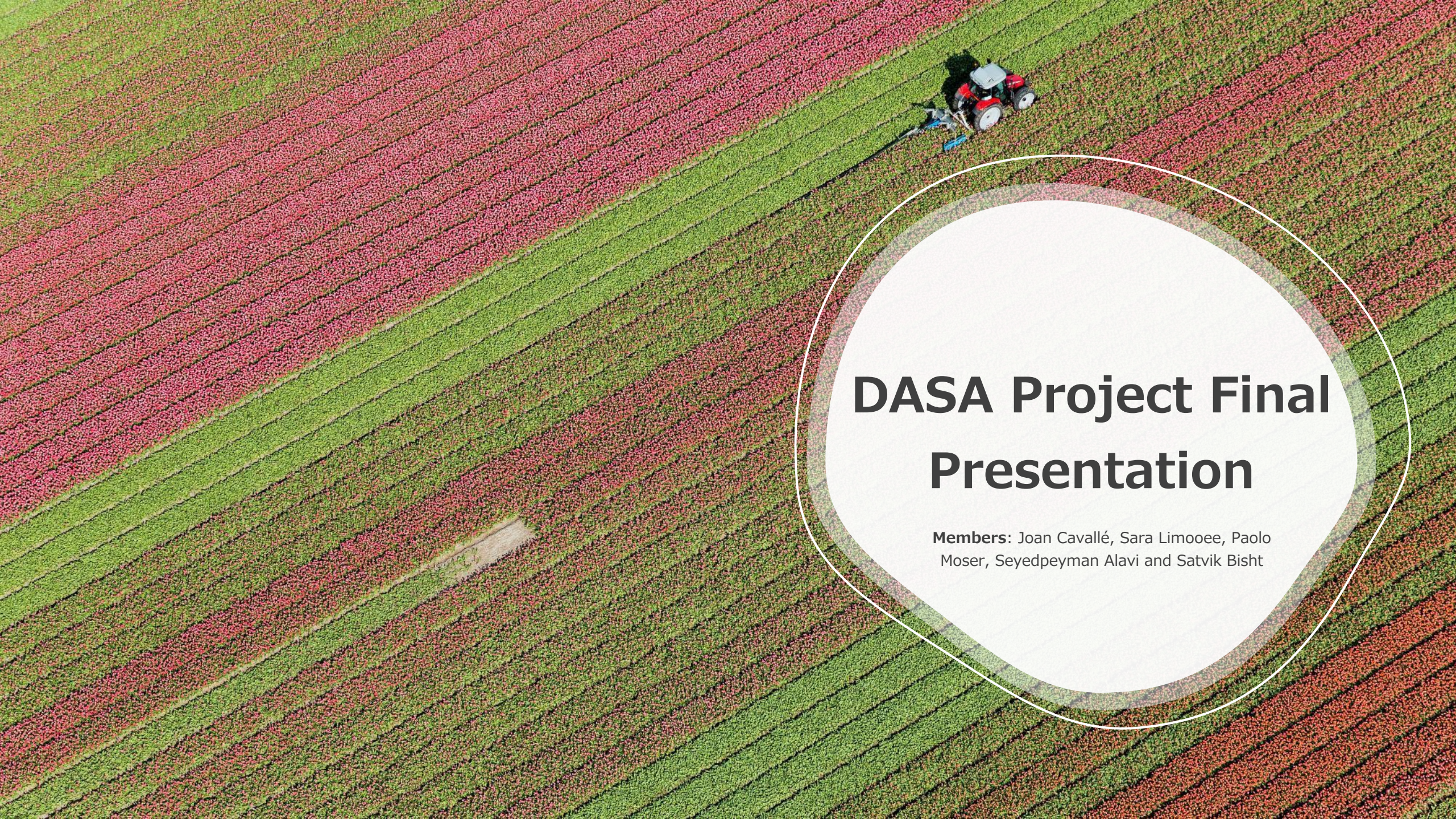


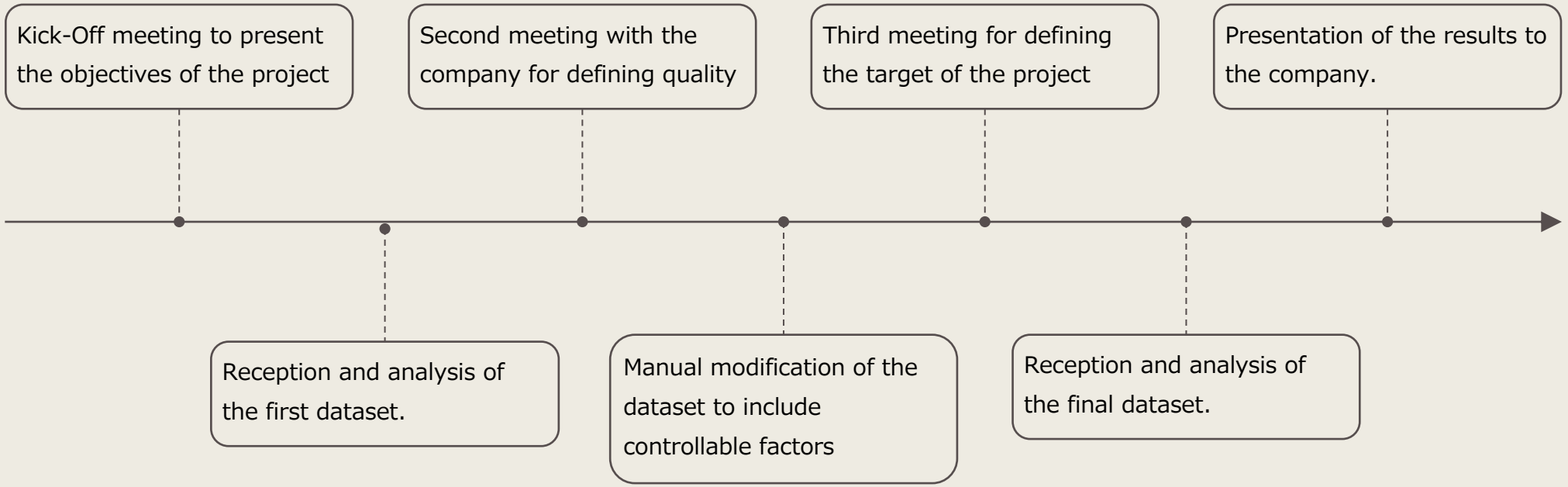
DASA Project Final Presentation

Members: Joan Cavallé, Sara Limooee, Paolo Moser, Seyedpeyman Alavi and Satvik Bisht

Objective of the project

Identify factors influencing **milk quality** to enable *Latteria company(?)* to **increase its revenues.**

Steps


Kick-Off meeting to present the objectives of the project
Second meeting with the company for defining quality
Third meeting for defining the target of the project
Presentation of the results to the company.
Reception and analysis of the first dataset.
Manual modification of the dataset to include controllable factors
Reception and analysis of the final dataset.

Data structure

Timestamp

1

farm	Date of the analysis	spore	proteine	cellule somatiche	carica batterica	grasso	crioscopia	lattosio	urea	caseine	quality
-	-	-	-	-	-	-	-	-	-	-	

Text Integer Float Integer

2

farm	zone	nº of cows	height	breed
-	-	-	-	-

Text Integer Text

3

farm	Milking robot	Age of farmer >45
-	-	-

Boolean

4

farm	Welfare 2020	Welfare 2021	Welfare 2022
-	-	-	

Float

Milk quality

- It is calculated based on the level of:
    1. Fat
    2. Caseins
    3. Spores
    4. Somatic Cells
    5. Bacterial Load

farm	Date of the analysis	**spore**	proteine	**cellule somatiche**	**carica batterica**	**grasso**	crioscopia	lattosio	urea	**caseine**	**quality**
-	-	-	-	-	-	-	-	-	-	-	

- It adds/subtracts points depending the level of them
- The milk quality function is used to define the price per liter that they pay to the farmers

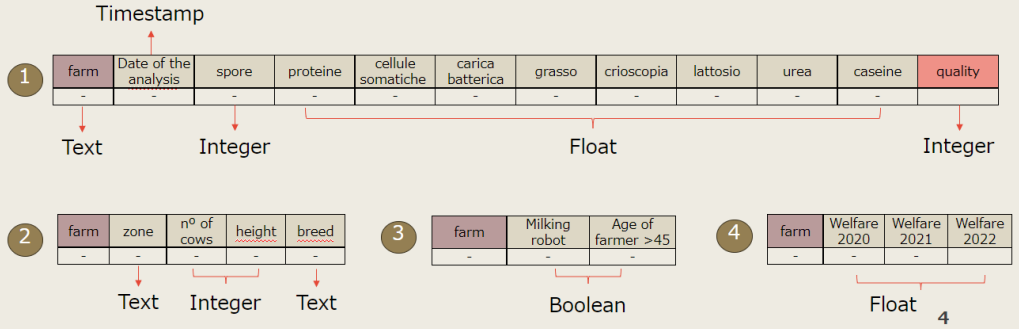
Preliminary tasks on the data

1. DATA CLEANING

- Invalid data (out of range features)
- Duplicated rows
- Farms with low samples (not representative)

2. DATA STANDARDIZATION

- Farm names different across the multiple datasets, we had to standardize them
- Different breed names
- Fuzzywuzzy library



3. MANUAL DATA INSERTION

- Number of cows/farm
- Altitude of the farm
- Age of the farmer
- Geographical zone in which they belong
- Predominant cow breed/farm
- Quality computation

Target definition

1

STUDY IF HEIGHT AFFECT QUALITY

- Technique: Statistical analysis

2

STUDY WHICH MILKING AREAS HAVE BETTER QUALITY

- Technique: K-Means, Hierarchical Clustering

3

STUDY IF THE NUMBER OF COWS AFFECT THE QUALITY

- Technique: Statistical analysis

4

WHAT IMPORTANT FEATURES AFFECT QUALITY

- Technique: Clustering

5

STUDY IF THE COW BREED AFFECTS QUALITY

- Technique: Statistical analysis

6

MACHINE LEARNING FOR CLASSIFYING MILK QUALITIES

- Technique: Logistic regression, SVC

7

STUDY IF THE MILKING ROBOT AFFECTS QUALITY

- Technique: Statistical analysis

8

SOMATIC CELLS DEPENDS ON NUMBER OF COWS

- Technique: Statistical analysis, K-Means

9

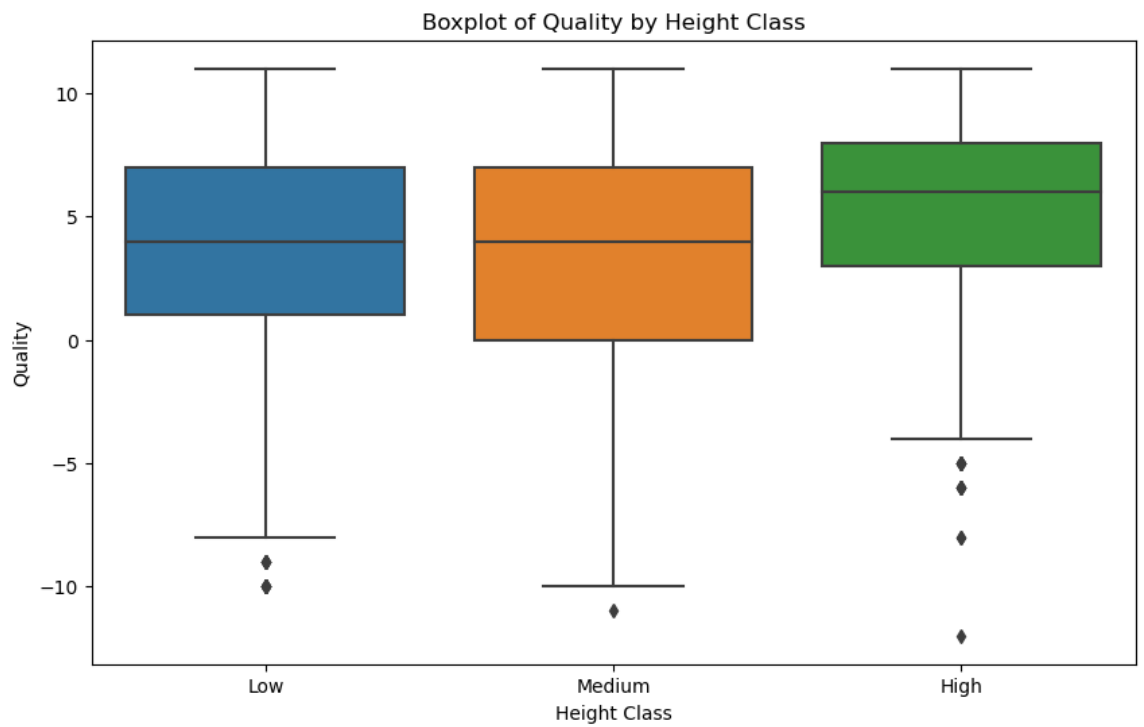
STUDY THE WELFARE OF THE COWS

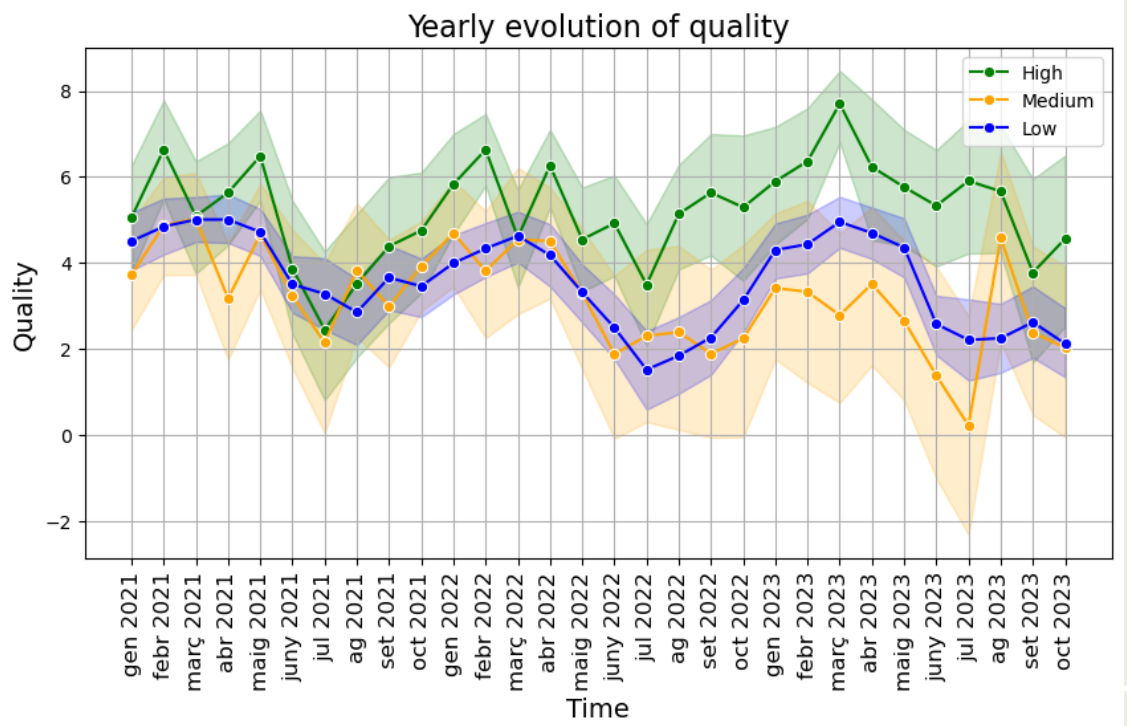
- Technique: Statistical analysis

Results

1 Height effects on quality

Height Class
1. Low = 0-500m
2. Medium = 500-1000m
3. High = 1000-1500 m





Result: Height statistically affects the quality of the milk in any season

- Checked using **Welch's t-test**: p-value = 0; t-statistic = 7.942

1 Height effects on quality

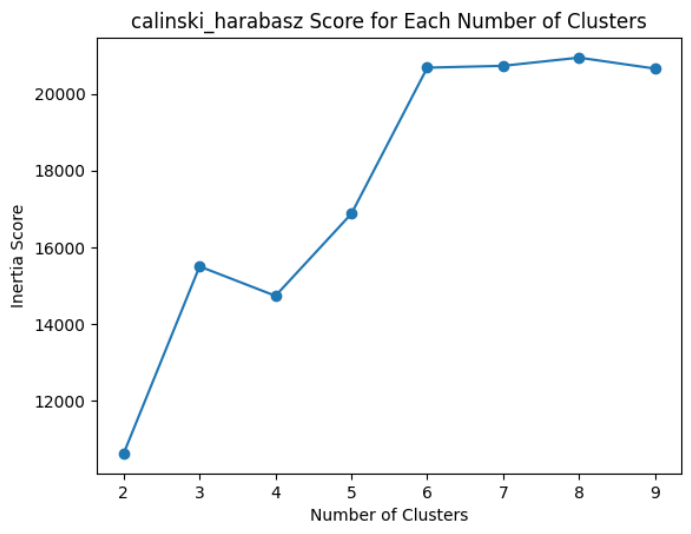
Height Class
1. Low = 0-500m
2. Medium = 500-1000m
3. High = 1000-1500 m

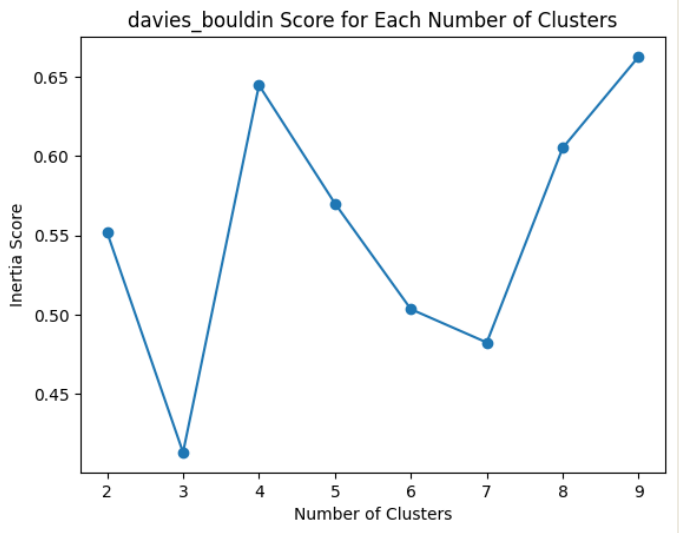
Effect on the farmer's revenues

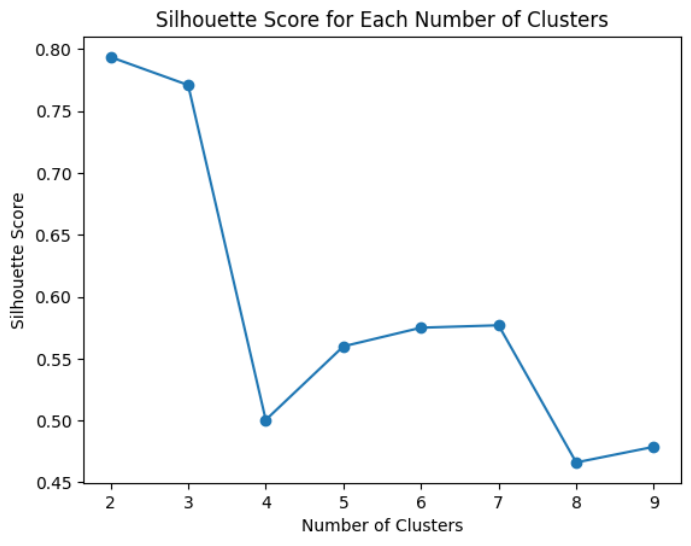
- Average earnings/liter

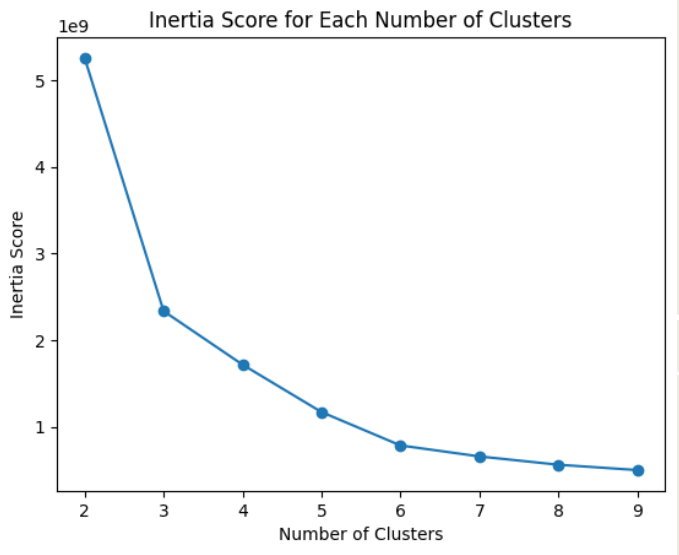
	Low	Medium	High
Revenues/liter	XXX €/l	YYY €/l	ZZZ €/l
Percentual increase	0.16%	0%	**8.8%**
Assuming: - annual production of 350,000 l/farm - Revenues/liter on medium height is 1€/l			
Total revenues	350,405 €	350,000 €	**380,775 €**

Optimal value of K for clustering





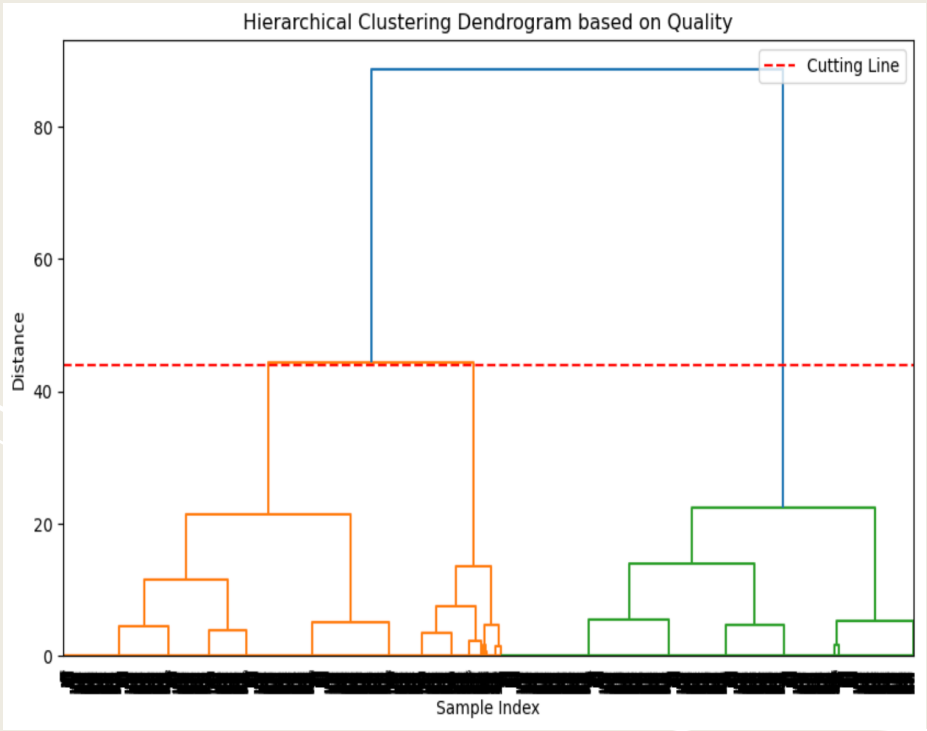


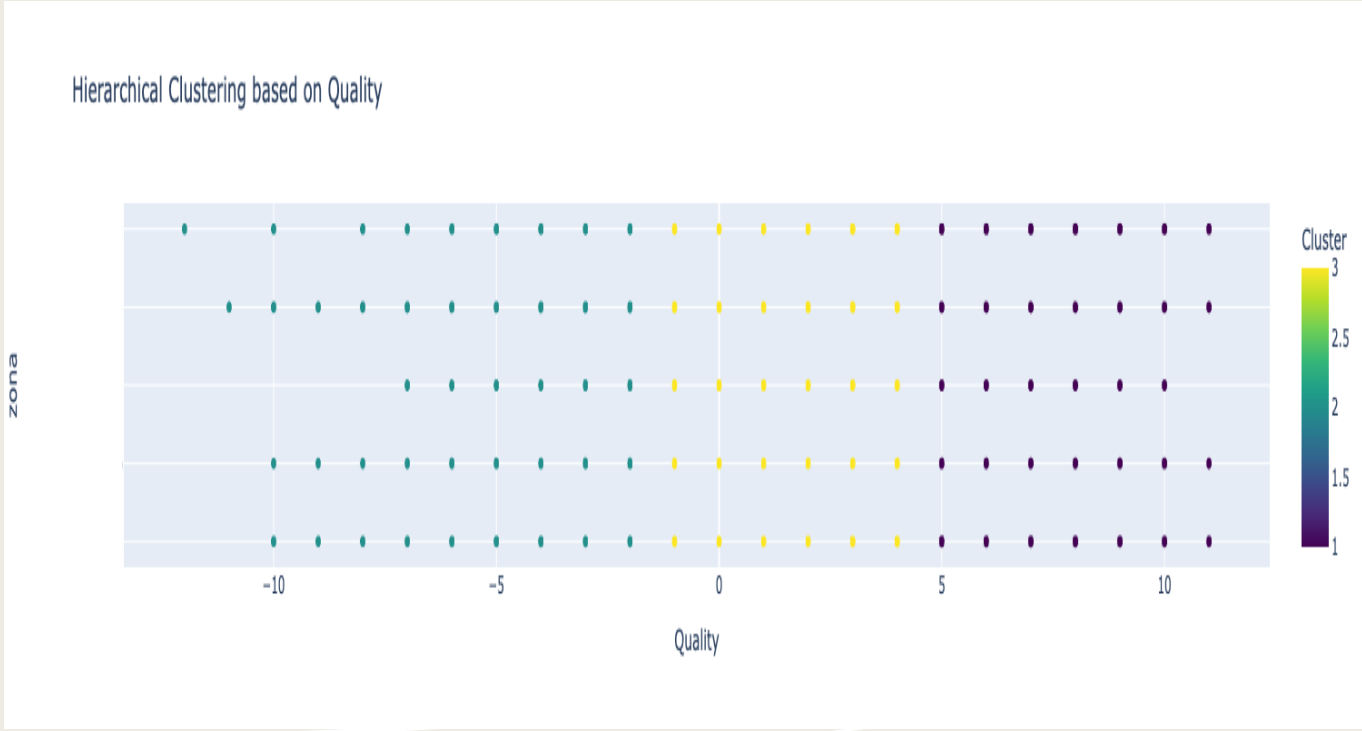


2 Area with better quality of milk

Hierarchical Clustering : 1st Algorithm we used to get the three different clusters

GOOD: 5 to 11
NORMAL: -1 to 4
BAD: -12 to -2

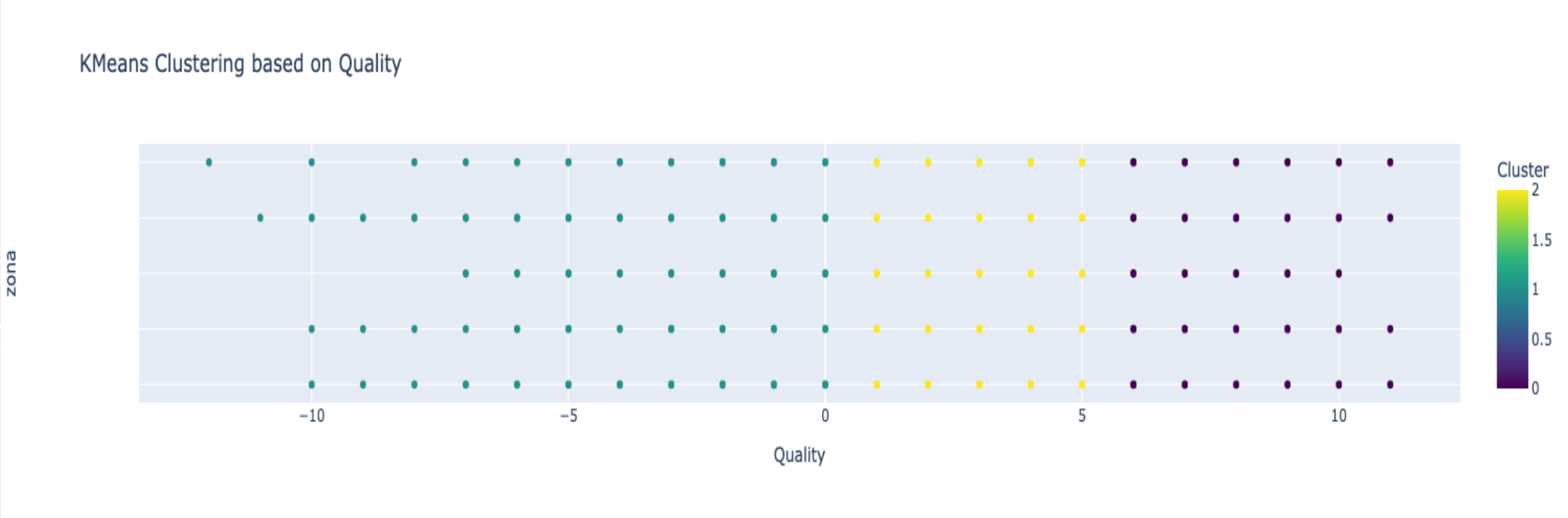




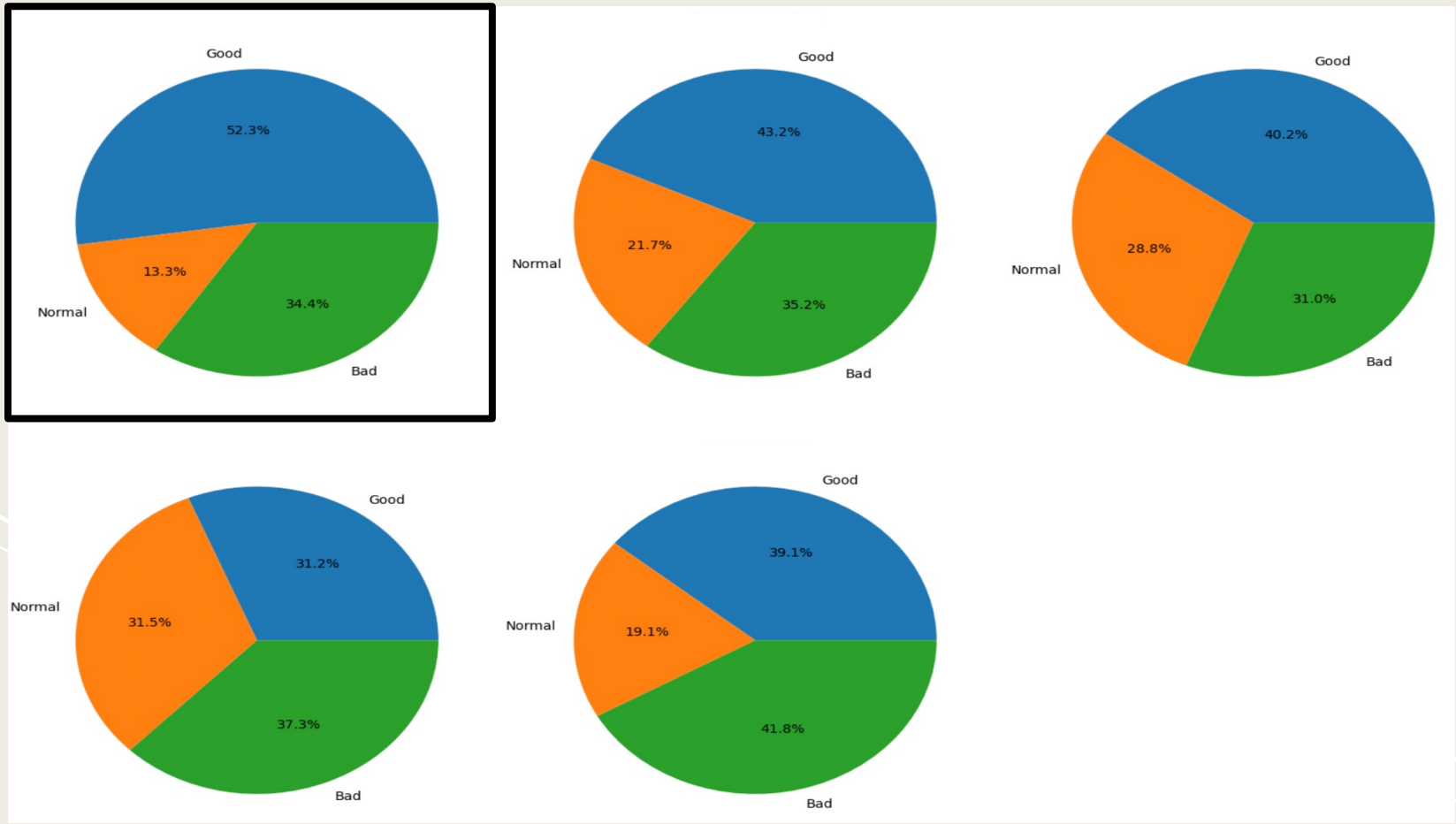
2 Area with better quality of milk

K-Means : 2nd Algorithm we used to get the three different clusters

GOOD: 6 to 11
NORMAL: 1 to 5
BAD: -12 to 0



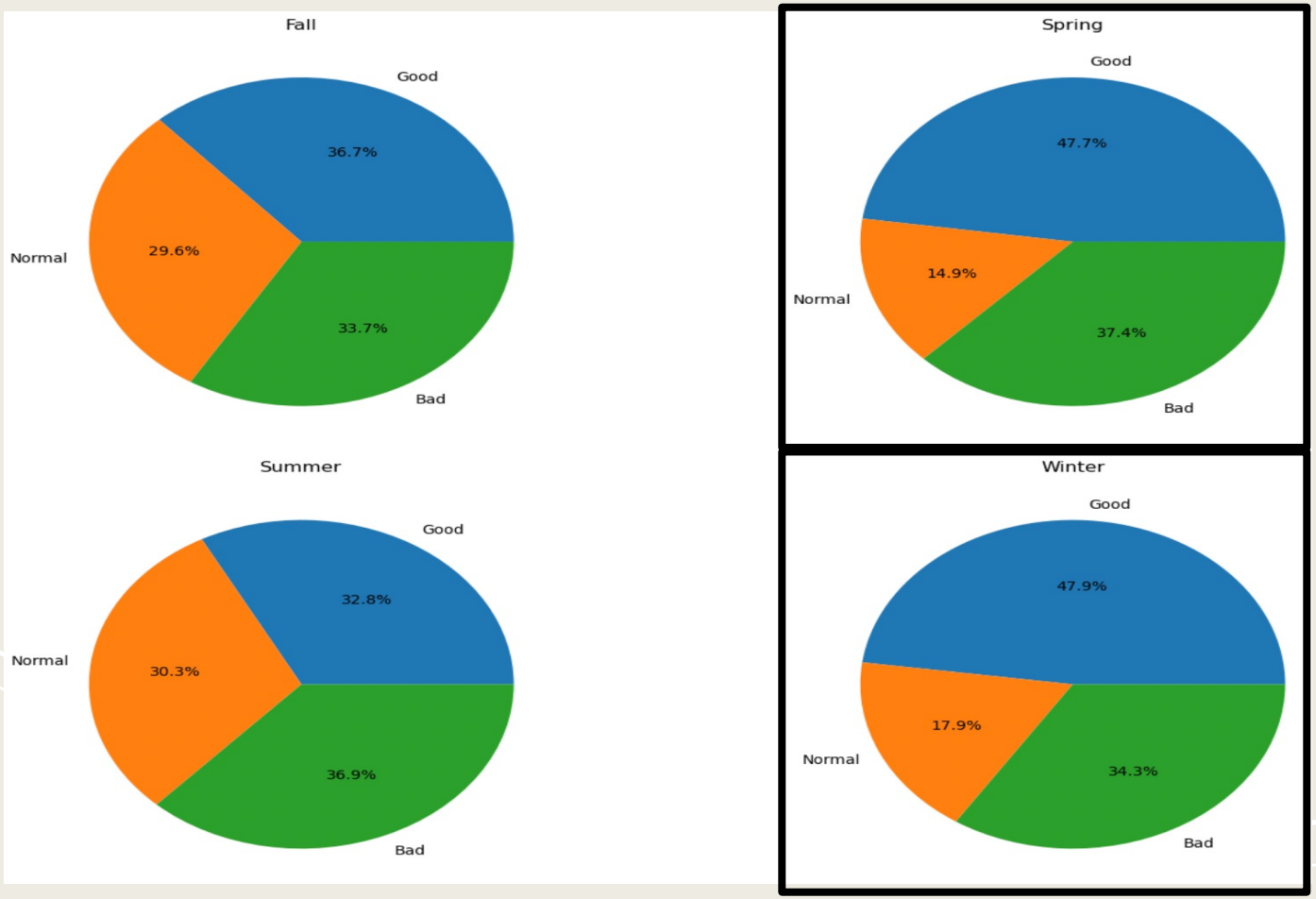
② Area with better quality of milk



Result: (Selected Zone?) has the best quality of milk.

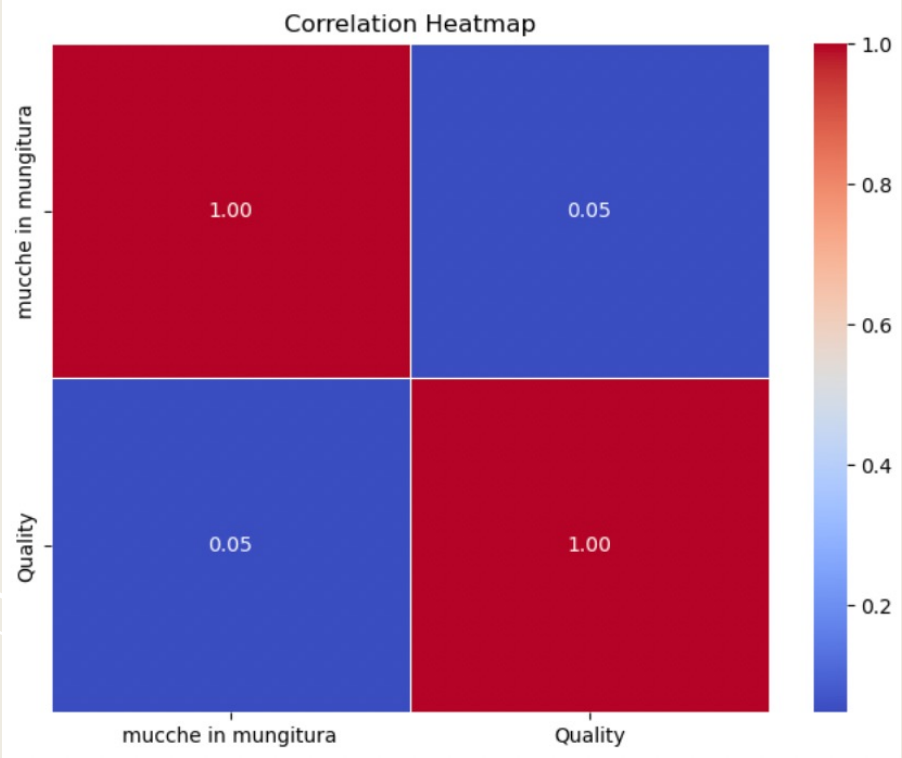
- More than the half of each farms produce good milk

2 Quality of Milk based on Seasons



Result: Spring and Winter has the best quality of milk

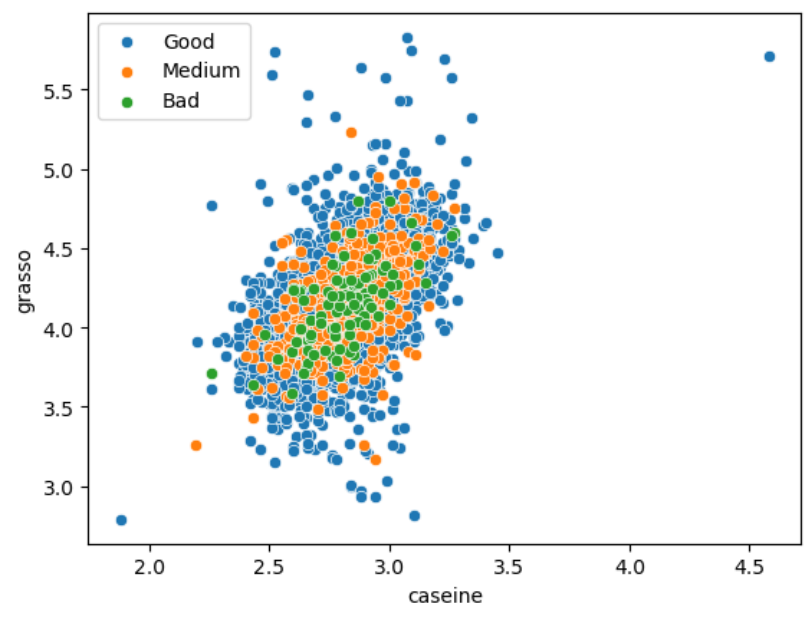
3 STUDY IF THE NUMBER OF COWS AFFECT THE QUALITY

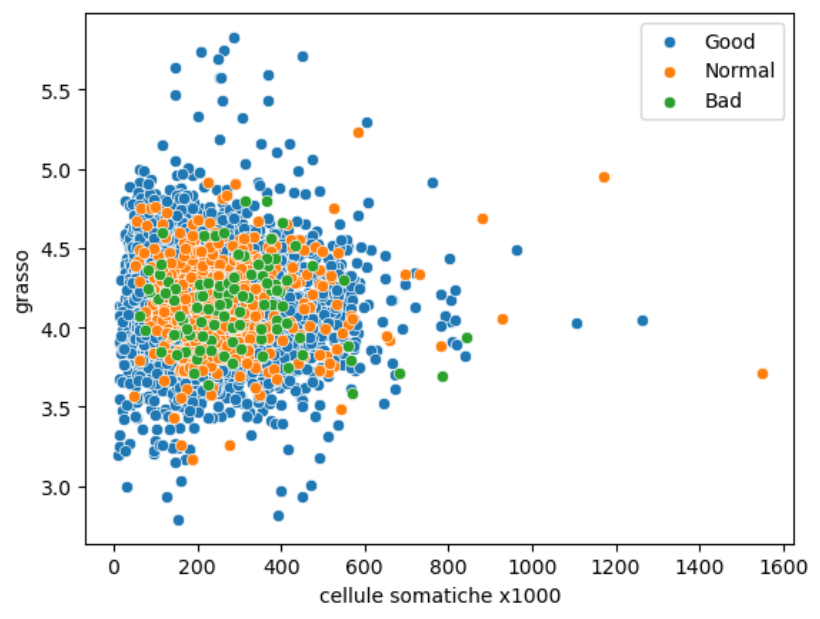


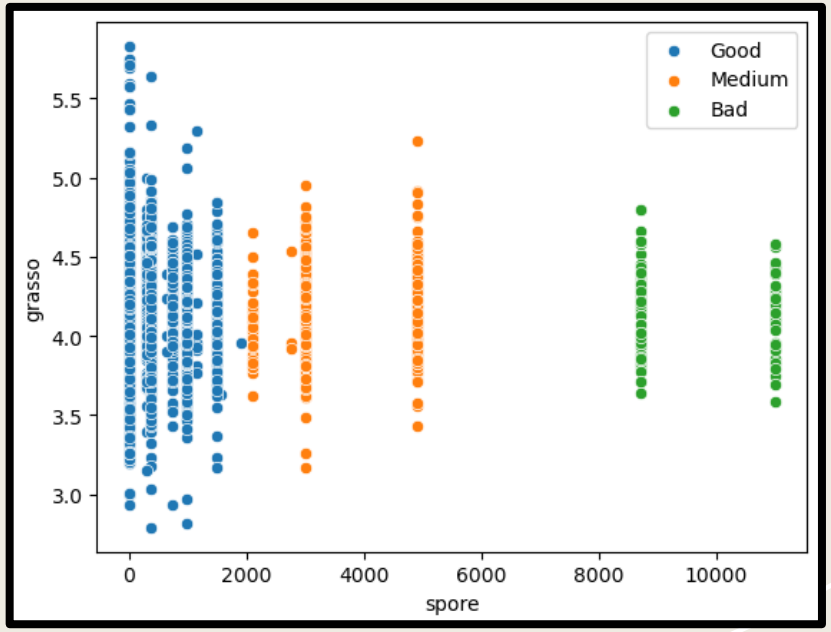
Result: The number of cows does not affect the quality

4 What important features affect quality

Based on the clustering explained before

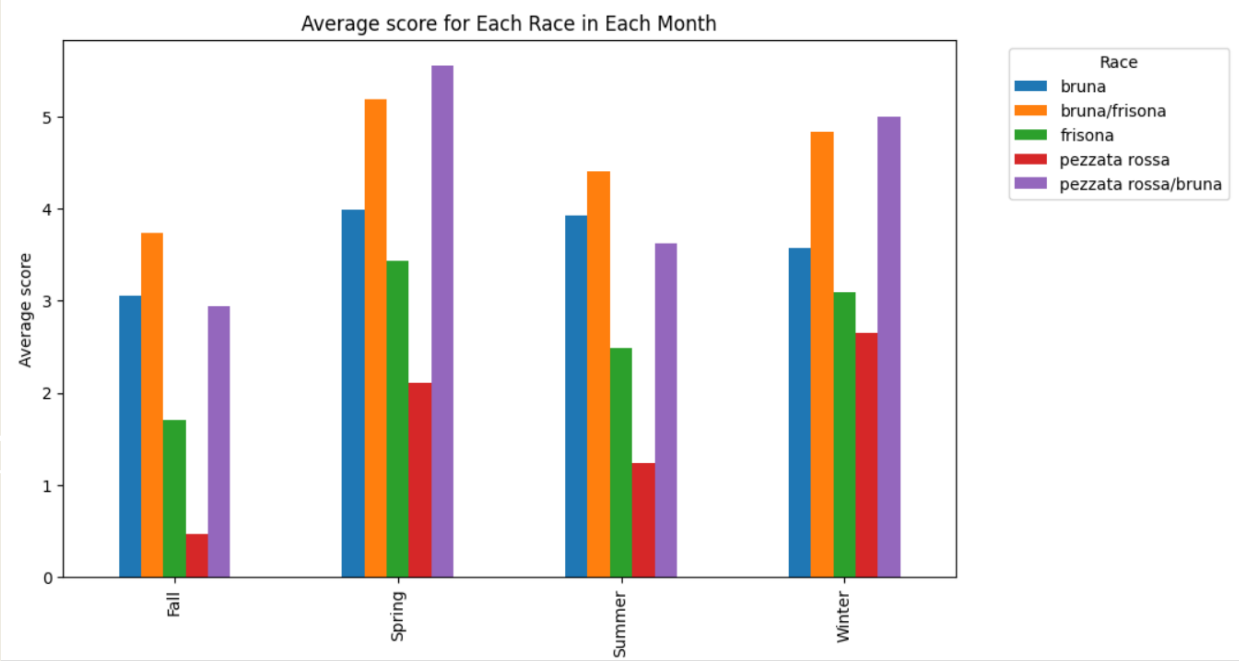


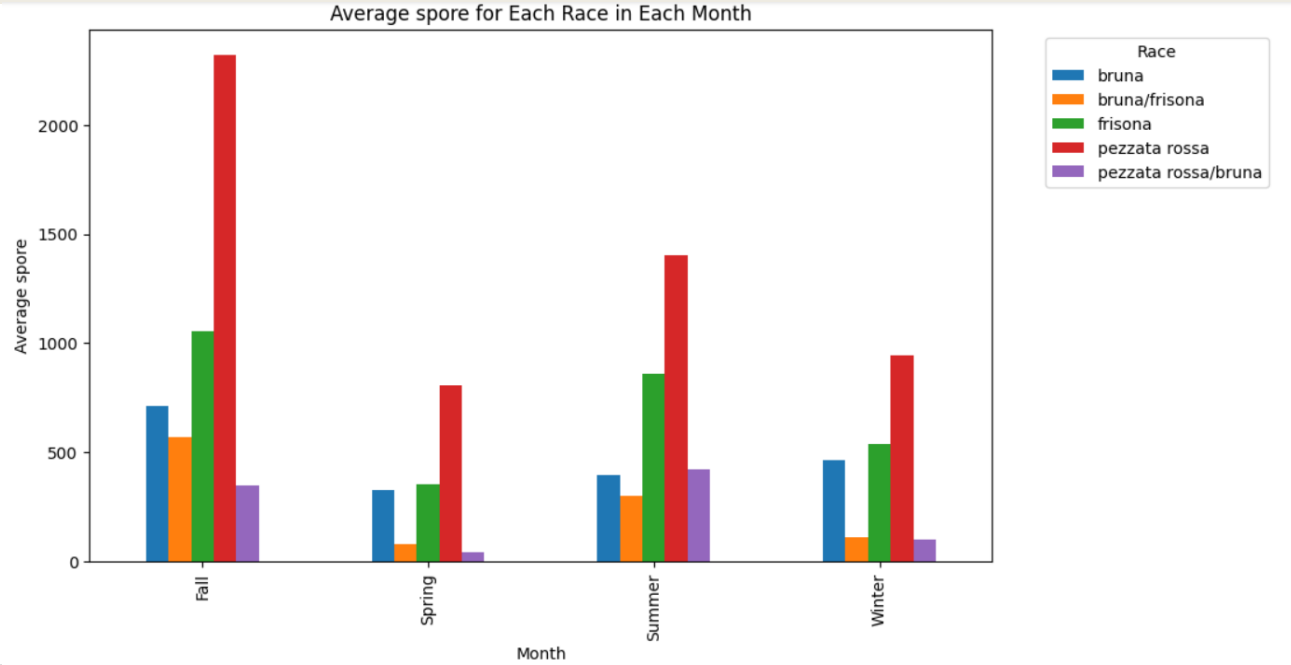




Result: We understood that the spores significantly affect quality

5 Does Breed Affect the Quality of Milk?





6 Machine Learning for Classifying Milk Qualities

Preprocessing Steps:

- **Removing Outliers**
- **Onehot Encoding to Nominal Features**
- **Extract Season and Day of Week and assign a number (Ordinal Features)**
- **Apply Min-Max Scalar to Numerical Features**

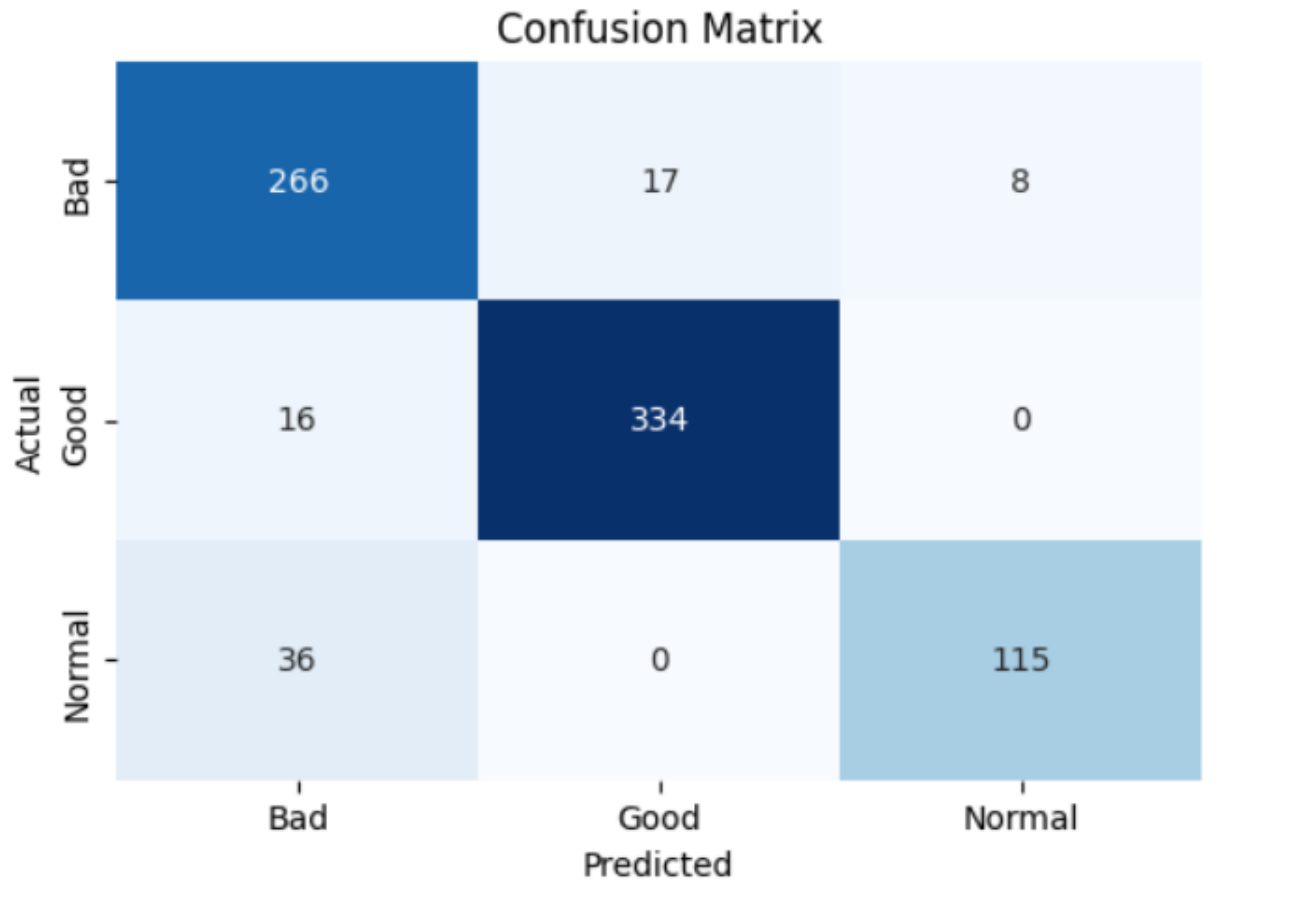
6 Machine Learning for Classifying Milk Qualities

Training Procedure:

- **Choosing two of the most popular classification models, named SVC and Logistic Regression**
- **Choosing GridSearch Cross Validation due to hyper-parameter tuning.**
- **Training these two classification algorithms with different combination of hyper-parameters.**
- **Choosing the best one, whose accuracy was greater than the others on the testing set.**

6 Machine Learning for Classifying Milk Qualities

Best Parameters: {'C': 10, 'gamma': 'scale', 'kernel': 'linear'}

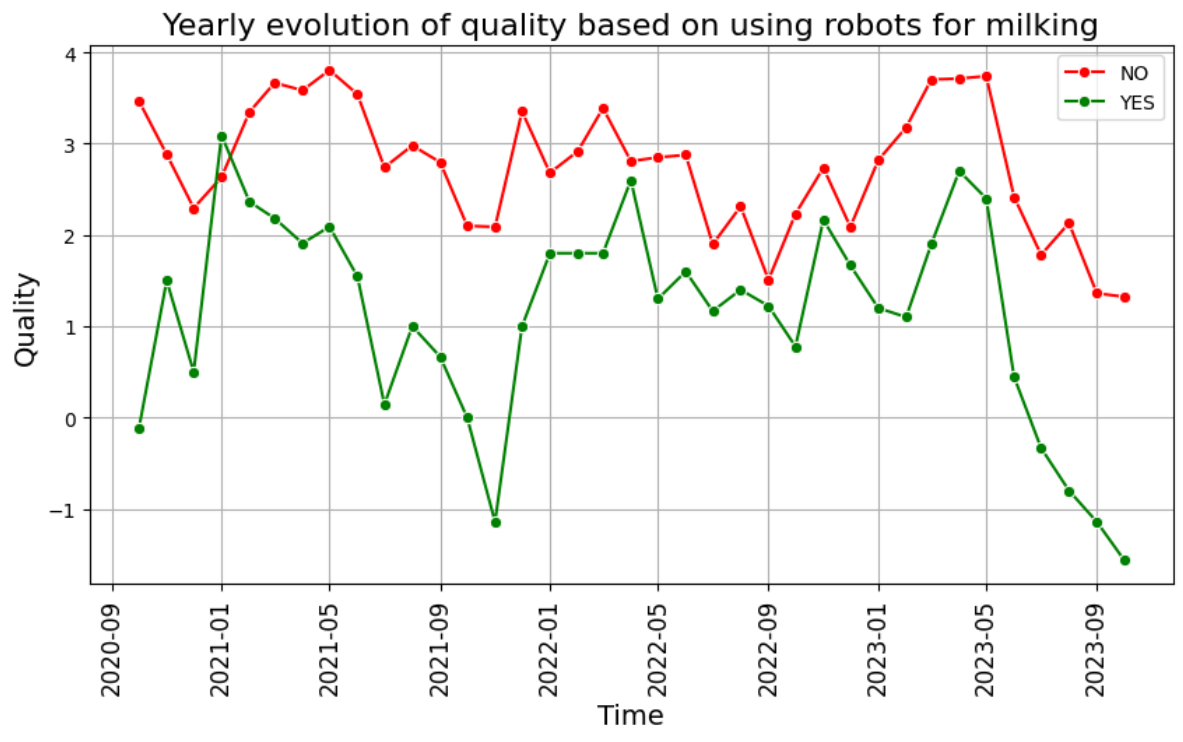


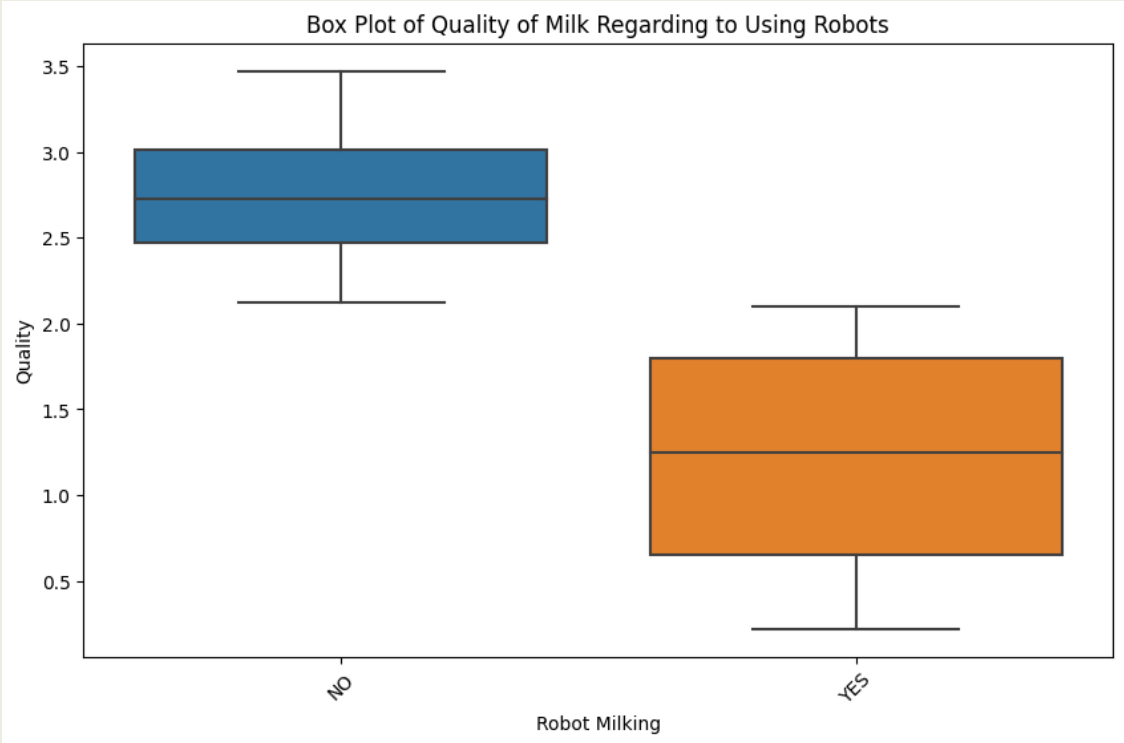
Classification Report:

	precision	recall	f1-score	support
Bad	0.84	0.91	0.87	291
Good	0.95	0.95	0.95	350
Normal	0.93	0.76	0.84	151
accuracy			0.90	792
macro avg	0.91	0.88	0.89	792
weighted avg	0.91	0.90	0.90	792

7 Milking robots' effects on quality

Use of milking robot Class
1. **0 = No**
2. **1 = Yes**

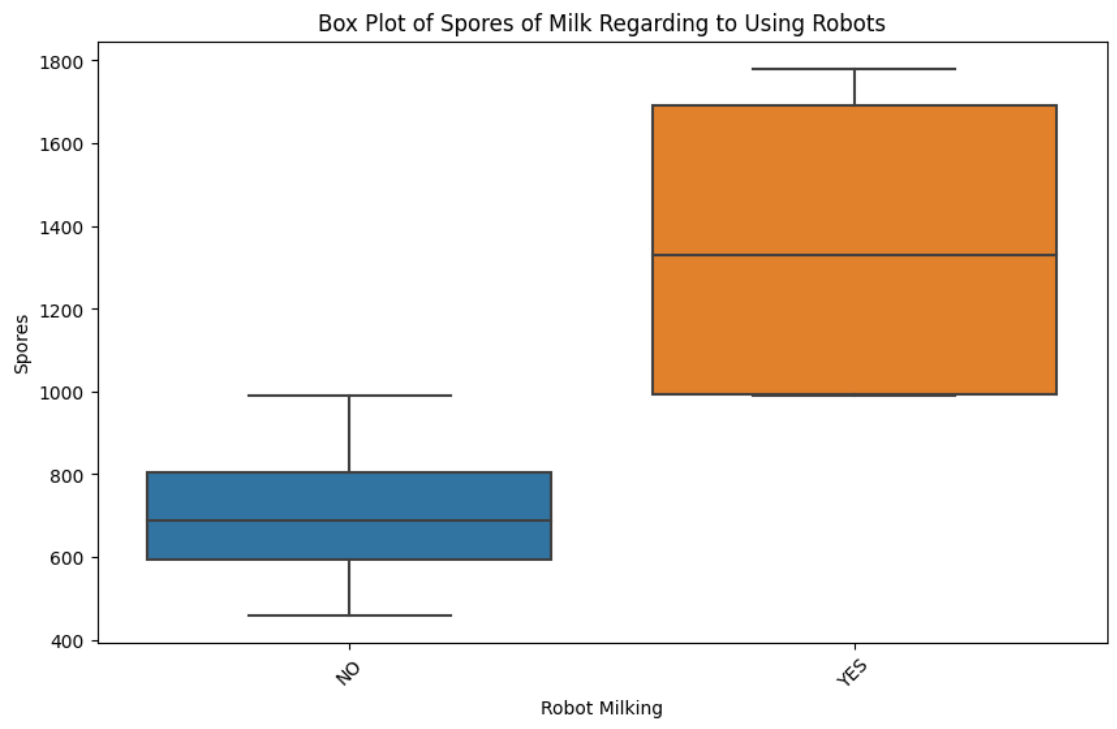


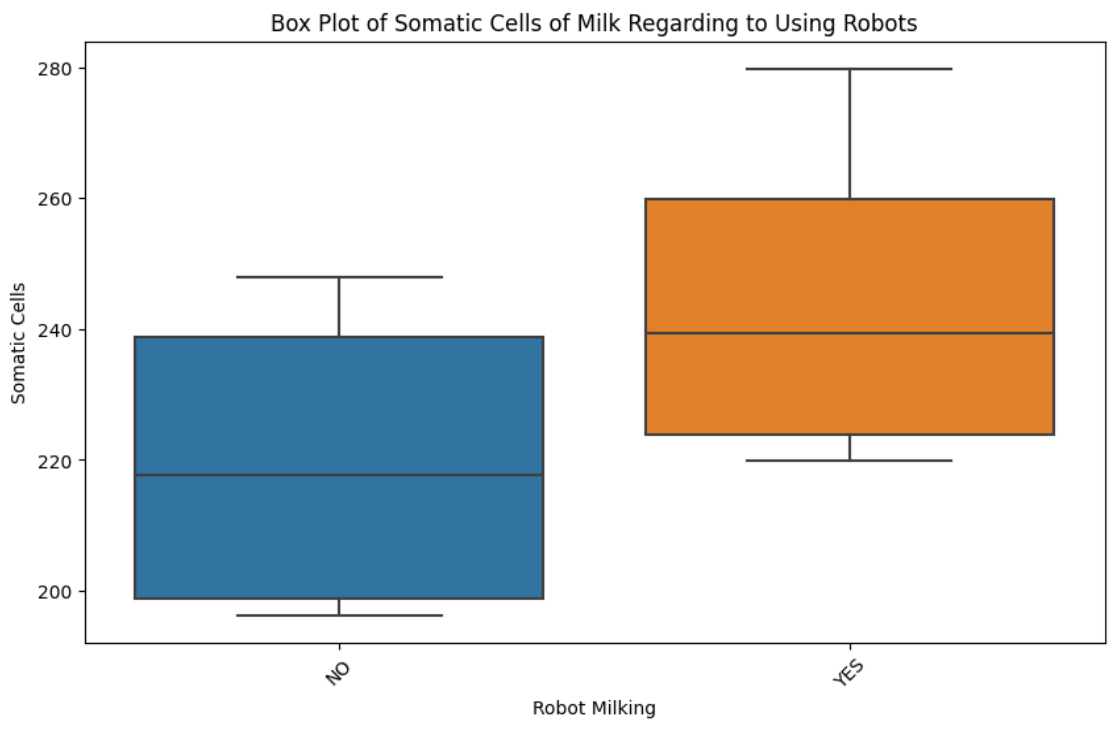


Result:

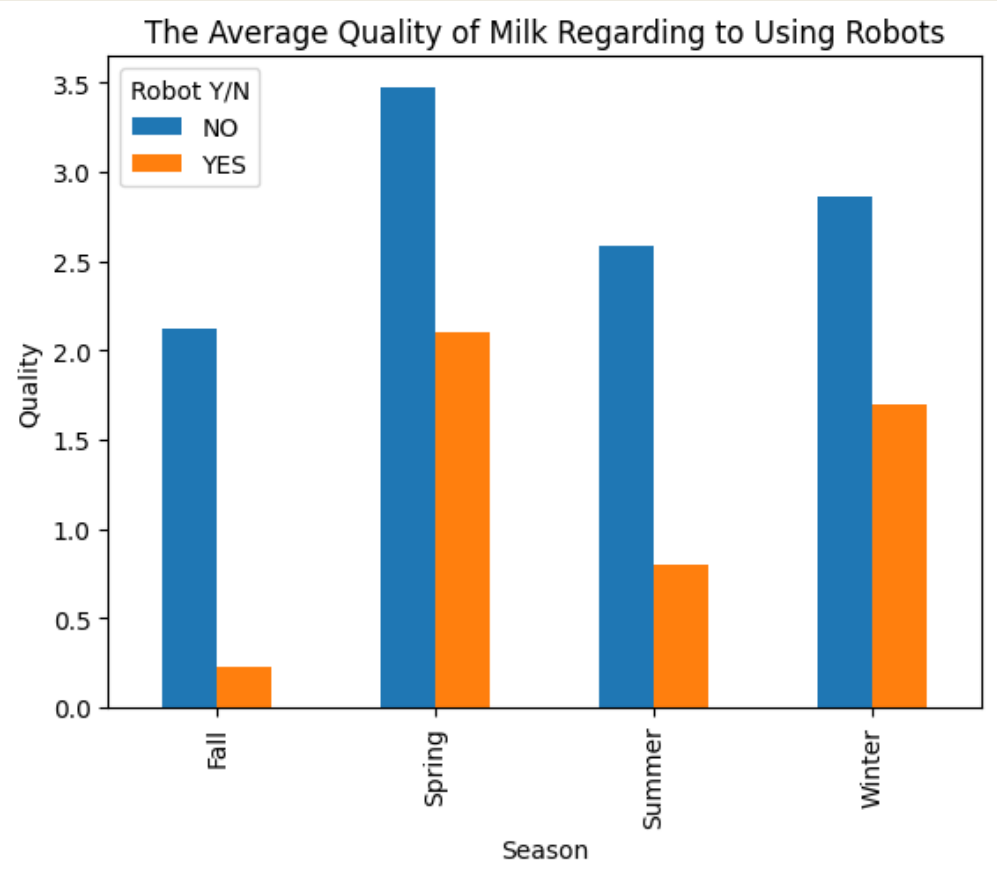
In general, using robots for milking decreases the quality. But why?

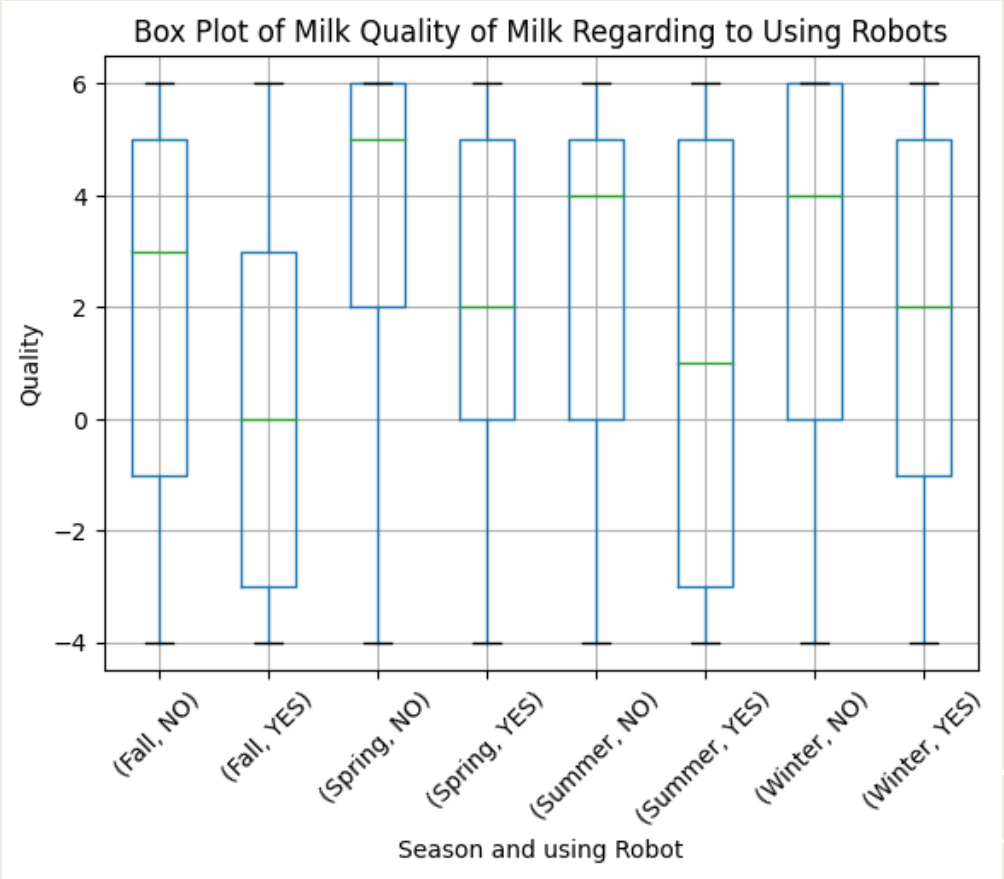
7 Milking robots' effects on spores/somatic cells



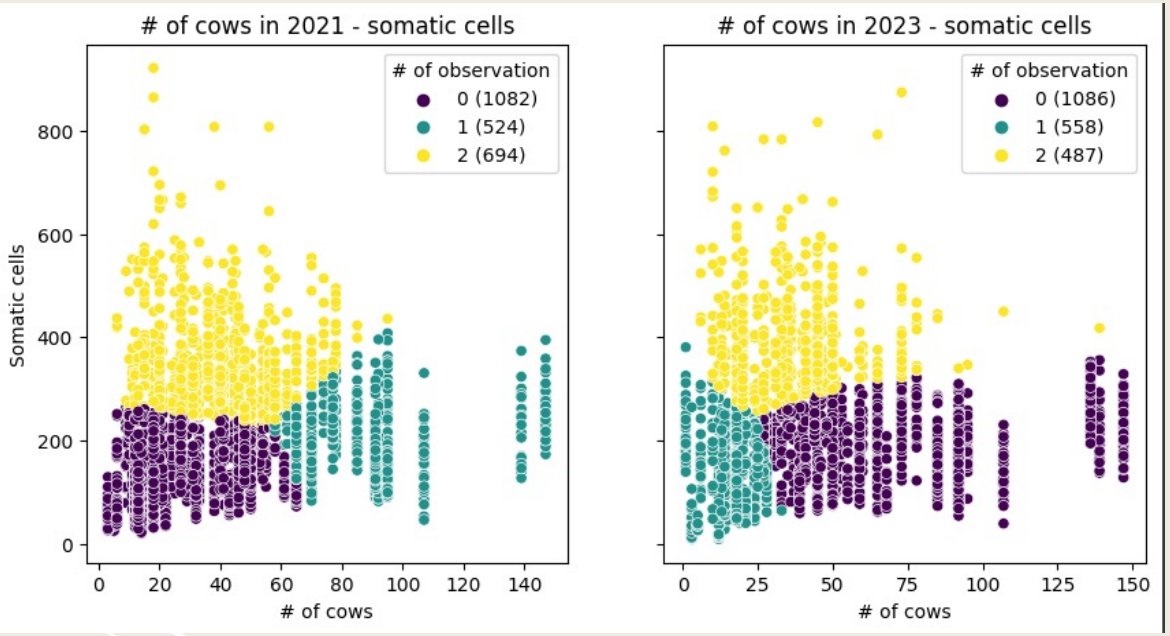


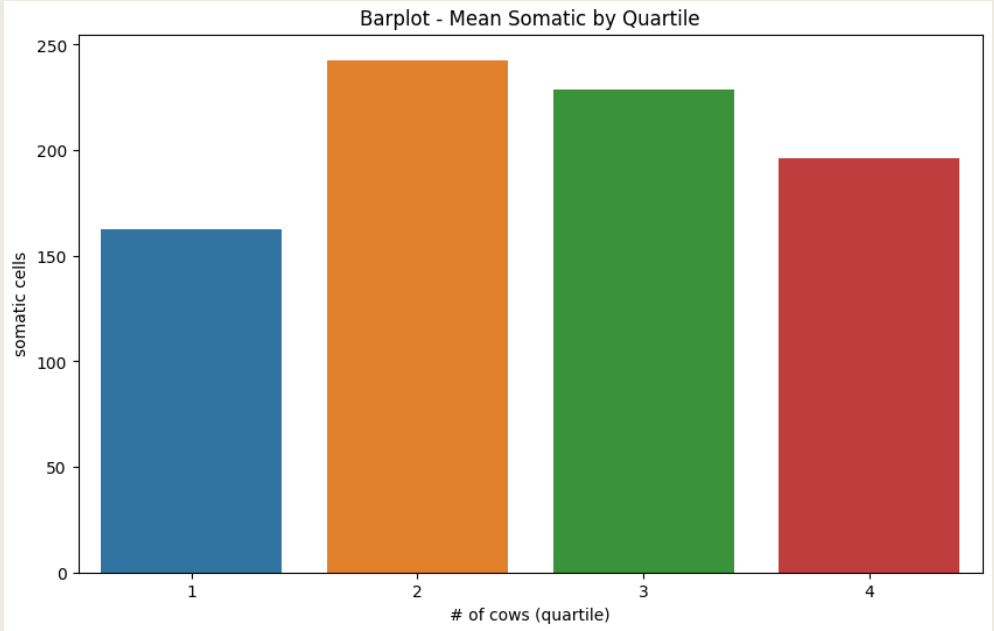
7 Milking robots' effects on quality (seasonly analysis)





8 Somatic cells depends on number of cows?





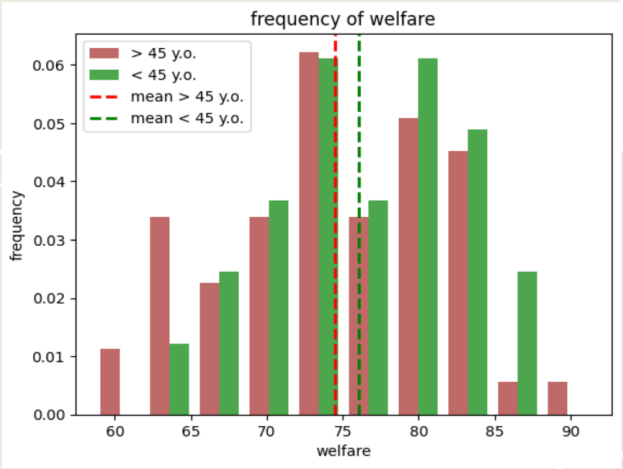
Result: the number of somatic cells very weakly (quite negligible) tends to decrease when the number of cows increase

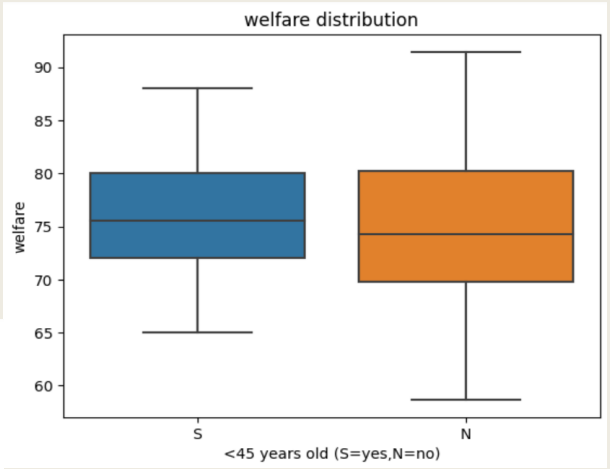
Correlation Coefficient: -0.02264

9 Analysis on the welfare of the cows

⚠ This analysis was conducted on a limited amount of data

Does the age of the farmer influences welfare?





Result: age of the farmer doesn't have a significant impact on the welfare of the cows

Result: welfare of the cows have a weak positive correlation with milk quality.

Pearson correlation coefficient (mean of 4 years analysis): 0,317

P-value: 0,04

Summary of the results

1

STUDY IF HEIGHT AFFECT QUALITY

- Farms with higher heights produce better quality

2

STUDY WHICH MILKING AREAS HAVE BETTER QUALITY

- (zone?) has the best quality of milk.

3

STUDY IF THE NUMBER OF COWS AFFECT THE QUALITY

- Weak or negligible correlation

4

WHAT IMPORTANT FEATURES AFFECT QUALITY

- Spores are significantly affecting the quality

5

STUDY IF THE COW BREED AFFECTS QUALITY

- Worst breed: Prezza Rossa

6

MACHINE LEARNING FOR CLASSIFYING MILK QUALITIES

- SVC

7

STUDY IF THE MILKING ROBOT AFFECTS QUALITY

- using robots for milking decreases the quality

8

SOMATIC CELLS DEPENDS ON NUMBER OF COWS

- Very weak or negligible linear dependencies

9

STUDY THE WELFARE OF THE COWS

- Welfare is not influenced by age of the farmer and quality

THANK YOU FOR YOUR ATTENTION