Business Intelligence

Now and beyond

Muhammad W. Hijazi

Version 4.0.0

November 17, 2015

Advanced Analytics

Table of Contents

Introduction

Our unique solution of BusinessAdvanced Analytics consists mainly of two modules:

a. Building On Line Analytical Processing (OLAP) Multi-Dimensional 2 Enterprise-Level Cubes. Operational and statistical. Details will follow.
b. Building 10 Data Mining Models to cover almost all kinds of Machine Learning advanced analytics.

Online Analytical Processing

What is OLAP? OLAP is more than an acronym that means Online Analytical Processing. OLAP is a category of software tools that provides analysis of data stored in a database. With OLAP, analysts, managers, and executives can gain insight into data through fast, consistent, interactive access to a wide variety of possible views. Stated another way, OLAP is a category of applications and technologies for collecting, managing, processing, and presenting multidimensional data for analysis and management purposes. A widely adopted definition for OLAP used today in five key words is: Fast Analysis of Shared Multidimensional Information (FASMI).

- Fast refers to the speed that an OLAP system is able to deliver most responses to the end user.
- Analysis refers to the ability of an OLAP system to manage any business logic and statistical analysis relevant for the application and user. In addition, the system must allow users to define new ad hoc calculations as part of the analysis and report without having to program them such as ‘What-if’
- Shared refers to the ability of an OLAP system being able to implement all security requirements necessary for confidentiality and the concurrent update locking at an appropriate level when multiple write access is required.
- Multidimensional refers to a concept that is the primary requirement to OLAP. An OLAP system must provide a multidimensional view of data. This includes supporting hierarchies and multiple hierarchies.
- Information refers to all of the data and derived data needed, wherever the data resides and however much of the data is relevant for the application.

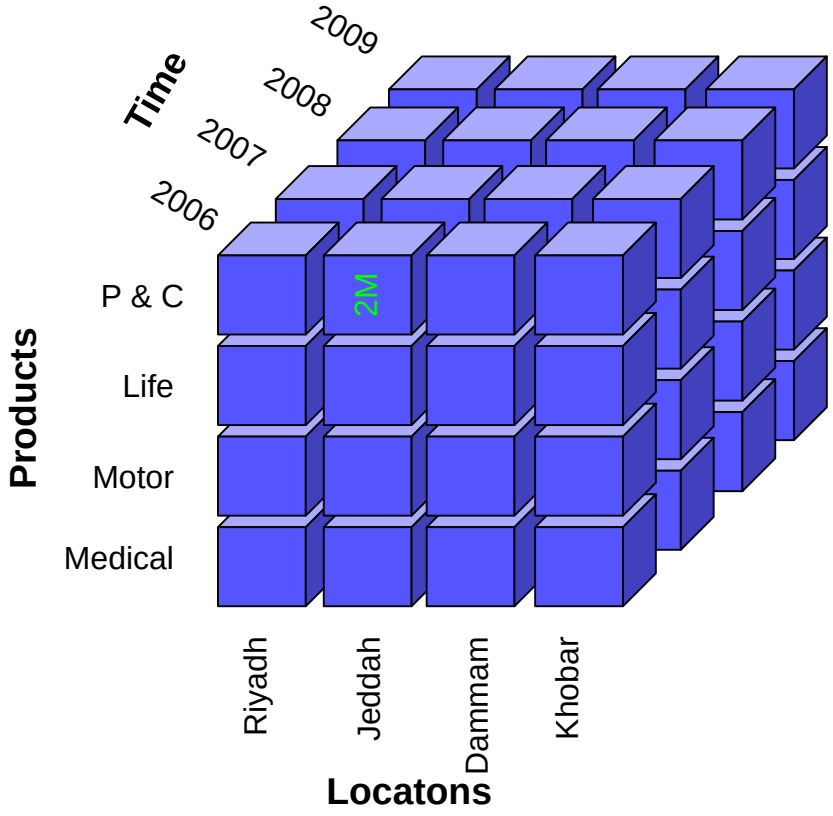
1.1 Cubes

What is an OLAP Cube? As you saw in the definition of OLAP, the key requirement is multidimensional. OLAP achieves the multidimensional functionality by using a structure called a cube. The OLAP cube provides the multidimensional way to look at the data. The cube is comparable to a table in a relational database.

The specific design of an OLAP cube ensures report optimization. The design of many databases is for online transaction processing and efficiency in data storage, whereas OLAP cube design is for efficiency in data retrieval. In other words, the storage of OLAP cube data is in such a way as to make easy and efficient reporting. A traditional relational database treats all the data in a similar manner. However, OLAP cubes have categories of data called dimensions and measures. For now, a simple definition of dimensions and measures will suffice. A measure represents some fact (or number) such as cost or units of service. A dimension represents descriptive categories of data such as time or location.

The term cube comes from the geometric object and implies three dimensions, but in actual use, the cube may have more than three dimensions.

The following illustration graphically represents the concept of an OLAP cube.



In figure 1, time, products, and locations represent the dimensions of the cube, while '2M' represents the measure. Recall that a dimension is a category of data and a measure is a fact or value.

Three important concepts associated with analyzing data using OLAP cubes and an OLAP reporting tool are slicing, dicing, and rotating.

1.2 Slicing

A slice is a subset of a multidimensional array corresponding to a single value for one or more members of the dimensions not in the subset. For example, if the member Actuals is selected from the Scenario dimension, then the sub-cube of all the remaining dimensions is the slice that is specified. The data omitted from this slice would be any data associated with the non-selected members of the Scenario dimension, for example Budget, Variance, Forecast, etc. From an end user perspective, the term slice most often refers to a two- dimensional page selected from the cube.

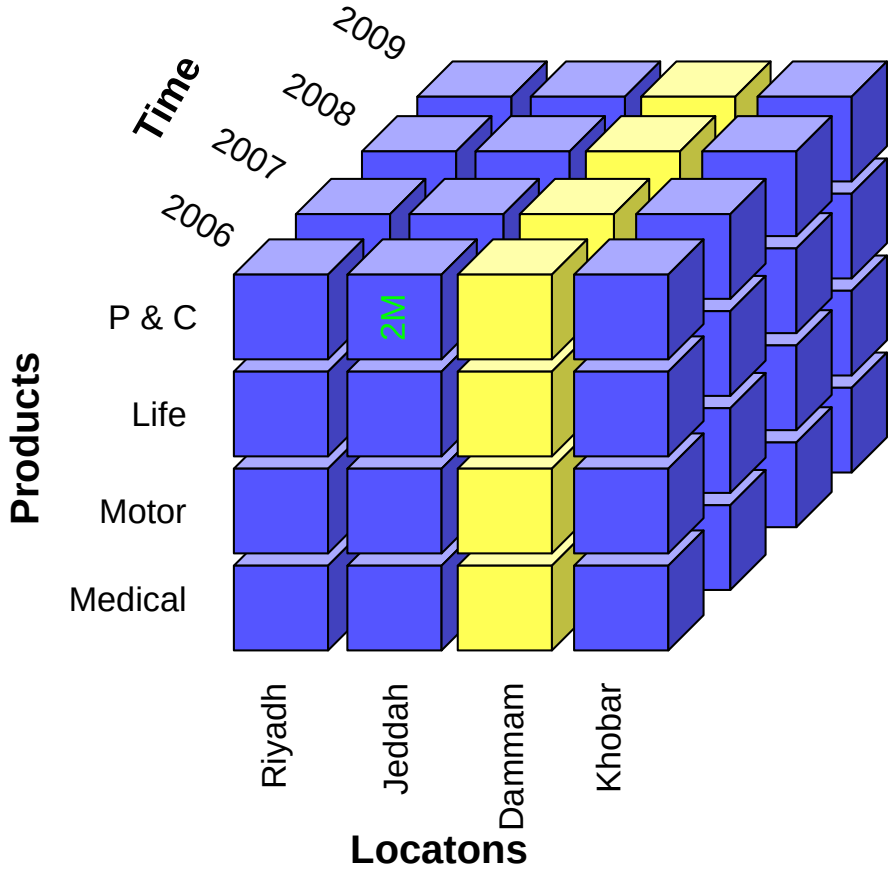
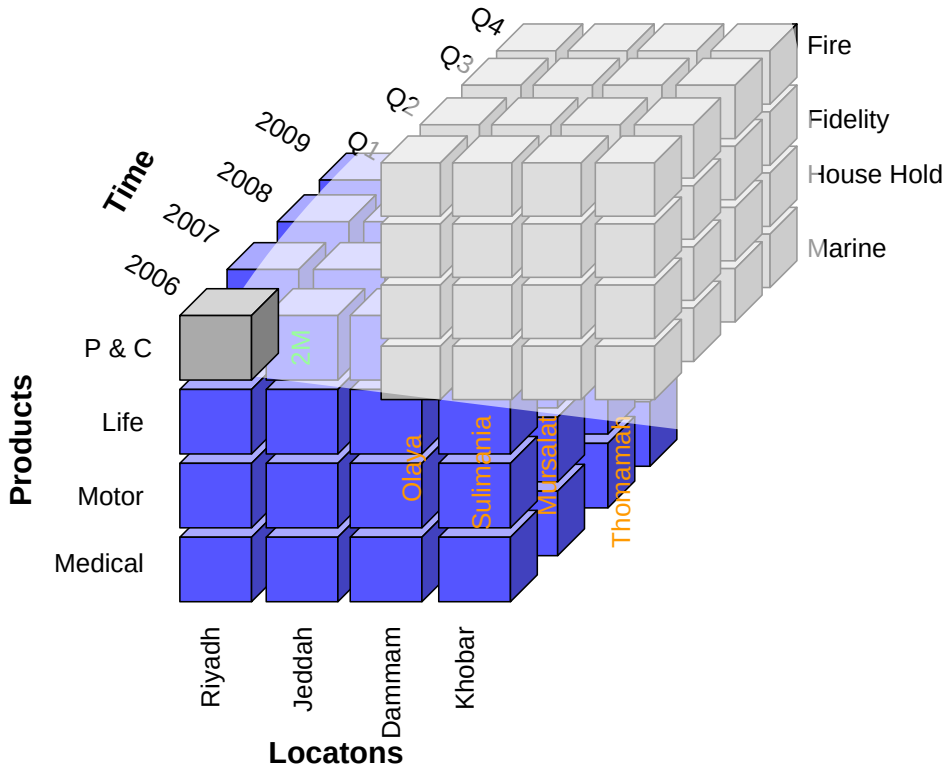


Figure 2 illustrates slicing the location Dammam. When you slice as in the example, you have data for Dammam for the years and products as a result. Stated another way, you have effectively filtered the data to display the measures associated with the Dammam Location.

1.3 Dicing

A related operation to slicing is dicing. In the case of dicing, you define a subcube of the original space. The data you see is that of one cell from the cube. Dicing provides you the smallest available slice. Figure 4 provides a graphical representation of dicing.



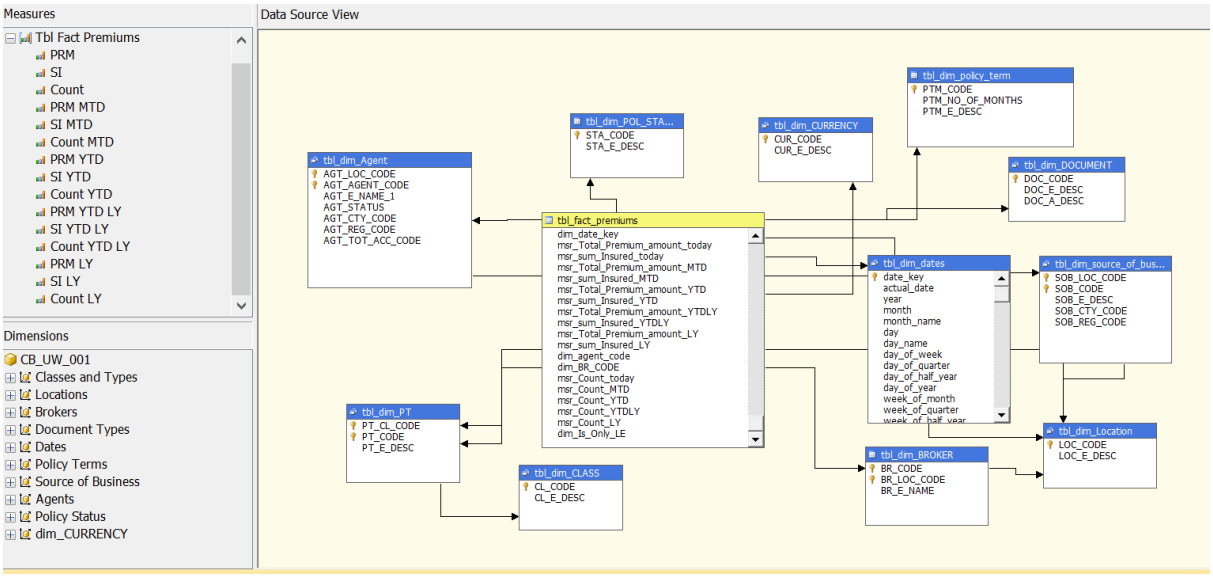
1.4 OLAP Proposed Cubes:

The most efficient way to brows an OLAP cube is using Excel. However, the proposed dashboard can also be linked to the OLAP cube and browse it fully.

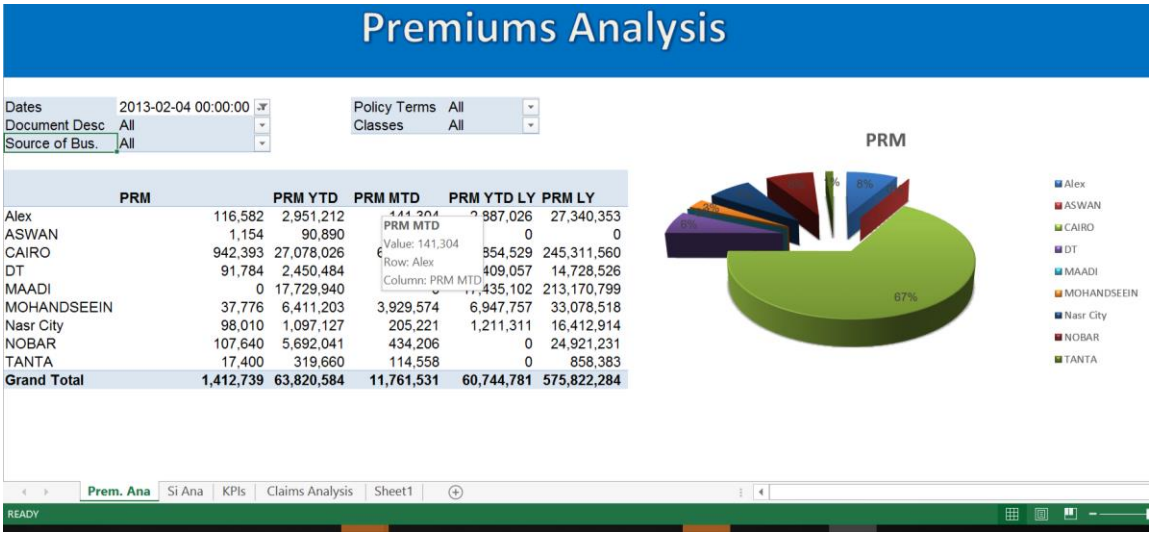
Below are example of the 2 proposed cubes:

1.4.1 UW Cube:

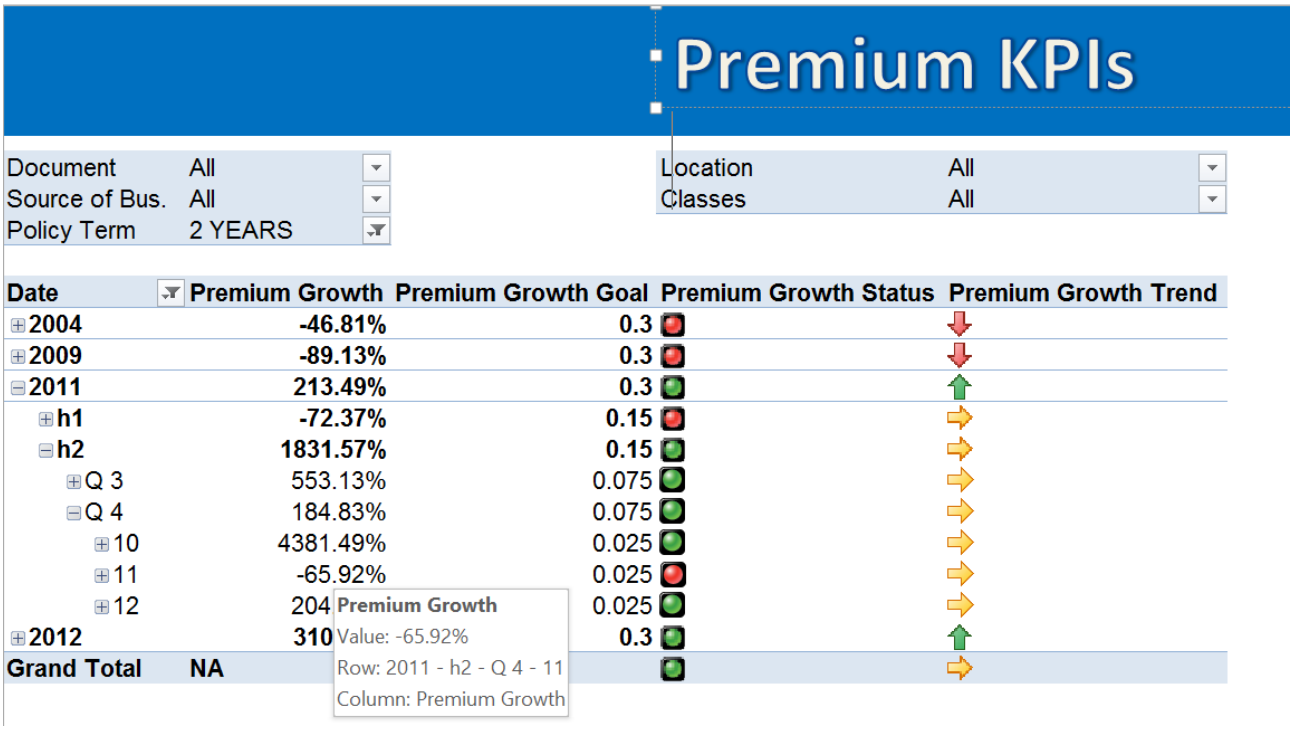
The cube star-schema structure:



The User Interface:

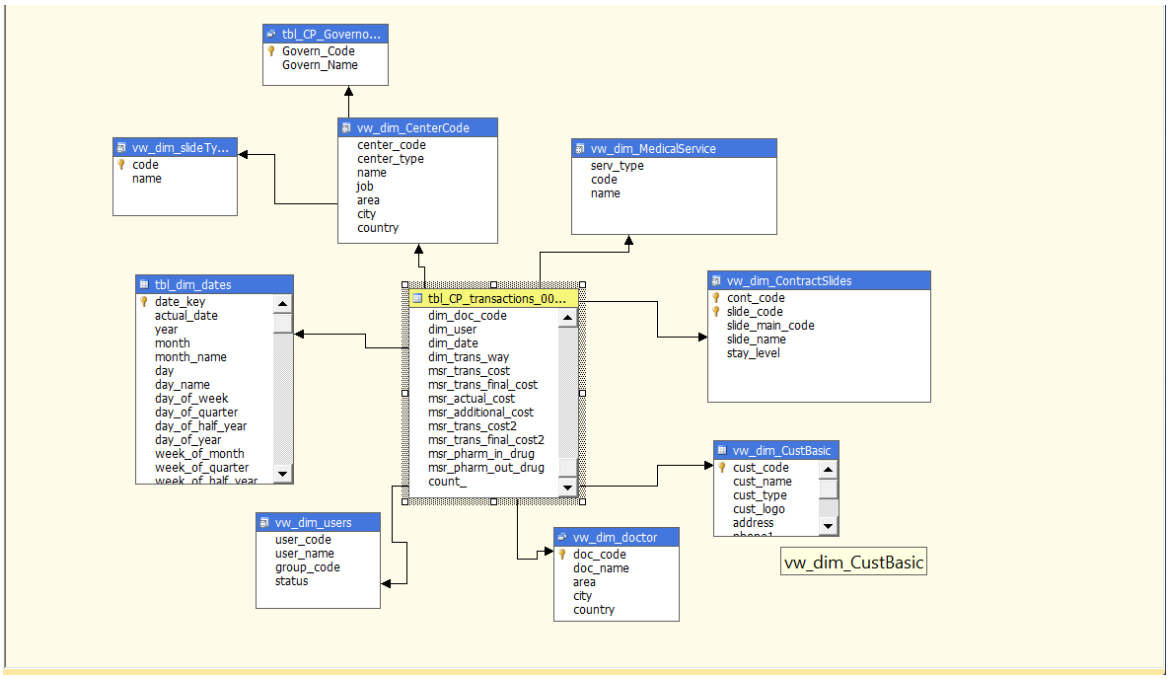


Premium KPIs Date wise, could be for class wise, location wise, .. etc. Just drag any dimension from filter area to row area … re-calculations take no time!



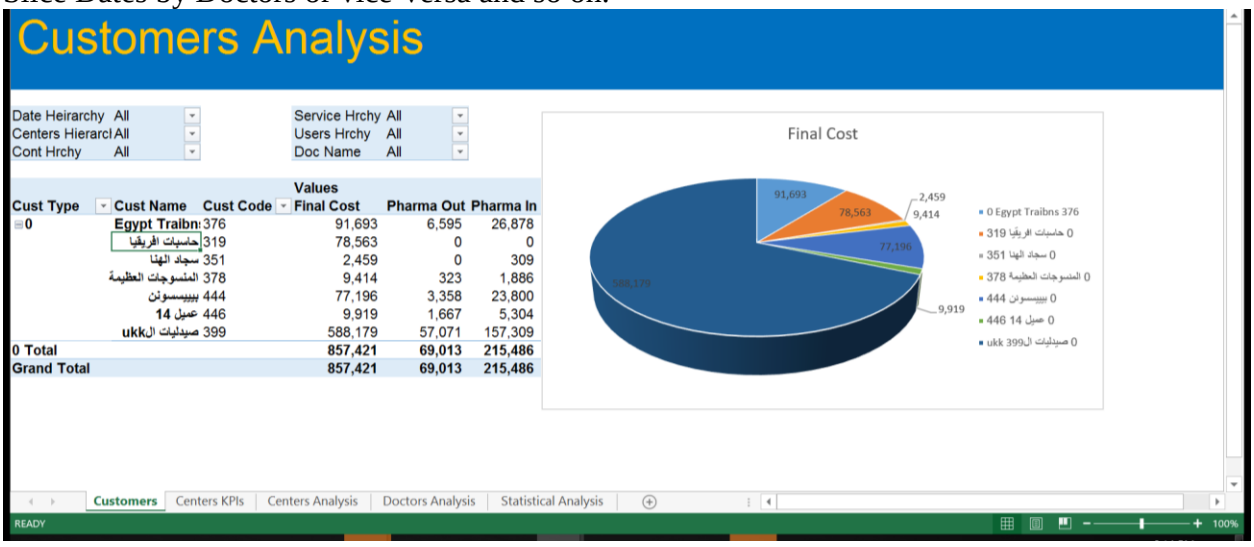
1.4.2 Claims Cube:

Cube Snow-Flake Schema:

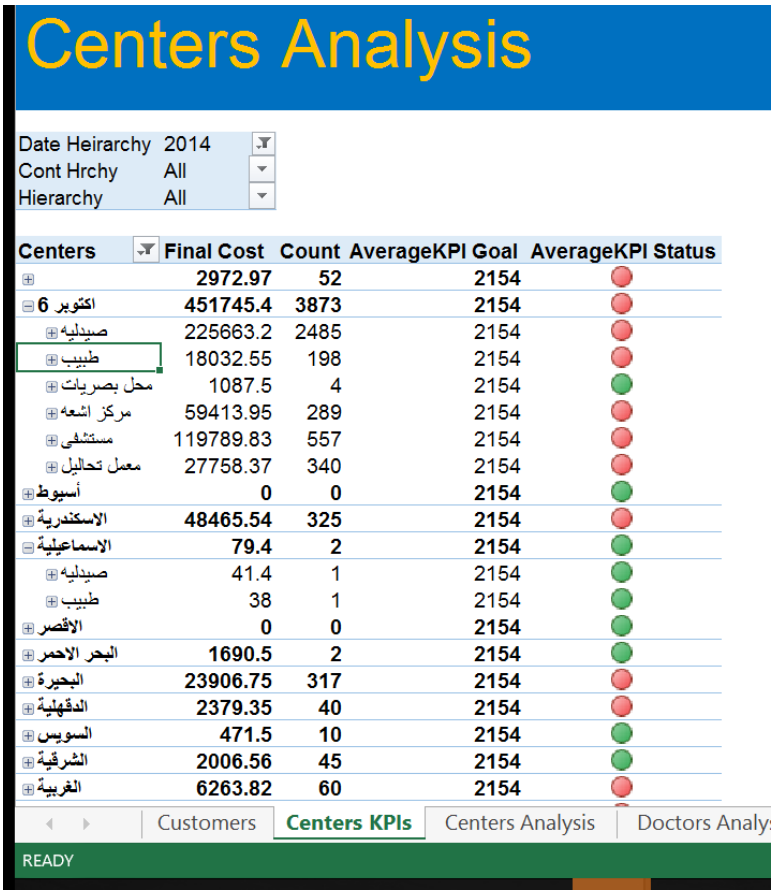


1.4.3 Claims Interface:

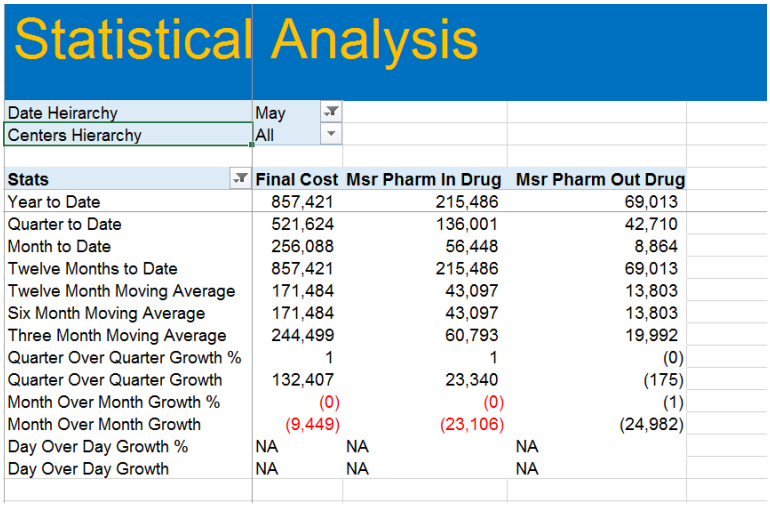
You below pivot table is connected to the above server cube allowing you to pivot, drill down, slice and dice data as you like. You can build your own hierarchy of business parameters, eg. Slice Dates by Doctors or vice-versa and so on.



Medical Centers KPIs:



Online Trend and Statistical Analysis:



Data Mining

Data mining is the process of exploring large quantities of data in order to discover meaningful information about the data, in the form of patterns and rules. Different algorithms are applied to discover critical patterns and hidden factors that would lead to business opportunities. In designing the Enterprise Data Warehouse, an important considerations must be considered if the enterprise are willing to use Data Mining technologies.

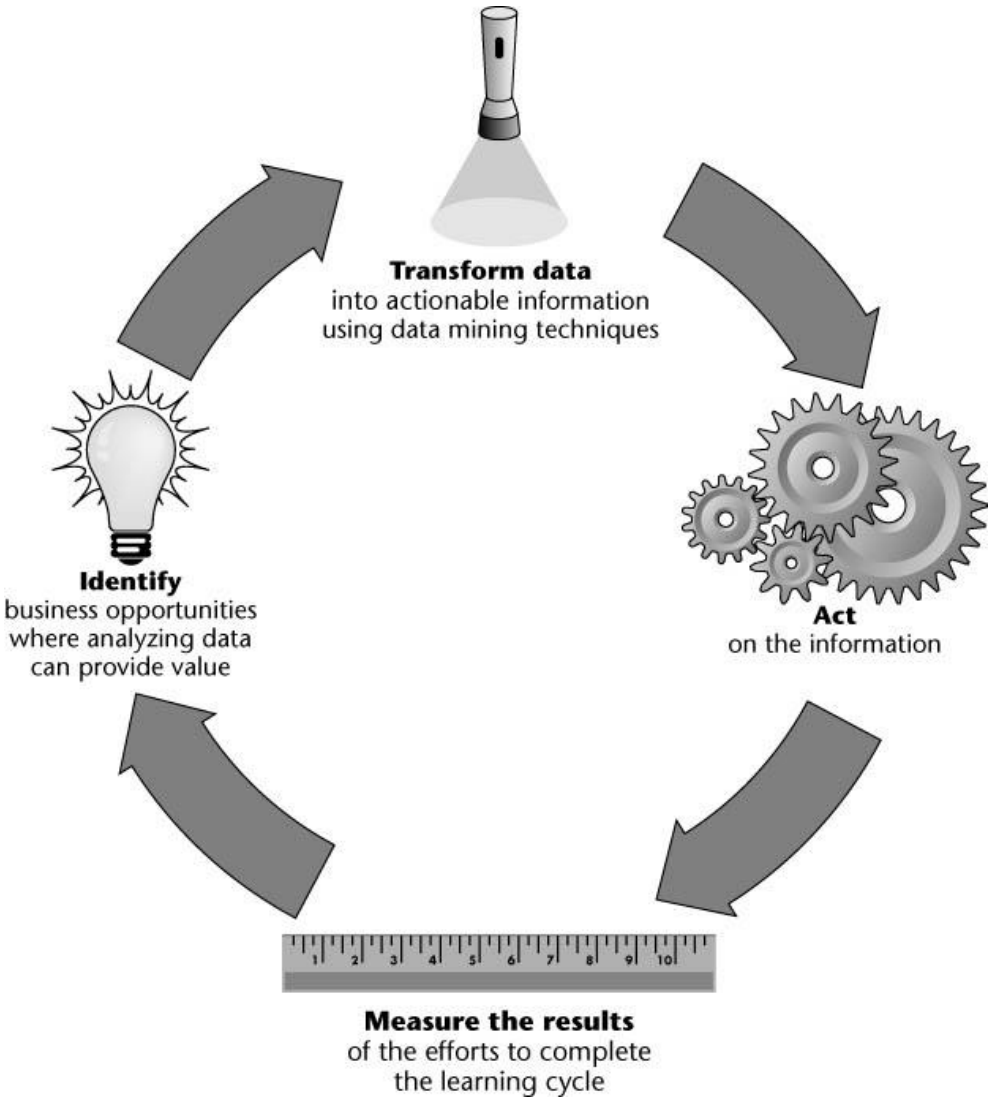
DM begins by selecting an area of interest and a goal. You might know exactly what you are looking for, but setting a goal at the beginning, say identifying cress-selling opportunities, is very important. If you don't set a goal, it is possible that you will invest time and effort that yields results that are not usable, or are trivial or incorrect. What Data Mining will help you do is weed through all the possible factors to find patterns, saving a tremendous amount of trial and error.

Data Mining is also known **Knowledge Discovery**, in that you discover knowledge that is hidden in the data. In the business environment, the end result of this newfound knowledge is some kind of policy ore action that improves profits. This policy or action is not as simple as impact a price increase has on the top line. Instead, it may be a better retention of existing customers that allows steady growth or at least stops sales erosion.

Data Mining is an excellent complement to **OLAP** analysis. Although OLAP makes it possible, even easy, to view data in all kinds of new ways and learn a great deal about the underlying business processes, it is still bound by the basic limitation that it will return no more than exactly what you ask of it. Data Mining breaks that barrier by applying technique that can create a new model of the data that might not have been considered before. So, OLAP is a user-directed activity, while algorithms control data mining.

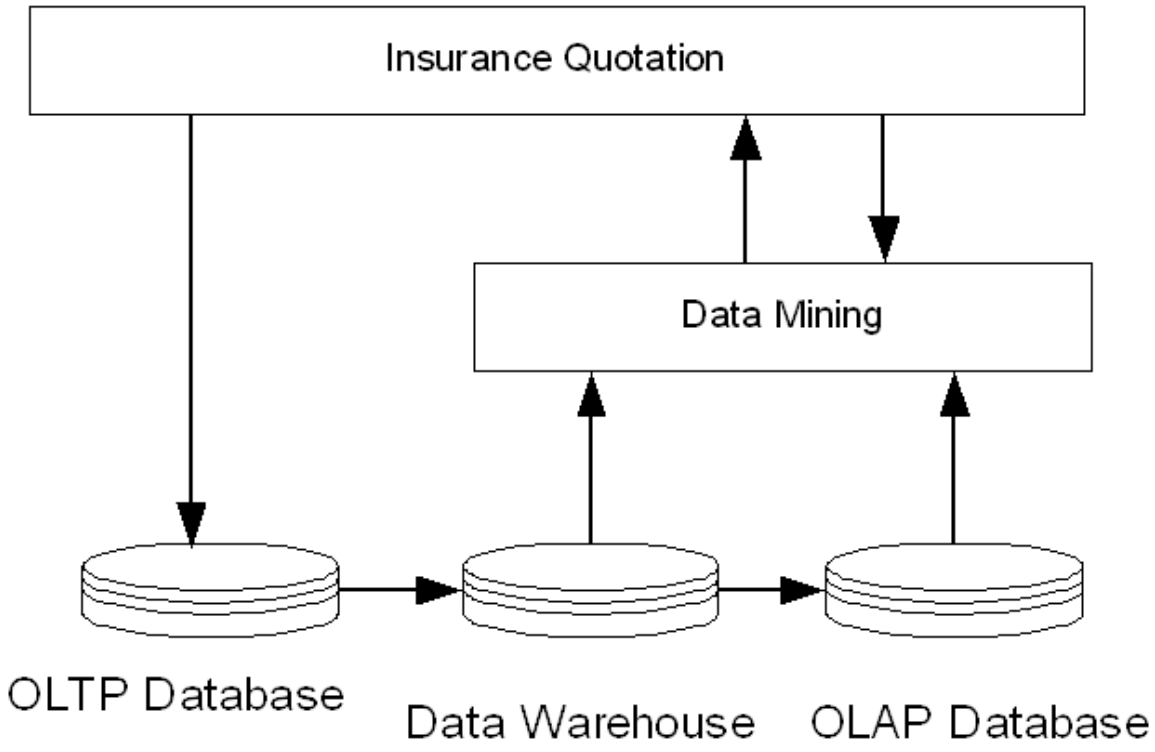
Data Mining vs. Statistical Analysis

Data Mining starts with general goal that you want to learn more about. This is in contrast to statistical analysis, in which you form a formal hypothesis and conduct various statistical tests to determine whether the hypothesis can be rejected.

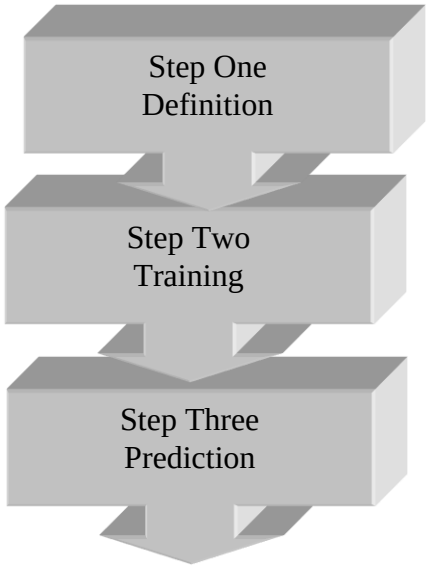


Operational II

Here the operational system is in closed loop with the knowledge discovery platform of the data warehouse so that the learning modules get updates from every day business, meanwhile support the operations in ongoing adaptive business validations.



Data Mining Process:



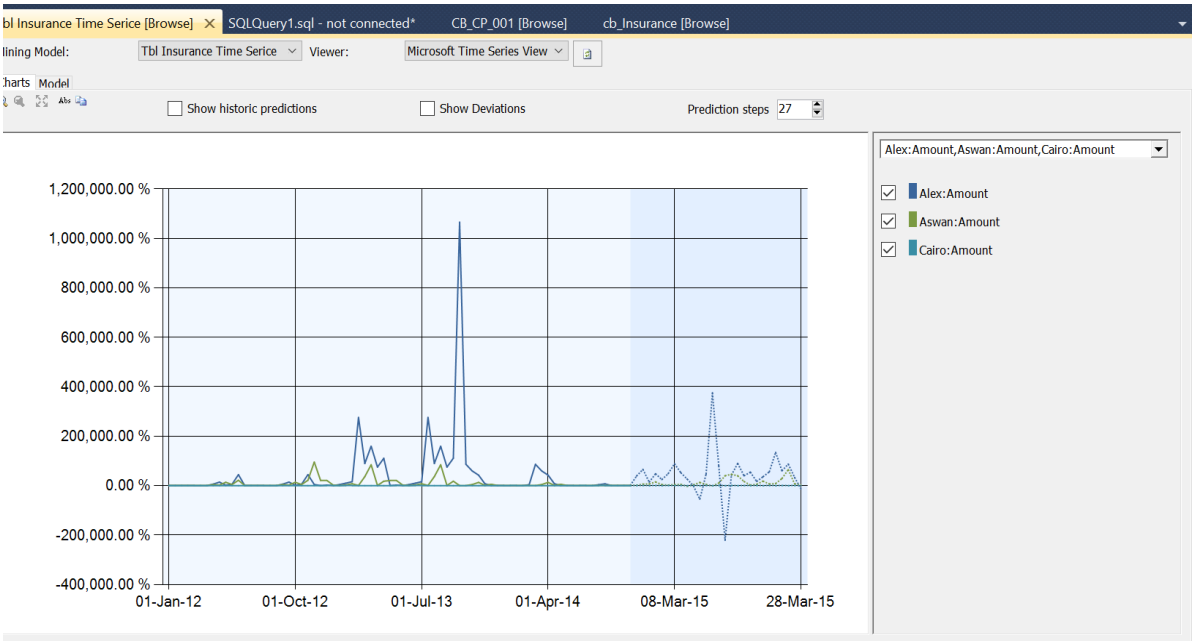
Our Proposed Solutions:

As explained above, successful data mining models should answer business needs and problems. As such, 3adda team will need to listen to your different business areas where advanced analytics are required and then suggest the best fit solution. However, the following are just suggestions, guidance and ideas drawn from our experience in Businessadvanced analytics solutions for last 20 years.

1.5 Machine Learning and Basic Operations:

1.5.1 Predictive

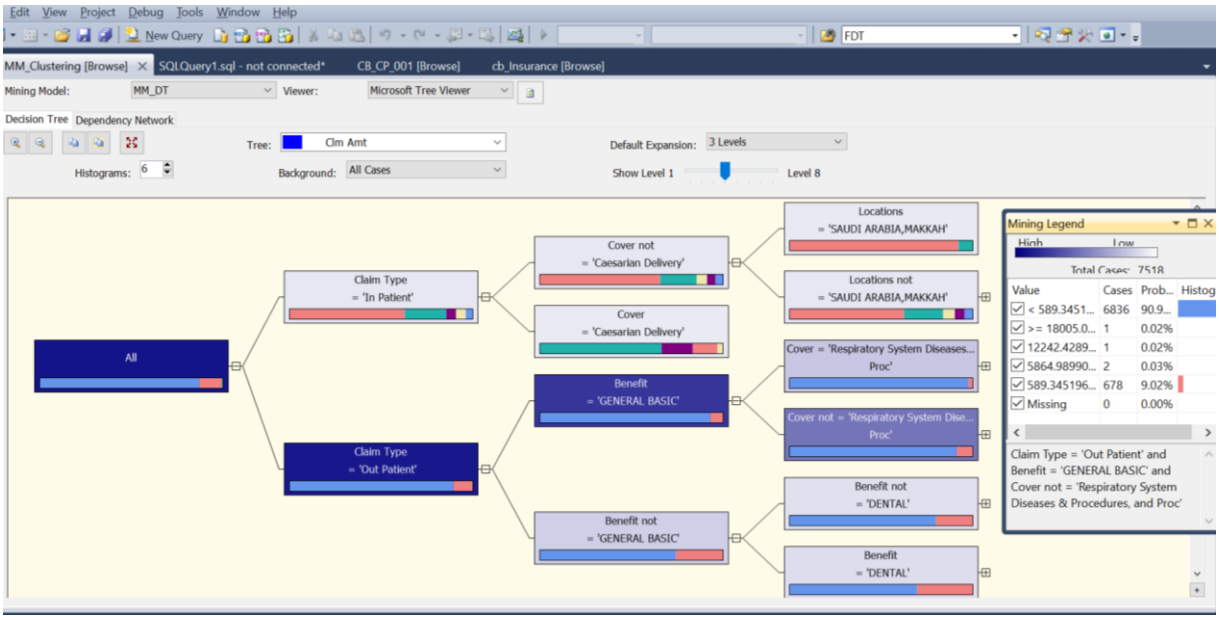
1.5.1.1 Regression



We are now in April, how sales amount can be predicted for the coming year? This time series regression algorithm answers you!

1.5.1.2 Classification

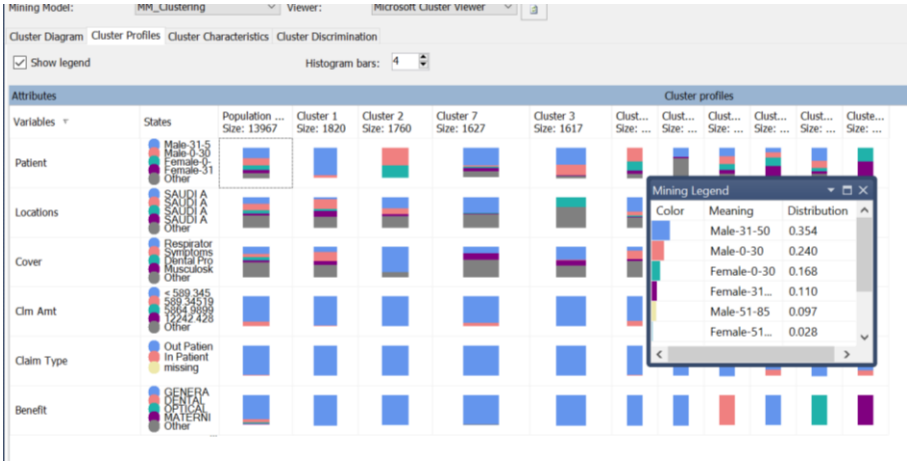
Here, through the Decision Tree Algorithm, you can see where claim amount are more concentrated in specific nodes and how the machine learning technology classifies your data based on our design of the model.



1.5.1.3 Collaborative Filtering

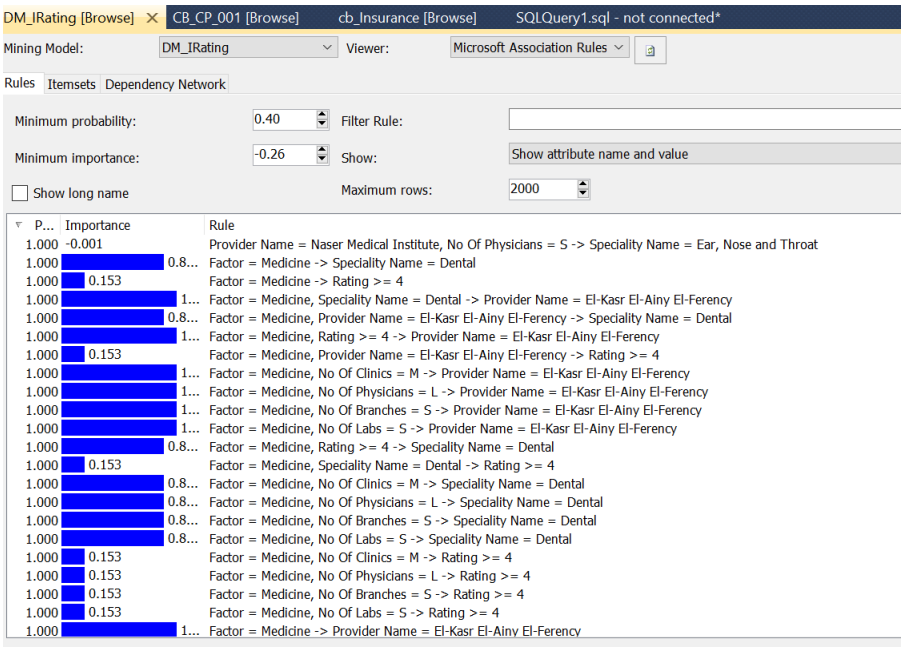
1.5.2 Descriptive

1.5.2.1 Clustering / Similarity Matching



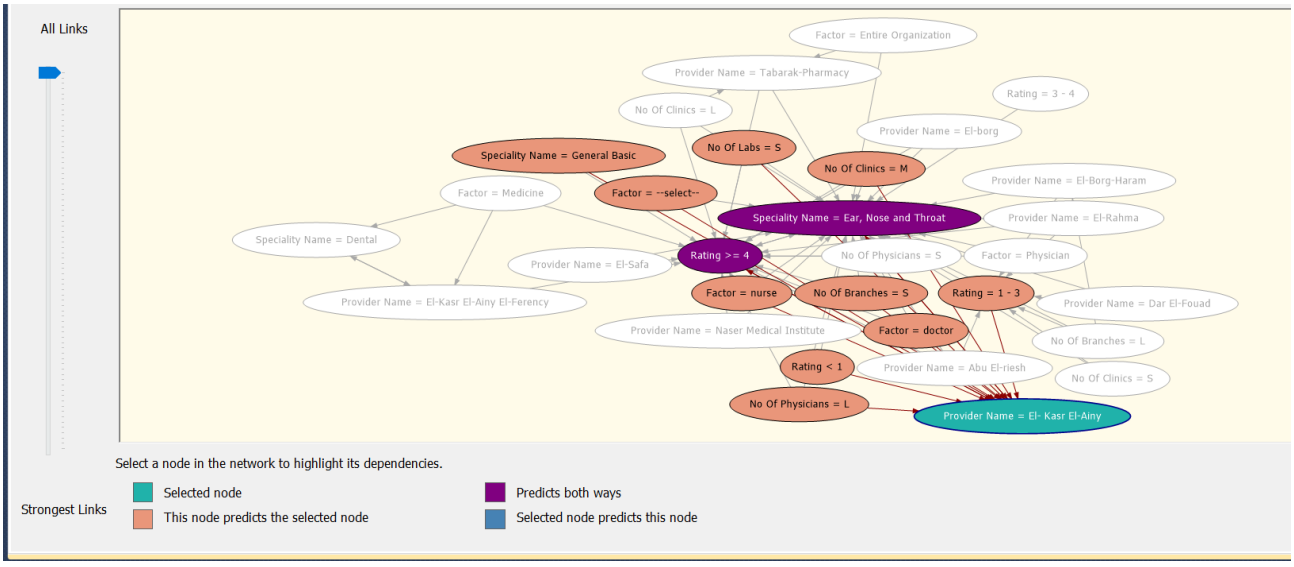
Here we analyse how population of data is clustered drawing a lot of learning and understanding in a non-traditional way. You will see your business in totally different ways, giving you a lot of conclusions which will deffinitly impact your decision making very positively, yet intelligently.

1.5.2.2 Association Rules and Variants



The mining model above build on Association Rules algorithms to know what are the result if some business parameters are jointly found in the transactions. Many conclusions can be drawn from such associations in Sales, Marketing, UW, Claims and Risk Management.

How different business parameters when grouped together with certain value for each one are affecting each other and what can predict which can be found in the same model as shown below in Dependency Network:



1.5.2.3 Deviation Detection

Data mining task in which the goal is to build a model that describes the most significant changes in the data from previously measured or normative values.