**Target/Questions**

I want to use my data in order to

* estimate the fish population (i.e. understand that there is a deviation between the calculated fish number/average weight and the real ones)
* adjust feeding, based on the real fish population and the behavior of the fish
* predict a disease or another problem that will happen soon
* improve production efficiency by understanding the relationships between the parameters and their effect on the production result
* compare my performance to the one of other similar companies (benchmarking) and verify the validity of the findings of the previous bullet

The first 3 bullets are targets that have to do with the current situation and the management of the farm for the next period. The last two bullets is related both to current as well as long-term improvements

**Datasets**

There is a dataset for each running or closed cage. It has the attributes described in “Periodic Datasets” tab of “RequirementsGathering AquaSmart v0 1 - bm 20150220.xlsx”. There are three main groups of information

* KPIs that show the performance of the cage (output variables, blue cells) – **Group 0**
* Attributes of the fish population that exists in the cage (like hatchery, Fry CV, Stocking Year and Month, etc- gray cells). Those attributes do not change over time (with two exceptions, if the fish are vaccinated or not and the grading code) - **Group 1**
* Daily data of the fish. The dataset will contain either the information from the stocking of the fish until today or the information from the day the fish have been transferred to this cage until today). There will be a record for each day. There are 7 groups of information, shown in different colors (e.g. orange for the feeding-related data, light green for the net and density related, etc.) - **Group 2**

So the dataset is in the form master (KPIs and population attributes) – detail (daily records)

Each time it is submitted for analysis it will contain

* Data that are the same with the previous analysis (for example, the fish were stocked 1/1/2014 and the previous analysis was done on 15/1/2015, therefore the dataset contained the records from 1/1/2014 – 15/1/2015)
* New data that were produced from the previous analysis until today

Of course, for the new cages it will be completely new datasets

For each attribute

* There is a column that shows if it will be available for data mining at the global level.
* There are two groups of columns that show if the attribute exists currently in the data that the companies keep or it will exist in the future)
* There is a column for the companies to provide the importance they think this attribute has

**Analysis to be performed on the datasets**

We want to study

* The relationships between the parameters of **Group 2.** Some examples are provided in “interaction between parameters.xlsx”. For example, we want to know if the fish density influences the daily mortality
* The impact of the parameters of **Group 1** and **Group 2** to the KPIs of **Group 1**. We want to know the more important parameters, in order to focus our effort for improvement. I think there is a question here regarding the way we are going to consider the daily parameters. For example, we want to study the effect of the fish density to the daily mortality, the total mortality and the Economical FCR. How are we going to take into account the daily fish density regarding the LTD Econ. FCR? The input will be the daily values or just the average? In other words, we want to study the effect of a daily value to an LTD (Life to Date) Value
* The impact of the **change rate** of parameters of **Group 2** to other parameters of the same group or to the KPIs of **Group 0**. For example, does the mortality increase when temperature changes fast? So we are interested not only to the values but also to the rate they change.
* The effect of various parameters on the final distribution of the fish in harvest categories when harvested. Question: do we need to add another set of attributes to Group 1 that will be the distribution of the fish to harvest categories, for the closed units ??

Also, we want to identify missing fish or differences in the average weight (the first question) by looking at the data. For example, if there is a deviation between model and actual feeding for the last x days and if the behavior of the fish is that they are always at the bottom of the cage (not a real example) and if the fish density is OK and then most probably we have less fish than the calculated number. Input required: we need the users to provide a good set of rules on how to identify differences in the fish number or average weigh. What do they examine now, how they make conclusions.

And finally, we want the system to help users to identify problems in order to be able to react quickly. For example, if there is

* High deviation between model and actual feeding and/or
* high number of mortalities or quick increase of mortalities
* strange fish behavior
* big differences between estimated and actual average weight in samplings
* big number of holes the divers repair

Then it means that there is a problem. If for example the fish are losing appetite and the number of mortalities is increasing it is likely a disease is going to appear. Again, we need input here: we need to understand the way the users think. What they look first, what they do next, how they understand that something is happening.