



**StatSoft®**  
**Business White Paper**

***STATISTICA Data Miner  
in the Insurance Industry***

---

**U.S. Headquarters: StatSoft, Inc. • 2300 E. 14th St. • Tulsa, OK 74104 • USA • (918) 749-1119 • Fax: (918) 749-2217 • info@statsoft.com • www.statsoft.com**

Australia: StatSoft Pacific Pty Ltd.  
Brazil: StatSoft South America  
Czech Republic: StatSoft Czech Rep. s.r.o.  
France: StatSoft France

Germany: StatSoft GmbH  
Hungary: StatSoft Hungary Ltd.  
Israel: StatSoft Israel Ltd.  
Italy: StatSoft Italia srl

Japan: StatSoft Japan Inc.  
Korea: StatSoft Korea  
Netherlands: StatSoft Benelux BV  
Poland: StatSoft Polska Sp. z o. o.

Portugal: StatSoft Iberica Ltda.  
Russia: StatSoft Russia  
Singapore: StatSoft Singapore  
S. Africa: StatSoft S. Africa (Pty) Ltd.

Spain: StatSoft Espana  
Sweden: StatSoft Scandinavia AB  
Taiwan: StatSoft Taiwan  
UK: StatSoft Ltd.

## Overview

With the lack of competition due to protective regulations, insurance companies have been slow to adopt data mining tools commonly used in other industries. However, deregulation and growing competition have insurance companies moving to more towards customer-centric operations, and the need for such tools has become vital to the industry.

Firms in the insurance industry maintain enormous amounts of data about the activities and preferences of their customers. They must convert this large amount of raw data into intelligence about customers, markets, and competitors, in order to develop new products and services to meet customers needs. By recognizing subtle patterns within terabytes of data, *STATISTICA Data Miner* allows companies to predict customers' behaviors and responses before they happen, along with identifying hard to find yet costly events such as fraud, delinquency, and noncompliance.

## Areas of Application

### Claims analysis

*STATISTICA Data Miner* helps users understand subtle business trends in claims, which would have been otherwise difficult to spot. OLAP tools can be used to analyze and drill down to the detailed level for a better understanding of these trends.

### Predict which customers will buy new policies

*STATISTICA Data Miner* provides the insurance firm with reporting, tracking, and analysis tools to identify trends. Sequential pattern mining functions are powerful and can detect sets of customers associated with frequent buying patterns.

### Claims estimation

OLAP tools can also be used to arrive at better claim estimates by analyzing the claims data across geographies and customer segments.

### Identify behavior patterns of risky customers

Insurance firms have huge databases of information used to adjust estimates of risk. By analyzing more data than previously possible and discovering groups of high risk customers, insurers are able to alter their premiums to reflect risk potential.

By using *STATISTICA Data Miner* to find patterns and predict likely behavior, companies can also identify people who lie on applications or are more likely to engage in dangerous or illegal activities.

### **Identify fraudulent and abusive billing practices**

Claims fraud is a significant and costly concern, costing property and casualty insurance companies several billion dollars annually. Losses due to fraud have increased dramatically in the past ten years. Despite actions by insurance companies, a large amount of fraud remains undetected.

*STATISTICA Data Miner* helps the insurance company anticipate and quickly detect fraud and take immediate action to minimize costs. Through the use of sophisticated data mining tools, millions of medical claims can be searched to spot patterns and detect even subtle variations in billing practices, by analyzing above normal payoffs along different factors like geographical region, agent, and insured party.

## **Other Uses**

*STATISTICA Data Miner's Associations Rule* module may be used to analyze claim forms submitted by patients to a medical insurance company. Every claim form contains a set of medical procedures performed during one visit. Using the Associations Rule module, the insurer will be able to find relationships among medical procedures that are often performed together.

*STATISTICA Data Miner's STATISTICA Generalized EM and k-Means Clustering* module may be used for customer segmentation, by grouping the entire customer base into clusters, identified on the basis of various demographic and psychographics factors.

*Regression* modules are used to predict one variable from the value of one or more other variables. Insurance risk analysts use regression when setting the rates of premiums; for example, the average value of a claim will be estimated from variables such as age gender of policy-holders.

