POLITECNICO MILANO 1863

M.Sc. Computer Science and Engineering
Data Analysis for Smart Agriculture

Data Analysis On An Eggs Farm

Davide Canali	10674880	davide1.canali@polimi.it
Matteo Cordioli	00000000	matteo.cordioli@polimi.it
Federico Camilletti	10619856	federico.camilletti@polimi.it
Shakiba Shahidiani	00000000	shakiba.shahidiani@polimi.it

December 12, 2022

Contents

1 Introduction

In this study, we are going to analyze data from an eggs farm near Mantova to see if it's possible to improve both animal welfare and farmer revenue.

The farm under analysis is [//TODO INSERT NAME] and has around 40'000 chickens that produce organic eggs. We have the data starting from 2014, the production of eggs is divided into cycles lasting about 13 - 15 months each. We have 5 complete cycles and the current 2022 cycle. The first 2 cycles (called X and Y) are non-organic which means the chickens are treated differently from the last 4 cycles (Z, A, B, C) which are organic.

Upon talking with the farmer we focus our attention on 3 main topics which involve:

- Understanding the mortality between different cycles and organic with non-organic.
- Improve the welfare of the chickens.
- Quantify the monetary loss when a chicken dies at the start of the cycle.

2 Data Collection

The farmer collected the data thanks to sensors and by inspecting the barns each day. Before cycle A the data were only collected in paper spreadsheets and had to be imported in a digital format.

3 Data Cleaning

Before starting to analyze the data we cleaned them. In all the cycles we performed these next steps:

- Removed tail and head (20 rows each more or less).
- Adjusted feed and water consumption to correct typos and switchboard errors.
- Adjusted eggs production when % of laid when above 100%, distributed in the days before assuming that those eggs were not collected.
- Adjusted death count to match the delta of death at the end of the cycle.
- Checked that the difference between the total amount of eggs produced and the number of eggs sold was less than 0.6%.

Specifically, in cycle A we also adjust a series of missing data regarding the water consumption adding them following the cycle distribution.

//TODO Teo add what u did in cycles Y and X

4 Analysis of Each Cycle

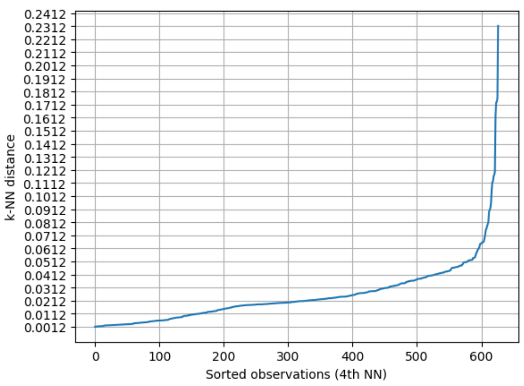
4.1 Cycle A

Arrival date	#Chickens	Frist Laid	End of cycle	Organic	#Eggs
19/7/2018	42.009	1/9/2018	19/5/2020	Yes	20.208.086

Significant features:

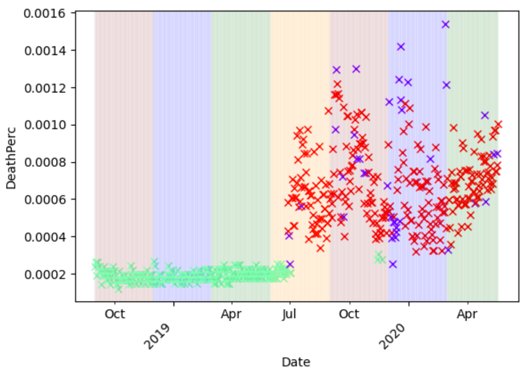
- Average Percentage of Laid Eggs each day: 83,8%
- Average Percentage of Death each day: 0,044%
- Average Temperature during the cycle: 12,99 °C
- Average Humidity during the cycle: 74,55%

Cluster:

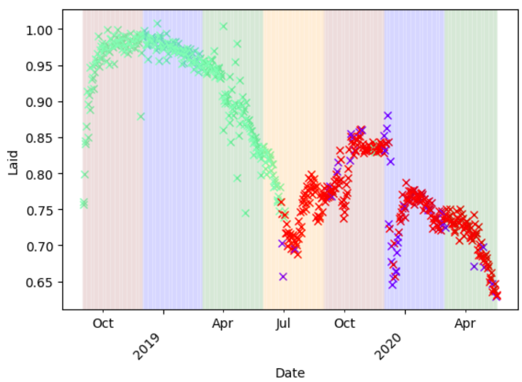


k-NN distance to find ϵ

Density scatter plot merged



Percentage of death



Percentage of laid eggs

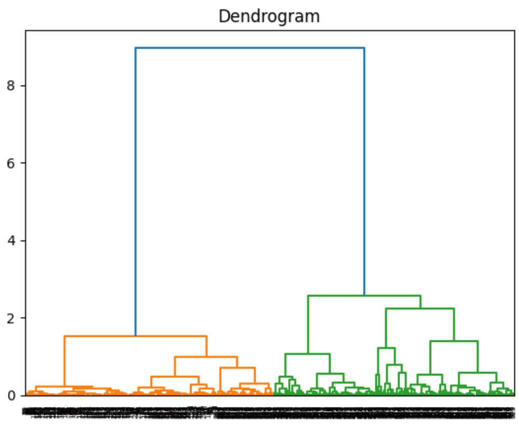
4.2 Cycle B

Arrival date	#Chickens	Frist Laid	End of cycle	Organic	#Eggs
09/08/2020	42.098	24/09/2020	02/05/2022	Yes	18.392.640

Significant features:

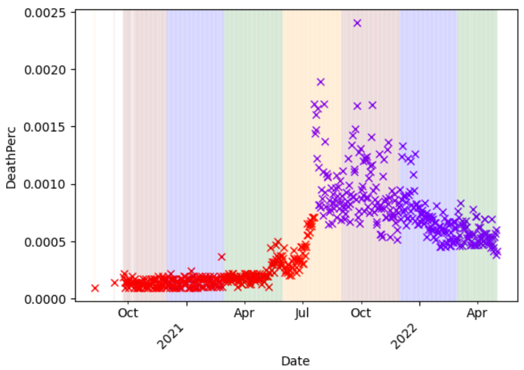
- Average Percentage of Laid Eggs each day: 81,39%
- Average Percentage of Death each day: 0,0494%
- Average Temperature during the cycle: 11,81 °C
- Average Humidity during the cycle: 73,24%

Cluster:

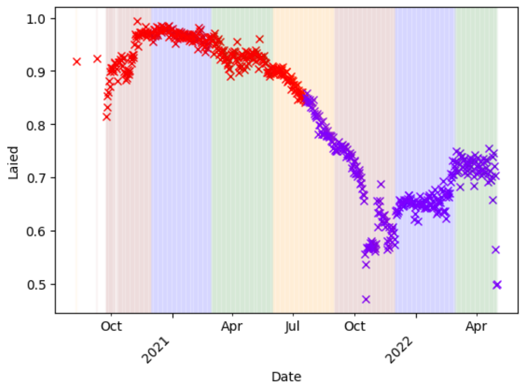


Dendogram

Hierarchical cluster scatter plot



Percentage of death



Percentage of laid eggs

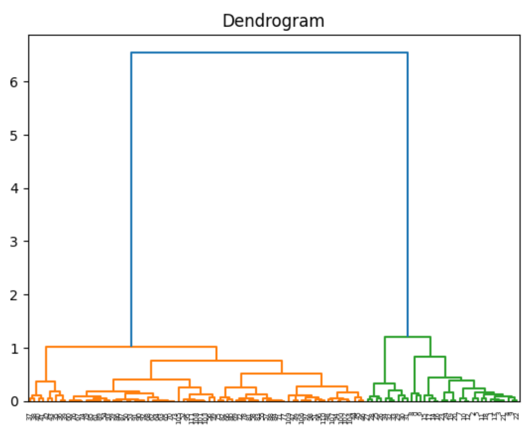
4.3 Cycle C

Arrival date	#Chickens	Frist Laid	End of cycle	Organic	#Eggs
20/06/2022	42.098	08/08/2022	In progress	Yes	In progress

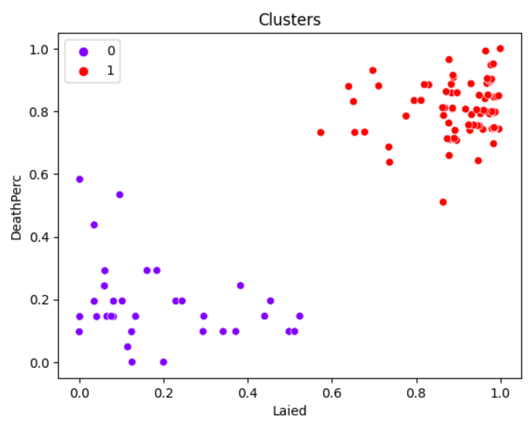
Significant features:

- Average Percentage of Laid Eggs each day: 55,85%
- Average Percentage of Death each day: 0,043%
- Average Temperature during the cycle: 18,61 °C
- Average Humidity during the cycle: 70,05%

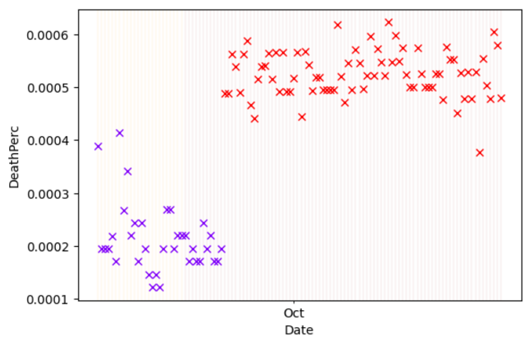
Cluster:



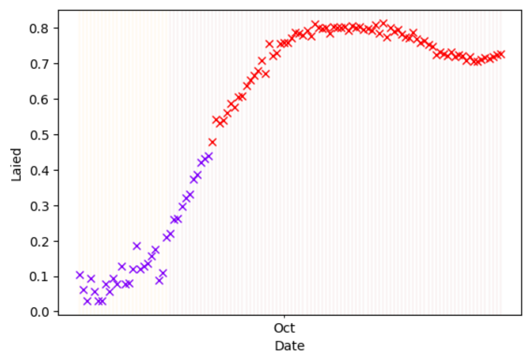
Dendogram



Hierarchical cluster scatter plot



Percentage of death



Percentage of laid eggs

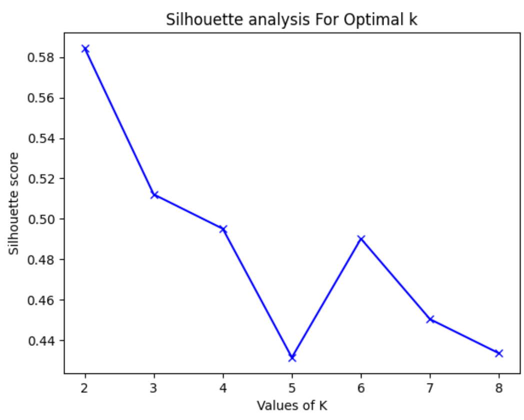
4.4 Cycle X1

Arrival date	#Chickens	Frist Laid	End of cycle	Organic	#Eggs
20/01/2014	33.743	18/03/2014	08/07/2015	No	12.375.840

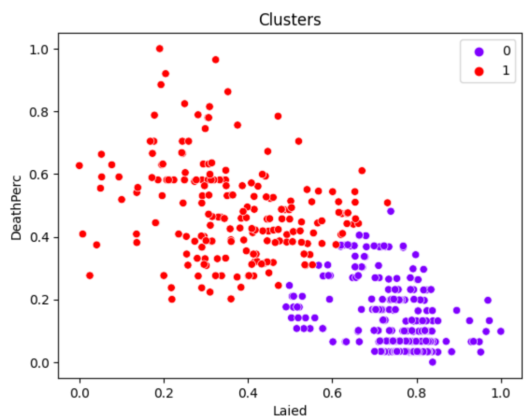
Significant features:

- Average Percentage of Laid Eggs each day: 80,69%
- Average Percentage of Death each day: 0,032%
- Average Temperature during the cycle: 15,34 °C
- Average Humidity during the cycle: 72,83%

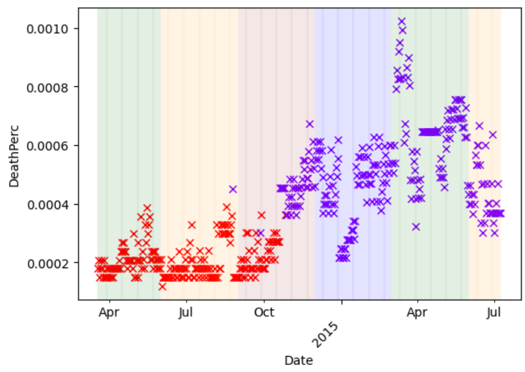
Cluster:



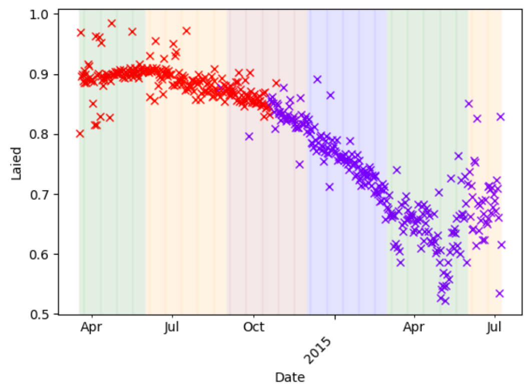
Silhouette



K-Means cluster scatter plot



Percentage of death



Percentage of laid eggs

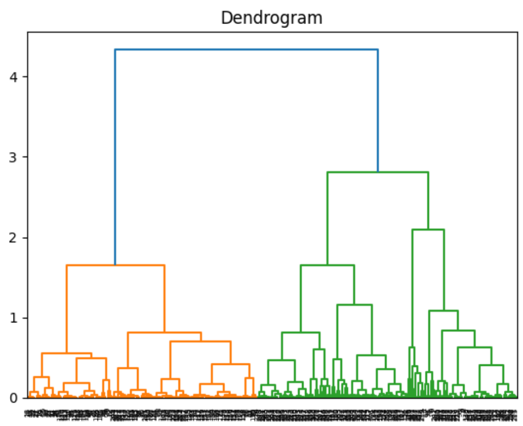
4.5 Cycle X2

Arrival date	#Chickens	Frist Laid	End of cycle	Organic	#Eggs
26/05/2014	23.898	15/07/2014	21/06/2015	No	7.558.799

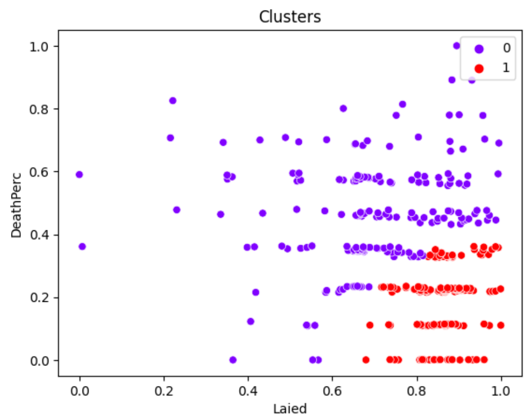
Significant features:

- Average Percentage of Laid Eggs each day: 95,36%
- Average Percentage of Death each day: 0,021%
- Average Temperature during the cycle: 13,76 °C
- Average Humidity during the cycle: 76,04%

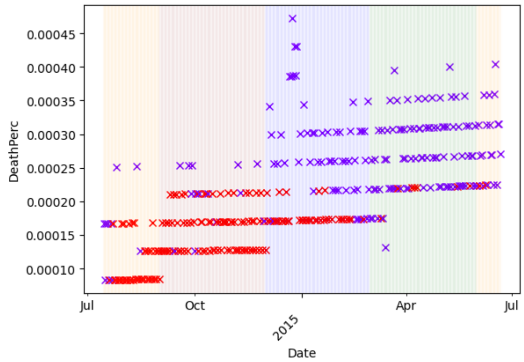
Cluster:



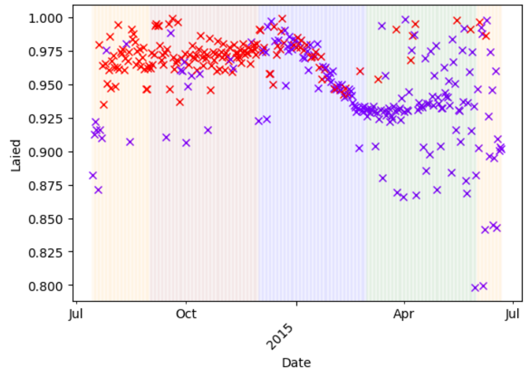
Dendogram



Hierarchical cluster scatter plot



Percentage of death



Percentage of laid eggs

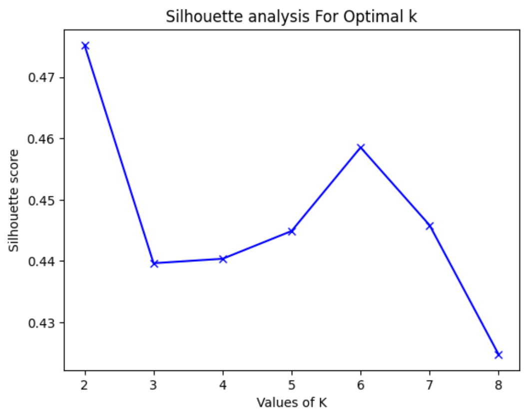
4.6 Cycle Y

Arrival date	#Chickens	Frist Laid	End of cycle	Organic	#Eggs
11/08/2015	57.346	05/10/2015	27/09/2016	No	16.759.240

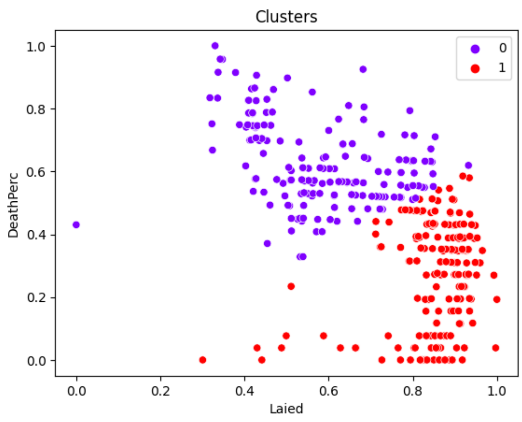
Significant features:

- Average Percentage of Laid Eggs each day: 83.70%
- Average Percentage of Death each day: 0,022%
- Average Temperature during the cycle: 14,20 °C
- Average Humidity during the cycle: 74,54%

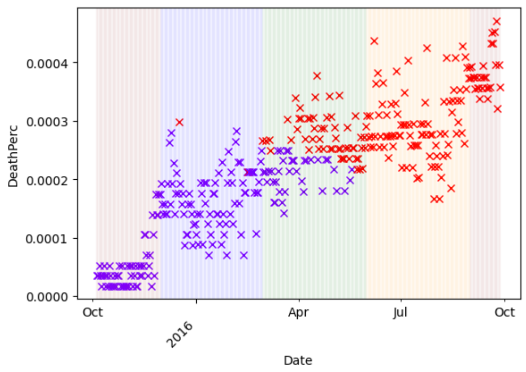
Cluster:



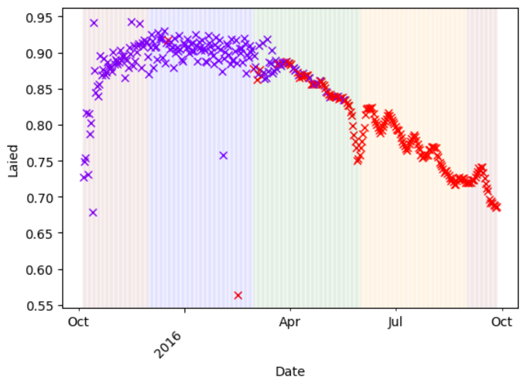
Silhouette



K-means cluster scatter plot



Percentage of death



Percentage of laid eggs

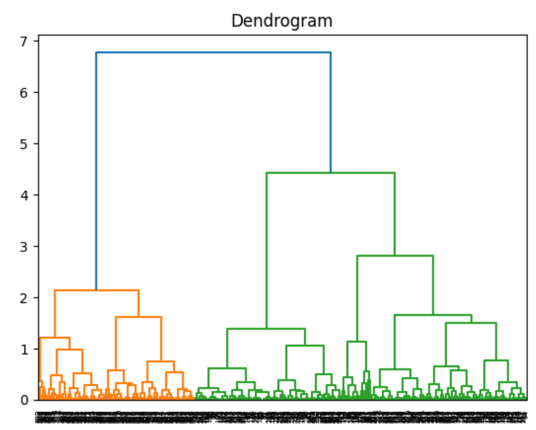
4.7 Cycle Z

Arrival date	#Chickens	Frist Laid	End of cycle	Organic	#Eggs
17/11/2016	42.130	08/01/2017	27/05/2018	Yes	17.721.240

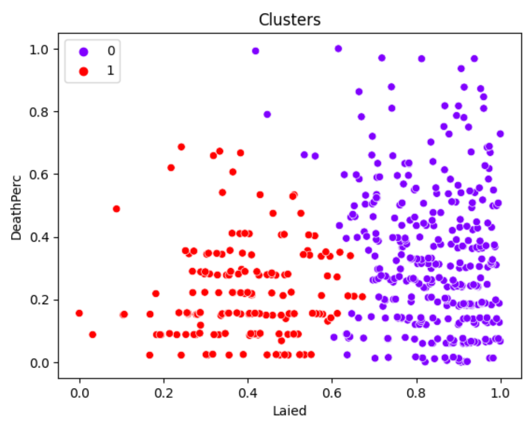
Significant features:

- Average Percentage of Laid Eggs each day: 87,78%
- Average Percentage of Death each day: 0,021%
- Average Temperature during the cycle: 13,07 °C
- Average Humidity during the cycle: 74,27%

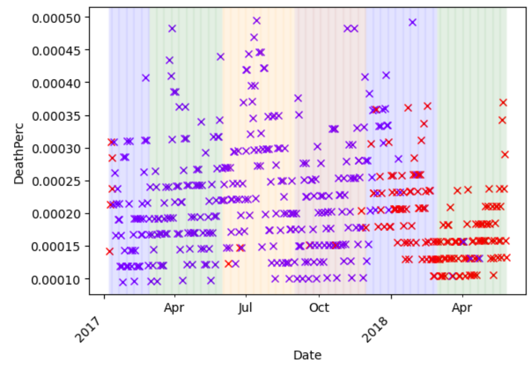
Cluster:



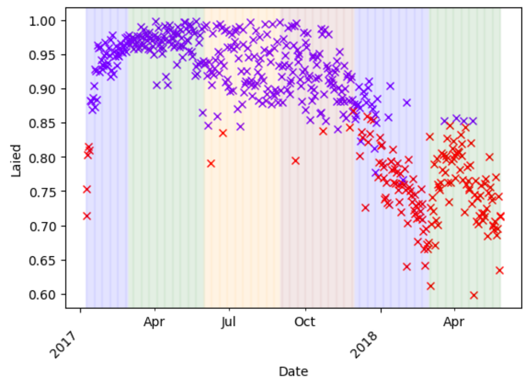
Dendogram



Hierarchical cluster scatter plot



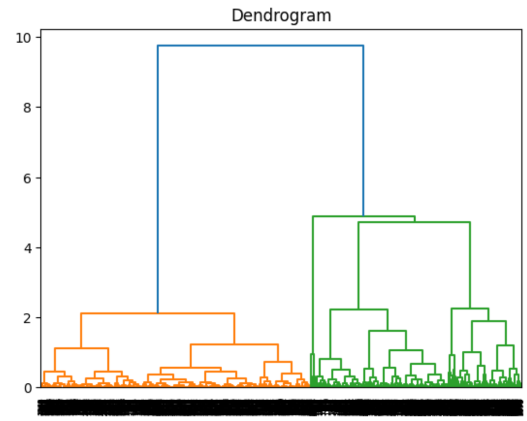
Percentage of death



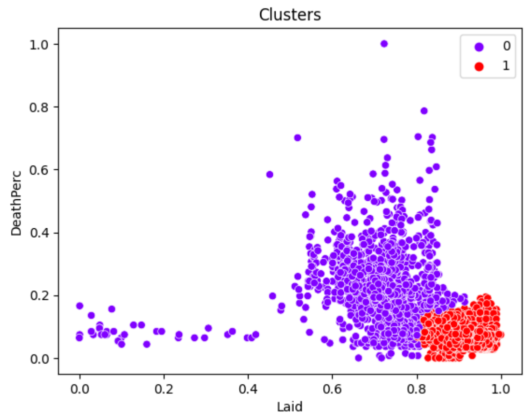
Percentage of laid eggs

5 Common features

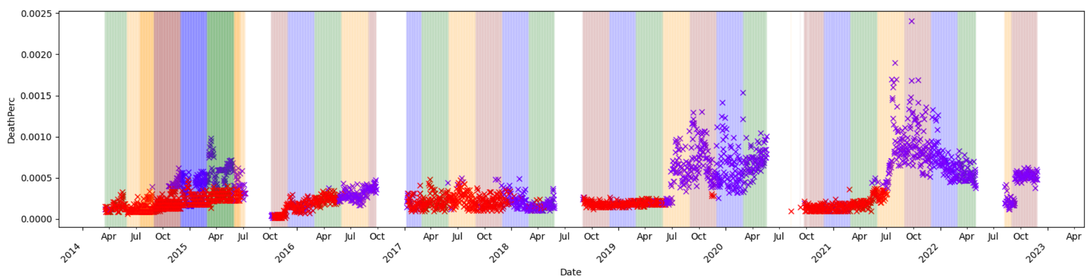
In this section, we will discuss what we discovered by comparing all the cycles together. After seeing that each cycle can be clustered in two parts, those clusters also divide the cycle into two distinct sections in a plot based on date. We tried to apply the same idea in a dataset composed of all the cycles together.



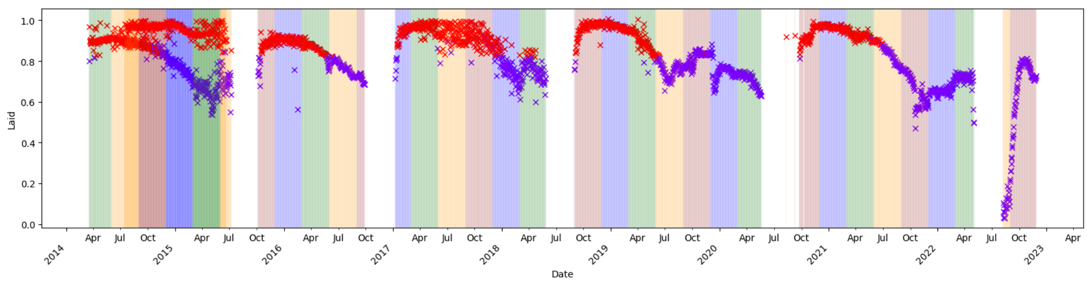
Dendogram



Hierarchical cluster scatter plot



Percentage of death

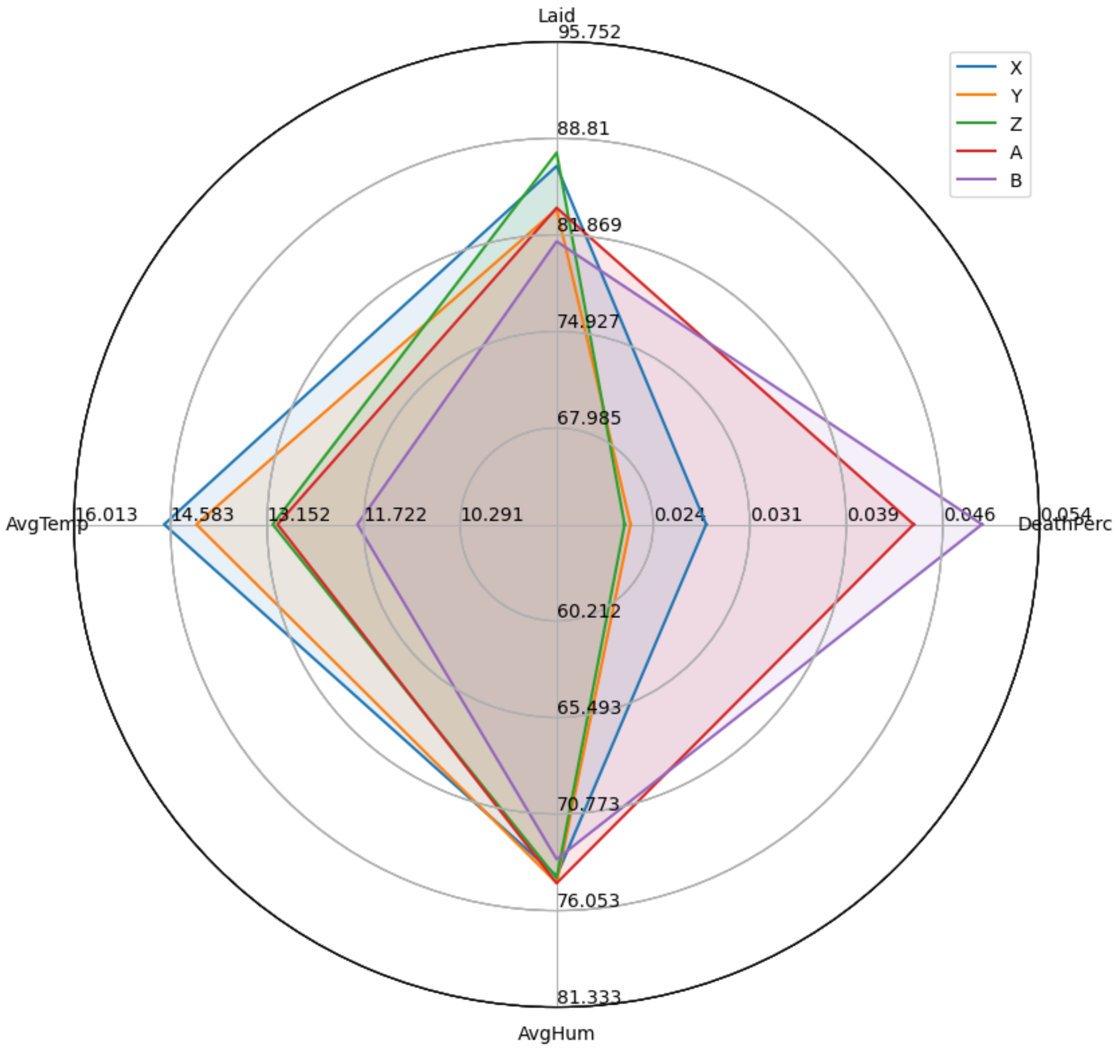


Percentage of laid eggs

Looking at the clusters we find that indeed the division of each cycle is respected. We find also that the only one that didn't respect this division is cycle C, which could be expected since it just started, but the interesting thing is that the whole cycle C is clustered as cluster 1 which is the one that identifies the end of every other cycle so we can see how bad is performing this new cycle both in death and laid rate. This can show how the ban on beak-cutting affected the welfare of the chickens, with the beak the violence between animals increased a lot and not only increase the death rate but increased a lot the stress of the chickens reducing their productivity.

To better compare the different cycles we made a spider plot with 4 features:

- Average Percentage of Laid Eggs each day
- Average Percentage of Death each day
- Average Temperature during the cycle
- Average Humidity during the cycle



6 Organic vs non-organic cycles

One of the main differences in the cycles under analysis is the breeding method. Cycles A, B, C and Z are grown with organic farming, a method regulated by the European Union with the regulation 589/2008 modified from text 2168/2017 while the cycles Y, X1 and X2 are non-organic. We can summarize the main differences:

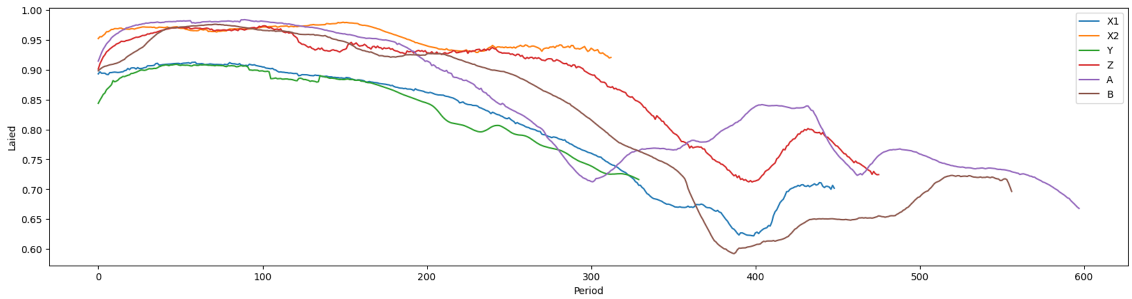
Charachteristics	Organic	Non-organic
Chickens/m^2	6	9
Outdoor access	✓	✓
Feed requirements	✓	×

Given the differences in the breeding methods, we analyze if there are implications in the cycles' behavior. Since the non-organic ones have no information on feed and water consumption, we can't go deep into the analysis because of the lack of inputs. For this reason, we will focus mainly on the outputs like the percentage of eggs laid and deaths.

One of the most important outputs is the productivity of a cycle expressed as EggsProduced#Chickens. In the following results, we exclude cycle C since it is in progress. Furthermore, we round the data to the second digit.

Cycle	Organic	Productivity	Breeding days	Productivity/Breeding days
A	✓	481.04	627	0.77
B	✓	436.90	586	0.75
Z	✓	420.63	505	0.83
X1	×	366.77	478	0.77
X2	×	316.29	342	0.92
Y	×	292.25	359	0.81

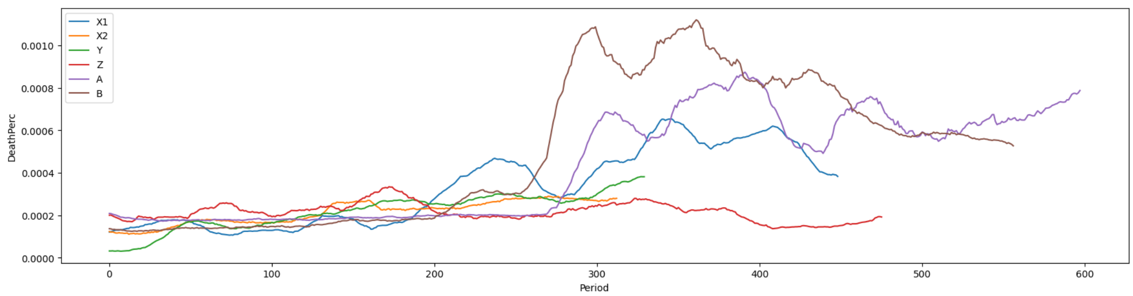
We can notice that non-organic cycles have a lower EggsProduced/#Chickens rate than the organic ones, but considering the days of breeding the non-organic cycles are better. The medium ProductivityBreeding days of an organic cycle is 0.78 while the one of non-organic is 0.84.



Beyond the productivity in terms of eggs produced there is another important aspect to consider which is the chickens' health. Since it is a difficult feature to value, we decided to consider the chickens' death percentage on the number of chickens as a health indicator.

Cycle	Organic	Deaths on chickens	Breeding days	Daily deaths average
A	✓	0.24	627	16.18
B	✓	0.25	586	18.07
Z	✓	0.10	505	8.53
X1	×	0.14	478	10.10
X2	×	0.07	342	4.93
Y	×	0.07	359	12.03

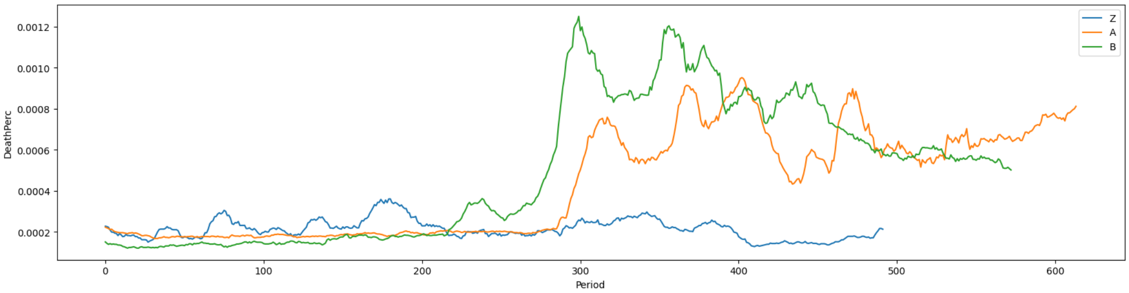
It is evident that in organic cycles the deaths on chickens' number is much higher than in the non-organic ones. So we decided to better understand the reason of this behavior. This question could have a different ways to be answered or a mix of them: - Different feeds quality - Different water consumption - ... So we decided to look for correlations:



7 Death-season correlation

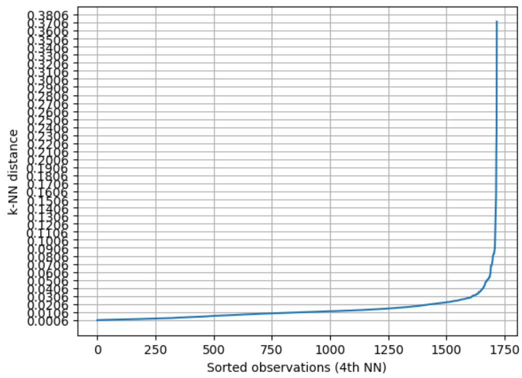
In this section, we will analyze how the period when the chickens are introduced into the barn affects the mortality rate. This effect will be more visible in the organic cycles where the chickens are more fragile due to a lack of nutrients.

First, we plot the Death Rate of all the organic cycles (Z, A, B).

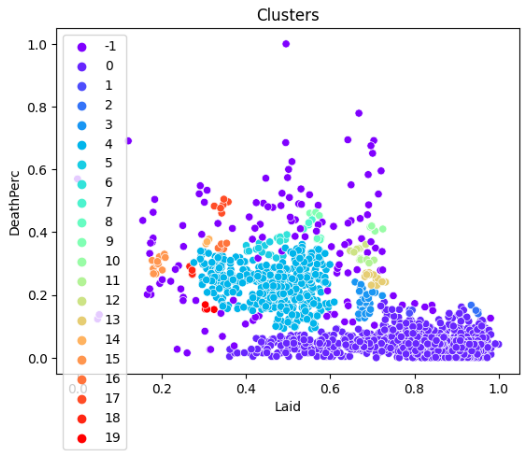


It's clear that the Z cycle outperforms the other two by a lot, so we will analyze what are the differences between those 3 cycles.

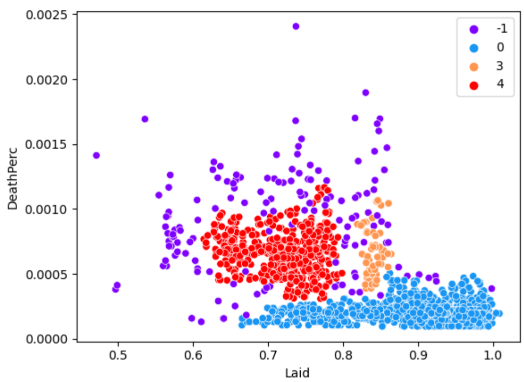
The number of chickens at the start of the cycle is identical (~ 42.000), and the feed in an organic farm is standard and checked so it doesn't change between those cycles. The difference in mortality cannot be explained with just "a lucky cycle" so we tried to analyze those graphs deeper.



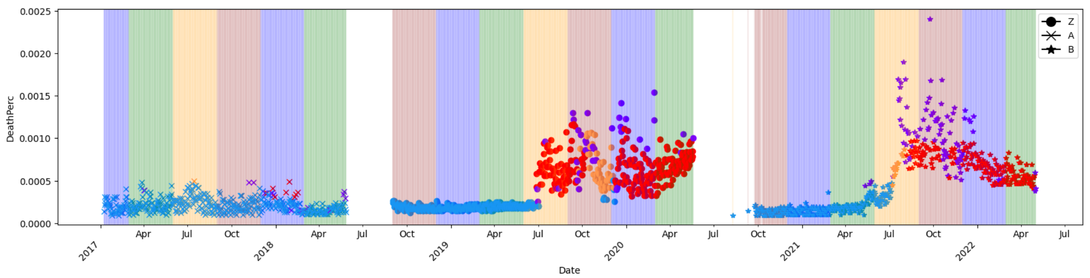
k-NN distance to find ϵ



Raw Density



Density Cleaned



Percentage of death

8 Economic results

9 Data processing: shakiba