ABSTRACT

Company Background - Leimberg Information Services, or LISI, is a niche data collection company that creates competitive advantage. Data collection can be very costly if done manually. CLE (Law) and CFP (Finance) credits are needed to filter on specific characteristics for matches LISI does not currently have.

The Team was able to quickly gather insights from more data collection following the ERD model and survey results - is needed to filter on specific characteristics for matches LISI does not currently have.

The foundation of LISI Listens is existing content, providing a starting point for LISI’s future data consisting of three main tables containing unique IDs for appropriate table joins, and four sub-tables containing more detailed information regarding more granular customer information. All tables in this recommended ERD have IDs and keys that assist in proper joins to allow for ease of reporting and analytics.

METHODS AND MATERIALS

For further analysis, we identified more methods and materials to aid LISI in better data collection and data analysis.

1) LucidCharts and Excel – Data Dictionaries and Entity Relationship Diagram (ERD)
2) Tableau/Table Prep and Excel - Data Cleaning and Visualization
3) ChatGPT – Google Sheets and Excel Extension for Text Generation
4) Python – Unsupervised K-means Clustering Algorithm with Term Frequency Inverse Document Frequency (TF-IDF) Vectorizer for Industry Analysis

RESULTS

Data Modeling and Design - Since data collection can be very costly if done wrong, we summarized all current data into data dictionaries so that we can note major flaws such as excessive redundancy, poor usability, and a lack of integrity. To help mitigate, we created new dictionaries and an ERD of potential future-state datasets containing a richer combination of data elements. The ERD below shows a starting point for LISI’s future data consisting of three main tables containing unique IDs for appropriate table joins, and four sub-tables containing more detailed information regarding more granular customer information. All tables in this recommended ERD have IDs and keys that assist in proper joins to allow for ease of reporting and analytics.

Industry Analysis - LISI appeals to other industries complementary to estate and tax law which can be areas of growth. However, only company name is available which can be difficult to analyze. Therefore, categories needed to be created to be able to assign each customer to an industry based on the company name. After generating text descriptions of the company names with ChatGPT, we carried out NLP methods of preprocessing the text, running it through a TF-IDF vectorizer, and feeding the text into a K-means clustering algorithm, k=10. We utilized the vectorizer to extract features from the clusters of the top 10 terms prevalent to help name the clusters and see the industry groupings. Sub industries were also identified from the clusters to note specializations of the customers within their industry. This information and machine learning process will allow LISI to continue to analyze customer characteristics and build other products and services as they continue to grow and collect more data.

DISCUSSION & CONCLUSIONS

Revise Current Data Structure for Better Data Governance - Data is representative, but we are missing the data quality. Without strong data governance, LISI puts themselves at risk for privacy breaches, data misuse, higher data management costs, poor data quality, and, ultimately, a loss in competitive advantage.

Use Customer Surveys to Reveal Future Growth Opportunities - The attitude of how people react and interact with products/services, as well as those who choose to do so, can very effectively reveal future growth opportunities. Use surveys to assess current operating performance and plan for future improvement.

Better Customer Email Communications - To avoid displeasing customers, LISI should employ the email delivery and subscription system, particularly for customers who do not opt for any subscription options and receive all emails. Additionally, it is important to format and accurately record useful client information during data collection, based on our future data recommendations, which can facilitate efficient response to client requests.

Expand The Institute - Prioritize program materials in areas generating the most interest based on customer dataset analysis and industry analysis. Consider offering certificates or credentials globally to individuals interested in the LISI Institute to expand the company’s reach. Provide professional continuing education credits such as CLE (Law) and CFP (Finance) credits to retain existing clients and potentially attract those who have discontinued their participation in the program.

LISI Network – Internal Network of LISI Professionals - New contact sharing service for individual and company customers to request or be matched to freelance work, job positions, expert witness testimony, and more with LISI answering inquiries with applicable candidates. The ERD model and data governance and collection is needed to filter on specific characteristics for matches LISI does not currently have such as industry, degree, and certificates. Methods and materials from the industry analysis will help maintain updated information for this service that will provide another revenue stream for LISI.