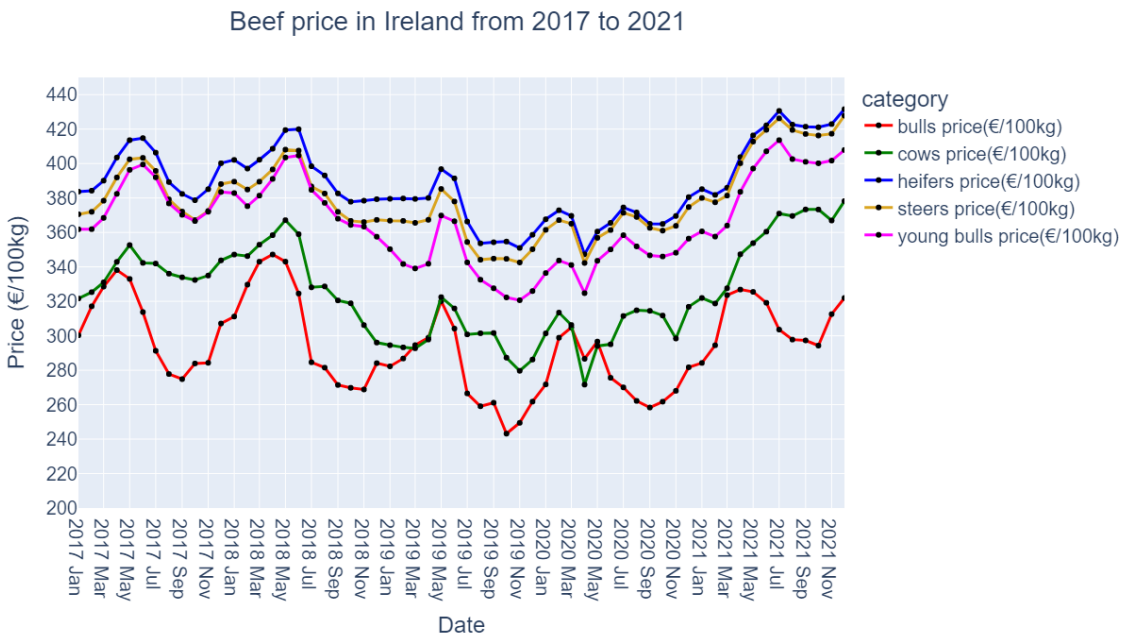
In this section, some of the interactive graphs were discussed. It is worth mentioning that the basic style of the visualisation was based on the average age of farmers and their gender distribution, as reported by the Central Statistics Office (2021). According to this report, in 2020, the average age of a farmer in Ireland was 54 years old, and more than 86% were male. Thus, the target audience for visualisations is older individuals. To effectively communicate information to this demographic, we highlighted a few key design considerations:

First, we suggested using high contrast colours for the graphs, to make them easier for older individuals to read. Additionally, the font size of the title should be large to improve readability.

Second, we recommended using simple graph styles that are easy to understand, such as line charts and bar charts. This will help older individuals to quickly grasp the information being presented.

Lastly, we noted that research shows that the top five colours that older men prefer are blue, red, green, black and pink. Therefore, it is important to take into consideration these colour preferences when designing the visualisations (Eldertech, 2017).

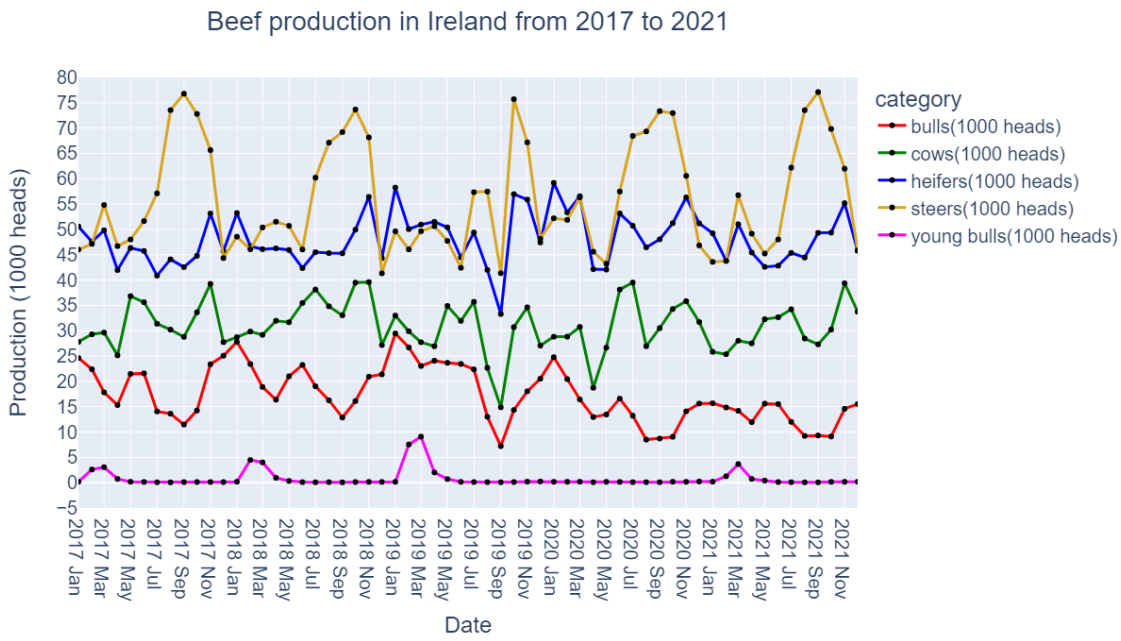
Figure 10: Beef prices in Ireland from 2017 to 2021



Overview:

- Heifer price > Steer price > Young bull price > Cow price > Bull price
- The prices of all five categories of beef tend to display similar patterns of fluctuation, with the top three categories showing the most correlation
- A trend of rising prices between March and May was observed each year
- Critical thinking: though weather condition was not considered in this study, the trend above may be the result of harsh winter conditions, as the largest price increase in the dataset occurred in March 2018, which coincided with a heavy snowstorm in Ireland. Similarly, the increase of beef prices after 2021 could also be attributed to impact of Covid pandemic

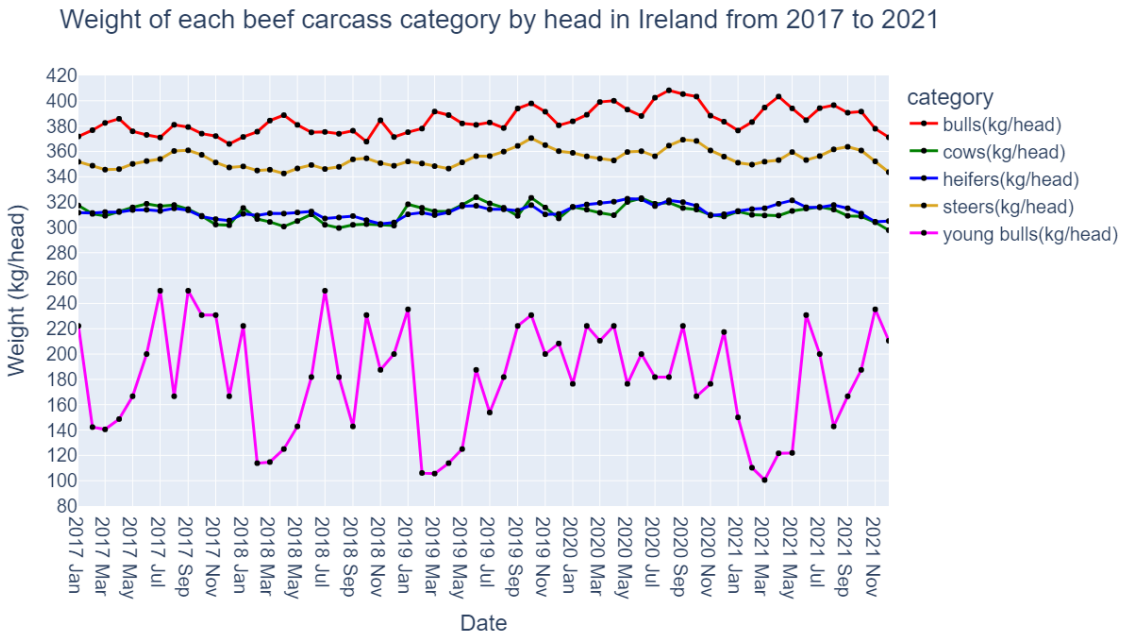
Figure 11: Beef production in Ireland from 2017 to 2021



Overview:

- Steer production > Heifer production > Cow production > Bull production > Young bull production
- September every year has the highest production of steers, however, March every year has the highest production of young bulls

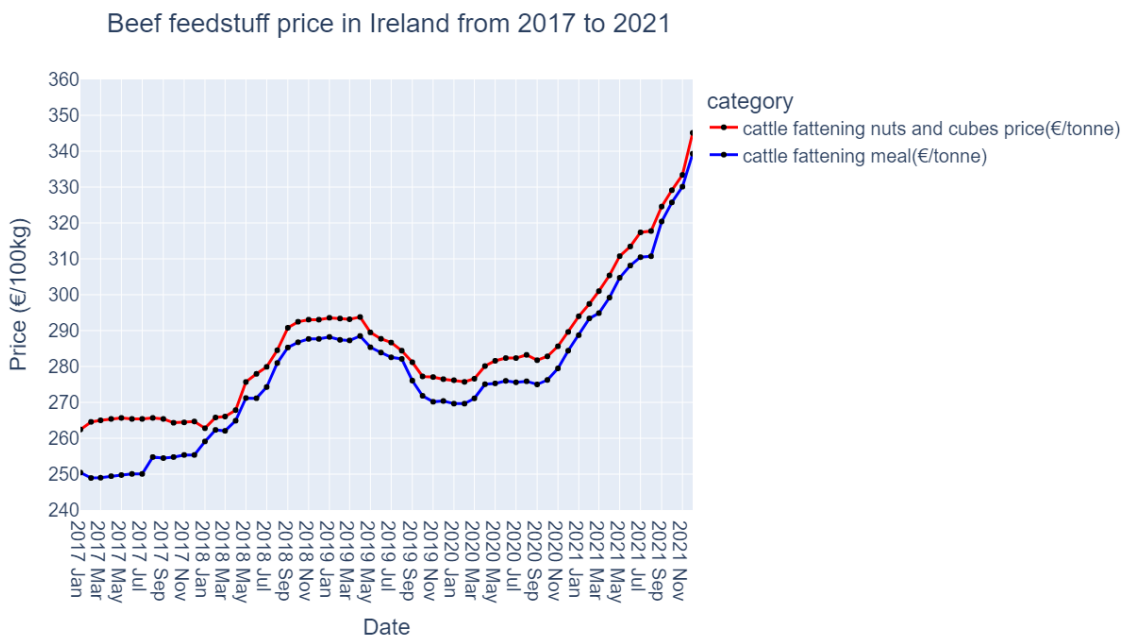
Figure 12: Weight of the beef per head in Ireland from 2017 to 2021



Overview:

- Bull weight > Steer weight > Heifer weight ≈ Cow weight > Young bull weight
- Bulls get a weight increase around March and September every year

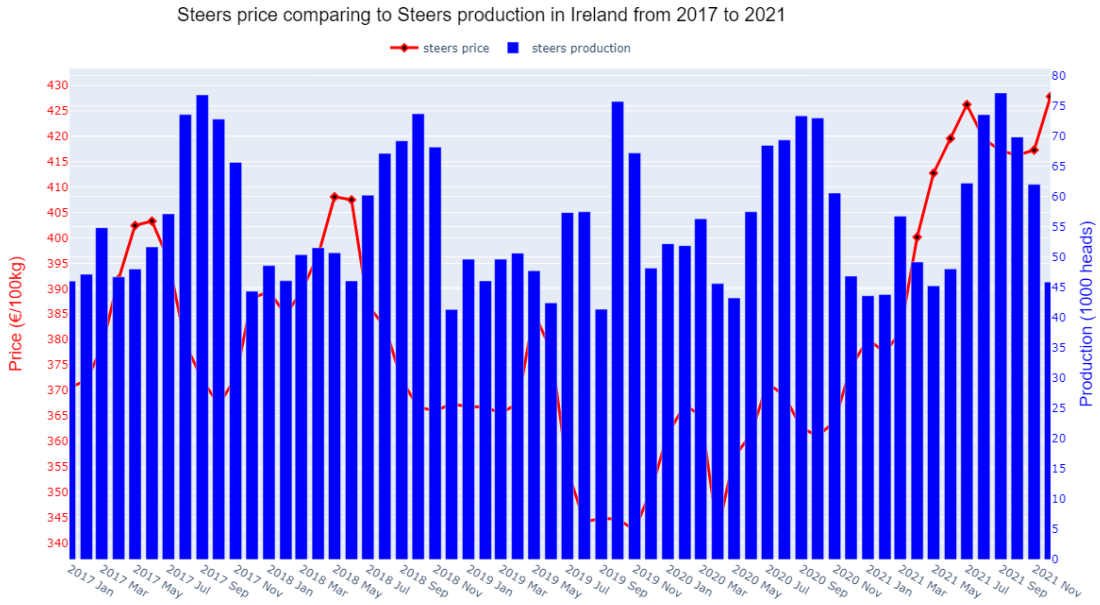
Figure 13: Beef feeding price in Ireland from 2017 to 2021



Overview:

- The beef feeding prices show a constant rising tendency, especially after September 2020
- The prices of two beef feedstuffs show high positive correlation

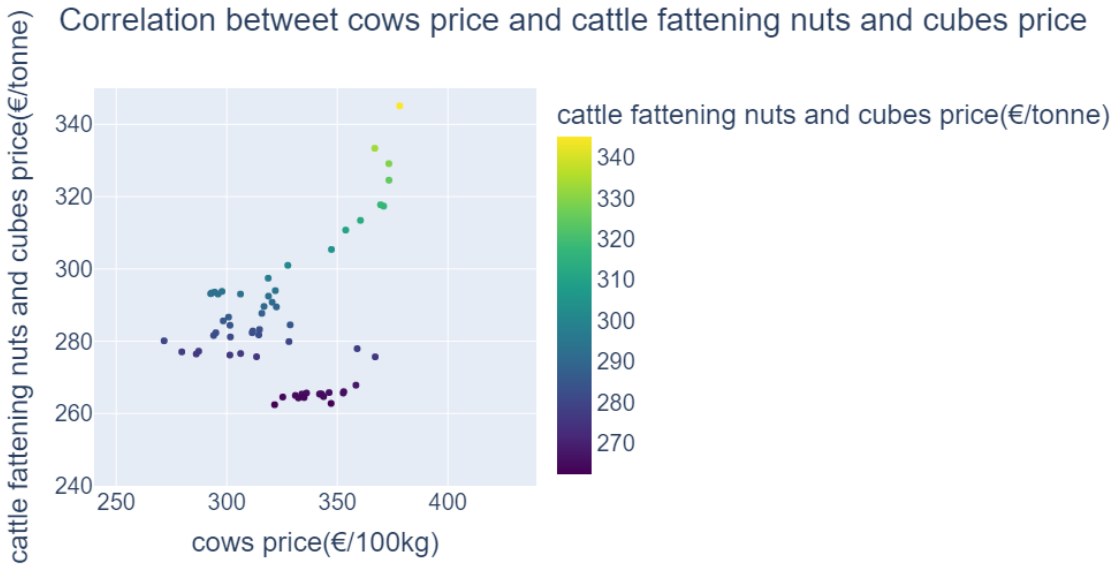
Figure 14: Relation between steer prices and steer production in Ireland from 2017 to 2021



Overview:

- This graph was present since it shows the "best" negative correlation between prices and production, e.g. September 2017, September 2019
- Generally speaking, the correlation between prices and production is not obvious (see Jupyter notebook)

Figure 15: Scatterplot showing correlation between cow prices and cattle fattening nuts and cubes prices



Overview:

- The correlation tends to be more obvious when the cow price gets really high
- Possibly the recent constant increasing in feeding prices is causing the cow prices to reach its highest
- The tendency mentioned above is also displaying in other beef categories (see attached)

Figure 16: Cow price in Ireland and Spain from 2017 to 2021

Overview:

- Cow prices in Ireland are higher than Spain
- The price difference is more obvious from January 2017 to May 2018 and from November 2020 to December 2021
- Similar tendency of fluctuation is showing

Figure 17: Heifer price in Ireland and Spain from 2017 to 2021

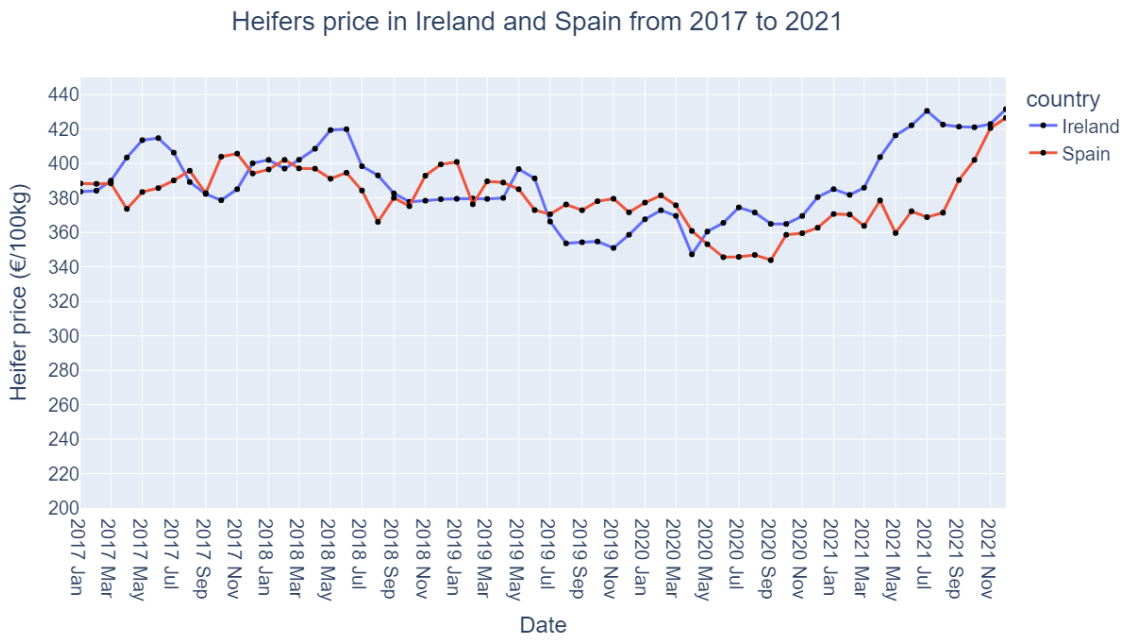
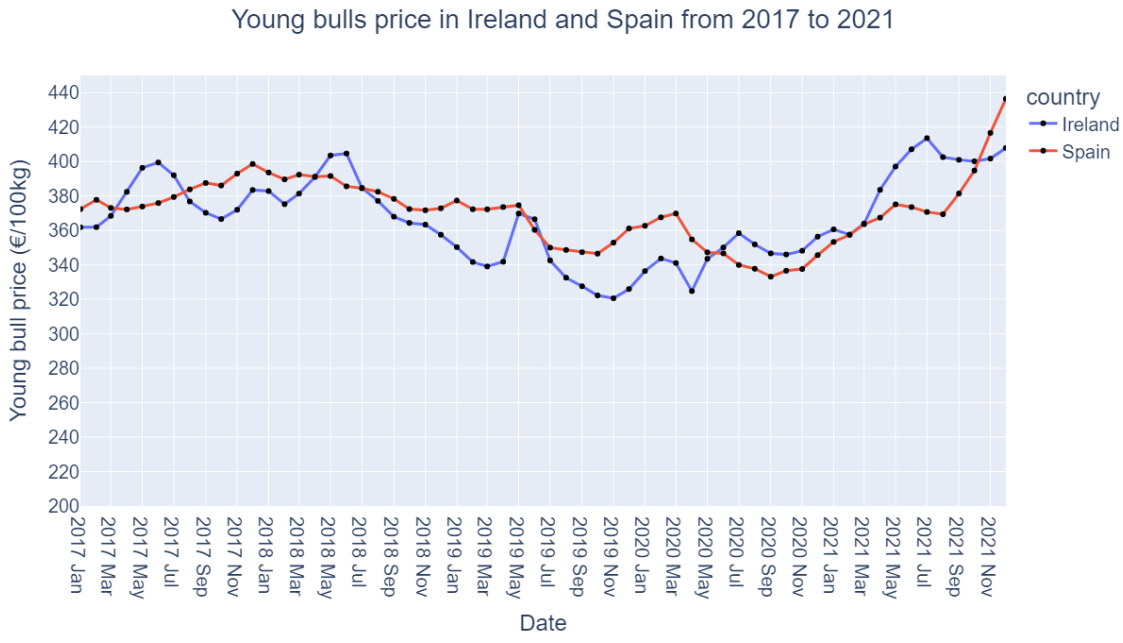


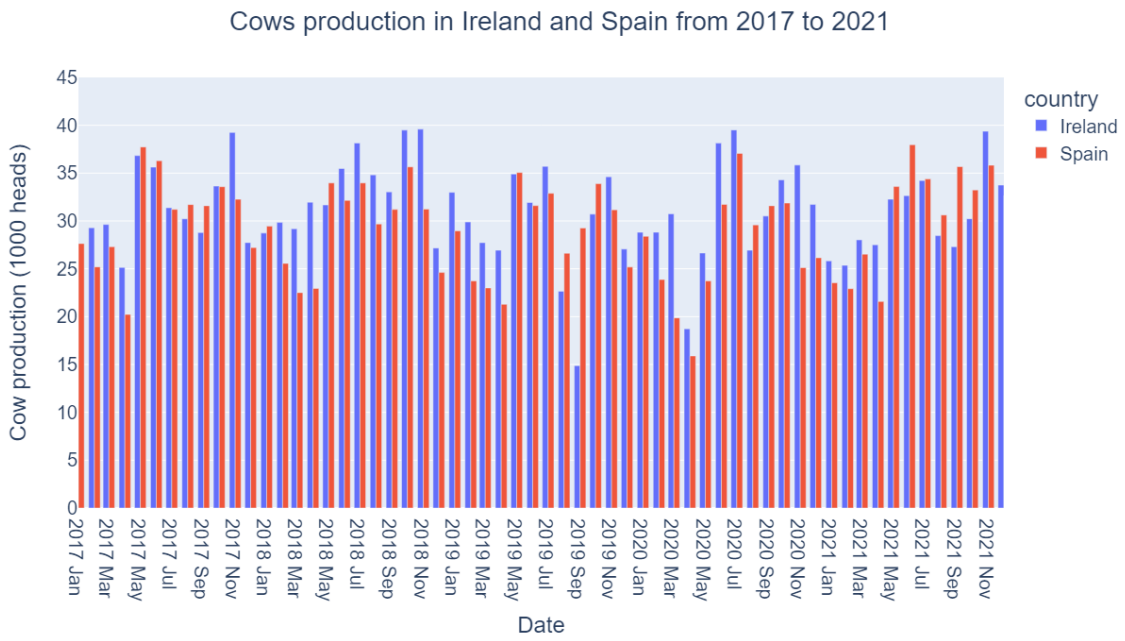
Figure 18: Young bull price in Ireland and Spain from 2017 to 2021



Overview:

- Figure 17 and Figure 18 demonstrate that the price differences of heifer and young bull between Ireland and Spain are insignificant

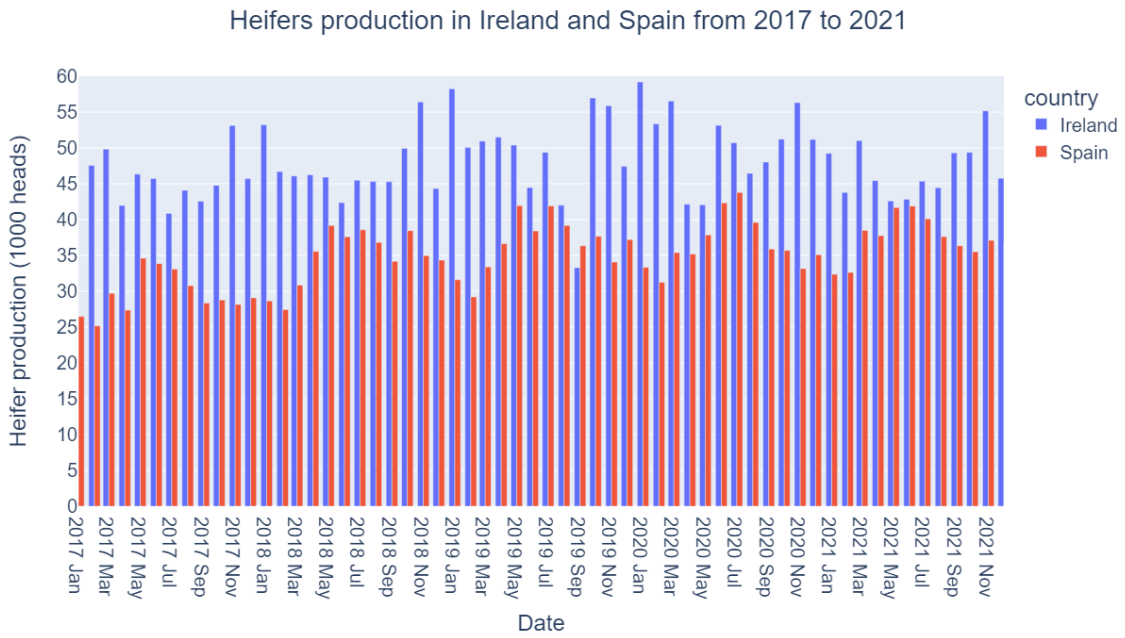
Figure 19: Cow production in Ireland and Spain from 2017 to 2021



Overview:

- Cow production in Ireland is higher than Spain in general
- Both countries show increasing in production around July each year

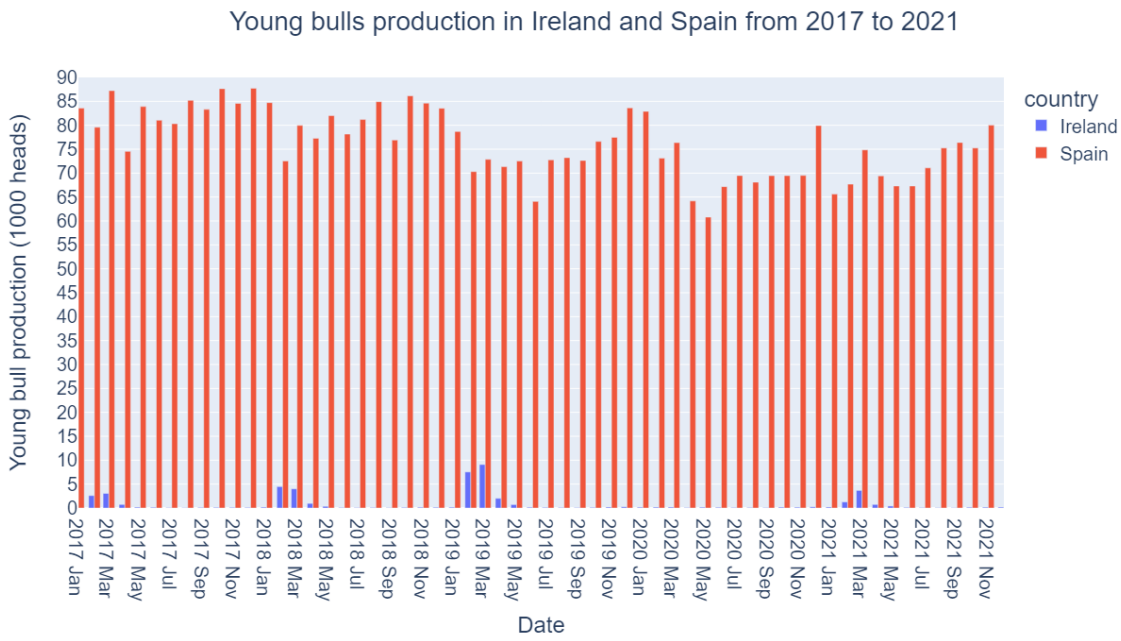
Figure 20: Heifer production in Ireland and Spain from 2017 to 2021



Overview:

- Heifer production in Ireland is much higher than Spain

Figure 21: Young bull production in Ireland and Spain from 2017 to 2021



Overview:

- There is way more young bull production in Spain than Ireland
- Spain has high young bull production across the whole years

Figure 22: Weight of cow per head in Ireland and Spain from 2017 to 2021

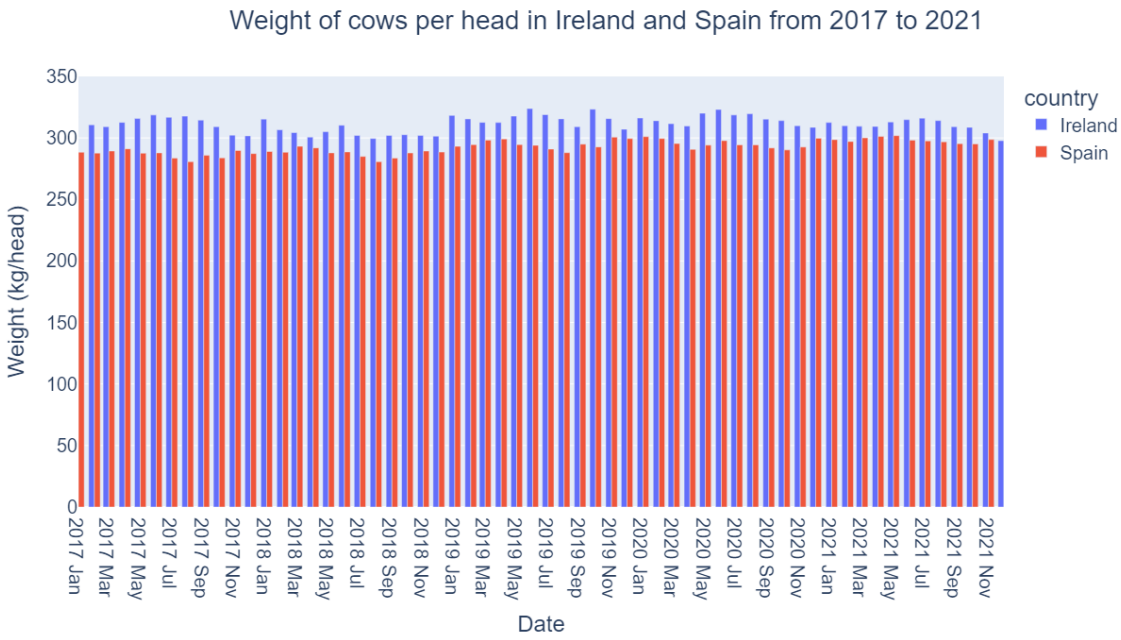


Figure 23: Weight of heifer per head in Ireland and Spain from 2017 to 2021

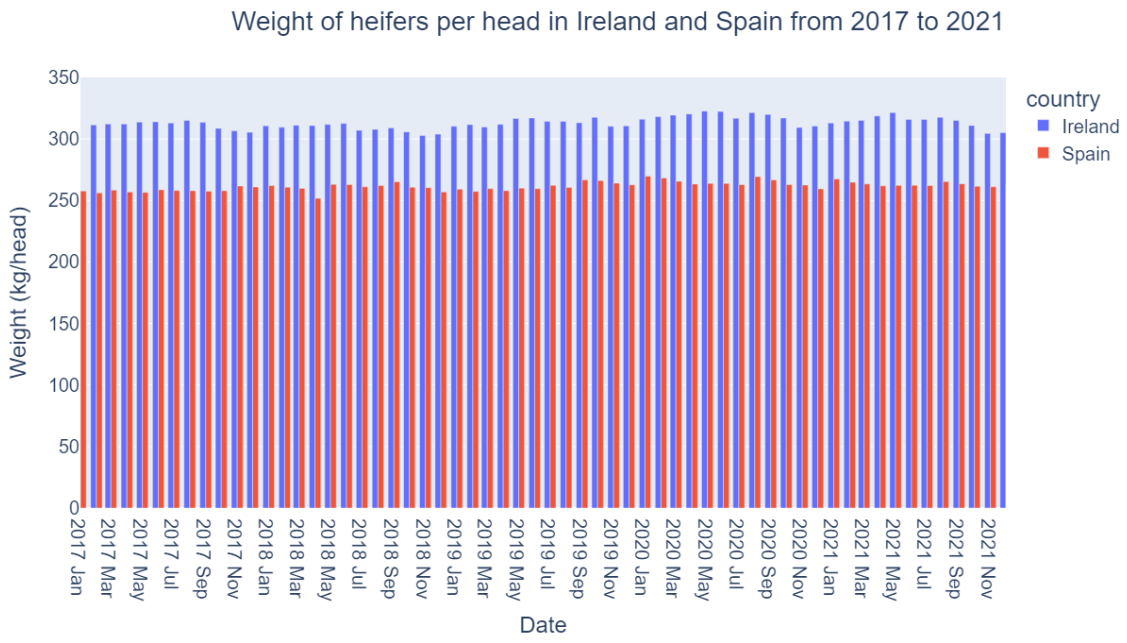
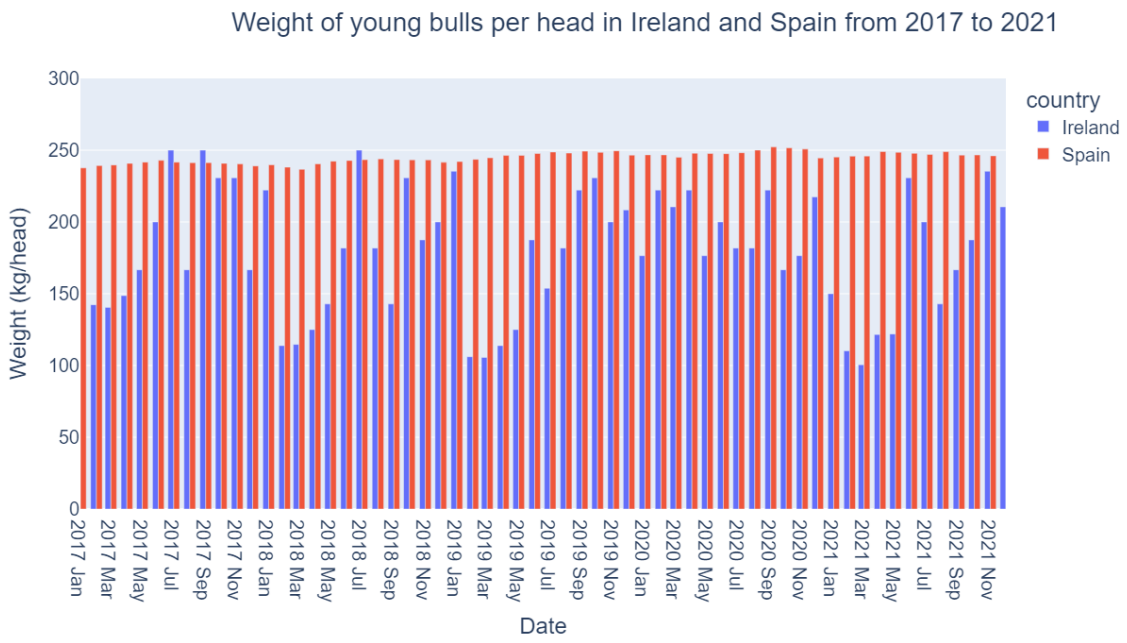


Figure 24: Weight of young bull per head in Ireland and Spain from 2017 to 2021



Overview:

- Figure 22-24 demonstrate that the weight of cow and heifer in Ireland is higher than Spain, however, the weight of young bull in Spain is higher than Ireland

6. Machine Learning

In our study, we employed two supervised machine learning techniques to train and test our data in order to make predictions on bull beef prices. The data utilized for training was sourced from the dataset "ireland_beef" and our goal was to predict bull prices based on the other features within the dataset, with the exception of the date and country variables

6.1 Linear Regression

In the Linear Regression model, we used the StandardScaler function to normalise our data. The dataset was then split into a 75% training and 25% testing partition. In order to optimise the model, we practiced GridSearchCV and tuned the hyperparameters to find the best combination that maximised the accuracy. Our final model achieved an accuracy of approximately 60% using 12 optimal features, as detailed in the Jupyter notebook.

6.2 Random Forest

The Random Forest model performed extremely poorly, reaching at most an accuracy of 6%, even after practicing GridSearchCV to tune the hyperparameters. As a result, this model is not recommended for training on this dataset, as demonstrated in the accompanying Jupyter notebook.

7. Sentimental analysis

Sentiment analysis, also known as opinion mining, is a field of natural language processing that aims to identify and extract subjective information from source materials, such as text or speech (Liu, 2012). The primary goal of sentiment analysis is to determine the attitude, sentiment, or emotional state of the text's author or speaker with respect to a particular topic or subject.

The goal of the sentimental analysis was to gather and analyse recent public opinions on beef prices using the Twitter API. To accomplish this, we requested and collected over 80 tweets through the API. We then applied various text processing techniques, such as converting all words to lowercase, removing punctuation, and removing stopwords and so on.

Finally, we applied a logistic regression model to fit the processed data, utilising either the TfidfVectorizer or CountVectorizer function, and achieved an accuracy of

approximately 94%. This suggests that the logistic regression model was able to effectively classify the sentiment of the tweets.

8. Conclusion

In summary, through our study, we found that hypothetically there was no significant difference in cow and heifer prices between Ireland and Spain. However, there was a possibly significant difference in young bull prices between the two countries. Additionally, we observed an annual increase in beef prices around March for all beef categories, and noted that beef feeding prices have been on the rise since 2017, with a particularly steep increase after 2020. Our study also revealed that Ireland had higher beef production for cow and heifer compared to Spain, but that young bull production was rare in Ireland. Finally, we found that Linear and Logistic Regression models were effective for machine learning and sentiment analysis, achieving good accuracy, whereas the Random Forest model was not suitable for our data.

- **Version control link:** https://github.com/alecfei/beef-price-analysis

References

Einar Martin, A. (2021) *Parametric versus nonparametric tests*. Available at: https://ledidi.com/academy/parametric-versus-nonparametric-tests (Accessed: 8 January 2023).

Branco, P.A.B. (2020) *Food Production and Agriculture*. Canada: Delve Publishing.

Central Statistics Office (no date) Databases - CSO - Central Statistics Office. CSO. Available at: https://www.cso.ie/en/databases/ (Accessed: 8 December 2022).

Central Statistics Office (2021) *Demographic Profile of Farm Holders - CSO - Central Statistics Office*. CSO. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-coa/censusofagriculture2020-preliminaryresults/demographicprofileoffarmholders/ (Accessed: 5 January 2023).

Croxton, F.C. (1905) 'Beef Prices', *Journal of Political Economy*, 13(2), pp. 201–216.

Department of Agriculture, Food and the Marine (2022) *EU Beef Carcase Classification Scheme*. Available at: https://www.gov.ie/en/collection/bc95b-eu-beef-carcase-classification-scheme/ (Accessed: 17 December 2022).

Eldertech, M. (2017) 'Designing Technology for Seniors - Color in User Interfaces for Elderly People', *User Interfaces for Seniors*, 25 March. Available at:

https://eldertech.org/color-in-designing-technology-for-seniors/ (Accessed: 4 January 2023).

European Commission (2022) *CAP overview*. Available at: https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview_en (Accessed: 10 December 2022).

European Commission (2021) *EU country factsheets*. Available at: https://agriculture.ec.europa.eu/system/files/2022-04/agri-statistical-factsheet-ie_en_0.pdf (Accessed: 15 December 2022).

Liu, B. (2012) *Sentiment Analysis and Opinion Mining*. Cham: Springer International Publishing (Synthesis Lectures on Human Language Technologies). Available at: https://doi.org/10.1007/978-3-031-02145-9.

Kahn, L. and Cottle, D. (2014) *Beef Cattle Production and Trade*. Collingwood, Vic: CSIRO PUBLISHING.

Reif, F. (2008) *Fundamentals of Statistical and Thermal Physics*. 56946th edition. Long Grove, Ill: Waveland Pr Inc.