



Case Study

Data Analysis

Problems

The data repository provided by the client was about five years long, and was accumulated on a daily basis. The client wanted DART to analyze the data and arrive at the key trends, insights and indicators that will help the client in forecasting farm supplies better and result in better informed decision making in the organization. The client required the following trend analysis from the data:

- Yield (No. of nuts per month) of the farm and location trend.
- Harvesting frequency change.
- Button nut shedding – The reduction in the number of nuts in case of the farms where harvesting has not happened.
- Impact of the nut position on

productivity for the subsequent years.

- Impact of rainfall on productivity.

Methodologies

DART Analysts surveyed the selected farms on daily basis. The number of farms surveyed per month was 120.

First, cluster analysis is done to identify homogenous clusters in the data. This eliminates the inputs which are out of range. The purpose of cluster analysis was to discover a system of organizing observations into groups where members of the groups share properties in common. Then, in order to forecast the farm supplies, dependent and independent variables are identified. It is found that “Expected Nuts in Next Plucking” is the dependent variable and the following are the independent variables:

No. of trees	Each farm has a fixed number of trees
Number of Nuts Harvested	Plucking of nuts (harvesting) happens once in two months. The total number of nuts harvested for each farm is given under 'Number Of Nuts Harvested'.
Rainfall data	Rainfall measurements
Last Plucking Month	Month when the nuts were last plucked
Last_Plucking_Year	Year when the nuts were last plucked
Next Plucking Month	Month when the nuts are expected to be plucked next
Next Plucking Year	Year when the nuts are expected to be plucked next
Button_Nuts	Nuts grow in bunches, with a new bunch springing every one month. The newest bunch is called the button bunch and the average number of nuts per tree for a farm is called the 'Button_Nuts'
Matured_Bunches	The average number of matured bunches per tree is given in the column 'Matured_Bunches'.
Matured_Nuts	The average number of matured nuts per tree in all these bunches put together is given in 'Matured_Nuts'.

Using these variables a multiple regression model is developed which can be used for forecasting the “Expected Nuts”. Thereafter, the hypothesis is tested using the Anova table.

Results

The results found were conveyed to the client and the client was very thrilled with DART's analytical abilities. The client could formulate its strategy for the later years based on the findings as given by DART.



DART's Data Analysis Services

DART uses various analytics software to perform data analysis. Analysts perform deep data mining, data analysis, and online research to present useful information from such accumulated data. The captured data is then converted into information and knowledge. DART has a lot of experience in data mining and surfing large amounts of data and sifting out relevant information.