

# **Market Basket Explorer 1.0 User's Guide**

**Beta Version**

**September 2009**

**License Terms:**

SENSORLYTICS MARKET BASKET ANALYZER (THE PROGRAM) AND ASSOCIATED MATERIALS (DOCUMENTATION AND DATA SETS) ARE PROVIDED TO YOU "AS IS," WITHOUT WARRANTY. THERE IS NO WARRANTY FOR EITHER THIS PROGRAM OR ASSOCIATED MATERIALS, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. YOU ASSUME ALL RISK ASSOCIATED WITH THE USE OF THIS PROGRAM OR ASSOCIATED MATERIALS. SENSORLYTICS LLC ASSUMES NO LIABILITY FOR ANY DAMAGES, EITHER DIRECT, OR INCIDENTAL INCURRED AS A RESULT OF YOUR USE OF THIS PROGRAM OR ASSOCIATED MATERIALS.

Sensorlytics LLC (the Licensor) hereby grants you the following rights, subject to the restrictions specified in this License:

- 1) You may run the program for your own use, or for the use of your organization provided that:
  - (a) The use of the program is not associated with the provision of consulting services, or other services of a similar nature, to any third party or parties outside of your organization.
- 2) You may copy and distribute the program and supporting documentation provided that:
  - (a) You do not modify the program or associated materials in any way.
  - (b) Your distribution of the program includes all of the associated materials.
  - (c) No payment (directly or indirectly) is received for your distribution of the program or associated materials to a third party.

If you or your organization are interested in alternative licensing terms, please contact Sensorlytics LLC for further information.

## Market Basket Explorer 1.0 Overview

Market Basket Explorer (MBE) is a proprietary analytics tool developed by Sensorlytics LLC for the purpose of identifying sales relationships between individual products within a given product line. It accepts data on customer purchases and will analyze this data looking for product combinations which occur more frequently than would be expected by chance alone. It provides the ability to screen such combinations based on various statistical measures, and to both interactively display the results and to write them back to a file for further analysis and investigation.

For additional information on the data analysis techniques used by MBE, consult the Sensorlytics White Paper, *Identifying Exploitable Product Synergies in Sales Data.*, document #09005A.

## Minimum System Requirements

Windows XP or Vista, 1GB RAM, 800x600 display, 10MB free hard disk space.

## Installation

After downloading the 'zipped' file MBE10B.ZIP, decompress it to a directory of your choice. Once the executable file MBE10B.EXE has been decompressed, it may be run by double-clicking on the file icon without any further installation. If you have received MBE on a CD-ROM (in unzipped form), you may either copy the files to a directory of your choice, or may run it directly from the CD-ROM drive.

## Included Files

In addition to the program file MBE10BD.EXE, the zip file MBE10BD.ZIP also contains a number of additional files (including this one). These included files are:

File	Description
MBE10B.EXE	Executable program file (demo version)
MBE10B_Manual-D09006A.PDF	This manual (PDF file)
AffinityAnalysis-D09005A.PDF	White paper describing analysis techniques used by MBE
DATASET01.CSV	Example of market basket data lacking deliberate affinity relations
DATASET02.CSV	Example of market basket data with X-Y relations
DATASET03.CSV	Example of market basket data with X-Y and X-Y-Z relations
DATASET01A.CSV	Similar to DATASET01.CSV, but with non-uniform product distribution
DATASET02A.CSV	Similar to DATASET02.CSV, but with non-uniform product distribution
DATASET03A.CSV	Similar to DATASET03.CSV, but with non-uniform product distribution

## Analysis Capacity

MBE is limited by the following capacity constraints:

Constraint	Limits
Maximum unique item types	2,000
Maximum Customers	2,000
Maximum Total Items	4,000,000

## Data Preparation

MBE is intended for use with a spreadsheet program (e.g. Microsoft Excel™), and uses comma-separated-variable (CSV) file formats for input and output. Input data files are prepared in the following format:

```
<PRODUCTDATA>
Product1, Product1_Description
Product2, Product2_Description

<BASKETDATA>
CustomerA, Product1, Product2, Product3,..... ProductN...
CustomerB, Product1, Product2, Product3,..... ProductN...
CustomerC, Product1, Product2, Product3,..... ProductN...
```

The first section is indicated by the line ‘<PRODUCTDATA>’, which is used to provide text descriptions of products. Each of the following line is the product identifier, a comma, and a text description. This section is optional. Product identifiers may be alpha-numeric strings.

The next section begins with the line ‘<BASKETDATA>’ and describes customer purchases. Each customer’s ‘basket’ must be placed on a single line and may *not* be continued to the following line. The first item is a customer identifier, followed by a comma-separated list of which products that customer has bought. The following is an example of a short basket file with both product and basket data sections. Items may be enclosed in quotes, but quotes will be ignored. Commas should not be used as parts of labels or descriptive text as even those appearing within quotes will be interpreted as item delimiters.

```
<PRODUCTDATA>
156345, "YELLOW CLUCK JUMBO EGGS"
175923, "HAPPY BROWN COW FARMS MILK"
236545, "HAPPY BROWN COW FARMS CHEDDAR BLOCK"
674544, "MOMS KITCHEN SUPERLOAF BREAD"

<BASKETDATA>
CUST_1, 156345, 175923, 674544
CUST_2, 156345, 236545, 175923
CUST_3, 156345, 674544, 175923
```

The <PRODUCTDATA> section may be omitted if sufficiently descriptive alphanumeric labels are used. For example:

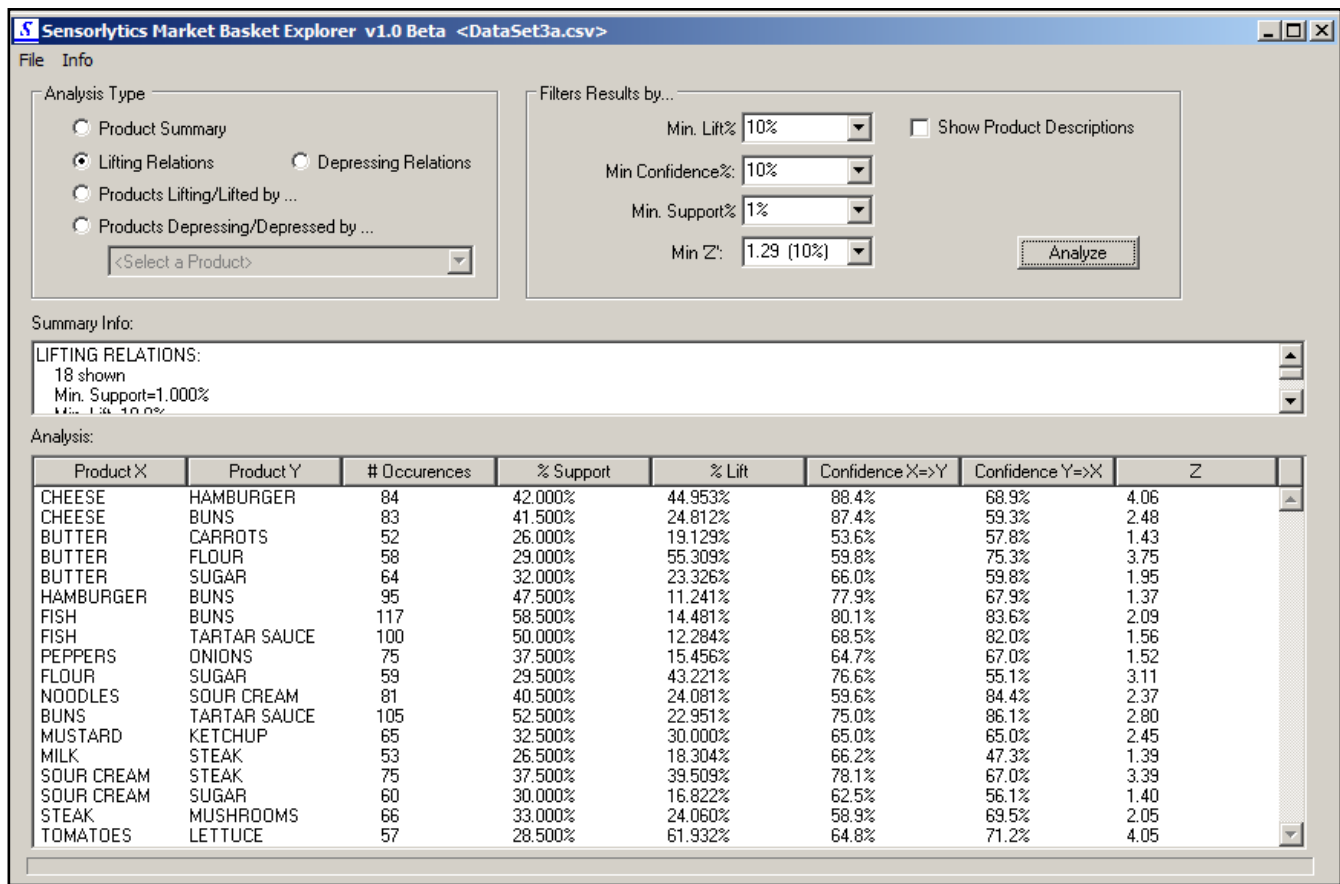
```
<BASKETDATA>
"Customer-Jim", "EGGS", "MILK", "BUTTER"
"Customer-Mary", "MILK", "CHEESE", "BREAD"
"Customer-Ralph", "BEER", "CHEESE", "STEAK"
```

Blank and duplicate purchases are ignored. Blank lines are also ignored. The order in which items are listed is also irrelevant. For example, the following two data lines are equivalent:

```
"Jim", "MILK", "", "BUTTER", "", "MILK", "EGGS"
"Jim", "EGGS", "MILK", "BUTTER"
```

## MBE Screen and Operations

The following sections will refer to controls on the MBE main screen, shown below:



## Loading Data into MBE

To load a CSV data file, select FILE>LOAD DATA from the menu bar. You will be prompted to select a file. After the file loads, the filename and information will be displayed in the 'Summary Info' box.

## Analysis Types

MBE provides five analysis types:

- 1) Product Summary - list of products and purchase frequencies.
- 2) Lifting relations - list of all product pairs with positively correlated purchase rates.
- 3) Depressing relations - list of all product pairs with negatively correlated purchase rates.
- 4) Products Lifted by... - list of products whose purchases positively correlate with those of a specified product
- 5) Products Lifted by... - list of products whose purchases positively correlate with those of a specified product

To perform an analysis, select the desired analysis from the ANALYSIS frame and click on the <ANALYZE> button. When the analysis is complete, the results will be displayed in a table in the analysis window at the bottom of the screen. Note that in order to be displayed, a relation must meet the current filter criteria, so it is possible to have an analysis run and not display any results - see the FILTER OPTIONS section for more details. Also note that only the first 2000 results encountered will be displayed, so for many analyses it will be necessary to experiment with the filter settings to obtain potentially useful results.

## Displaying Results

Results are automatically displayed after performing an analysis. To refresh the display, click on the <ANALYZE> button. Only results meeting the filter criteria are displayed. See the FILTER OPTIONS section for more details.

One can adjust the column widths of the display area to aid in viewing results. Clicking on the column heading sorts the results by that column, but in alpha-numeric, as opposed to strictly numeric order.

## Filter Options

Because analyses can identify *very* large numbers of product relationships, and only a small minority of those relationships may be either statistically significant or particularly interesting, MBE provides a means of filtering displayed and exported results. The following filter criteria are available in the 'Filter Results by' box:

**Min Lift%** - Only display relationships that possess a minimum percentage of lift. This is useful to filter out relations that do not occur often enough to be potentially interesting from a commercial standpoint. Applies when seeking lifting relations

**Min Depression%** - Only display relationships that possess a minimum percentage of depression. This is useful when trying to find products whose sales are mutually exclusive. Applies when seeking depressing relations

**Min Confidence%** - Only show relations where there is at least this probability that buying one item implies buying the other. Useful in quantitatively linking the sales of different items. Only applies when seeking lifting relations.

**Min Support%/Min Expected Support%** - Only display relationships that occur in a minimum percentage of baskets. This is useful in helping to determine if a relation is potentially interesting from a commercial standpoint. When seeking lifting relations this filters on actual support. When seeking depressing relations this filters on *expected* support based on independent sales rates.

**Min Z** - Only display relationships that have a 'Z' higher than this rating. This is mainly useful in screening out relationships that are unlikely to be statistically significant. In large data sets with large numbers of relationships, there will often be many relationships with high 'Z' ratings as a result of random chance. Note that statistical significance does not automatically imply practical significance!

Show Product Descriptions - checking this box will include the product description data from the <PRODUCTDATA> section of the input file in analysis results. This can make analyses more readable, but also more verbose.

If you change filter options, the display doesn't automatically update. To update the display you must click on the <ANALYZE> button.

## Exporting Results

Once you have an analysis completed, you can export the results to either a text file with fixed-width columns (\*.TXT) or a comma-delimited (\*.CSV) file. To do this, go to the menu bar and select either 'FILE>SAVE AS TXT' to export a text file or 'FILE>SAVE AS CSV' to save a CSV formatted file.

## About the Sample Data Sets

The sample data sets included in the distribution package are synthetic data that are essentially random. Each of the products was included at approximately a uniform probability (the non-'a' data sets) or with a randomized probability (the 'a'-data sets). In addition, certain relations have been deliberately emphasized in each of the data sets as shown by the following table:

Set	Base Distribution	Order-2 Relations	Order-3 Relations
Dataset01.csv	uniform	-	
Dataset01a.csv	random		
Dataset02.csv	uniform	<Steak-Mushrooms> <Mustard-Ketchup> <Lettuce-Tomatoes> <Peppers-Onions> <Bacon-Eggs>	-
Dataset02a.csv	random		
Dataset03.csv	uniform		<Fish-Buns-Tartar Sauce> <Sour Cream-Steak-Noodles> <Hamburger-Cheese-Buns> <Sugar-Butter-Flour>
Dataset03a.csv	random		

Loading and analyzing these data sets can be helpful in building confidence in the use of MBE. It can also provide some insights into some of the challenges of finding relations in data, an even more challenging task if you *don't* know what they are *a priori*, as is the case here.

***About Sensorlytics***

*Marketing, Communication and Analytic Services for the Sensor and Technology Industries*

Sensorlytics LLC is a consultancy that provides services to help sensor and technology companies achieve their marketing goals. Our primary service offerings include:

Market Research & Analysis  
Technical Communications  
Product Definition  
Analytic Services

For more information on what we do, the services and resources we provide, and how we might be able to assist you, please visit our website at:

**[www.sensorlytics.com](http://www.sensorlytics.com)**

**Disclaimer:**

Sensorlytics LLC makes no warranty of any kind, expressed or implied, with regard to the contents of this document. The authors and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of the furnishing or use of the information contained herein.