HOW CLUSTER ANALYSIS IMPROVED TERRITORY MAPPING FOR ONE OF THE OLDEST PERSONAL LINES INSURERS IN THE U.S.
EXL’s client is among the oldest and largest insurers in the US, with annual revenues of over $5 billion. Like most personal lines insurers, its auto and home insurance division uses geographic risk as a primary rating variable for calculating policy premium rates. This ensures that customers living in the same neighborhood pay similar insurance premiums, as they are likely to experience the same geographical risks to their automobiles and property.

Over the years, the insurer had defined rating territories that grouped together geographical areas with homogenous risk profiles. However, the definitions for these territories used subjective information, such as agent feedback or insufficient claims information, which was not very credible and did not appropriately reflect the geographic risks.

The client approached EXL to review its old territory definitions for one of the states where it operates and define new contiguous territories using a data-driven technique that captures historical risk exposures. The goal was that the new territories would display a lower risk variance within their boundaries. This would help the company offer better-tailored premiums more aligned with the actual risk profiles of their customers.

Context
EXL helps its clients solve their difficult challenges with the concept of Digital Intelligence. This approach utilizes deep domain expertise and analytics capabilities to examine large data sets in order to understand the context of the problem and how to solve it.

As the client’s support and availability towards problem solving and project execution would be minimal, EXL’s performance would be evaluated based on the thoroughness of its own approach and methodology, the diligence demonstrated in data preparation, and the accuracy of the analysis.

“The goal was that the new territories would display a lower risk variance within their boundaries. This would help the company offer better-tailored premiums more aligned with the actual risk profiles of their customers.”
To kickstart the project, the client provided EXL with historical data of underwritten insurance policies and incurred claims transactions in the state under study, and the EXL team commenced work on territory mapping. The aspects of the project are as follows:

I. Client Data Preparation:
• Integration of claims and policy data at policy level
• Identification of building blocks based on data densities
• Data summarization at building block level
• Geocoding to identify latitude and longitudes for building blocks

II. Adding Variables from External Data Sources:
• Web scraping to extract demographic data at building block level
• Leveraging industry policy and loss data
• Integration with internal data set to create master data set

III. Variable Creation and Normalization
• Binning of numerical variables using percentile distribution and creation of dummy variables
• Calculating % losses and loss amounts by claim reasons
• Multiple weighting of latitudes and longitudes
• Calculating credibility weighted pure premiums and claim frequencies using principle of locality

IV. Preliminary Data Analysis
• Extended data dictionaries using EXL proprietary univariate analysis
• Policy and loss distributions
• Data profiling and loss trend analysis

V. Variable Selection to Reduce Noise
• Correlation analysis
• Principal Component Analysis
• LASSO (Least Absolute Shrinkage and Selection Operator)

VI. Unsupervised Machine Learning and Cluster Profiling:
This involved assessing the clustering tendencies of the data, identifying the optimal number of clusters, profiling and visualizing the clusters, and comparing the results of multiple clustering algorithms.

• Partition Methods
  • K-Means
  • Partition Around Medoids
  • Clustering for Large Applications (CLARA)
• Hierarchical Clustering
  • Agglomerative Nesting
• Selecting the optimal number of clusters using Silhouette Width and Elbow Curve
• Visualization of clusters to test for contiguous territories
• Profiling of clusters for tagging of risk severity and types

The team finalized 11 groupings of variables in the auto category and nine such groupings in the home category. These groupings included variables such as incurred losses for factors such as hail, wind, fire, theft, vandalism, and accidents; as well as geographical co-ordinates, credibility weightings for premiums and claims, customer earnings, number of owned properties, and home value.

Orchestration
Having finalized the variables and the methodology to be used for the analysis, the core team orchestrated the right people and digital technologies to drive meaningful business outcomes. This included using R/RStudio and Tableau to extract demographic data using R scripts and populate the identified data fields in the master dataset. Then the team used several clustering methods to study the similarities in the characteristics of the data sets.

Outcomes
By the end of the project EXL had identified 18 clusters for the home insurance category, and 15 clusters for the auto insurance category. These clusters were used to define distinct new contiguous territories, which when compared to the existing territory definitions had a lower risk variance within their boundaries.

The reduction of the total variance observed within the new territory definitions was to the tune of 45% in the home insurance category, and 30% in the auto insurance category. Further the clustering analysis helped identify four high catastrophic risk zones for home insurance, and two such zones for the auto insurance.

The new territory definitions were presented to the client for optimal allocation of sales agents, and for easy segregation of high risk and low risk zones. The incorporation of demographic risks in the territory definitions gives the client a far clearer picture on premium ratemaking, and the reusable processes and tools that EXL has created can be used to replicate the approach used in Illinois across other states as well.
How Cluster Analysis Improved Territory Mapping for One of the Oldest Personal Lines Insurers in the U.S.

Client Challenge
- Outdated territory mapping information on geographic risks across Illinois
- Existing data not credible and based on subjective inputs
- Current premium ratemaking did not factor demographic information

Context
- External datasets and publicly available information identified for inclusion of demographic variables in the master data
- Structured methodologies used to identify and group the right number of variables to reduce noise and generate better results

Orchestration
- R/RStudio and Tableau used for extraction of demographic data, data analysis and machine learning
- Multiple clustering methods to analyze the similarities in the master data characteristics
- Clustering tendencies of data studied, and an optimal number of clusters identified, profiled, and visualized
- Results of multiple clustering algorithms compared

Outcomes
- 18 clusters for home insurance and 15 clusters for auto insurance identified
- Analyzed clusters used to define distinct new contiguous territories
- Reduction of the total variance observed within the new territory definitions was to the tune of 45% in the home insurance category, and 30% in the auto insurance category
- Multiple high catastrophic risk zones identified
- Reusable processes and tools enable replication of solution approach across other states

To find out more about how EXL uses Digital Intelligence to help companies transform, visit www.exlservice.com.