Leimberg
Information Services, Inc.

Carnegie Mellon University
Tepper School of Business

LEIMBERG INFORMATION SERVICES
CARNEGIE MELLON UNIVERSITY TEPPER SCHOOL OF BUSINESS
MSBA CAPSTONE PROJECT FINAL REPORT

ALICIA CARTER, ERIN MURPHY, DEANNA SWINERTON,
TIFFANY WANG, DANIEL WEISS

Table of Contents

Background - Problem Statement

Leimberg Information Services, or LISI, is a niche content provider for Estate and Tax attorneys, CPAs, and financial professionals. LISI offers newsletters, webinars, podcasts, courses, and other content to its customers who range from small business tax attorneys to the IRS. LISI is under new ownership and is growing rapidly. They want to reorganize their operational and strategic model to provide space for sustainable and traceable growth.

LISI's main problem is that its previous owners ran the company as though it were a hobby. LISI mostly ran itself in that professionals would provide content to Leimberg, Leimberg would proofread and edit newsletters as needed, and they would then email the newsletters out to a large email list. Under their original ownership, LISI did not attempt to commercialize, and was run under a suboptimal management style. This model worked for what Leimberg wanted to do - provide time-sensitive, important information to subscribers. But for profitable and sustainable growth, this is not the appropriate operational model LISI's new owners can continue with. LISI's new owners strive to grow the company and monetize the information and data they have, but the framework for that growth and tracking was never put into place. They also want to expand into other untapped industries and learn about what areas they can grow and improve upon within and outside of LISI. They need help restructuring their operational model to ensure that they are tracking their products' usage, maintaining their products to the standards that their customers expect, and keeping up with what is going on in their industry and customer base. LISI's needs are very wide-ranging. Below we will outline the work we have done for LISI in the past four months and our overall recommendations for the company.

Three Solutions for Growth

Our solution focuses on three distinct avenues for growth. First, LISI should launch a podcast that is built on its extensive archive of recordings but grows as new content emerges. This mobile-friendly approach will reach a younger audience as well as enhance the experience of existing subscribers. Second, LISI needs to accelerate its newly created product: The Institute. With at most a handful of users despite compelling learning content, the path to growth is all about marketing, and perhaps partnerships with various accreditation organizations. Third, we propose the development of a networking tool that makes the most of its subscribers' expertise and LISI's customer information. This will be a completely new revenue stream for LISI.

Exploratory Analysis Shows Initial Customer Characteristics

One of the first steps taken to begin analysis for these solutions was to understand the current state of the variables and what was contained in them. To do this, we created data dictionaries to document the current state of the three data sets provided: Newsletter Data Export, CUSTOMDLOGB01&02, and Contact Export (Appendix 1).

Contact Export

The contact export dataset includes LISI's entire email list of applicable professionals that continues to grow. These customers may be newsletter subscribers or not; therefore, this is another area of the business that can be leveraged for continued growth. Those who receive emails but are not subscribers could potentially convert to paying customers, while those who are subscribers can receive a more enriched experience with more information being sent to them. However, this feature is not being efficiently used. The email cadence can be more structured and purposeful to show LISI has a benefit rather than a burden. This dataset currently contains over 80,000 records with over 50% who have opted out. The rationales for customers opting out of receiving data communications have been classified into distinct categories to facilitate subsequent analysis. Upon examining the contact export dataset, it becomes evident that a significant proportion of opt-out rationales, approximately 73%, are unsubstantial, while negative feedback regarding excessive email volumes accounts for around 26.5% (Appendix 2, Exhibit 9). Thus,

there is a need to enhance both the email delivery mechanism and subscription system, as customers who do not choose any subscription options receive all emails, which triggers their displeasure. While pursuing new client acquisition avenues, it is imperative to update this current business component.

Lister Export

The Lister Export within the Newsletter Data Export dataset provided was identified to be the most helpful with exploring components of LISI's subscription customers. These customers subscribe to their newsletters and make up the largest portion of their customers. Therefore, they contributed to the analysis to understand customer characteristics to be able to acquire more. This data initially had 7,992 rows with 16 variables (Exhibit 2, Appendix 1) and data cleaning was carried out using Tableau Prep software and Microsoft Excel. Data cleaning steps involved coding the state and country variables to be all spelled out since they were mixed with some abbreviations along with removing records that were repetitive, had missing values, and out of scope. After the data was in a proper state, Tableau was used to gain further insights.

Growth - LISI has experienced a spike in new subscribers over the last few years. However, growth has slowed in 2022 as fewer people started a subscription in 2022 than in the past few years before. Specifically, fewer customers joined in 2022 than in the previous year, 2021. Overall, we believe our solutions will help increase the growth to its 2021 level and remain steadily increasing. The graph displays this by showing the amount of customers gained in each year based on subscription start data. This does not include customers who have canceled their subscriptions or the net number of customers.
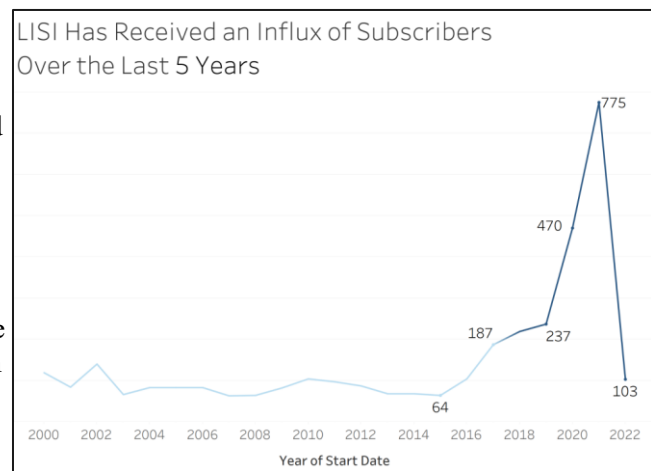
Subscriber Location - After the cleaning stages, 7,785 unique subscribers were identified based on subscriber email found all around the United States. This shows the wide range of reach LISI has and its ability to attract customers from all regions of the United States. The top five states are displayed below and there is a strong foundation of presence of LISI. There are many people in the large cities of these states and in between throughout the country where the current company presence can further grow.
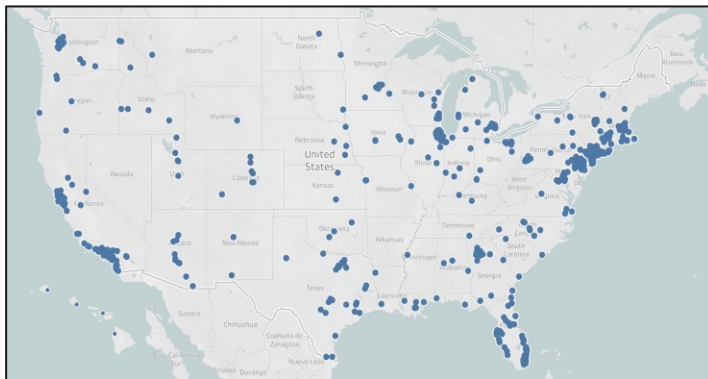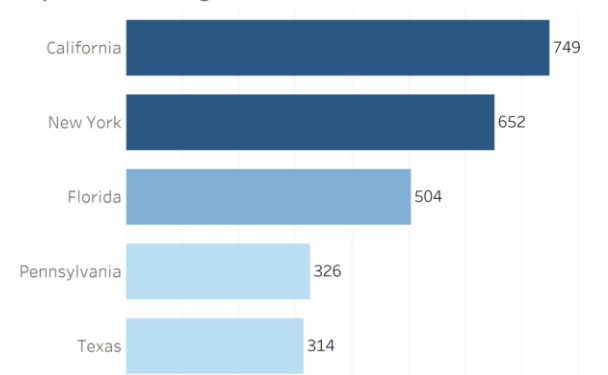
CUSTOMDLOGB01 & CUSTOMDLOGB02
These datasets contain information on transactions of LISI's other products and services such as webinars, ebooks, and finance/tax tools. Data cleaning for the CUSTOMLOG datasets included combining 01 and 02 into one and removing duplicates. With this, all the items contained in both datasets can be included together in the analysis. Variables that went through more specific cleaning steps included To Email Address and From Email Address. Some records in these fields contain company-specific emails and transactions for business purposes of distributing funds, making payments, or issuing refunds. These transactions were out of scope and deleted. After this data was in a proper state, Tableau was used to gain further insights.

Top Items - There are over 2,000 distinct items offered since LISI publishes new content often but also keeps past items, such as webinars, as archives to continue being bought. The LISI Tool Access is the next top item after the newsletter when all the other services were analyzed, and many webinars and books also place in the top 10 items (Appendix 2, Exhibit 1). These items already have high popularity and prove to be desirable to LISI customers. Therefore, these items can be leveraged for content to potentially provide for The Institute and Podcasts to attract others.

Transaction Location - Location analysis on the transactions also shows LISI has a large international presence for their other products which could provide another area of growth (Appendix 2, Exhibit 2). Specifically, it would be worth exploring translating the content to other languages, starting with Spanish because it is spoken in Panama, Mexico, and Colombia–top countries after the United States (Appendix 2, Exhibit 2).

Subscriber Industry Analysis Provides More Customer Insights

Another variable that can help identify other characteristics of customers is the industry they work in. LISI has a wide presence of complementary industries of law, finance, banking, and more which work within the estate and tax planning space. However, LISI does not currently record the industry its customers work in or have a large amount of other variables to help identify other characteristics of subscribers other than the company name found in the lister export. Therefore, to begin this analysis, industry variables needed to be created to be able to categorize each customer into an industry based on the company name. A subcategory field should also be created in order to provide more details on their specialty within the industry. LISI is aware they have a large following among estate law, tax law, and finance professionals, but, as the company continues to grow, there may be other industries or sub industries that become attracted to their products and services they are not aware of. Knowing the industries can help LISI continue to expand into the industries most represented but also into industries that begin to have a growing presence. This analysis will be helpful in implementing the solutions stated earlier and can be applied to other data sets, such as the customer export, to gain more insight on all other customers.

Manual Industry Selection
We first looked at the company names and manually assigned industries to customers by filtering on specific words in the company name or email. If the company name was blank, the tail end of the email address was then analyzed. Keywords such as law, legal, financial, bureau, trust, wealth, and tax were first searched in the company name to create general groupings of customers. In the email, keywords such as CPA, law, tax, and more were identified to add those with blank company names to the groupings. This process was used to categorize 4,506 records with 3,442 remaining with Law ranking the top industry followed by CPAs, Financial Services Firms, and Wealth Management Firms (Appendix 2, Exhibit 3).

Industry was next expanded with Industry Details which can provide a more specific subcategory within the industry where customers specialize. This was also done for the 4,506 records where there were large

differences within the industry showing different specializations. For example, within the financial services industry, there are customers who are financial planners or consultants while those who are CPAs are individuals, retired individuals, or work for a company. The banking industry is also broken into investment firms, trust companies, and other general banking companies. This expanded view allows more detail to be added about the different specializations within the industry LISI customers have (Appendix 2, Exhibit 3).

This manual process allowed us to gain a better understanding of the data, where the customers come from, and choose appropriate industry tags. However, with many records remaining and the goal of growth in mind, this process can become time-consuming and difficult. The remaining records also did not have any clear keywords to identify for categorization so at first glance it was not evident what industry they worked in. The variable limitation in the dataset made this task more difficult. To be able to deepen this analysis and provide LISI with a process they could use again and apply to other datasets, we explored other data avenues and models.

ChatGPT
In order to continue to analyze and aid in a more efficient process, the Open AI platform ChatGPT was used to generate text descriptions of the companies customers work for to provide more data to be used in the industry analysis. "ChatGPT is an artificial intelligence (AI) chatbot developed by OpenAI and launched in November 2022" (Use chatgpt in google, n.d.). It combines an AI model, GPT-3, and dialogue conversation style in an answer and response chat format. It was created using reinforcement learning from human feedback and optimization methods to optimize responses (OpenAI, 2022). It has many uses to aid in productivity, data analysis, and research across all industries. However, limitations exist since it was trained on data up to October 2021. It can only return information it was given up until that time so the outputs can only be as good as its inputs. Therefore, ChatGPT can sometimes write plausible sounding but incorrect answers or the same answer to different prompts if they are not phrased differently enough (OpenAI, 2022). These, along with the common limitations of AI models like bias and inaccuracy, are also present. Though limitations exist, ChatGPT is continually being improved, has many applications, and can be a great tool for simple data or business tasks for companies looking to grow.

ChatGPT can be used through the OpenAI website or directly in Google Sheets and Google Docs to aid in many data analysis tasks such as data cleaning, generating tags, and generating text descriptions. Business tasks of generating email communications, summarizing meetings or notes, and translation can also become more efficient with the use of ChatGPT (Use chatgpt in google, n.d.). To use this tool, we created an OpenAI account and followed Stanislas Marion's steps to install the "ChatGPT in Google Sheets and Docs" app and connect to the account through an API key (Marion, 2023). The API key is a unique id that is specific to the ChatGPT account. This tool can also be used as an extension in Microsoft Excel in a similar way. This process was tested out in Excel following Antonio Nakić-Alfirević's similar steps of installing the extension and connecting to an OpenAI account with an API key (Antonio Nakić-Alfirević, 2023). The industry analysis was further carried out with these extensions where Google Sheets proved to be slightly more efficient. The following equation referencing the company name in column/cell A2, for example, follows the same format for the formulas used to integrate ChatGPT.

=GPT(CONCAT("write a one sentence company industry description about", A2))

The comparable excel formula would be

=GPT.Prompt(CONCAT("write a one sentence company industry description about", A2,0.5,0))

The following are some output examples from these formulas:

	A	B	C	D	E
1	Company Name	GPT Industry Description	Customer Email	Manually Assigned Industry	Industry picked with GPT description
2	Wells Fargo	Wells Fargo is a financial services company that provides banking, mortgage, investment, and insurance services to individuals and businesses.	msidoresq@gmail.com	Banking	Banking
3	Conner & Associates, PC	Conner & Associates, PC is a professional services firm that specializes in providing tax, accounting, and consulting services to individuals and businesses in various industries.	kconner@connercpa.com	CPA	Financial Services
4	DON GARY PLLC	DON GARY PLLC is a law firm that specializes in providing legal services to clients in the areas of business law, real estate law, and estate planning.	don@dgarypllc.com	N/A	Law

The first two records - at first glance - are clearer in the industry this customer may work in from being a more well known company (Wells Fargo) or having a key word in the email (cpa). The GPT description is in line with what was manually selected for the first two while helping identify the third one which was initially unclear. However, to continue the assessment for all records we would have to read the descriptions to manually assign. The ChatGPT extension also has the ability to generate a category or tag for a cell of text from pre-specified labels with the GPT_Tag in sheets or GPT.Tag in excel formulas. The remaining categorization therefore could be completed this way in a short amount of time if needed and if the text is known to be accurate. However, a more automated approach with machine learning was used next to further aid in this analysis. This approach can also be used to analyze any future customer-related variables for which LISI chooses to collect data.

Text Clustering with K-Means Algorithm
K-means Clustering is an unsupervised machine learning algorithm which is used to label unlabeled data. It has various uses for many types of data, but we decided to use it to assist the industry analysis, specifically clustering text inputs from the company description. Since company description is currently not available in LISI's dataset, we used the ChatGPT text generated from the prompts stated earlier in Google Sheets and Excel. Once we had the text data, we were able to use Python and begin modeling. This is a process that can be followed in the future once the descriptions are collected more accurately or can be followed to model any other variables LISI wants to analyze.

Clustering was carried out in Python by first preprocessing the text. Text preprocessing is necessary to remove punctuation and special characters and convert all letters to lowercase so it can be fed into the model and recognized more clearly with less noise. These types of preprocessing steps were taken but others include stemming, removing stop words, and expanding contractions. The ChatGPT descriptions were preprocessed and formed into a list referred to as documents. This is a very common process called natural language processing (NLP) which has many model applications.

For this analysis, we applied NLP more to clustering with a Term Frequency Inverse Document Frequency (TF-IDF) method for transforming text into model inputs. This vectorizer transforms the text into numerical vectors with weights based on the frequency of the words so the algorithm can fit it for prediction. Other text vectorizers that can be used for NLP models include bag of words, n-gram, word embedding, word2vec, and BERT. TF-IDF was chosen based on its ease of use and accuracy without becoming overly complicated and time-consuming. However, it does not take into consideration the meaning and context of the words which may be more beneficial for other tasks. In this analysis, the overall occurrence of words in the description is what we needed to help categorize the companies and customers into industries, so TF-IDF was applicable for this task. Other vectorizing methods could also be explored to further build on this algorithm with future data.

Once the text was preprocessed, the list of documents was run through the TF-IDF vectorizer with an addition in the code to remove stopwords at this step. The documents were used as the input variable X into the K-means clustering model. Overall, this clustering model takes the input X to produce Y outputs of k groupings, Yk. Each output (Y1, Y2, …,Yk) represents one cluster and each record, in this case company, is assigned to a cluster. This is beneficial because it can take many inputs and group them all at once to identify potential industries to assign. For this analysis, we fit the clustering model on the first set of data where we manually assigned from earlier and then used it to predict the cluster for the remaining records still needing an industry assignment. However, predetermined labels are not needed for this type of model since it is unsupervised and could run over all records at once. We wanted to see how the algorithm grouped the first set of companies based on company description then used those grouping on the remaining records. We were also able to output the top terms from the clusters but utilizing the vectorizer established earlier and extracting feature information. We printed a list of the clusters and the top 10 terms within them to be able to name the clusters of what industry they represent (Appendix 2, Exhibit 7). This is how we were able to see the groupings and compare them between each other more easily rather than looking at the cluster number.

We then compared the model clusters to the manual categories and chose a final set of industries and updated the original visuals. The clusters ended up grouping more categories together we manually separated such as some Banking and Finance terms being grouped together more in cluster 7 and Wealth Management and Investment Services being grouped more together in cluster 1. However, this provides more insight into the possible subcategories of specialties within the industry. For example, we manually created the financial services industry to hold other specializations of wealth management, tax services, and CPAs. So, even though there are multiple clusters for financial services, this shows that there is a distinction among the financial service companies with different specializations within the cluster which can help identify sub industries among customers as well. Overall, this analysis strengthened our outcomes from our manual process but provided a scalable process with machine learning. The exhibit on the right shows the 10 clusters and their size based on the number of companies grouped.

Companies in Each Cluster

Cluster Names	Companies
Law - Estate	922
Banking & Finance	698
Financial Services - Wealth & Invest	581
Law - Corporate	489
Law - Family	402
Financial Services - Tax & Accounting	374
Banking - Trust	171
Other	138
Insurance	129
Government	46
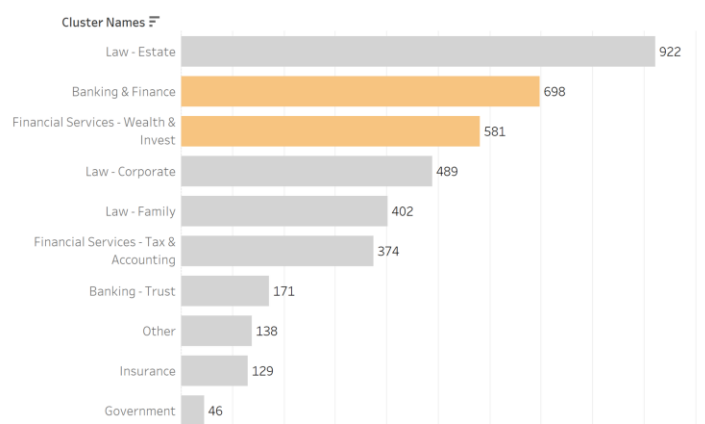
Updating the Manual Industry Assignments

After receiving industry categorizations for all records and combining them with the information from the manual analysis, the visualizations were run again to view more succinct and clean categories (Appendix 2, Exhibit 5 and 6). Any customers whose company names were blank or ChatGPT could not return a description for were removed. From the graph above, we still see law being the largest cluster and top industry. After clustering we also see that there is a stronger presence of Banking and Finance as a secondary top industry. When expanded out, the clustering analysis grouped investment and wealth management services together in one cluster based on the description, so this subcategory under financial services was updated and the manual ones related to these areas were added (Appendix 2, Exhibit 6). Some industries with a smaller presence include insurance, government, nonprofits, and more. These industries could be an opportunity to attract additional customers through establishing more content for financial services that is not already available such as planning, consulting, and investing. Overall, these insights from the industry analysis can guide LISI into other areas for growth and provide a process for analysis that can be continued.

Law Industry Analysis

As seen above, the law industry is the most common characteristic among LISI customers. Since LISI has a large following of law professionals, it would be beneficial to understand what type of law they specialize in to better understand their customers. The Google Sheets GPT formula was also used to describe the type of law to create a law industry detail text variable based on the company name and then a shortened version for the overall industry detail (Appendix 2, Exhibit 8). If ChatGPT could not find specific information due to a generic law description of "full service" or "various industries" listed then it would widely return "Corporate Law, Tax Law, Intellectual Property Law " as a default. Therefore, with better quality descriptions a clustering model could also be run separately on the law industry professionals to analyze the different types of law specialization that are represented among LISI customers. We were able, though, to use Python to preprocess the law description text and program more visualizations to help understand the most popular law fields from the ChatGPT-generated law descriptions.

All Law Descriptions - We see above the typical law specialties expected such as corporate, tax, and estate planning. Other areas displayed that may show an increase in industries LISI attracts is family, real estate, and intellectual property law. We also saw all of these terms listed in the top terms of some law clusters of "Law" and "Law - Family". Among all of these, the words ``variety" and "wide range" are also prominent which further shows LISI's ability to appeal to law firms that specialize in various types of law.

Short Law Descriptions - The law short variable created can also be visualized in this way (Appendix 2, Exhibit 9). Since "Corporate Law, Tax Law, Intellectual Property Law" was widely used by ChatGPT and estate law and planner were also very prominent, these terms were excluded from the next word cloud to

see the other specific law specialties more clearly. It is with this shortened version of the law description we begin to see smaller law specialties represented such as immigration, criminal, employee, injury, and healthcare. If the description from GPT is accurate, these could potentially become future areas of law LISI can create content to expand into.

Industry Analysis Conclusions
ChatGPT is a new and very powerful AI tool that can be used in many capacities by business but could interfere with potential model accuracy if used on such a large scale since it can only access data up to October 2021. For example, many companies are given an industry description of software companies or technology/IT which seems slightly off. When some of these records were searched with Google search engine, they had other results of law offices or financial services firms. Therefore, some records assigned with ChatGPT may require more due diligence or be updated with future versions of the tool continually being worked on by OpenAI. It is recommended this tool be primarily used to help fill in the gaps of missing data fields in the future as LISI builds better data governance methods to aid growth.

Overall, this analysis was useful to categorize the remaining records more efficiently and update the visualizations with more data while also having a second check on the manual assignments. The ChatGPT extension in Google Sheets was specifically used to generate more data to analyze based on the company name of the newsletter subscriber customers. The extension can also be applied in Microsoft Excel but carrying it out to Google Sheets allowed the formula to load faster and have fewer errors. The Google Sheets files were eventually converted to Excel files to upload into python.

These industry categorizations chosen are a starting point and can be updated with categories LISI deems more applicable. Therefore, this process can be used in the future once the company descriptions and more data on customers and their characteristics are collected. Ideally, the customer would select their industry, describe their specialty, and type in a description for their company when signing up for a subscription or emails so this information is captured. Or, LISI can send out emails encouraging people to update their account information and input this information to build the database with this new information. This information will become instrumental in the growth of LISI and can be applied to subscriber data and the customer export data of the full email list. More data allows for the company to carry out further customer analysis and track trends in customer characteristics that they see on the rise. Industries, and other characteristics, can be analyzed quarterly or annually, for example, to identify any shifts in the customers they are attracting. Then, clustering as described can be utilized to update industry categorizations and any other variable groupings as LISI continues to grow.

Applying Analysis

LISI Listens - Podcasts
LISI's direction for podcasts is clear: convert old webinar files to audio form and continue the process as new webinars emerge (Appendix 5). There is also an opportunity to experiment with shorter-form content, but we believe that the type of person who is interested in such detailed content is not looking for snippets. Plus, younger audiences are more likely to be comfortable with podcasts as a form of media consumption, and this demographic is the business justification for podcast creation (as opposed to a simple value-add for existing customers).

The foundation of LISI Listens is existing content, but we expect that industry experts will continue to eagerly contribute to LISI's knowledge base. Also, we expect the podcast to be highly differentiated from any competitors due to the depth of coverage. As such, trying to extrapolate general podcast trends to LISI Listens seems unhelpful and perhaps even misleading. Nevertheless, it will be important to monitor common podcast metrics and fine-tune as needed over time.

The Institute

By conducting an analysis of the datasets provided, we are able to examine their interests and preferences based on their responses. As the program is still in its early stages, we can prioritize the provision of materials in areas that generate the most interest, such as Tax Planning, Law, and Financial Planning. This approach enables us to improve the program environment and expand its reach. With the LISI Institute being available online, there are opportunities to offer certificates to individuals globally, providing significant benefits to the company. We have examined customer distribution locations across the United States and overseas (Appendix 2, Exhibit 2 and 10) and the opt out reasons which provided an expressed desire to obtain CLE/CE credits (Appendix 2, Exhibit 9). Therefore, providing the necessary credentials or certificates that our clients require starting in the top states represented among customers may be advantageous in retaining existing clients and potentially attracting those who have previously discontinued their participation in the program.

The industry analysis also offers more insight into what to offer for The Institute to attract more customers. The financial service industry is the next most represented field among LISI customers (Appendix 2, Exhibit 5). Therefore, CFP credits may be another type of continuing education course to offer to grow this product.

LISI Network - Internal Contact Sharing and Network Service

Our initial solution to aid in new customer acquisition by partnering with established expert networks has evolved into a more personal and controlled approach. We now believe LISI can connect to their customers more by offering a new internal network service through their website. The industry analysis showed that LISI serves many complementary roles to estate and tax law that can be leveraged as another avenue of growth. Sometimes a law professional needs a CPA or a CPA needs a law professional. Since LISI customers are also groups/companies, this service would be used by companies to send job openings or request freelance and temporary workers. It can also be utilized by law professionals to find expert witnesses for court testimonies. On an individual level, customers can include their skills and certifications in their accounts if they are looking for a job or freelance work. This new LISI Network can be offered along with the subscription, be an add on to the subscription with a monthly fee, or be a one-time fee. For non-subscribers, this can still be a beneficial service that they could pay for as well and marketed to the entire email base LISI has from the customer export dataset.

LISI would be the bridge between customers and companies and receive inquiries about contact information of people or positions that would match the request. To be able to do this, the current data structures outlined in the data dictionaries should be updated. Appendix 3 displays a more ideal data structure LISI can follow in the future that will help execute this new service. For example, the Future Datasets - All Contacts (Appendix 3, Exhibit 2) can first be filtered for those who opt into LISI Network and pay for it. Then, LISI can respond to inquiries with a list of applicable candidates based on further filters using the other future data sets (Appendix 3, Exhibit 3 and 4). This is where the industry analysis will be beneficial in action since Industry and Industry Detail in this data set would be best to filter by when trying to provide specific candidates for the internal network inquiries. Other fields we suggest to include in the datasets to also aid with this service for inquiry filtering are Certification and Degree fields. More details on the future data structures will be provided in the following section.

Current Data Dictionaries

The Team was able to quickly gather insights for basic analysis but had to generate text data to aid in a deeper analysis. If LISI wants to truly connect with their subscribers, LISI needs to emphasize better data collection. The following are some apparent flaws with the current datasets that are worth noting.

- *Minimal data controls mean maximum back-end work.*
  There are too many freeform fields where customers can enter any information in any format. Analysts would have to review, interpret, and address each case individually. This may not be too big of an issue for smaller databases but is indeed a major issue for larger databases. This is something that LISI should prepare for today if their goal is to expand.

- *Data redundancy works by costing more while providing less.*
  The datasets contain duplicate and unnecessary variables and many values among records are inconsistent such as differing address formats, missing item IDs, or a majority of the field being blank. Database storage can be very costly, so developers and designers always seek to eliminate redundancy by saving room for a greater variety of information. Different information is the key because it provides different perspectives and ultimately greater business decisions.

- *The current data fields are not enough for meaningful analytics.*
  Many fields in the current data are unclear and do not hold additional information to be analyzed. The data is also not connected by unique ids which prevents variables in other datasets to be used in another deeper analysis. As LISI thinks about expansion, the company should also think about what additional wisdom would be useful to the decision-making process. Example, the customer's year of birth. This information could be used to understand whether the current marketing strategy is reaching the intended customer segmentation.

Future Datasets Can Be Structured with ERD Models
Many of the data structure issues can be eliminated with a proper data model and configuring Entity Relationship Diagrams. The datasets, and associated data dictionaries, in their current states can be viewed in Appendix 1. As discussed above, there are no unique keys or identifiers that allow for connecting datasets together to gain helpful insights. To fix some of these issues stated, potential future state data sets were created as well as an ERD to assist in visualizing the connections between future datasets (Appendix 3).

The potential future datasets we are proposing consist of three main tables and four associated detail tables. The three main tables are as follows:

- *All Contacts* - This dataset contains contact information for all contacts LISI has in their possession, whether a subscriber or not
- *Subscribers* - This dataset contains more detailed information for all customers that are subscribed or have been subscribed at some point in time.
- *Purchase Details* - This dataset contains all purchasing information for all LISI products.

The four associated detail tables are detailed information from the main three tables in a more organized fashion, to ease reporting off of them:

- *Company Details* - This dataset contains all company details for companies that LISI contacts work for. This dataset can be created either through collecting from contacts at time of subscription, or through web scraping.
- *Industry Details* - This dataset contains more detailed information about the customer's industry. This dataset can be created either through collecting from contacts at the time of subscription, or through web scraping.
- *Email Details* - This dataset contains all emails a LISI customer is subscribed to by customer ID.

● *Interest Details* - This dataset contains details regarding the customer's interests.

This proposed future state may not be exhaustive, but we believe it is a good starting point to point LISI in the right direction in terms of what data we believe LISI should be capturing and the most appropriate ways to store that data.

Overall Recommendations for Growth

Prior to starting this capstone, we had an initial hypothesis which claimed that each component of our proposed solution (defining expert networks, building LISI Listens, and shining light on the Institute) would increase customer engagement. We hoped that this plan would, subsequently, help LISI gain more subscribers. Although these ideas could not yet be physically implemented, we still believe that the information gained through our work will help LISI implement these new products and services and achieve smooth growth.

1. **Begin having conversations about Data Governance.**
   Data governance is how a company manages the availability, usability, integrity, and security of their data. Without strong data governance, companies put themselves at risk for privacy breaches, data misuse, higher data management costs, poor data quality, and ultimately, a loss in having a competitive advantage. Additionally, data is representative of credible decision-making. Data governance is the bridge between intuition-based decisions and data-based decisions.

2. **Perform a mass review of all current datasets.**
   Considering the flaws with the current datasets, we believe that LISI is vulnerable to the repercussions from poor data governance. To mitigate, we first suggest a regrouping of the data, similar to how we composed the attached data dictionaries. The majority of the data was separated into two dictionaries: 1) Subscriber Information and 2) Contact Information. Each of these was designed to control data noise by facilitating clean data inputs. As a result, the tables going forward would only store data from which we believe that LISI would most benefit. Such inputs include a combination of basic information (i.e., customer name, email address, subscription start date) and new/insightful information (i.e., industry, industry detail, reasons for opting out) that can be used for advanced analytics.

   By scrubbing the current data tables for unusable information, these refined data dictionaries will serve as a single summary of all captured information in terms of meaning, relevance, and quality. Paired with stricter data controls (i.e., dropdown selections, format controls), data collection going forward will be more credible.

3. **Conduct a customer/contributor survey to reveal future growth opportunities.**
   Administering surveys is an underestimated data collection strategy. Although it can be slow, the insights gained can be very valuable because they provide a snapshot of current operating performance. Specifically, the attitude of how people react and interact with the products/services, as well as those who choose to do so, can very quickly reveal future growth opportunities. Here, it is also important to note that the data comes from the individuals who directly impact your goals so their feedback is influential to the longevity of LISI.

   The survey questions that the Team suggests are included in Appendix 4. The questions were carefully chosen to not only confirm that current LISI operations have been fruitful, but also act as promotional opportunities for some upcoming projects (i.e., content in other languages, more audio content, contact sharing for an internal network, the institute).

4. **Consider Distributing Certificates and Licenses Through The Institute**
   To gain momentum, LISI should consider a distribution of certificates and/or licenses through the Institute. This can be done with either an accredited route (which offers four levels of accreditation: programmatic, institutional, regional, and national) or a non-accredited route (Learn.org, n.d.).

   Because of the type of information distributed, certificates most likely need to be accredited where courses must meet specified standards around academic quality, ethics, and experience which are set by, for example, the American Bar Association (Learn.org, n.d). This route can be costly and time-consuming, but, because accredited certificates are more trusted and recognized, more people are willing to pay. There is also the option of non-accredited certificates. Non-accredited certificates are less expensive and carry less liability for education quality. On the downside, because of the lack of standards, employers and customers tend to put less weight on the non-accredited certificates. On the plus side, having some type of certificate and/or license offered for courses taken is an attractive incentive to customers as opposed to offering nothing.

   Based on our analysis of current data sets, we recommend researching licenses to give accreditations in California, New York, and Florida, as those are the three states that capture the majority of LISI's customer base as of now. We also recommend exploring continuing education credits for other industries, such as the CFP for financial service professionals. Ultimately, a customer survey and further analysis will aid in collecting customer preferences for this product.

5. **Introduce a LISI Network**
   After capturing more data such as data outlined in Appendix 3, the LISI Network described earlier can be implemented. This new internal network service can be offered along with newsletter subscriptions, be an add-on to subscriptions with a monthly fee, or be a one-time fee. For non-subscribers, this can be a paid service marketed to the entire email base. Processes and tools from the industry analysis outlined can further aid in maintaining updated industry and industry detail for this service and also be applied to analyzing other information in the future datasets such as degrees, birthdates, and other demographic and psychographic information. Implementing an internal network such as this will create another stream of revenue and also increase user traffic on the new and improved LISI website. With the increased traffic, LISI can begin to utilize tools such as Google Analytics for further data analysis.

6. **Utilize ChatGPT to Aid in Data Analysis and Business Operations**
   ChatGPT is a new and very powerful AI tool that can be used in many capacities by business but could interfere with data collection accuracy if used on such a large scale since it can only access data up to October 2021. Though limitations exist, ChatGPT is continually being worked on to update and has many applications and can be a great tool for simple data or business tasks for LISI as they grow. It is recommended this tool be used more as it continues to be updated but for now primarily be used to help fill in the gaps of missing data fields as LISI builds better data governance methods. ChatGPT can also be used as a business operations tool to summarize text, translate text, and create new email communications - all of which can aid in general business needs and summarizing/translating current content.

7. **More focused and purposeful email communications**
   To avoid displeasing customers, LISI should enhance 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.

Appendix 1: Current Data Dictionaries

Exhibit 1: Contact Export

Var #	Field Name	Description	Values
1	Email address	Customer contact email	Emails
2	First name	Customer First Name + Middle Name	names
3	Last name	Customer Last Name	names
4	Company	Customer Company Names	Names
5	Job title	Customer Current Job Title	job title
6	Email status	customer subscription status	Active/Unsubscribed
7	Email permission status	TBD	Express/Implied/blanks
8	Confirmed Opt-Out Date	Customer Opt-Out Date	Date
9	Confirmed Opt-Out Source	TBD	A/C
10	Confirmed Opt-Out Reason	Customer's reasons for opt-out	reasons comment

11	Phone - home	Customer Home phone number	phone number
12	Phone - work	Customer Work Phone Number	phone number
13	Street address line 1 - Home	Customer home street address	address
14	City - Home	Customer Home Address City	city
15	State/Province - Home	Customer Home Address State	state
16	Zip/Postal Code - Home	Customer home address Zipcode	zip code
17	Country - Home	Customer Home Address Country	country
18	Email Lists	Email groups customers signed up for	Email lists from LISI
19	Notes	Customer Notes/Comments	Mostly left blank except for two customers asking to get back to the email lists
20	Source Name	Customer info source	"Added by you"
21	Created At	Customer info create date	date
22	Updated At	Customer info update date	date

Exhibit 2: Lister Export

Var #	Field Name	Description	Values
1	Name	customer full name	first name last name:Edit
2	Start Date	TBD	dates from 2000-2022
3	email	customer contact email	emails, one phone number
4	address	customer address	full address, just building name, just street name, numbers
5	city	customer city	some caps, some lower case, some made up?
6	state	customer state	state names, state abbreviations, 60606
7	zip	customer zip	3 digits, 4 digits, 5 digits zip, state abbreviations, "M5K1A2", zip-####
8	phone	customer phone	range from 3 digits to 11 digits, include ext, with/without dash, "John Hancock"
9	fax	customer fax	0, n/a, none, include (), periods, with/without area codes
10	company	customer company of employment	dates, years, company names, blanks
11	sponsor	customer account sponsor	individual, company, school, etc.

12	plan	type of subscription plan	values 1-9, KC
13	List Name	list of subscriptions	abbreviations of subscriptions (Digest, ecl, fpp, etc.)
14	Stripe Cust ID	hexadecimal customer ID	cus_***********
15	Date Inserted	date record was created	dates within years 1900, 2022, and 2023, blanks
16	Date Restored	TBD	dates within years 2022 and 2023

Exhibit 3: Stripe

Var #	Field Name	Description	Values
1	id	customer UID	unique customer values
2	Description	TBD	random values: 2022-1036, GJRupert@reedsmith.com, mostly blanks
3	Email	customer contact email	unique customer values
4	Name	customer full name	unique customer values
5	Created (UTC)	TBD	dates from 2022-2023

6	Delinquent	TBD	TRUE, FALSE
7	Plan	hexadecimal price plan	price_************
8	Status	status of Stripe subscription	active, past_due, trialing, blanks
9	Cancel At Period End	customer decision to cancel subscription going forward	TRUE, FALSE
10	Account Balance	account balance in usd	four values, mainly 0
11	Currency	currency of payment	usd
12	Total Spend	accumulated spending of customer since time of subscription	values range from 0 to 9,664.49.
13	Payment Count	number of payments customer has made	values range from 0 to 12
14	Average Order	total customer purchase for that order	values range from 0 to 9,664.49.
15	Refunded Volume	amount usd refunded to customer	values range from 0 to 389.70.
16	Dispute Losses	total usd loss for LISI after payment dispute	values range from 0 to 381.60.
17	Business Vat ID	TBD	blank/unused column

18	old_id (metadata)	TBD	unknown UID

Exhibit 4: PNP

Var #	Field Name	Description	Values
1	orderid	order id	unique identifier for the order
2	plan	plan number (need translation table for this)	1 through 29, with blanks
3	purchaseid	TBD	blank except for one labeled "tester only"
4	name	name of the customer on the order	names
5	company	company the customer is associated with	company names
6	addr1	address details the customer is associated with	address (street and number)
7	addr2	address details the customer is associated with	address suite or unit
8	city	address details the customer is associated with	city
9	state	address details the customer is associated with	state
10	zip	address details the customer is associated with	zip

11	country	address details the customer is associated with	country
12	shipname	shipping details for the customer order	only populated for three records
13	shipaddr1	shipping details for the customer order	only populated for three records
14	shipaddr2	shipping details for the customer order	only populated for three records
15	shipcity	shipping details for the customer order	only populated for three records
16	shipstate	shipping details for the customer order	only populated for three records
17	shipzip	shipping details for the customer order	only populated for three records
18	shipcountry	shipping details for the customer order	only populated for three records
19	phone	cutomer phone number	phone numbers
20	fax	customer fax number	fax numbers
21	email	customer email	emails
22	billcycle	TBD	12:00:00 AM
23	startdate	TBD	dates

24	enddate	TBD	dates
25	monthly	TBD	integer values from .1 to 3111
26	lastbilled	date of the last bill to customer	dates
27	status	status of the customer	active/cancelled/pending
28	acct_code	all blank	blank

Exhibit 5: Group

Var #	Field Name	Description	Values
1	id	customer UID	unique customer values
2	purchaseid	purchase ID	Some are two digits longer than others, no idea why. Some Nulls
3	order_id	order ID	Some numeric, some hexadecimal, some null.
4	startdate	TBD	time stamps (to the second)
5	enddate	TBD	All null. So this worksheet is not updated if someone cancels. Since they're sponsored, seems to make sense that no cancellation by the named person
6	mode	TBD	12.95, 13.95, 14.95 or Null

7	status	Status of payment	Processing, null, ok
8	email	customer email	email addresses
9	address	most or all office addresses	addresses
10	city	city	cities
11	state	customer state	Some state abbreviations, some not abbreviated. Can merge. Some blank. At least one zip code
12	zip	zip code	zip codes (some 5 digits, some 9, even some 4 so probably foreign). At least one "MN"
13	phone	customer phone	10 digits, different formats
14	fax	fax number	mostly null
15	company	customer company	company names
16	plan	TBD	1,2,4,7,9,null
17	listname	possibly which emails customers have opted into	comma-separated values
18	name	customer full name	full names
19	signupdate	time of signup	time stamps (to the second)

20	paytype	TBD	Annual, comp, null/blank
21	occupation	occupation	as expected, and many Nulls
22	sponsor	sponsor	tends to be the same as company (sometimes the person, sometimes the company, sometimes both)
23	ccmail	TBD	email address
24	pnp	TBD	11202015 or null. From legacy PNP system
25	pt	TBD	y or null
26	sp	TBD	y or null
27	ptsp	TBD	y or null
28	firstname	customer first name	first name, many null
29	lastname	customer last name	last name, many null
30	mypass	TBD	adminaccessteam,adminaccesstom, null
31	stripeCustID	id for stripe payment system	alphanumeric id's starting with cus_L or cus_M. And nulls
32	DateInserted	date record was created	dates within years 1900, 2022, and 2023, blanks

33	restoredate	Date account was restored	All null

Exhibit 6: CUSTOMDLOG01&02

Var #	Field Name	Description	Values
1	Date	Date of Transaction	May 2017 to November 2022
2	Time	Time of Transaction	12 AM to 11 PM
3	TimeZone	Time Zone of Transaction	PDT and PST
4	Name	Customer Name, 19557 Unique Values	First Name Last Name ; Company Names ; Blanks
5	Type	Transaction Type, 22 Unique Values	Account Hold, Cancellation, Chargeback, Dispute, Refund, Subscription Payment, Website Payment, Card Deposit, Withdrawal, Mobile Payment, PreApproved, and more
6	Status	Status of Transaction	Completed, Denied, and Pending
7	Currency	Currency Purchase Made In	USD
8	Gross	Total Amount of Purchase	-$270,758.55 to $73,837.23
9	Fee	Fee of Purchase	-$57.42 to $14.47

10	Net	Net Purchase: Gross - Fee	-$270,758.55 to $73,837.23
11	From Email Address	Email of who made the purchase, Blanks are purchases to the company employee email howgoog2@gmail.com	2,017 Unique Values ; Blanks
12	To Email Address	Email of who payment is going to, majority is the company email howgoog2@gmail.com. Others contain customer emails are typically refunds while blanks are internal balancing.	692 Unique Values ; Blanks
13	Transaction ID	Unique ID of Purchase Transaction	66,067 Unique Values
14	CounterParty Status	TBD	UnRegistered ; Unverified ; Verified ; Blank
15	Shipping Address	Address of Customer where purchase being delivered	24,756 Unique Values ; Blanks
16	Address Status	Verification Staus of Customer Address	Confirmed ; Non-Confirmed ; Blank
17	Item Title	Name of LISI Product Sold	3,063 Unique Values ; Blanks
18	Item ID	Number Identifier of LISI Product Sold	2,297 Unique Values ; Blanks
19	Shipping and Handling Amount	Total Amount of Shipping and Handling of Transaction	0 ; 5.48 ; 7 ; Blank
20	Insurance Amount	TBD	0 ; Blank
21	Sales Tax	Sales Tax of Purchase	0 to 131.25 ; Blank

22	Option 1 Name	TBD	Audience Size ; Email Address ; Group Size ; Group Size, Group Size ; Price Options ; Registrant Email ; Registrants Email ; Select a Payment Option ; Blank
23	Option 1 Value	Value of the Option 1 Name Selection	3,797 Unique Values based on Option 1 Name Selection
24	Option 2 Name	TBD	Email Address, Group Name, Registrant Last Name ; Registrants First Name ; Blank
25	Option 2 Value	Value of the Option 2 Name Selection	10 Unique Values
26	Reference Txn ID	TBD	1,376 Unique Values ; Blanks
27	Invoice Number	TBD	270 Unique Values ; Blanks
28	Custom Number	TBD	71 Unique Values ; Blanks
29	Quantity	Quantity of Item Purchased in the Transction	0 to 21 ; Blank
30	Recipt ID	TBD	32,993 Unique Values : Blanks
31	Balance	TBD	269.66 to 27,2471.86
32	Address Line 1	Street Number, Street Number and Street Name	20,591 Unique Values ; Blanks
33	Address Line 2/District/Neighbor	City, State, Zipcode	2,956 Unique Values ; Blanks

	hood		
34	Town/City	Town or City Only	3,555 Unique Values ; Blanks
35	State/Province/Region/Country/Territory/Prefecture/Republic	State, Province, or Region Only	113 Unique Values ; Blanks
36	Zip/Postal Code	Zip Code or Postal Code Only	7,355 Unique Values ; Blanks
37	Country	Country Only	29 Unique Values ; Blanks
38	Contact Phone Number	Customer Phone Number	6,066 Unique Values ; Blanks
39	Subject	TBD	3,084 Unique Values ; Blanks
40	Note	Comments on the Transaction	135 Unique Values ; Blanks
41	Country Code	Standard Country Abbreviation	29 Unique Values ; Blanks
42	Balance Impact	Transaction Type for LISI	Credit or Debit
43	Buyer Wallet	Other Customer Purchase Type	Venmo ; Blank

Appendix 2: Analysis

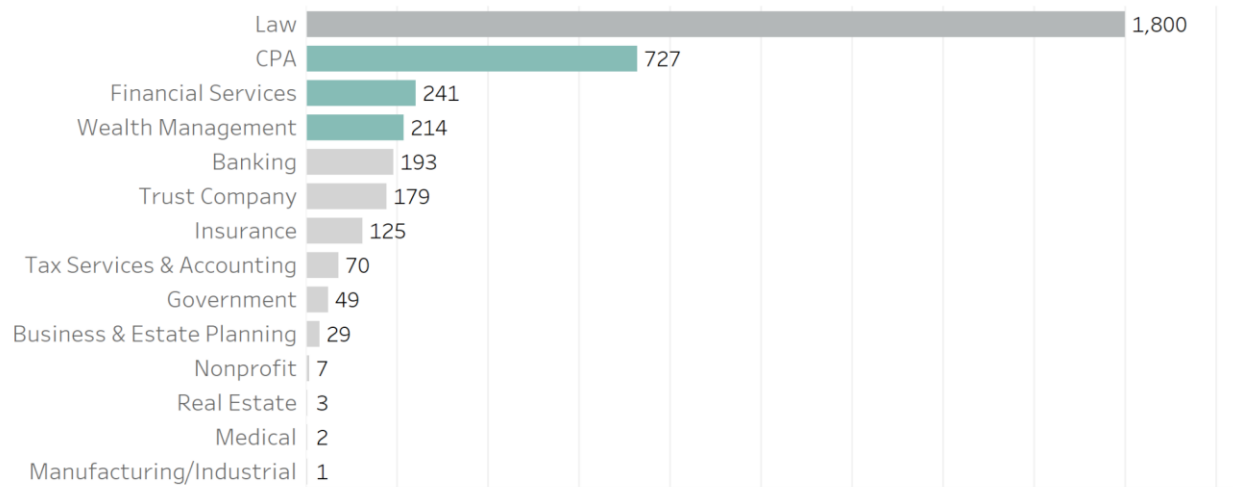**Exhibit 1: Top Items Purchased from Customer Transactions**

Exhibit 2: Location Analysis of Product Purchases Shows International Process for LISI

Exhibit 3: Manually Assigned Top Industries Customers Work in Based on Company Name and Email Address

Top Industries of Subsribers

Industry	Value
Law	1,800
CPA	727
Financial Services	241
Wealth Management	214
Banking	193
Trust Company	179
Insurance	125
Tax Services & Accounting	70
Government	49
Business & Estate Planning	29
Nonprofit	7
Real Estate	3
Medical	2
Manufacturing/Industrial	1

Exhibit 4: Manual Sub Industries Customers Specialize within the Industry Appear to Work In.

Industries and Sub Industries of Subscribers

Company Field	Other Company Details ⇥	
Law	All	1,800
CPA	Office/Firm	618
	Individual	71
	Office/Firm, Consulting	7
	Retired Individual	1
Financial Services	Other	205
	Planning	25
	Consulting	11
Wealth Management	All	214
Banking	Other	152
	Investment Services	41
Trust Company	All	179
Insurance	All	125
Tax Services & Accounting	Tax Services	51
	Tax & Accounting	6
	Accounting	6
	Consulting	5
	Retired	2
Government	All	49
Business & Estate Planning	All	29
Nonprofit	All	7
Medical	Healthcare Technology	1
	Consulting, Research	1

Exhibit 5: Initial and Remaining Customers per Industry After Clustering

Top Industries After Clustering

Company Field =

Company Field	Value
Law	1,936
Financial Services	1,225
Banking	380
Banking & Finance	155
Insurance	150
Other	55
Government	49
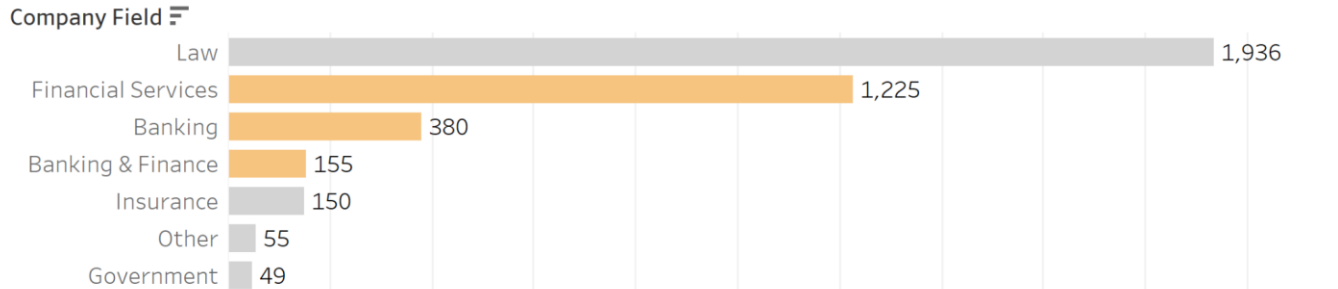
Exhibit 6: Initial and Remaining Customers per Industry and Sub Industry After Clustering

Sub Industries after Clustering

Company Field	Industry Detail	Value
Law	All	1,936
Financial Services	CPA	592
	Wealth & Investment	285
	Other	218
	Tax & Accounting	130
Banking	Other	194
	Trust Company	185
Banking & Finance	Banking and Finance	155
Insurance	All	150
Government	Government	49
Other	Business & Estate Planning	31
	Nonprofit	7
	Education	6
	Technology/IT	4
	Real Estate	4
	Healthcare Consulting	2
	Manufacturing/Industrial	1

Exhibit 7: Top 10 Terms in Each Cluster Used to Assign Cluster Names

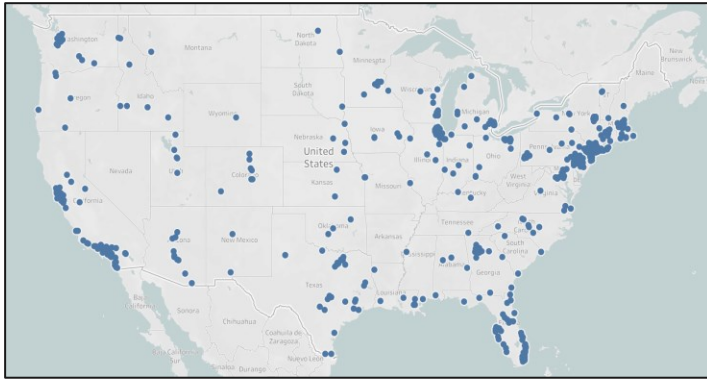| Top Terms Per Clusters				
Cluster 0: **Law - Estate**	**Cluster 1:** **Financial Services - Wealth & Invest**	**Cluster 2:** **Other**	**Cluster 3:** **Insurance**	**Cluster 4:** **Banking - Trust**
law	financial	law	insurance	trust
legal	management	entertainment	life	management
firm	wealth	media	mutual	wealth
specializing	investment	technology	financial	services
estate	services	focus	northwestern	company
areas	planning	industries	services	financial
business	company	service	businesses	institutions
providing	individuals	services	company	families
litigation	families	specializing	annuities	provides
clients	specializes	firm	individuals	individuals
Cluster 5: **Law - Corporate**	**Cluster 6:** **Law - Family**	**Cluster 7:** **Banking & Finance**	**Cluster 8:** **Financial Services - Tax & Accounting**	**Cluster 9:** **Government**
real	family	industry	accounting	irs
industries	lawyers	services	tax	agency
law	adoption	financial	consulting	federal
clients	child	banking	professional	united
healthcare	divorce	leading	various	states
estate	custody	solutions	businesses	barbano
corporate	related	range	industries	tax
services	matters	bank	services	provider
firm	specialize	products	firm	accounting
legal	representation	company	cpa	leading

Exhibit 8: ChatGPT Formulas and Sample Output from Law Industry Analysis

Google Sheets	=**GPT**(CONCATENATE("if the company ", A1, "is a law firm, write a one sentence description of the law they specialize in")) =**GPT**(CONCATENATE("if the company ", A15, "is a law firm, write up to three types of law they specialize in separated by commas"))					
Microsoft Excel	=**GPT.Prompt**(CONCATENATE("if the company ", A1, "is a law firm, write a one sentence description of the law they specialize in", 0.5,0)) =**GPT.Prompt**(CONCATENATE("if the company ", A1, "is a law firm, write up to three types of law they specialize in separated by commas", 0.5,0))					
Column A	**Company Name**	**Company Field**	**Industry detail**	**ChatGPT Description**	**ChatGPT law description**	**Law Short**
1	Pillar Wealth Advisors, LLC	Financial Services	Wealth Management	Pillar Wealth Advisors, LLC is a financial services company that specializes in providing comprehensive wealth management solutions to high net worth individuals and families.	Pillar Wealth Advisors, LLC is not a law office or firm. They specialize in providing wealth management and financial planning services.	not a law firm
2	Kauai Estate Law, LLC	Law	Estate Planning, Probate, Trust Administration	Kauai Estate Law, LLLC is a law firm specializing in estate planning, trust administration, and probate law services for individuals and families in Kauai, Hawaii.	Kauai Estate Law, LLLC specializes in estate law.	Estate Planning, Probate, Trust Admin
3	Brackett & Ellis PC	Law	Corporate Law, Tax Law, Intellectual Property Law	Brackett & Ellis PC is a law firm that specializes in providing legal services to clients in various industries, including real estate, healthcare, energy, and construction.	Brackett & Ellis PC is a law firm that specializes in a wide range of legal services, including corporate law, real estate law, tax law, intellectual property law, and litigation.	Corporate Law, Tax Law, Intellectual Property Law

Exhibit 9: Word Cloud of the Short Law Descriptions Displays the Top Law Specialties Following Corporate, Tax, Estate Planning, and Intellectual Property

Exhibit 9: Contact Export Dataset Reasons for opting out of emails

OPT-OUT REASONS

Reason	Count
No Reason	60721
Negative (Received too many)	22194
Changed Address	214
Changed Address	214
Retired	128
Death	35
Price NotAffordable	21
CE credits	19
Business Closed	13
Product Issues	11
No Time	4

Exhibit 10: Location Analysis of Subscribers

Top 5 Subscribing States

State	Subscribers
California	749
New York	652
Florida	504
Pennsylvania	326
Texas	314
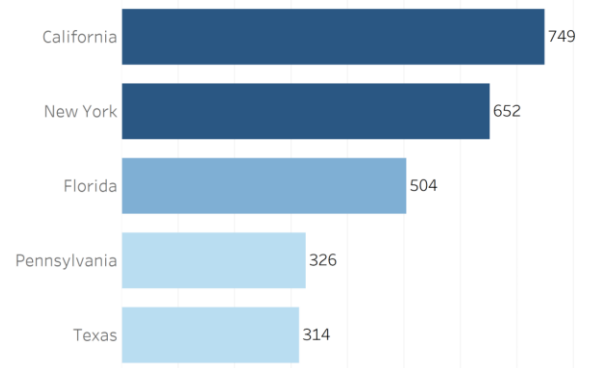
Appendix 3: Future Data Dictionaries and ERD

Exhibit 1: Future Data Set - Subscribers

Var #	Field Name	Format	Description	Values
1	cust_id	NUM	Unique identifier for every individual customer	CUS0001
2	birthday	DATE	Customer's birthdate	44927
3	group_acct_ind	CHAR	Indicator of whether or not customer is part of a larger group account	Y/N
4	sub_dt	DATE	Date the customer subscribed	44927
5	sub_cancel_dt	DATE	Date the customer cancelled	44927
6	sub_status	CHAR	Status of the customer's subscription	Active/Suspended/Delinquent/Cancelled/etc.

Exhibit 2: Future Data Set - All Contacts

Var #	Field Name	Format	Description	Values
1	cust_id	NUM	Unique identifier for every individual customer	CUS0001
2	first_nm	CHAR	First name of customer	John

3	last_nm	CHAR	Last name of customer	Doe
4	email_address	CHAR	Email address of customer	johndoe@lawoffice.com
5	company_name	CHAR	Company customer works for	John Doe Law
6	job_title	CHAR	Customer's Profession/Job Title	Attorney
7	industry	CHAR	Main Category/Field the customers line of work falls under selected by customer. Only offer specific ones and avoid allowing other/fill in	Law Nonprofit Trust Company Financial Services Wealth Management Tax Services & Accounting Banking Insurance Manufacturing/Industrial Business & Estate Planning Real Estate Medical Government
8	lisi_network_ind	CHAR	Indicator of whether or not customer is using LISI network	Active/Unsubscribed
9	join_dt	DATE	First date customer subscribed to LISI emails	44927
10	email_preference	CHAR	Description of customer's current email preference	Subscribed/Notsubscribed/Unsubscribed/NA

11	out_out_dt	DATE	Date the customer opts out of emails	44927
12	opt_out_rsn	CHAR	Customer's reasons for opting out of emails	Too many Emails, Too pricey, Unable to find wanted source, no longer needed the services, Changes of Emails, etc.
13	degree	CHAR	Highest level of degree customer has earned	Highschool, Bachelores, Masters, Doctorate, or PhD
14	degree_detail	CHAR	Name of degree	Doctorate in Estate Law
15	certifications	CHAR	Text fill in of any professional certifications or licenses the customer has	CPA, CFP, etc
16	phone_number	CHAR	Customer phone number	XXX-XXX-XXXX
17	delinquent_ind	CHAR	Indicator of whether or not a customer's account is delinquent	Y/N
18	account_balance	INT	Current Account balance of the customer	450

Exhibit 3: Future Data Set - Company Details

Var #	Field Name	Format	Description	Values
1	company_name	CHAR	Name of the company the customer works for, captured at time of subscription	Law Office of John Doe
2	company_address_ln1	CHAR	Company address line one, captured either at subscription or through web scraping	10 Main St
3	company_address_ln2	CHAR	Company address line two, captured either at subscription or through web scraping	Unit 1
4	company_city	CHAR	Company city, captured either at subscription or through web scraping	Boston
5	company_st	CHAR	Company state, captured either at subscription or through web scraping	MA
6	company_country	CHAR	Company country, captured either at subscription or through web scraping	USA
7	company_zip_cd	CHAR	Company zip code, captured either at subscription or through web scraping	00000
8	company_phone_number	CHAR	Company phone number, captured either at subscription or through web scraping	XXX-XXX-XXXX

Exhibit 4: Future Data Set - Industry Details

Var #	Field Name	Format	Description	Values
1	industry	CHAR	Main Category/Field the customers line of work falls under	Law, Nonprofit, Trust Company, Financial Services, Wealth Management, Tax Services & Accounting, Banking, Insurance, Manufacturing/Industrial, Business & Estate Planning, Real Estate, Medical, Government
2	industry_detail	CHAR	Details regarding this industry	Law:Type of Law, Nonprofit:Area of Nonprofit, Trust Company:Planning, Consulting, etc, Financial Services:Investment, Retirement, etc, Wealth Management:Life, Health, Consulting, etc, Tax Services & Accounting:Consulting, Planning, etc, Banking:Strategy, Consulting, Wills, etc, Insurance:Type of Real Estate, Manufacturing/Industrial:Technology, Research, Healthcare, etc, Business & Estate Planning:IRS, Dept of Justice, etc, Real Estate:Type of Real Estate, Medical:Technology, Research, Healthcare, etc, Government:IRS, Dept of Justice, etc

Exhibit 5: Future Data Set - Purchase Details

Var #	Field Name	Format	Description	Values
1	purchase_id	NUM	unique ID for the individual purchase	PUR0001
2	cust_id	NUM	Unique identifier for every individual customer	CUS0001
3	invoice_dt	DATE	date the item was invoiced	44927
4	purchase_dt	DATE	date the item was purchased	44927
5	charged_amt	INT	dollar amount charged for the item	120
6	currency	CHAR	currency the item was purchased in	USD
7	item_purchased	CHAR	item that was purchased	List of LISI products
8	item_price	INT	price of the item	120
9	discount_prct	INT	percentage the item was discounted	0.35
10	discount_code	CHAR	code used to discount the item	LISIDISC10

Exhibit 6: Future Data Set - Interest Details

Var #	Field Name	Format	Description	Values
1	cust_int_id	NUM	Unique ID for each customer interest	CUSINT0001
2	cust_id	NUM	Unique identifier for every individual customer	CUS0001
3	interest	CHAR	Interest selected from value list when subscribing on LISI website	Steve Leimberg's Elder Care Law Planning Steve Leimberg's Estate Planning Newsletter Steve Leimberg's Employee Benefit and Retirement Planning Steve Leimberg's Business Entities Newsletter Bob LeClair's Finance and Markets Letter Asset Protection Planning Newsletter Charitable Planning Newsletter Income Tax Planning Newsletter International Tax Planning Newsletter Financial Products Planning Newsletter TaxTalk News Newsletter Steve Leimberg's State Cases and Legislation Weekly Digest LISI Network LISI Listens The Institute: Courses and Certifications

Exhibit 7: Future Data Set - Email Details

Var #	Field Name	Format	Description	Values
1	cust_email_id	NUM	unique ID for each customer active email list	CUSEM0001
2	cust_id	NUM	Unique identifier for every individual customer	CUS0001
3	email_list	CHAR	Email groups customers signed up for	Newsletter emails LISI distributes

Exhibit 3: ERD Diagram - Link to access ERD in LucidChart: https://lucid.app/lucidchart/fddab42b-41eb-4bd6-8b11-282f6ac0950a/edit?viewport_loc=-284%2C-12%2C2215%2C1065%2C0_0&invitationId=inv_e2def650-474f-4e04-94db-5c33d3465931
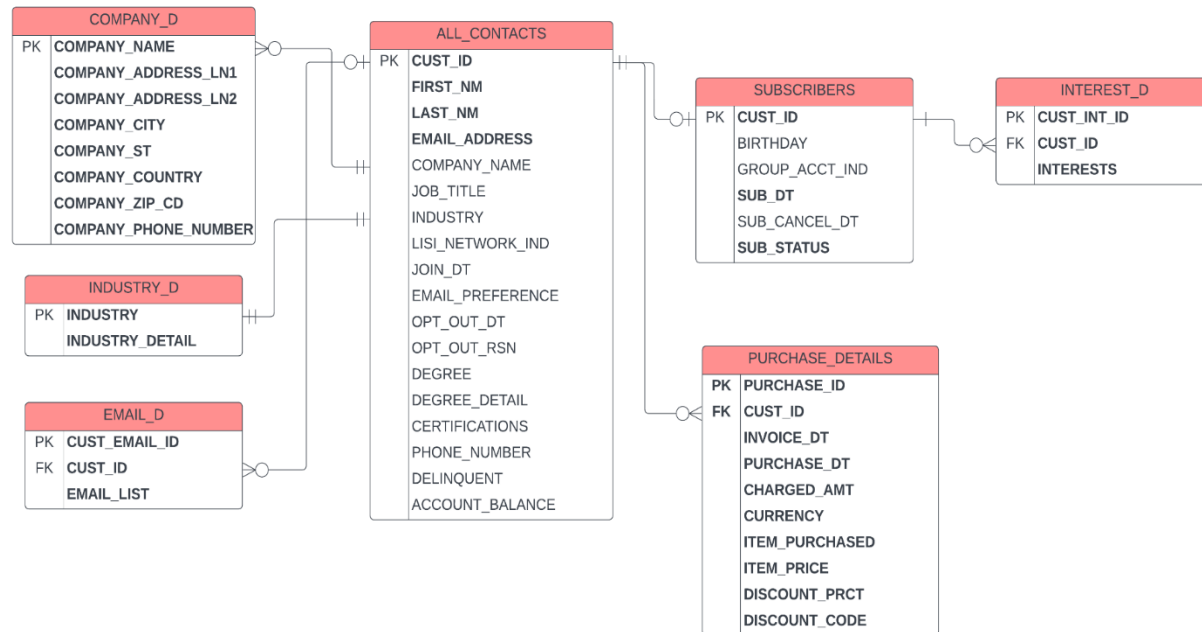
Appendix 4: Potential Survey Questions

Questions About LISI Currently
- How did you hear about LISI? *Multiple selection*
- What attracts you to LISI? *Multiple selection*
- Have you purchased the following? *Multiple selection: ebooks, webinar, tool kit, etc*
- What information are you most interested in? *Multiple selection: Financial Planning, Tax Planning, etc*
- How do you feel about the cadence of our emails?
  Single choice: too little, too much, just right
- Other Comments? *Fill In*

Questions About New Offerings
- What other information and topics do you wish LISI provided or had more of? *Fill in*
- Would you be interested in contact sharing for networking and freelance opportunities? *Yes or No*
- If yes, would you still be interested if the contact sharing had a cost per month or one time opt in price? *Multiple selection Per Month, One Time, No Cost*
- Are you interested in audio versions of webinars or ebooks? *Yes or No*
- Would you like LISI content in another language? *No ; If yes fill in*
- Do you have any licenses or certifications that need continued education for renewals? *No ; If yes fill in which ones, how much do you currently pay, and how often?*

Appendix 5: LISI Listens Additional Information

Below is an informal compilation of materials that we hope will assist with your implementation of the LISI Listens podcast. 1) Advertising partners. 2) Things to keep in mind during the process. 3) Potential platforms for using AI to convert content to podcast form. We have not vetted these platforms, which were outputs from a series of ChatGPT prompts.

Suggested advertising partners:

- Cybersecurity firms and/or VPNs: listeners work with sensitive data and might be interested in good security solutions.
- DoorDash etc.: listeners work long hours and might want easy delivery to their office. Also HelloFresh (related).
- LifeLock and other identity protection services: high net worth clients might want this peace of mind.
- TurboTax etc.
- Thermospas/meditation apps/anything for stress reduction
- SimpliSafe (goes to peace of mind)
- LinkedIn
- Grammarly: for personal use or even company-wide mandate
- Dry-cleaning pick-up and delivery services (local chains)
- Wine subscription services

- Dog-walking services

The following numbered plan was constructed with the help of ChatGPT:

1. Market research: Conduct a thorough market research to understand the target audience for the podcast, their preferences, and the current podcast landscape.
    1. Study the competition: Analyze the current podcast landscape and identify the most popular podcasts that relate to specialized industries.
    2. Assess audience preferences: Determine the preferred length, format, and frequency of podcasts among the LISI Listens audience. This information can be gathered through surveys. Is an hour the right length?
    3. Determine the need for the podcast: Understand the reasons why the target audience would listen to a podcast (instead of a webinar) and how the podcast can meet their needs and add value to their lives.
    4. Identify potential advertising partners: Advertising partners and sponsors who would be interested in reaching the LISI's demographic.
    5. Determine distribution channels: Identify the most effective distribution channels for the target audience, including social media, email, and other platforms. Maybe just include a link to the podcast in the newsletter.
2. Content review: Does all of the webinar content make sense in podcast form? Is there any content where the visuals are too important to remove?
3. Recording and editing: Determine whether there is a need for an editing process. Just copy the audio file from the webinar, or is there more to it? Ensure that the audio quality is high.
    1. Content adaptation: Removing visual aids and adjusting the length.
    2. Audio quality improvement: Removing background noise, equalizing the volume, and ensuring consistent sound quality throughout.
    3. Enhancing engagement: Editing can help to make the podcast more engaging by adding music, sound effects, and other elements that can help to keep the listener's attention.
    4. Clarity and conciseness: Editing can help to make the content clearer and more concise, ensuring that the message is delivered effectively to the target audience. Maybe there is AI technology that can process the audio and smooth out any "um" or related filler words.
    5. Consistency in formatting: Editing can help to maintain a consistent format throughout the podcast, including consistent pacing, tone, and structure. Maybe a LISI employee kicks off each podcast with a brief intro.
4. Podcast hosting and distribution: Choose a suitable hosting platform and distribute the podcast to various platforms such as iTunes, Spotify, and Soundcloud.
5. Analytics and optimization: Monitor the podcast's performance and engagement metrics, and make changes as needed to optimize the content and marketing strategy.
    1. Download numbers: How many people listen?
    2. Subscriber numbers: Break down by podcast only vs. success conversions to more LISI content.
    3. Listener engagement: Metrics such as play rate, completion rate, and average time spent listening can be useful.
    4. Audience demographics: Different from typical LISI demographic?

5. Feedback: Ratings, reviews, and comments. Use them.
6. Social media engagement: Metrics such as likes, shares, and comments on social media can help to gauge the popularity and impact of the podcast on social media platforms.
7. Website traffic: Data on website traffic from the podcast, including referral sources and pages visited, can help to understand the impact of the podcast on LISI's website. Do people look you up after they hear the podcast?

Potential software solutions for transforming webinars to podcasts (from ChatGPT):

1.   Descript - Descript is an all-in-one audio and video editing software that uses AI to automatically transcribe, caption, and edit audio and video content. It's a great option if you want to use AI to help you transcribe and edit your webinar content into a podcast.
2.   Anchor - Anchor is a free podcast creation platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want a simple, free option for transforming your webinar content into a podcast.
3.   Audacity - Audacity is a free and open-source audio editing software that can be used to record and edit audio content. It's a good choice if you want a powerful, free option for transforming your webinar content into a podcast.
4.   Alitu - Alitu is an online podcasting platform that uses AI to automatically clean up audio recordings, add intros and outros, and publish the podcast to major platforms. It's a great option if you want an easy-to-use platform that does most of the work for you.
5.   Adobe Audition - Adobe Audition is a professional audio editing software that's widely used in the media industry. It's a good choice if you want a powerful and versatile option for transforming your webinar content into a podcast.
6.   Hindenburg Journalist - Hindenburg Journalist is a professional audio editing software that's designed for radio journalists. It's a good choice if you want a professional and easy-to-use option for transforming your webinar content into a podcast.
7.   Auphonic - Auphonic is an online audio post-production and distribution platform that uses AI to automatically improve audio quality, normalize loudness, and add music and sound effects. It's a great option if you want an easy-to-use platform that does most of the work for you.
8.   Zencastr - Zencastr is an online platform that allows you to record high-quality audio interviews remotely. It's a great option if you want to transform your webinar content into a podcast that features guest speakers.
9.   Cleanfeed - Cleanfeed is an online platform that allows you to record high-quality audio interviews remotely. It's a great option if you want to transform your webinar content into a podcast that features guest speakers.
10.   Cast - Cast is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.
11.   Buzzsprout - Buzzsprout is a podcast hosting and distribution platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.
12.   Blubrry - Blubrry is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

13.   Spreaker - Spreaker is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

14.   Podbean - Podbean is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

15.   Simplecast - Simplecast is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

16.   Libsyn - Libsyn is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

17.   Transistor - Transistor is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

18.   Squadcast - Squadcast is an online platform that allows you to record high-quality audio and video interviews remotely. It's a great option if you want to transform your webinar content into a podcast that features guest speakers and video content.

19.   Podomatic - Podomatic is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

20.   Megaphone - Megaphone is a podcast hosting and analytics platform that offers an easy-to-use interface for recording and editing audio content. It's a good choice if you want an all-in-one solution for transforming your webinar content into a podcast and hosting it.

21.   Reaper - Reaper is a professional audio editing software that's widely used in the music industry. It's a good choice if you want a powerful and versatile option for transforming your webinar content into a podcast.

22.   Ferrite Recording Studio - Ferrite Recording Studio is a mobile audio editing app that allows you to record and edit audio content on your phone or tablet. It's a good choice if you want a mobile option for transforming your webinar content into a podcast.

23.   Pro Tools - Pro Tools is a professional audio editing software that's widely used in the music and media industries. It's a good choice if you want a powerful and versatile option for transforming your webinar content into a podcast.

24.   GarageBand - GarageBand is a free audio editing software that's available on Mac computers. It's a good choice if you want a free and easy-to-use option for transforming your webinar content into a podcast.

25.   Audio Hijack - Audio Hijack is a Mac-only software that allows you to record and edit audio content from any source on your computer. It's a good choice if you want a versatile option for transforming your webinar content into a podcast.

26.   Loopback - Loopback is a Mac-only software that allows you to route audio between applications on your computer. It's a good choice if you want to record and edit audio content from multiple sources for your podcast.

27.   Sound Forge - Sound Forge is a professional audio editing software that's widely used in the media industry. It's a good choice if you want a powerful and versatile option for transforming your webinar content into a podcast.

28.   TwistedWave - TwistedWave is a Mac-only audio editing software that's known for its easy-to-use interface. It's a good choice if you want a simple and intuitive option for transforming your webinar content into a podcast.

29.   Ocenaudio - Ocenaudio is a free audio editing software that's available on Windows, Mac, and Linux. It's a good choice if you want a free and easy-to-use option for transforming your webinar content into a podcast.

29.   WavePad - WavePad is a free audio editing software that's available on Windows, Mac, and mobile devices. It's a good choice if you want a free and user-friendly option for transforming your webinar content into a podcast. WavePad allows you to record and edit audio content, apply various effects and filters, and export your finished product in a variety of formats. The software also offers noise reduction features to help improve the quality of your recordings. Overall, WavePad is a great option for those who want a free and accessible way to transform their webinar content into a polished podcast.

<u>References</u>

[Antonio Nakić-Alfirević]. (2023, February 10). Chat GPT in Excel [Video]. YouTube. https://www.youtube.com/watch?v=8z0gtMbx08A

Learn.org. (n.d.). What's the Difference Between Accredited & Non-Accredited Degree Programs? Learn.org. Retrieved April 25, 2023, from https://learn.org/articles/What_is_the_Difference_Between_an_Accredited_Degree_Program_and_a_Non
-
Accredited_College_Degree.html#:~:text=If%20an%20individual%20wants%20to,their%20education%20or%20their%20integrity

Marion, S. (2023, April 9). *Get started with GPT for sheets*. GPT for Work. Retrieved April 18, 2023, from https://gptforwork.com/gpt-for-sheets/get-started

OpenAI. (2022). Introducing chatgpt. Introducing ChatGPT. Retrieved April 18, 2023, from https://openai.com/blog/chatgpt

Use chatgpt in Google Sheets and DocsSupports all models: GPT-4 (if you have access), chatgpt (GPT-3.5-turbo) and GPT-3. ChatGPT for Google Sheets and Docs. (n.d.). Retrieved April 18, 2023, from https://gptforwork.com/