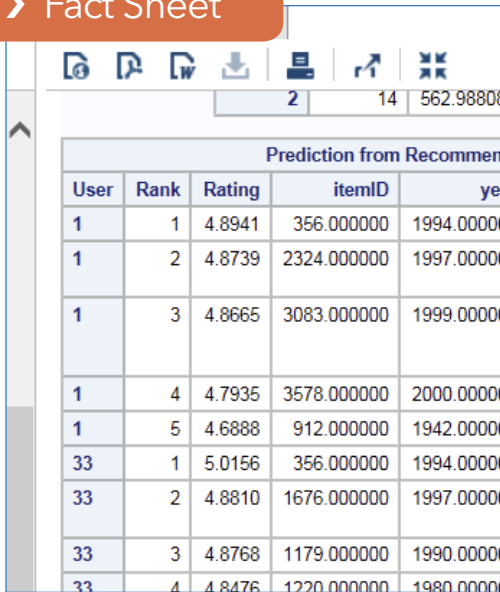


## > Fact Sheet



| Prediction from Recommendation |      |        |             |           |
|--------------------------------|------|--------|-------------|-----------|
| User                           | Rank | Rating | itemID      | year      |
| 1                              | 1    | 4.8941 | 356.000000  | 1994.0000 |
| 1                              | 2    | 4.8739 | 2324.000000 | 1997.0000 |
| 1                              | 3    | 4.8665 | 3083.000000 | 1999.0000 |
| 1                              | 4    | 4.7935 | 3578.000000 | 2000.0000 |
| 1                              | 5    | 4.6888 | 912.000000  | 1942.0000 |
| 33                             | 1    | 5.0156 | 356.000000  | 1994.0000 |
| 33                             | 2    | 4.8810 | 1676.000000 | 1997.0000 |
| 33                             | 3    | 4.8768 | 1179.000000 | 1990.0000 |
| 33                             | 4    | 4.8476 | 1220.000000 | 1980.0000 |

### What does SAS® In-Memory Statistics do?

SAS® In-Memory Statistics provides a single interactive programming environment for analytical data preparation, variable transformations, exploratory analysis, modeling, integrated model comparison and scoring.

### Why is SAS® In-Memory Statistics important?

It's a fast, powerful and customizable in-memory programming language that lets multiple users concurrently and interactively analyze large amounts of data. This results in greater analyst productivity and turn-on-a-dime creativity.

### For whom is SAS® In-Memory Statistics designed?

It's designed for experienced statisticians, data miners, data scientists, engineers, researchers, biostatisticians and scientists who need to analyze large and complex data.

# SAS® In-Memory Statistics

A single interactive programming environment for analytics

Big data. Messy data. Structured and unstructured data. Unfortunately, it's becoming more difficult to perform the fast analytical processing needed to deliver instantaneous insights from so much raw and unorganized information. It's also the case that multiple software products and rare data scientist talent are needed to manage the variety of tasks in the analytical life cycle. Getting insights from large, disordered data in a timely manner requires a different approach.

SAS In-Memory Statistics gives multiple users the ability to simultaneously analyze huge volumes of data. This, combined with very powerful analytical techniques, provides an unprecedented way to tap into big data to quickly derive insights for high-value analytical decisions. Dramatically reducing model development time means more models can be put in action sooner.

This solution also minimizes processing time and scales to meet enterprise needs. You can handle more data, more users and even more complex issues.

## Benefits

- **Delve deep into big data for fast, accurate insights.** Apply proven state-of-the-art statistical algorithms and machine-learning techniques to find the best answers. You can explore and use multiple analytic approaches to reveal insights and make fact-based decisions.
- **Increase productivity for your data scientists.** Multiple users can concurrently and interactively analyze big data using the fast, in-memory analytical programming language. Prepare, manipulate, transform, explore, model, access and score data.
- **Take advantage of a scalable environment.** Until now, statisticians and data scientists had to piece together different programming languages or products to access, prepare, model and score data. And when it came time to operationalize models, the software couldn't scale. No more. From data manipulation and exploration to model building and deployment, our solution is proven, tested and accurate - and can scale to your production environment.
- **Avoid unnecessary, multiple passes through the data.** Our in-memory infrastructure eliminates costly data movement and persists data in memory for the entire analytic session. This significantly reduces data latency and provides rapid analysis at lightning-fast speeds.

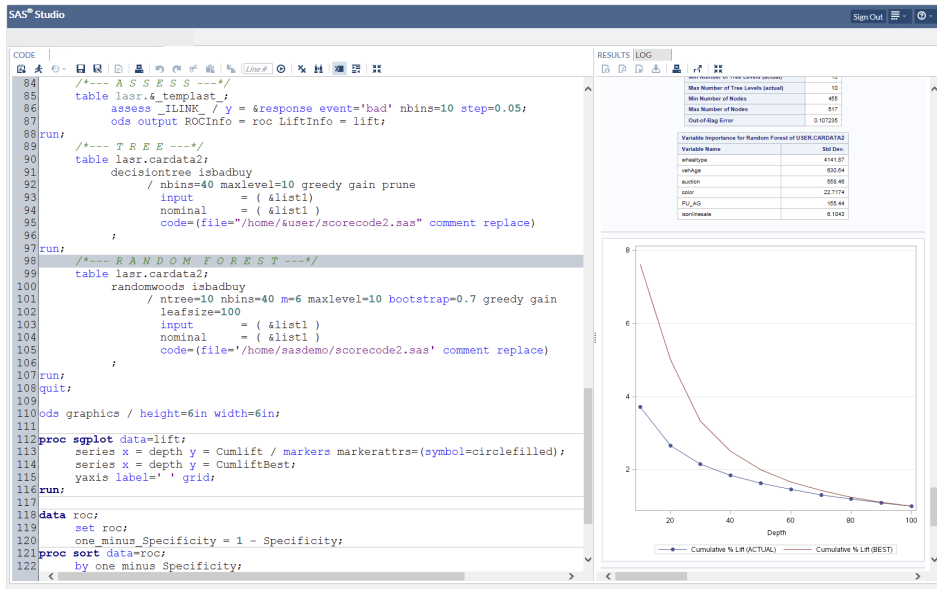
## Overview

SAS In-Memory Statistics is a single, interactive programming environment where multiple users can simultaneously access and manage data, transform variables, perform exploratory analysis, build and compare models, and score models. It uses in-memory analytic process-

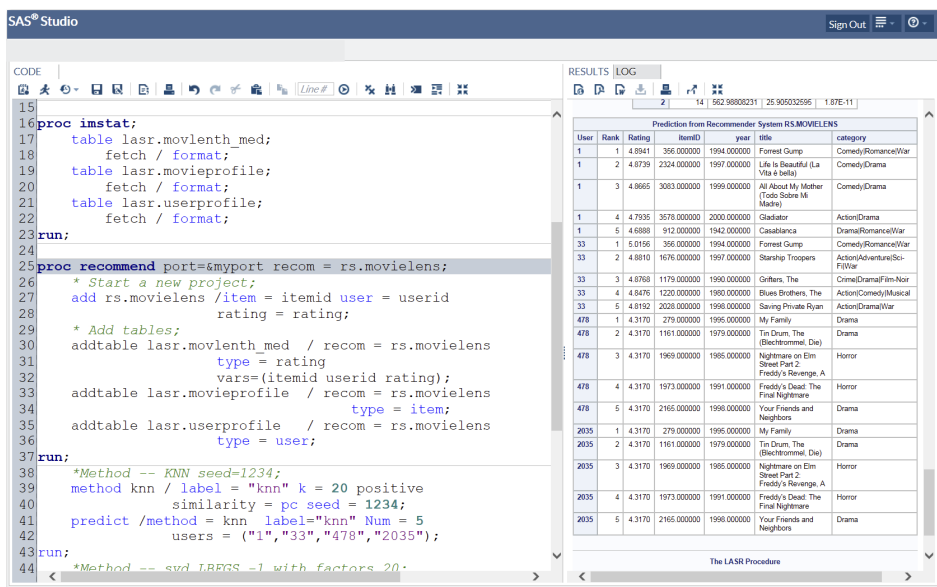
ing so results are produced nearly instantaneously.

State-of-the-art statistical and machine-learning techniques, including regression, clustering, neural networks, decision trees, random decision forests, etc., help identify new patterns, trends

and relationships in complex big data. The interactive language and built-in actions enable you to explore multiple modeling approaches and reach optimal solutions faster than other programming languages like MapReduce and R. This in turn drives better strategies for improving business processes.



A wide variety of statistical algorithms and machine-learning techniques, like this random decision forest approach, are available in a highly interactive programming environment.



Generate personalized, meaningful recommendations in real time with a high level of customization.

## Interactive programming

Multiple users can concurrently and interactively analyze large amounts of data by interactively submitting SAS code through the web browser developer environment. Instant responses are provided for complex problems. The large level of interactivity gives users the flexibility to quickly explore different paths and ask difficult what-if questions that can generate higher returns.

## In-memory analytical processing

Get fast analytic computations that are optimized for multiple passes across a distributed cluster. This dramatically shortens model development time so more models can be developed in less time and put in action sooner. SAS In-Memory Statistics uses a "read once, analyze repetitively" in-memory computational paradigm for faster interactive ad hoc analysis to prepare the data for modeling.

## Data persisted in memory

This solution requires loading the data in parallel only once to support truly interactive discovery and problem solving using a full spectrum of analytical methods. Gain speed and reduce latency because data is held in memory.

## Analytical data manipulation

Access and manipulate data, transform and create variables and perform exploratory analysis. Users can explore and prepare their data more efficiently and are able to shorten the time for model development and deployment.

## Statistical algorithms and machine-learning techniques

Uncover patterns and trends faster than ever with a huge breadth and depth of analytical techniques. This wide range of analytical algorithms and processes can reveal patterns, anomalies, key variables and relationships, leading ultimately to new insights and better answers faster. Only SAS offers an integrated suite of statistical algorithms and machine-learning techniques that is unmatched in the industry, delivered to you in a single programming environment.

## Text analytics

Analyze your unstructured (and structured) data using a wide variety of text analysis techniques. Combining structured data with text data uncovers previously undetected relationships and adds even more predictive power to models.

## Model development

Quickly create, evaluate and compare multiple statistical models. You can explore and use multiple analytic approaches to reveal insights and make high-impact decisions. In real-time, you can see the effect of changing a component of a model – for example, adding a temporarily computed variable, changing a filter condition, adding a variable selection or changing a variable from continuous to classification.

## Recommendation system

Generate personalized, meaningful recommendations in real time with a high level of customization. Collaborative filters, matrix factorization, hybrid models and affinity analysis are used to build customized recommendation systems that present the next best offers, etc.

# Key Features

## Interactive in-memory programming

- Performs all mathematical calculations in memory.
- Uses a dynamic group-by processing operation to compute and process results for each group, partition or segment without having to sort or index data each time.
- Provides a new web-based interface, SAS Studio, for SAS programmers.
- Interactive programming language supports submitting, retrieving results and then submitting more statements on the fly.
- You chain together analytical tasks as a single in-memory job without having to reload the data or write out intermediate results to disks.
- You can update source tables with new column transformations, filter rows and perform group-by processing.

## Analytical data preparation

- Intelligent partitioning by a variable(s) across the cluster for more efficient data access. You can repartition or unpartition the data at any time.
- Derive new temporary tables and promote them for use by other analysts.
- Data manipulation for subsetting, filtering, joining and promoting tables, and computing columns.
- Data access patterns and processing methods such as group filtering, partitioning, data ordering within partitions, and delete/undelete operations.
- Data sources can be defined using update, append, set operations, filter, derive column, and aggregate statements.
- Perform transformations on input data - missing value imputation, outlier transformation, functional transformation, and binning - yielding output tables and/or score rules.
- Export ODS results tables for client-side graphic development.

## Statistical algorithms and machine-learning techniques

- Classification trees and regression trees.
- Forecasting.
- General linear models.
- Generalized linear models.
- Logistic regression.
- Neural networks.
- Hypergroups.
- Random decision forests.
- k-means clustering.
- Density-based spatial clustering.
- Association rule mining.

## Descriptive statistics

- Distinct counts.
- Box plots.
- Pearson's correlation.
- Crosstabulations, including support for weights.
- Contingency tables, including measures of associations.
- Parallel by-group processing.
- Histograms with options to control binning values, maximum value thresholds, outliers and more.
- Multidimensional summaries in a single pass of the data.
- Percentiles for one or more variables.

## Key Features (continued)

To learn more about SAS In-Memory Statistics, download white papers, view screenshots and see other related material, please visit [sas.com/in-memory-statistics](http://sas.com/in-memory-statistics).

- Summary statistics such as number of observations, number of missing values, sum of nonmissing values, mean, standard deviation, standard errors, corrected and uncorrected sums of squares, min and max, and the coefficient of variation.
- Kernel density estimates using normal, tri-cube and quadratic kernel functions.

### Model assessment

- Supports generating model comparison summaries, such as lift charts, ROC charts, concordance statistics and misclassification tables, for one or more models.

### Text analytics

- Parsing and stemming.
- Start and stop lists.
- Term and document frequency.
- Matrix factorization (singular value decomposition).
- Entity extraction and resolution.
- Topic projections of the document.

### Model scoring

- Generation of SAS DATA step code.
- Score statement for applying scoring logic to training, holdout and new data.

### Recommendation system

- Interactive RECOMMEND procedure. All algorithms can be run interactively in memory.
- Interactively apply a filter to develop recommendations for specific populations.
- Project-based to support loading user, items and rating tables into memory.
- Cold starting for new users without any history based on weighted averaging.
- Slope-One fast regression-based model commonly used as a simple benchmark.
- k-nearest neighbor, including cosine, adjusted cosine and Pearson's correlation.
- Matrix factorization with options for loss functions, regularization factors, optimization methods and more.
- Clustering of users and/or items using other attributes, including term frequency and inverse document frequency weights.
- Hybrid or ensemble models.
- Define a holdout set of users and ratings for training and validation evaluation.
- Predict action for scoring one or more new users or a table.

**PROC Recommend Movie Recommendations**

| Obs | UserID        | Rank | Rating | Title  | Year | Category                          |
|-----|---------------|------|--------|--|------|-----------------------------------|
| 1   | Mary Clarke   | 1    | 6.07   | Indiana Jones and the Last Crusade             | 1989 | Action Adventure                  |
| 2   | Mary Clarke   | 2    | 5.54   | Star Wars: Episode V - The Empire Strikes Back | 1980 | Action Adventure Drama Sci-Fi War |
| 3   | Mary Clarke   | 3    | 5.43   | Alien  | 1979 | Action Horror Sci-Fi Thriller     |
| 4   | Mary Clarke   | 4    | 5.40   | Pretty Woman                                   | 1990 | Comedy Romance                    |
| 5   | Mary Clarke   | 5    | 5.30   | Stand by Me                                    | 1986 | Adventure Comedy Drama            |
| 6   | John Thompson | 1    | 4.70   | Fugitive, The                                  | 1993 | Action Thriller                   |