Data Cleanse and Analysis Example

The purpose of this case study is to detail typical results of our efforts to cleanse and normalise our clients' Customer Data as well as to offer insights and recommendations for marketing efforts to their existing customers.

Base Data Review

The customer data set is reviewed for informational purposes and to identify basic data issues such as no address information or out of jurisdiction customers.

In this example 1,000 records had base issues that excluded them from the data cleanse process. This was done because 200 were International customers (outside of North America) and 800 had no addresses largely because they were transferred to other business/dealers.

Base Address Counts

Jurisdiction	Count
USA	9,900
Canada	100
International	200
No Address	800
Grand Total	**11,000**

Detailed lists of these records are provided digitally.

Address Validation and Correction

Our address validation software is designed to produce results that meet with USPS and Canada Post requirements to achieve the lowest postage rates. Below is a list of the changes and issues identified.

Address Cleanse Results

Change Reason	Count	Percent
NO CHANGE	7,800	78.0%
CUSTOMER MOVE - UPDATE ADDRESS	2,080	20.8%
MANUAL REVIEW	20	0.2%
NO ADDRESS - FOREIGN MOVE	100	1.0%
Grand Total	**10,000**	**100%**

Results description by group:

- No Change – Current Records are correct and do not need to be updated.
- Customer Move – Address Update - The vast majority (often over 20%) are related to customers that have moved. We are able to identify their new addresses.
- Manual Review - Some records require a review. MLS manually corrects these records.
- No Address/Foreign Move – These records cannot be updated and typically this is due to customers moving out of jurisdiction.

Digital lists of these records are provided so the client can update their records.

Data Normalisation

Customer data is normalised. This includes updating the customer information into the same structure (name order, capitalisation, etc). Multiple spelling of the same OEM is updated. For example: BRP, Bombardier, Can-am, Brp, bombardier are changed to the BRP.

Duplicate Record Analysis

The next stage of our process is to run a series of queries to identify duplicate records. This is done for two purposes. To identify potential issues in the database and to identify multiple purchase customers. Some examples of this are as follows:

- Same Customer (Name and Address) with the same VIN
- Same Customer (Name and Address) with the same Brand (OEM)
- Multiple Customer at the same Address

This is done to avoid redundant communications to the same customer or household.

Results and Analysis

After removing undeliverable records and duplicate records, as defined above, 9,115 records remain.

OEM	Count	Percent
BOMBARDIER	3,000	32.9%
YAMAHA	2,000	21.9%
KAWASAKI	1,000	11.0%
HONDA	1,000	11.0%
SUZUKI	800	8.8%
POLARIS	600	6.6%
TRAILER	300	3.3%
KYMCO	200	2.2%
BUELL	100	1.1%
TRITON	100	1.1%
TRIUMPH	15	0.2%
TOTAL	9,115	100%

This allows specific brands to be targeted.

Below highlights the count of customer records by the age of their vehicle.

Year	Count
2003	200
2004	200
2005	200
2006	200
2007	300
2008	400
2009	400
2010	400
2011	700
2012	900
2013	1,000
2014	1,175
2015	1,450
2016	1,250
2017	340
Grand Total	**9,115**

This table details the number of units sold in each calendar year.

Year Sold	Count
2010	500
2011	800
2012	1,000
2013	1,000
2014	1,500
2015	2,050
2016	2,065
2017	200
Grand Total	**9,115**

Age of the vehicle and when it was sold are important considerations when developing the marketing retention strategy. We also determine which units were sold as new and used units.

At this point we develop a customised customer retention strategy for our clients. The focus is to bring existing customers back for additional purchases.