

WHITE PAPER

Better Marketing Starts with Better Data

Improve Data Quality for More Accurate Analysis with Alteryx and Melissa



alteryx

melissa
GLOBAL INTELLIGENCE

ABSTRACT

Organizations are under more pressure than ever to gain accurate contact data for their customers. When your consumer base ranges from Los Angeles to Tokyo, it can be challenging. Poor data quality has a critical impact on both the financial stability as well as the operations of a business. Verifying and maintaining vast quantities of accurate contact data is often inefficient and falls short of the mark. According to IBM, the yearly cost of poor data quality is estimated at 3.1 trillion in the U.S. alone¹.

Melissa's Global Address Verification and Predictive Analysis for Alteryx are the tools your business needs to grow. Achieve marketing success and empower your strategy, while reducing the cost of doing business overall.

Using advanced analytics tools in Alteryx, it's easy for a business to identify which customers to target with new products. Customer retention is just as important as gaining new ones. Therefore, businesses should strive to engage existing customers with whom they have built a rapport. We will explore a few ways to identify the right targets and how to use Melissa's Global Address Verification to cleanse addresses: saving time; money; and resources.

I. INTRODUCTION

Undeliverable-as-addressed mail (UAA) is costly for businesses. According to the DMA, postage accounts for 24% of the overall costs associated with a mail piece. When a piece is undeliverable, for whatever reason, the entire mail piece is lost – and that includes the cost of lost opportunities. These costs can be considerably higher if a business has many international customers, as international postage is usually more expensive.

On average, according to the USPS®, 6.4 percent of the pieces in every direct-mail campaign are UAA and may never reach their intended destination. What if businesses could prevent this previously undeliverable mail from the errors that send it awry?

II. CORRECTIVE ACTION

What corrective measures can be taken to maintain data quality? Investing in address cleansing is a primary example. Improve customer relations and satisfaction when mail deliverability is maintained like a well-oiled machine. Melissa Global Address Verification cleanses and standardizes addresses for over 240 countries. Global Addresses Verification is offered both on-premise and as a web service. We will focus on the web service and show how easy it is to access through Alteryx.

One benefit of choosing a web service is routine updates of fresh data. The second benefit is usability. Within Alteryx, no programming experience is needed. Instead, control your data efficiently with intuitive drag-and-drop tools.

III. RUNNING A WORKFLOW IN ALTERYX

Alteryx predictive analytics tools, such as Logistic Regression, easily identifies existing customer targets.

The Logistic Regression tool shows the probability of an event happening (given certain conditions) unlike Linear Regression; where the user creates a relationship between a dependent variable and one or more independent variables. Alteryx Logistic Regression tool is used when a dependent variable is binary categorical. Users may analyze either a single independent variable or many which they believe “caused” the outcome of the dependent binary string. The primary difference between Linear and Logistic Regression is computing the probability: given one or more independent variables.

Unlike Linear Regression computing, the results are more complicated. Therefore, it’s not built natively in Microsoft Excel®. We will focus on interpreting results to increase business intelligence and decision making.

Together, the Alteryx Logistic Regression tool and Melissa’s Global Addressing Verification run inside an Alteryx workflow. Using a sample CSV file, TABLE I presents historical sales data for existing customers after the release of a product:

TABLE I
TARGET AND PREDICTOR VARIABLES

GENDER	AGE	AMT_SPENT	RESPONDED
1	27	80100.54	No
1	43	41415.21	Yes
1	26	50214.45	No
1	80	10520.96	No
0	49	25051.01	No

Our table shows predictor variables of Gender, Age, and Amount Spent. The target variable is seen in the “Responded” column; which must be a binary string.

The Predictor variable is data that the user believes has caused the outcome of the target variable. After reading in the CSV file, we will select the Logistic Regression tool; as well as select the predictor and target variables.

After running the workflow, results may be viewed through the three output streams. The O output will give the likely percentage of the event happening. In this case, what is the likelihood of current customers purchasing the new product based on historical predictors and target variables? The R output offers a Logistic summary report. This includes a summary and plots of the intercept and the coefficient of the variables. The I output displays an interactive dashboard of supporting visuals that allows you to zoom, hover, and click to explore the results in more detail.

IV. CLEANSE THE ADDRESS WITHIN ALTERYX

The introduction covered the cost associated with incorrect address data. The Melissa Global Address web service macro uses Alteryx tools to access the Melissa web service. In this way, it cleanses contact data to reduce undeliverable mail.

USPS data shows over 30 percent of the data is returned because of a missing secondary address². This can be as small as missing an apartment or suite number.

TABLE II
ADDRESS FROM CSV FILE

Address1	ID	Address2	Locality	AdminArea	Postalcode	Country
718 Rugby	100650		Lancaster	SC	29720	USA
beach nive	104787		puerto la cruz			Venezuela
Fichtenwe	107047		Vechelde	niedersach	38159	DE
192 SISON	109772	SIHOME C	Vientiane			LA

From the output of the Logistic Regression tool in Alteryx, customers who are likely to purchase the product based on probability, may be filtered. The probability may also be narrowed into select categories - high, medium, and low - based on the O output from the Logistic Regression tool. After we have identified the desired customer targets, the next step is to cleanse the address data for those customers. The details of calling the web service within the macro will not be covered. The only requirement is that data connected must have a column for country and at least one column for addresses. Column names will be matched with the appropriate field in the macro.

Melissa includes an output tool to glean the results from the web service. The user may choose different outputs such as a database or write to a CSV file. Figure 1 shows the Alteryx workflow with the macro and Logistic Regression tool to identify customers and cleanse their associated addresses.

The screenshot displays the Alteryx Designer x64 interface for a workflow named 'LogisticRegressionGlobalAddress.yxmd'. The workflow consists of several tools: 'GVC_SampleFile.csv' (Input Data), 'Join' (Join), 'Filter' (Filter), 'Score' (Score), and 'Select' (Select). The 'Score' tool is configured with the formula: 'category = if ([Score_1] > .70 then 'high' elseif ([Score_1] < .7...'. The 'Select' tool is configured with the formula: '[category] != low'. The results pane at the bottom shows the workflow execution log, including the number of records read, scored, and filtered.

Tool	Message
Designer x64	Started running C:\Users\Michael\Desktop\Alteryx\Global Address\Global Address\Global Address\Global Address\GVC_SampleFile.csv at 3/24/2017 8:52:23 AM
Input Data (1)	50 records were read from "C:\Users\Michael\Desktop\Alteryx\Global Address\Global Address\GVC_SampleFile.csv"
Browse (2)	33 records
Score (4)	50 records were scored.
Browse (5)	50 records
Filter (7)	7 records were True and 43 were False
Browse (9)	7 records
Designer x64	Finished running LogisticRegressionGlobalAddress.yxmd in 9.0 seconds

There is only a small setup to perform the operation described above. Most of the work is done by the Global Address macro and the Logistic Regression tool in Alteryx. The results code from the output shows what changes have been made to each individual record. For example, the result code AC03 indicates a Locality Change, where the locality (city, municipality) name was added or changed. One record may have one or many result codes. All the results codes and their descriptions may be found in the Melissa Wiki. The user also has the ability to add and remove output columns within the macro.

V. CONCLUSION

In this example, we have reduced the customer list by 60 percent to narrow the target to the 40 percent who are most likely to make a purchase based on the predictive analysis of historical data from these customers. And, we've cleaned the customer list by adding missing suite information, correcting postal codes, and changing city or municipality names, among other operations.

Utilizing predictive analytics tools, such as Alteryx, in conjunction with Global Address Verification from Melissa, improves the effectiveness of marketing initiatives by better targeting customers most likely to buy, and cuts costs associated with undeliverable mail.

REFERENCES

1. Redman, Thomas C. "Bad Data Costs the U.S. \$3 Trillion Per Year," *The Harvard Business Review* – September 22, 2016
2. Office of Inspector General USPS, "Undeliverable as Addressed Mail" Audit Report - Report Number MS-AR-14-006, July 2014.

About Melissa

Melissa is a leader in data-driven solutions that help organizations leverage Big Data and People Data (name, address, phone and email) to unite customer insights, analytics, data quality and cross-channel marketing. We profile, cleanse, verify, enrich and consolidate data assets, providing more than 10,000 brands in over 20 countries with accurate, reliable and trusted information that can be utilized throughout the enterprise. For more than thirty years, our extended legacy in data quality, ID verification and data enhancements has earned the trust of organizations from around the world.

Request a free trial: www.melissa.com/gav

1-800-635-4772

About Alteryx

Alteryx (AYX) is a leader in self-service data analytics. Alteryx Analytics provides analysts with the unique ability to easily prep, blend and analyze all of their data using a repeatable workflow, then deploy and share analytics at scale for deeper insights in hours, not weeks. Analysts love the Alteryx Analytics platform because they can connect to and cleanse data from data warehouses, cloud applications, spreadsheets and other sources, easily join this data together, then perform analytics — predictive, statistical and spatial — using the same intuitive user interface, without writing any code.

Alteryx is a registered trademark of Alteryx, Inc. All other product and brand names may be trademarks or registered trademarks of their respective owners.