Article

Mining Open Government Data for Business Intelligence Using Data Visualization: A Two-Industry Case Study

Anne Gottfried [1],*, Caroline Hartmann [2],* and Donald Yates [3],*

1 College of Business, Department of Marketing, University of Texas, Arlington, Arlington, TX 76010, USA
2 College of Business, Department of Accounting, Texas A&M University-Commerce, Commerce, TX 75428, USA
3 Information Systems (IS) Adjunct Faculty, Louisiana State University of Alexandria, Alexandria, LA 71302, USA
* Correspondence: anne.gottfried@uta.edu (A.G.); caroline.hartmann@tamuc.edu (C.H.); yates@dnkdata.com (D.Y.)

Abstract: The business intelligence (BI) market has grown at a tremendous rate in the past decade due to technological advancements, big data and the availability of open source content. Despite this growth, the use of open government data (OGD) as a source of information is very limited among the private sector due to a lack of knowledge as to its benefits. Scant evidence on the use of OGD by private organizations suggests that it can lead to the creation of innovative ideas as well as assist in making better informed decisions. Given the benefits but lack of use of OGD to generate business intelligence, we extend research in this area by exploring how OGD can be used to generate business intelligence for the identification of market opportunities and strategy formulation; an area of research that is still in its infancy. Using a two-industry case study approach (footwear and lumber), we use latent Dirichlet allocation (LDA) topic modeling to extract emerging topics in these two industries from OGD, and a data visualization tool (pyLDAVis) to visualize the topics in order to interpret and transform the data into business intelligence. Additionally, we perform an environmental scanning of the environment for the two industries to validate the usability of the information obtained. The results provide evidence that OGD can be a valuable source of information for generating business intelligence and demonstrate how topic modeling and visualization tools can assist organizations in extracting and analyzing information for the identification of market opportunities.

Keywords: open government data; business intelligence; big data; data mining; latent Dirichlet allocation; data visualization; pyLDAVis; business strategy; private organizations

1. Introduction

The business intelligence (BI) market has grown tremendously in the twenty-first century resulting from the increased adoption of cloud services, and the growth of data analytics and internet enabled technologies. In 2016, global BI revenues reached $17.09 billion and are forecasted to reach $26.88 billion by 2021, a 9.5% increase in 5 years [1]. Business intelligence refers to a company's ability to gather, analyze and communicate information for the purpose of making strategic decisions [2]. It relies on technology and the use of processes and applications to extract and analyze data, which can then be used to obtain consumer insights, identify market opportunities, and create innovative solutions for sustaining competitive advantages in the marketplace. Private companies such as Amazon have used technology and data analysis to create business intelligence that has transformed the retail industry [3] while the pharmaceutical industry has used it to minimize distribution related problems and identify their most profitable products [4]. Business intelligence has also been gathered to assist in developing public policy [5,6].

Despite the tremendous growth in business intelligence, the use of open government created content (OGD) by private sector organizations to generate business intelligence is quite limited; mainly due to companies having very little knowledge about the availability of open data [7] and a lack of insight as to its benefits [8,9]. The establishment of the Open Government Directive by the Obama Administration in 2009 [6], and the Open Government Partnership in 2011 [7] has required governments to make their data accessible to the public and become more transparent and accountable for their initiatives. Accessibility to OGD provided by government agencies worldwide, enables external stakeholders to use the data to enhance the creation of innovative ideas and products [10,11] that can ultimately contribute to economic growth [7,11], and generate social value for its citizens [12]. For example, in the United States, the automotive and trucking industries have recently used OGD to make better informed decisions related to assembly plant productions and rate setting of diesel fuel consumption [13]. Another international organization has used OGD to assist in underwriting weather insurance for farmers [14]. A recent study by Jetzek et al. (2014) also highlights how OGD can be used to generate economic and social value through the lens of two dimensions: external stakeholders (private sector) and the public sector (focused mainly on social value). The authors find that the use of OGD drove change in consumer energy consumption and created social value for its citizens [8].

Given the benefits but lack of use of OGD to generate business intelligence, we examine how business intelligence can be generated through the use of OGD for the identification of market opportunities and strategy formulation; an area of research that is still in its infancy [15]. Using a two-industry case study approach (footwear and lumber), we use LDA topic modeling to extract emerging topics in these two industries from OGD, and a data visualization tool (pyLDAVis) to visualize the topics in order to interpret and transform the data into business intelligence. In doing so, we extend research in the field of topic modeling and visualization tools [6,16] and answer the call by scholars to use these methods to acquire and visualize information to transform data into business intelligence [17,18]. Additionally, we perform an environmental scanning of the environment for the two industries based on the topics identified to validate the usability of the information obtained for identification of market opportunities. The results provide evidence that OGD can be a valuable source of information for private businesses in generating business intelligence and highlight how the use of topic modeling and visualization tools can assist organizations in extracting and analyzing information for the identification of market opportunities, formulating marketing strategies to expand into future markets, and obtain a competitive advantage.

2. Business Intelligence and Open Government Data

Business Intelligence (BI) is a process where information from internal and external sources is collected through the use of a variety of software and fact-based support systems to assist organizations in decision making, strategy formulation, and achieving a competitive advantage [19]. Organizations use BI to gather information, analyze competitors and predict the behavior of consumers, competitors, suppliers, and markets. The importance of BI has grown exponentially with the availability of big-data and technological advances that have shifted the business intelligence arena from an IT-led system-of-record reporting to business-led agile analytics that assist organizations in formulating strategy [20]. The big data phenomenon (i.e., volume, variety, and velocity of unstructured and undefined data) has further impacted business intelligence by including fast, predictive, visual analytics and data science [17]. The result is that BI is now associated with (1) improved value-based decision-making [21], (2) connections to firm sustainability [22], and (3) agile analytics [17].

The type of information available can be a key element in an organization's ability to create value, generate business intelligence and formulate strategy. According to Park et al. (2010) the use of publicly available data can be a rich source of information since it provides a wide variety of data ranging from research findings, marketing information, and legal information [19]. Furthermore, as companies increasingly operate in foreign markets the ability to obtain data that assists in analyzing competitive markets is even

more critical. One source of publicly available data that can assist organizations in creating value and generating business intelligence is open government generated data (OGD). The dissemination of open government data (OGD) has been growing since 2011, when governments and civil society advocates came together to form the Open Government Partnership. Over 78 countries and governments have joined this organization that focuses on ensuring governments provide free information in a multitude of areas [23].

Today, the government is considered one of the largest creators and collectors of data, with more than one million datasets available to the public by governments worldwide [24], in a variety of domains [25] including data governance, education, health, extractive industries (oil, gas and mineral resources), traffic, weather, geographical data, and data on businesses and public sector budgeting [8,26]. One of the main benefits of this type of data is that it has already been collected for specific use by governments and has been paid for by taxpayers, thereby mitigating the economic cost that is assumed by private citizens to collect this data. Besides governmental oversight and monitoring of data quality standards, other additional benefits of OGD are the increased public value obtained through citizen participation and empowerment, and improved public relations and citizen attitude towards government [9,27]. A few challenges or barriers related to the use of OGD are difficulty for users in processing the information and the usability of the content available, since the information has typically been prepared by governments with a specific goal in mind [28]. Additionally, in order to adopt usage of OGD at the firm level, organizations need to have the support of top management [14,29]. The biggest barrier to the use of OGD however, seems to be the lack of knowledge by private sector organizations regarding the availability of open data [7] and insight as to its benefits [8,9].

Despite these challenges, OGD offers organizations the potential to unlock new innovative solutions that can incentivize entrepreneurship and generate economic value [8,9]. A review of the OGD literature suggests that the use of OGD can generate economic value and drive innovation within societies [28,30,31]. A survey of government employees and business professionals' use of OGD in Brazil revealed that this type of data promoted greater citizenship involvement [28]. Other researchers report economic value is generated by reducing costs, adding value to current services and products, generating new products and services and increasing data availability for competitiveness [26,27,32]. A recent study by Jetzek et al. (2014) was the first to use OGD to create economic and social value [8]. The authors performed a longitudinal six-year qualitative case study of a private company and found that the company was able to achieve a considerable reduction in consumer energy usage over six years by using OGD and a data driven innovation framework.

OGD can also lead to the generation of new and innovative business models and business intelligence. Two case studies exemplify how OGD is being used to create innovative business models in different sectors of the economy [33]. The Climate Company, based in San Francisco, used weather data along with public satellite data to create a climate platform that assists farmers in making operating and financing decisions, which ultimately affect their decision to underwrite weather insurance for farmers. Using government data on public transportation, Netcetera, a Swiss software company, created a free app called Wemlin that allows users in Switzerland to have precise information on the departure times at particular stops of their public transportation system.

Hughes-Cromwick and Coronado (2019) additionally contend that access to public data is a competitive advantage for existing and new businesses [13]. They provide examples of how government data can inform business decisions in the automotive and energy industries in the United States. For the automotive industry, they find that auto makers combine OGD with privately generated data to make informed decisions on the production rates at assembly plants and to gain support for the development of autonomous vehicles. Energy consulting companies also use government generated data as a starting point for market analysis. For example, data on diesel fuel consumption and price provided by the Energy Information Administration (EIA) are used for rate setting in the trucking industry. Government data can therefore be seen as a strategic asset for gathering business

intelligence and converting that intelligence into potential solutions for economic growth, especially as technological advances occur, and by permitting big data to be accessible by both public and private entities.

3. Use of Topic Mining and Visualization Tools to Gather Business Intelligence

The digital revolution and the development of new technologies to extract business intelligence (BI) from large amounts of data available in the environment is changing the business landscape. Profitable knowledge obtained from BI allows a firm greater ability to respond to changes and compete in the global economy. In this next section, the techniques used to extract BI for this study are discussed. An overview of Section 3.1 includes text mining, topic mining within text mining, various topic mining methods and the topic model method selected for this project latent Dirichlet allocation (LDA) is explained. In Section 3.2, the reasons for including pyLDAVis visualization tool in order to interpret the topic modeling results are highlighted.

3.1. Use of Latent Dirichlet Allocation (LDA) Method for Topic Modeling

Text mining is defined as the process of extracting useful, meaningful, and nontrivial information from unstructured text [34–36] and is ideally suited for extracting concepts out of large amounts of text in a variety of settings. It is this process that originated in computer science literature [37] and surfaced in the marketing field around 2001 as a tool to assess market structures [38,39]. Text analytics is another term used to describe the process of analyzing unstructured text, extracting relevant information, and transforming this information into useful business intelligence. Marketing studies utilizing text-mining analytics have mostly centered on a single entry such as a product or product feature to extract, analyze, and visualize information from consumer-generated open source containing large numbers of data entries [36]. Netzer et al. (2012) used the information retrieved to establish brand association relationships between the entries [36] while Linoff and Berry (2011), applied text-mining to customer-generated content to extract marketing, sales, and customer relationship management information [40]. More recently text-mining to extract intelligence has branched into social media [41] and other areas beyond product reviews (e.g., stock markets) [42].

Within the text mining field, topic modeling has become quite popular as it is a type of statistical modeling used for discovering abstract topics within a collection of documents. It can be described as a method for finding a group of words (i.e., topics) and recurring patterns from a large collection of information. In the last two decades five topic modeling techniques have been introduced: Latent Semantic Analysis (LSA); Probabilistic Latent Semantic Analysis (PLSA); Latent Dirichlet Allocation (LDA); Non-Negative Matrix Factorization (NNMF); and Leximancer [43,44]. LSA and NNMF were the first to be introduced and work on a "bag of words"(BOW) approach, and are non-probabilistic or algebraic in nature. Subsequently, probabilistic models such as PLSA and LDA came into existence that added the probability sense to this topic modeling technique. Of these two, LDA has a supervised approach for model learning which requires user training but results in neatly packaged results with topic labels. In contrast, PLSA uses an unsupervised approach (user training is not necessary) but it is very limited because it needs high quality data, which is not always readily accessible [45]. These four techniques require manual context analysis by the user however, the preferred method is LDA since it delivers more accurate results, allowing users to get valuable insights and make decisions that are data driven [45]. For a comprehensive review of these topic modeling techniques please refer to Kherwa and Bansal [43]. Leximancer, introduced in the early 2000s, uses statistics-based algorithms to extract meaning from documents and in contrast to the other techniques, provides a conceptual map of the main concepts and themes (content analysis), thereby significantly reducing the time required to analyze information [44]. The main disadvantage of Leximancer is that it is a costly commercial product that is not easily available to users, resulting in limited use by individuals and organizations.

In the marketing/BI literature and "big data" analysis, latent Dirichlet allocation (LDA) has become the most popular topic modeling algorithm used due to its applicability in a wide range of contexts [46,47], and its ability to analyze long lengths of documents [48]. The LDA method established by Blei et al. (2013) is based on topic modeling whereby a keyword list is sorted by relevance rankings related to the topic [49]. The general idea of LDA is derived from the premise that a document is a mixture of a few related topics [50]. The LDA method incorporates Bayesian statistics and machine learning to extract these topics based on a frequency of the occurrence of words and infers the relationship between them [49]. It does not attempt to evaluate the meaning of these topics; they are simply numbered therefore it is up to the user to assign meaning to the topics. An advantage of this technique is that it can be run in python, which is a free, open-source programming language that is available for anyone to use. As a result, it is a cost effective topic modeling tool and advantageous for organizations where cost may be an important factor [51].

The LDA approach was first used in business marketing literature to extract words from customer-generated product reviews in order to analyze customer opinions on products and brands [52–54]. In recent years, LDA text-mining has shifted from analyzing online product reviews to examining social media texts (e.g., Twitter and Facebook) [55] and other product content [56]. Hejazi Nia (2015) used this approach in examining infographic research; a type of information presentation that inbound marketers use [56]. The research demonstrated that the LDA method can allow the infographic designers to benchmark their design against previous viral infographics to measure whether a given design decision can help or hurt the probability of the design becoming viral.

Several authors have used the LDA text-mining method to capture business intelligence [42,57]. Mahajan et al. (2008) proposed a stock market analysis system that used LDA to analyze financial news and identify major events that impact the stock market [58]. The authors were able to predict whether the stock market would fall or rise, based on financial news items. Guo et al. (2017), used LDA to extract consumer-generated content containing customer service opinions voiced by hotel visitors throughout the purchasing lifecycle. The results transformed the way that visitors evaluate, select and share experiences about tourism [59]. Wang et al. (2018), proposed a new framework of applying online product reviews to analyze customer preferences for two competitive products [60]. Key topics of online reviews were extracted for two specific competitive products using LDA, demonstrating the competitive superiorities and weaknesses of both products. In summary, the use of LDA is extremely beneficial because it can be adapted to analyze large volumes of data in any context which can be used to make strategic managerial and marketing decisions.

3.2. Data Visualization Using pyLDAVis

Despite the successful use of LDA to extract data for business intelligence, users can have difficulty understanding the topics that are generated or visualizing the topic distributions, as well as interpreting the meaning of the topics. In order to interpret the topic modeling results, the use of a visualization tool such as pyLDAVis can be very effective and allows users to "see" what is being extracted [61]. The pyLDAVis is a web-based interactive visualization tool that explores topic−term relationships using relevance to present topic model results and permit the user to interpret the meaning of the topics. The visualization is intended to be used within IPython and Jupyter notebooks to flexibly explore topic−term relationships using relevance to better understand a fitted LDA model [61].

The pyLDAVis approach allows for a deeper inspection of the terms most highly associated under each individual topic yielding "information" which is more optimal for interpretation. The two-step methodology (LDA and pyLDAVis) has various advantages: (1) lowers bias by machine filtering and sorting of raw data versus human filtering and sorting of raw data, (2) saves filtering and sorting time, (3) yields not just "raw data" but more accessible "information" from the topics in their LDA topic model, (4) the visualization system allows users more flexibility to explore topic−term relationships by

using relevance to better understand a fitted LDA model [61], (5) produces a simple HTML technology file which is easily shareable, and (6) the HTML technology facilitates the development of networked relationships and rapid deployment of intelligence which is considered the adoption of best practices in the development of a common intelligence based agenda [62].

A search of the literature related to the use of pyLDAVIS technology for visualization of LDA output reveals that the use of this tool is extremely limited. Li et al. (2018) was the first to use LDA and pyLDAVIS to extract and visualize consumer preferences of iPhones using a Chinese online Q and A community blog [63]. The study collected almost 50,000 answers under the discussion topics of "iPhone 7" and "iPhone X". This was the first study to attempt extraction from a blog, which requires more advanced technology to handle the user-generated contents due to the comments being longer in length than the usual short-texted posts you find on microblogs [64]. Hagen et al. (2019) subsequently used LDA and LDAVIS to extract and examine petitioning data for policy-making purposes and found the tool to be of great value for visualizing and enhancing the interpretability of results. These two studies thereby provide evidence of the usefulness of using visualization tools in interpreting data.

4. Research Design and Case Selection

4.1. Case Study Approach and Industry Selection

Our case study approach is categorized as an illustrated case study [65]. Illustrative case studies are descriptive and utilize one or two instances to demonstrate what a situation is like. The limited number of cases facilitates data interpretation especially when the research is unfamiliar. When using the illustrative case method, it is recommended the cases selected adequately represent the situation yet not be too diverse., therefore we selected two well established industries in the consumer product industry.

In this research we combine the illustrated case study method with an action research approach. Action research is known as a holistic approach to problem solving and is the type of method chosen when the study involves (1) real situations; (2) the research is unfamiliar; and (3) the research requires flexibility [66]. Participants in an action research project are coresearchers and collaborators in the research process which allows for a greater role of the researchers in defining the issues addressed to include the perceptions within the context of the study. Research collaboration is important as it diminishes the ability of an individual researcher to control the processes and outcomes. In action research it is important to not only describe "what" was done but also "how" it was performed.

In our study, participants are business researchers with academic and practitioner experiences in business ownership. The role of each researcher is as follows: Researcher #1—facilitator and responsible for the communication and coordination of all participants; Researcher #2—synthesizer and responsible for the fostering of reflective analysis among the participants; Researcher #3—planner leader. All researchers were requested to report any personal biases up-front. None of the researchers had any financial or personal ties to the two industries selected. As common to qualitative research, documentation was kept on the data collection and analysis process to include each step of the coding of the data and how deductive logic was applied to the observations. All parties were consulted during the research process, decisions were made about the direction of the research and probably outcomes were collective, all researchers were able to influence the work, and the development of the work was communicated, visible and open to all.

One of the main limitations in case study and action research is the lack of validity and generalizability. One of the ways to strengthen the validity and generalizability is through triangulation of data sources and triangulation of research methods [67]. In our study we gathered our data in a two-step process: (1) extracting data and information from OGD using the LDA and LDA visualization techniques and (2) using the results of the first step to guide the direction of our environmental scanning using other data sources (i.e., academic industry journal articles, industry documents, and other open source industry specific

information). Throughout the project all researchers searched for patterns in the empirical material to determine consistent or inconsistent information obtained from each step.

We used a two-case study approach to identify the competitive markets for the lumber and footwear industry. The case study approach is appropriate when the area of study is relatively new and when the aim of the research is to observe a phenomenon within its context, and the impact of the context to the phenomenon is not evident [67]. As our study performs exploratory research on the usage of OGD to create business intelligence, which is a relatively unexplored area of research, the use of case studies is appropriate [68]. We specifically selected a two-case approach because it permits us to examine value creation from two different industries and provides a richer viewpoint of how OGD data can be used to create business intelligence [33]. The advantages of the two-case method are more robust findings, higher generalizability, and less bias than if only one case were analyzed. This approach is consistent with prior studies who use multiple case studies to examine the adoption of OGD [14,15,33,69].

Our selection of the lumber and footwear industries was based on the notion that they comprise two of the five leading global industries that rely on business intelligence for gaining a competitive advantage in their respective marketplaces [70]. Additionally, the McKinsey Global Institute (2013) identified the market of consumer products as an area where OGD can be used to deliver value by improving product design and manufacturing, enhancing store operations and increasing targeted marketing and sales towards consumers [71]. Selecting these industries thereby allowed us to compare two cases within the consumer products market to find differences and similarities in the way OGD was used, consistent with Kaasenbrood et al. (2015) [14]. Further research on these two industries indicated that they are markets in high demand, growing quickly, of importance in multiple countries, and typically involve private organizations thereby worthy of examining in our study.

The lumber industry is an industry that impacts economies worldwide as it is a commodity that is in high demand but that is limited in resources. Anecdotal research suggests that the United States has 5 percent of the Earth's population, yet consumes 28 percent of the earth's wood products; mostly for construction of homes and commercial buildings [72]. A study examining the value of this resource found that the availability and cost of this raw material, along with the ability to tailor this product to customers' needs, were most important in the lumber industry [73]. Given the importance of the lumber industry to the economy, the critical need for this product, and the effect the availability of raw materials and innovativeness of wood products can have on the competitiveness and profitability of this industry, we chose to include it in our study.

The global footwear market generated $207.6 billion in 2018 and is expected to grow by 3.8 percent by the year 2025, mainly due to rising consumer disposable income in the Asia Pacific market [74]. It is composed of the athletic footwear and nonathletic footwear categories; with the fastest growth in athletic footwear. According to Scott (2006) the global footwear market is highly competitive and fragmented with a few major players and numerous smaller players, including designers, marketers, manufactures and retailers [75]. The competitive forces leading to collaboration are so intensely present in footwear because agglomeration is critical to competitive success. The exponential growth of the athletic footwear industry worldwide and the highly competitive environment was a motivating factor in selecting it as the second industry in our study.

4.2. Sample

Our sample was obtained from marketing intelligence reports gained from a U.S. Government approved free open resource data site: the USA Government/Department of Commerce Website: [76]. https://developer.trade.gov/market-intelligence.html (accessed on 1 August 2018). The reports used in this research paper were drawn from [76] Market Intelligence API. Application Program Interface (API) is a set of software instructions and standards that allows machine to machine communication and can be compared to websites

that use widgets to share a link on Twitter or Facebook. API contains protocols and tools for application development of "live" public data [9]. The "live" element means developers do not have to write their own codes from scratch [77]. We use business intelligence gathered from API because it provides metadata for Country Commercial Guides and other market insight reports that are produced by International Trade Administration (ITA) trade experts. ITA commercial officers that are stationed around the world publish these authoritative reports in conjunction with Foreign Service officers from the State Department. The API only provides the metadata and links to the reports, not the reports themselves. The output format for this API is JavaScript Object Notation (JSON). This data set is updated daily (accessed on 1 August 2018).

The Market Intelligence API provides helpful information for learning about various products' potential in specific markets, which can be used for market-entry strategy formulation. Two report types are available and are categorized by country and industry: (1) Country Commercial Guides, and (2) Market Insight Reports. Businesses use these reports to augment other intelligence that are used internally or are provided to customers. Search tags can be used to pull only the country, region, and industry information of interest. The API also provides the published dates, and it is recommended to use the last_published_date field to order the information display with the most recent updates on top (see Appendix A).

4.3. Methodology

Given the recent call for further research on new methods of data visualization related to the business intelligence delivery process [6,19] and that very few studies have utilized pyLDAVIS [6,78], we combine the use of LDA for topic modeling with pyLDAVIS for data visualization in order to transform OGD data into business intelligence for identifying market opportunities. The LDA technique is used for document topic extraction as it attempts to do the following: (1) identify a set of topics; (2) associate a set of words with a topic; and (3) define a specific mixture of these topics for each document. LDA uses an unsupervised Bayesian learning algorithm to effectively capture topic specific dimensions and there are no assumptions about distribution or content of the topic dimensions. This tool can complete many steps of textual analysis with minimal human intervention, even labeling dimensions, and is generalizable as it is considered reliable in processing large and unstructured data in creating realistic meaning.

We initially identified the data sets from specific government-generated data content obtained from the International Trade Administration website. Articles with the keywords "lumber" and "footwear" were used. The actual articles were downloaded using the URL of the data and these were located into a single file which became the "corpus" (see Appendix B for an example). LDA was then used to extract themes (topics) related to the footwear and lumber industries. Each theme is composed of topic words which are sorted by relevant rankings that appear most frequently in the documents. Following Hagen (2018), we selected 30 topics (most relevant terms) to be displayed for each industry. The top three country rankings for each industry were then identified from the most relevant and highly associated terms for each country. Specific industry, country, and additional term association information for the top three countries were also reviewed. (See Figure 1 below for LDA Technique Model and Appendix C for steps taken to extract the data).

To assist in visualizing the results, we then used pyLDAVIS, a topic-modeling visualization tool. This tool permitted us to visualize the 30 topics that were ranked and the list of countries that were identified from the topics for lumber and footwear. The left side of the figure displays the topics in circles, and if you hover over a topic (indicated in red), a bar graph on the right side of the figure displays the key words based on frequency that are relevant to the lumber or footwear industry. The blue bar indicates the "corpus-wide frequencies of each term" and the red bar the "topic-specific frequencies of each term" (Sievert & Shirley, 2014, p. 68) [61].
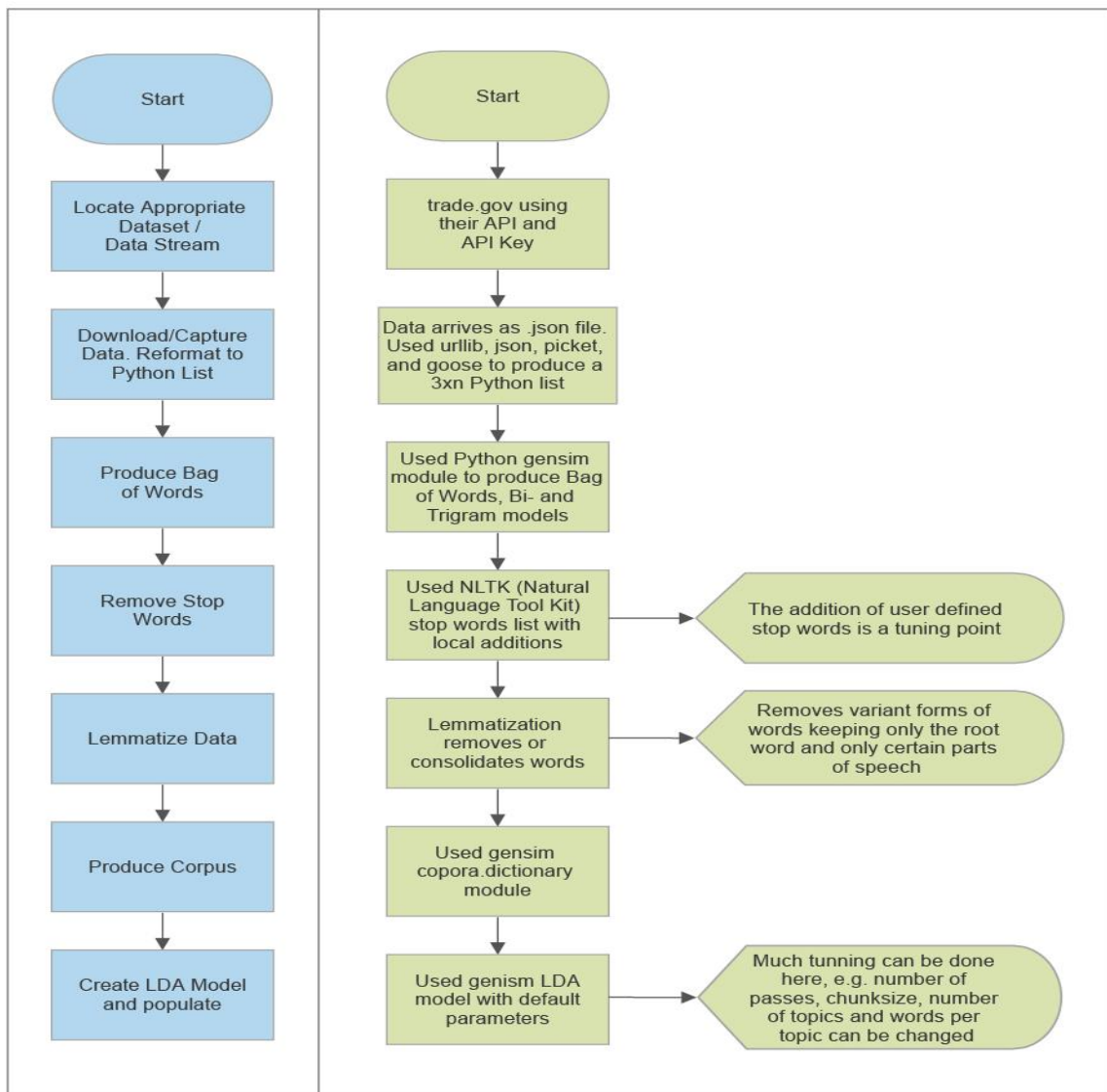
Figure 1. Overall model of LDA technique.

Figure 2, below, shows how the information related to lumber topics can be visualized and interpreted. Topic 2 indicated by the red circle, suggests that it is mainly about Poland since the keys "Poland" and "Polish" are found, and that Poland is considered the second highest ranking country where lumber is in demand. The bar graph shows how much of a specific key is located in a topic, indicating the key "soybean_meal" as it is found in almost no other topic, as represented by the red almost covering the entire bar. It is important to highlight that the interpretation of the data is facilitated by the use of the visualization tool. By examining the representations of relations between the topics, and the relevant words that represent a topic we are able to identify the size of the circles, the closeness or distance between the circles, as well as the level of importance of each term to a topic. Lastly, having identified the relevant topics and terms we performed an environmental scanning to validate the information obtained and identify relationships and possible marketplace opportunities (see Section 5).
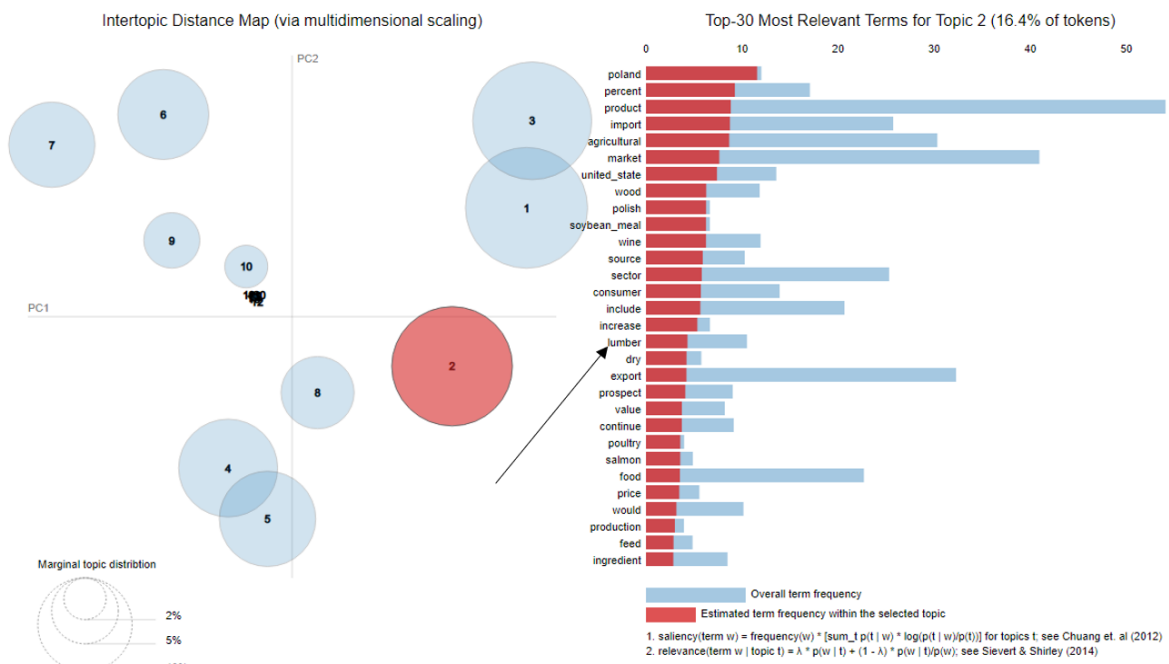
Figure 2. Lumber: Topic 2.

5. Results

5.1. Identification of Topics in the Lumber Industry

Using LDA to search for the key term "lumber" the top three countries associated with lumber and the top 30 most relevant terms (i.e., product, food, market, agriculture, etc.) associated with lumber were identified. Our analysis revealed that the top three countries for lumber, in order of importance, are: (1) Latvia, (2) Poland, and (3) Estonia. The results of the Lumber analysis can be seen in Figures 2–5 and illustrate how the exploratory data sets using "lumber' were converted into quantifiable perceptual maps based on key semantic terms using the LDA program. Figure 3 shows that Latvia (labeled Topic 1) is the top country associated with the lumber industry. It also shows that the key word "farm" is found mainly in this topic indicating a relationship between Latvia, lumber and farm(ing). Other key words of importance are "agricultural", "market", "export", and "opportunity" as they have a higher frequency rate within the topic. Figure 4 reveals there is a relationship between Estonia (Topic 3), lumber, the internet (several keys) and the US Department of Agriculture. A few other relevant words are "market", "trade", "food" and "product". Lastly, Figure 5 provides a representation of the various countries each topic represents, numbered by level of importance. These relationships were determined by visual inspection of each topic by hovering over the topic circle and looking for keys that were country specific.

Topic 1 is highlighted, and the inspection shows the only country key is "Latvia".

Topic 3 indicates that there is a relationship between Estonia, lumber, the internet (several keys) and the US Department of Agriculture.

The figure above shows which countries each topic represents. These relationships were determined by visual inspection of each topic by hovering over the topic circle and looking for keys that were country specific.
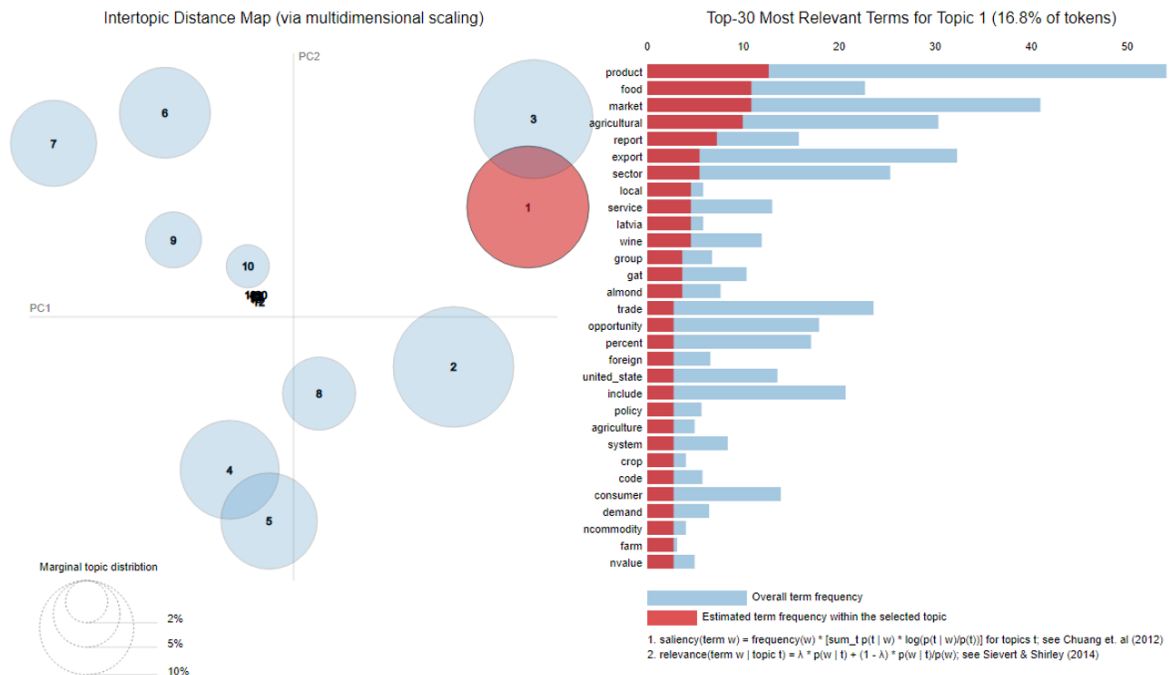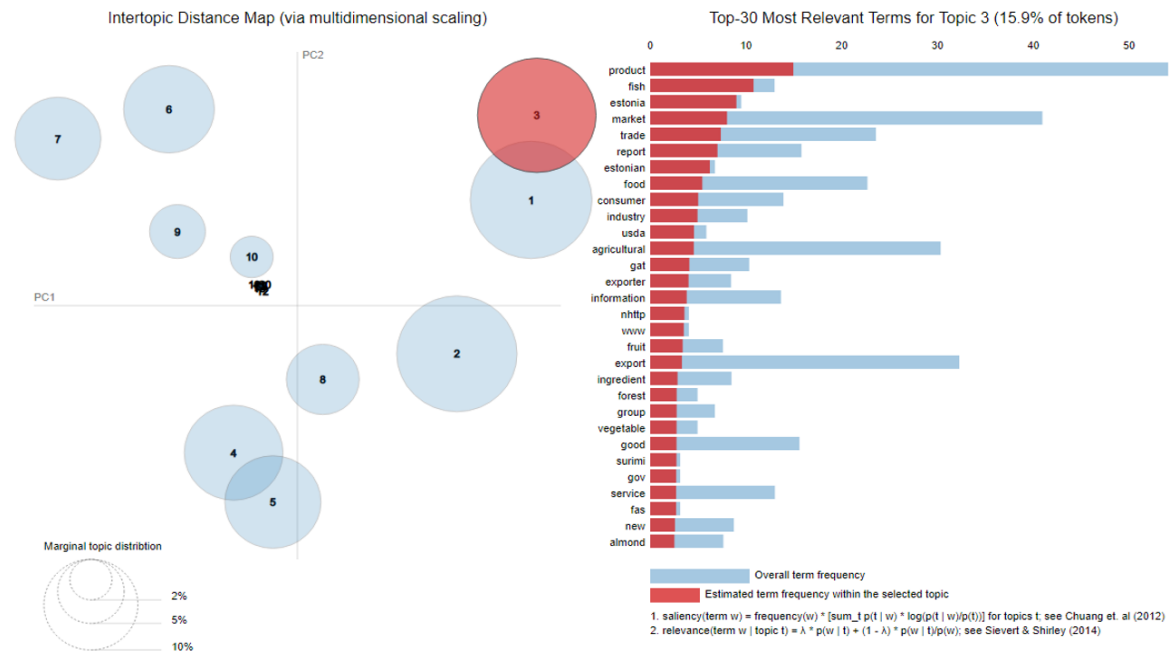
Figure 3. Lumber: Topic 1.
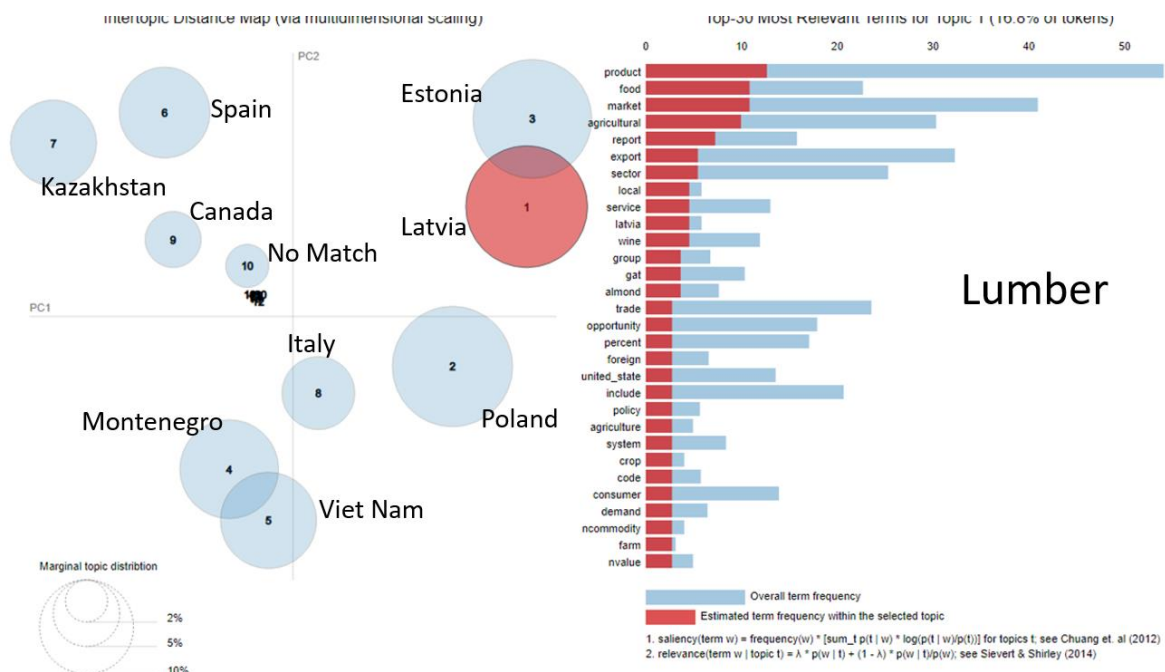

Figure 4. Lumber: Topic 3.

Figure 5. Lumber: Top 3 Countries.

5.2. Identification of Topics in the Footwear Industry

Using the key term "footwear" we identified the top 30 most salient terms (chain, Ukraine, equipment, etc.) as well as the top three countries where the footwear industry is of prime importance. Ukraine emerged as Topic (country) 1 while New Zealand appeared as Topic 2, and Mexico as Topic 3. Figure 6 shows the base map with no topics selected. On the left are the topics of which only 7 topics were "strong" enough to generate circles. The remaining 13 topics are clustered atop each other in the small black dot near the origin. The right side of the image displays the top 30 "salient" terms. Figures 7–9 provide information on the top three countries associated with footwear and their relevant terms.

Figure 7 indicates that Ukraine was selected as Topic 1 (country 1). The right side of the screen displays the salient terms rearranged by frequency within this topic. The red (dark) bars indicate the "relevance" of the term within the topic. The lighter bars indicate the overall importance of a term to all topics. In this image, the second most relevant term is "Ukraine" and the fourth is "Ukrainian". The dark bar (the importance of the term to the topic) almost completely covers the light bar (the term frequency for the entire document) indicating that this topic and the relevant terms are valid for Ukraine. Terms that were unique to the topic and Ukrainian footwear were "pharmacy", "distribution", "outlet", "Romania", "intellectual", "property", "fmco", "state" and "cargo". They suggest that distribution and property rights may be an influential factor for the footwear industry in Ukraine.

This figure shows the base map with no topic selected. On the left are the top 20 topics. Only 7 topics were "strong" enough to generate circles. The remaining 13 are clustered atop each other in the small black dot near the origin.

Figure 8 displays New Zealand as Topic 2 (country 2) and the key terms (as determined by dark bar coverage of light bar) of "sporting", "labeling", "equipment", "fitness", "sport", "label", "gym", "athletic" and "sub". These results are a strong indication that recreational footwear in New Zealand is of high importance. Interesting terms that suggest the need for further investigation into possible market opportunities are Dominican republic, Dominican, ad valorem, basis, tax. Figure 9 shows the LDAVis results for the third country of importance: Mexico (Topic 3) as indicated by the red circle. Highly relevant terms of "document", "documentation", "registry", "annex", "Salvador", and "certificate"

appeared. These terms were mostly unique to Mexico as shown by the coverage of the dark bar over the light bar and indicated possible legal or regulatory issues with the footwear industry in this country. Other interesting terms identified were "textile", "government", "price", "exporter", "complete" and "broker".
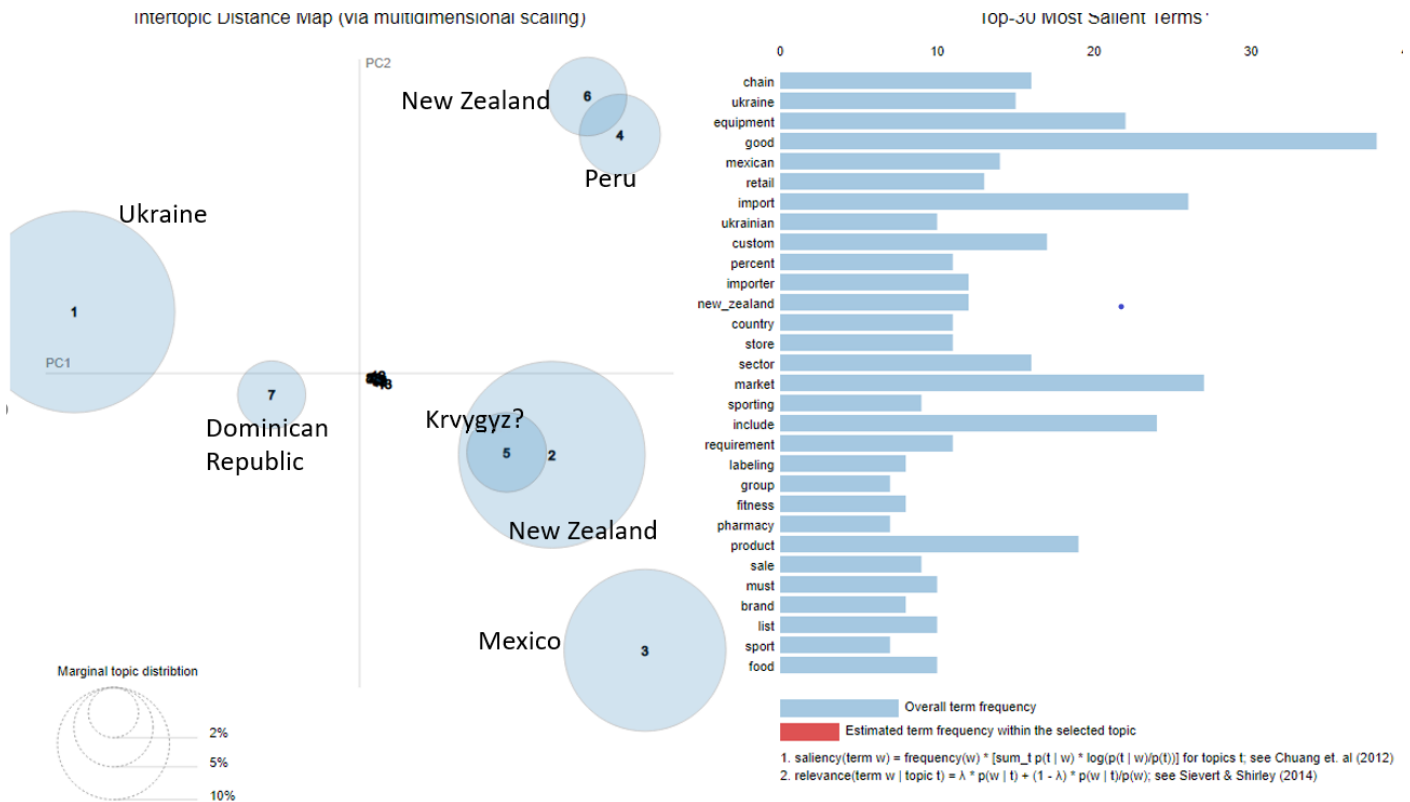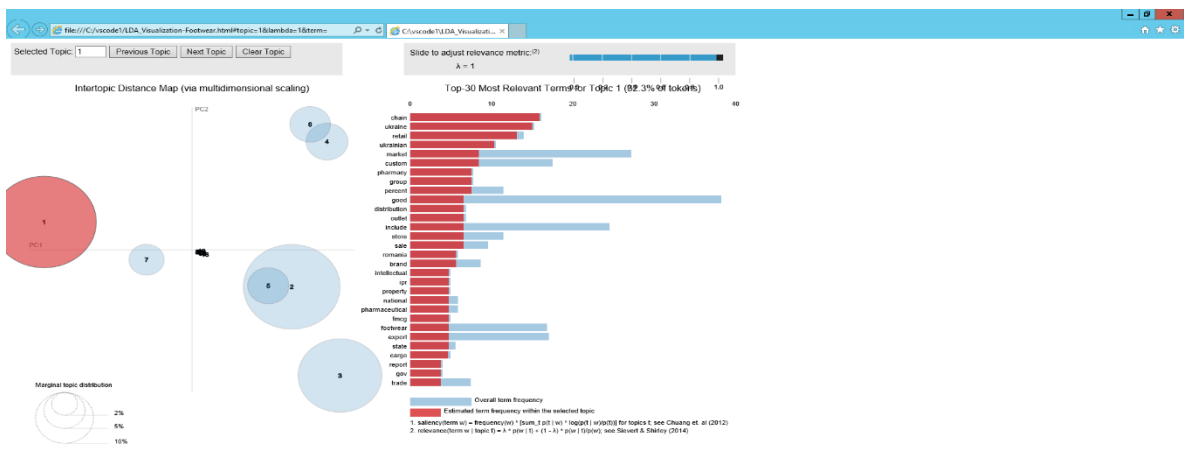

Figure 6. Footwear.

Figure 7. Footwear: Topic 1. This image shows the LDAVis screen with Topic 1 selected.
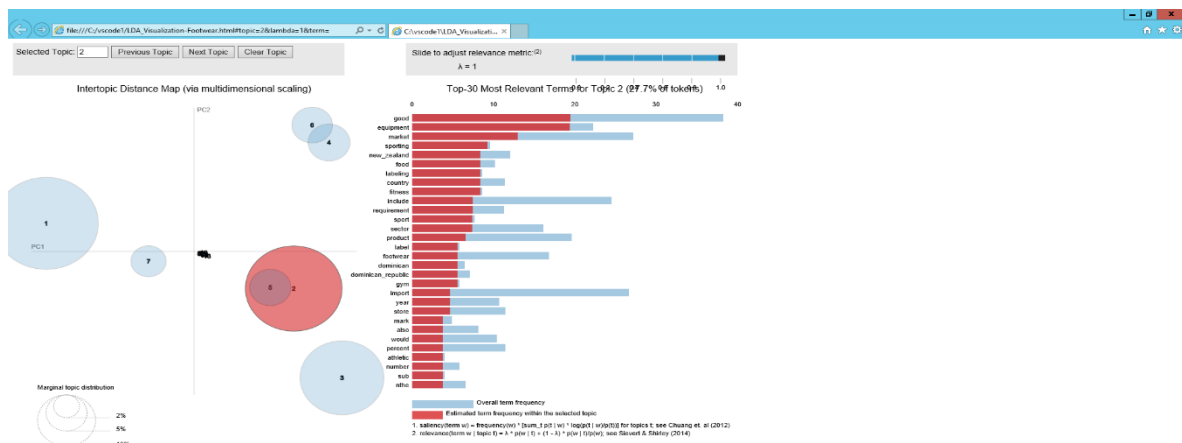
Figure 8. Footwear: Topic 2. This image shows New Zealand and key terms to the right.
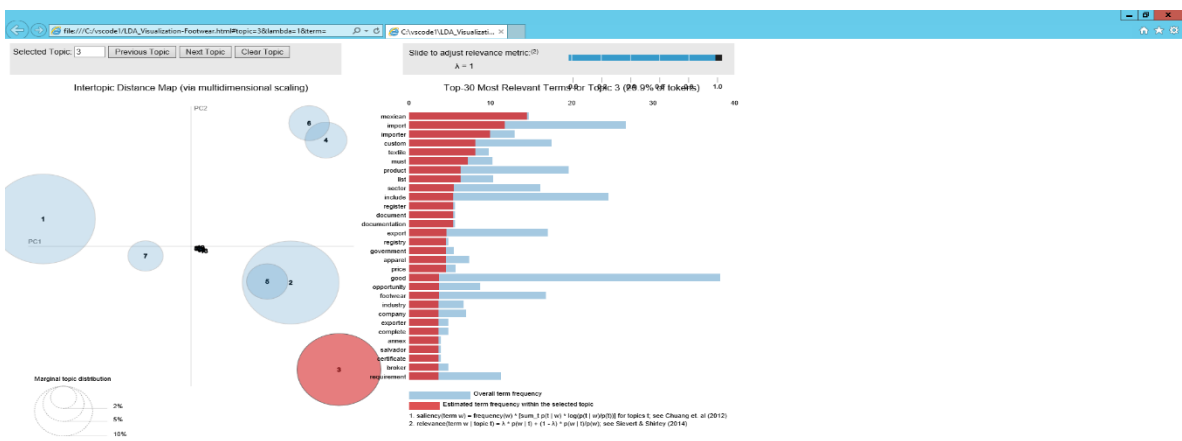

Figure 9. Footwear: Topic 3. This image shows Mexico as the third ranked country.

5.3. Environmental Scanning for Business Intelligence

The use of LDA and LDAVis allowed us to extract relevant terms related to the lumber and footwear industries and to identify the top three countries where these products are of high importance as well as other relevant terms associated with each country. In order to apply the information obtained through the use of these two tools and identify whether these countries could be potential competitive markets, we performed an environmental scanning of these countries and various terms identified using public sources as well as the CPI Index. The CPI index covers three main dimensions of global competitiveness and development: capacity to produce and export manufactured goods; technological deepening and upgrading; and dimensions of global world impact. Using the highly relevant terms that appeared for each of the top countries in the lumber and footwear industries we gathered additional detailed knowledge in order to assess the likelihood that each country could offer opportunities for organizations to gain a competitive advantage in these industries. The details of the assessments are discussed below.

5.3.1. Lumber Industry

Upon investigating the terms "lumber", "market", "export" and "farming" in Latvia, our research revealed that Latvia's forests have doubled in the past 100 years with half of these forests consisting of Scots pine and Norway spruce. Forests cover approximately 52% of the country's territory and the demand for lumber is growing exponentially due to a growth in the prefabricated wood (panel) modular house production industry. Production capacities are focusing on growing demand in North European countries. Latvia's

legislation on forestry is also among the strictest in Europe and firmly regulates wood harvesting. It has a well-developed wood processing industry therefore timber and wood products are among the country's most important exports. Latvia has one of the highest investment rates in Europe in the wood products markets due to its vast resources and competitive labor force with significant investments made in the production of particle board and oriented strand board [79].

Research on Poland revealed that forests cover about 30.5% of the country's land and are estimated to grow to 33 percent in 2050. Approximately 81 percent of the forest land is owned by public institutions and 19% by private owners. The country is the 10th largest world producer and the 4th largest world exporter of furniture. Total trade in forest products between the U.S. and Poland reached U.S. $42.6 million in 2016 and continues to grow [80]. IKEA, a large manufacturer of furniture in Europe, imports its wood from 50 different countries, however a large proportion comes from Poland, Russia, Lithuania and Germany [81].

Estonia shows that it has a booming economy and has become a haven for start-ups and new technology. Forests cover about 50% of the territory of Estonia and, despite its relatively small size, display a great variety in forest types. Two main types are forests growing on mineral soil (about 70%) and so-called swamp forests (about 30%), and the most common tree species are pine, birch and spruce. The forest industry is one of the most important sectors of the economy and ranks 6th in abundance of forestry coverage in Europe. Estonia is highly competitive in domestic and foreign markets, however in the last few years the domestic market has become stronger as a result of the furniture industry expanding locally [82]. Estonia is also famous for its e-solutions and digitalization. Estonian forest management companies use forestry software to conveniently manage forests, forest resources and plan various types of forest operations. The described systems can be used in an integrated manner which means that data and documents move between different parties (provider, carrier and client) digitally. Because people do not need to enter data repeatedly and manually, they save a lot of time and avoid potential errors. This increased use of technology provides Estonia with a competitive advantage [83].
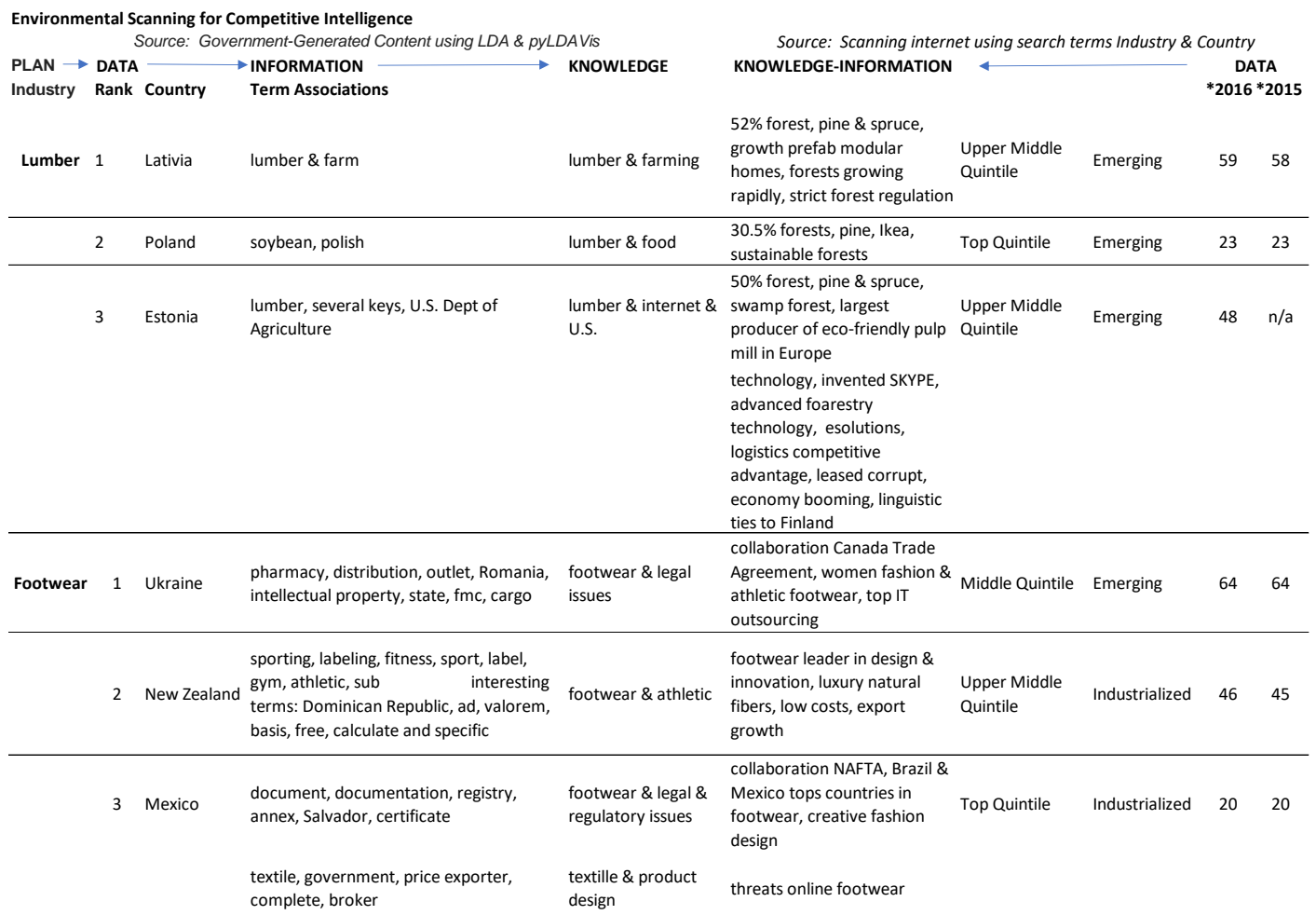
5.3.2. Footwear Industry

Having obtained additional knowledge on the three top countries identified within the lumber industry we proceeded to obtain further information of the top three countries in the footwear industry. Ukraine appeared as the top country of importance and further research reveals that the Canada-Ukraine Free Trade Agreement (CUFTA) has boosted Ukraine's economy by eliminating tariffs between these two countries [84]. Through the CUTIS program (Canada-Ukraine Trade and Investment Support (CUTIS) Project) Ukraine has benefited from increased exports to Canada in the clothing, footwear, furniture and IT services market [85]. Additionally, women are the main consumers of footwear in Ukraine and affordability is an important influencing factor for consumers. Within the athletic footwear market, brand imitations hold a strong presence in Ukraine even though multinational brands such as Adidas are beginning to enter the market [86]. Overall, this country provides many opportunities to generate business in the footwear market.

The highly relevant terms identified through LDA and pyLDAVIS related to New Zealand were centered around sports, fitness and athletic wear. Further research on New Zealand revealed that the labor market in New Zealand is more deregulated, flexible and less unionized than other countries in the region and ranks in the top 10 countries in the world for efficiency. New Zealand's workforce is highly educated, and there is in fact a surplus of skilled labor. The creative, digital and technology sectors are strengths. Many Australian organizations have already placed their operations in New Zealand, either directly, or with outsourcing companies. The speed and reliability of New Zealand's telecommunications infrastructure is world-class, as are data security and risk management frameworks. The World Bank rated New Zealand as the "easiest country to start a business" in its 2017 report [87]. New Zealand is also the world leader in the design and innovation

of luxury fibers for use in the fashion industry and have developed an industry worth over $100 million. Despite these accomplishments, imported clothing and footwear are often cheaper than New Zealand made goods and typically manufacturing costs are cheaper offshore, thereby presenting a good opportunity for competitors to enter the footwear market [88].

Research on Mexico revealed that this country is among the top 10 countries with the largest number of footwear exports globally. The footwear industry in Mexico employs around 80,000 people in 2800 factories and focuses on niche products that generate high sales figures nationally and internationally. Labor costs in Mexico are lower than other countries, such as the United States, Canada and Europe, permitting globally recognized brands such as Nike and Under Armour to form partnerships with local suppliers in Mexico to produce footwear. Of additional benefit is that technological advances, innovation and experience in designing footwear have transformed Mexico into a successful leader in the footwear manufacturing industry. The footwear industry in Mexico is undoubtedly a huge business opportunity for those who can enter the market [89].

Figure 10 provides a summary of the results of our environmental scanning for business intelligence for both the lumber and footwear industries.

Environmental Scanning for Competitive Intelligence

PLAN → DATA →→→→→ INFORMATION →→→→				KNOWLEDGE	KNOWLEDGE-INFORMATION ←←←←			DATA	
			Source: Government-Generated Content using LDA & pyLDAVis		*Source: Scanning internet using search terms Industry & Country*				
Industry	Rank	Country	Term Associations					*2016	*2015
Lumber	1	Lativia	lumber & farm	lumber & farming	52% forest, pine & spruce, growth prefab modular homes, forests growing rapidly, strict forest regulation	Upper Middle Quintile	Emerging	59	58
	2	Poland	soybean, polish	lumber & food	30.5% forests, pine, Ikea, sustainable forests	Top Quintile	Emerging	23	23
	3	Estonia	lumber, several keys, U.S. Dept of Agriculture	lumber & internet & U.S.	50% forest, pine & spruce, swamp forest, largest producer of eco-friendly pulp mill in Europe	Upper Middle Quintile	Emerging	48	n/a
					technology, invented SKYPE, advanced foarestry technology, esolutions, logistics competitive advantage, leased corrupt, economy booming, linguistic ties to Finland				
Footwear	1	Ukraine	pharmacy, distribution, outlet, Romania, intellectual property, state, fmc, cargo	footwear & legal issues	collaboration Canada Trade Agreement, women fashion & athletic footwear, top IT outsourcing	Middle Quintile	Emerging	64	64
	2	New Zealand	sporting, labeling, fitness, sport, label, gym, athletic, sub interesting terms: Dominican Republic, ad, valorem, basis, free, calculate and specific	footwear & athletic	footwear leader in design & innovation, luxury natural fibers, low costs, export growth	Upper Middle Quintile	Industrialized	46	45
	3	Mexico	document, documentation, registry, annex, Salvador, certificate	footwear & legal & regulatory issues	collaboration NAFTA, Brazil & Mexico tops countries in footwear, creative fashion design	Top Quintile	Industrialized	20	20
			textile, government, price exporter, complete, broker	textille & product design	threats online footwear				

*CPI Index 2018 - Report on Competitiveness and Industrial Development (Accessesd 1 June 2019)

Figure 10. Environmental Scanning for Business Intelligence.

6. Discussion and Conclusions

The purpose of our study was twofold: (1) to demonstrate the value of OGD data as a useful source of information for any organization; and (2) highlight the importance of

using topic modeling and visualization technologies to extract and visualize data to assist in generating business intelligence. We illustrated how the extraction and visualization of data on lumber and footwear industries can reveal valuable information related to potential market entry opportunities. We also highlighted the usefulness of performing environmental scanning to validate the information extracted. This process permitted us to delve deeper into the various characteristic of each topic and identify differences and similarities between them so as to assess what market opportunities might be available in these countries to help formulate strategic decisions.

Our first step involved obtaining OGD data for the lumber and footwear industries and extracting that data into useful information through the use of LDA for topic modeling and pyLDAVIS for visualization. The result was the identification of the top three countries where the lumber and footwear industries are of prime importance. The tools also highlighted important relevant terms that can be important for identifying opportunities in these two industries. We subsequently used the information obtained to perform an environmental scanning analysis so as to identify potential market opportunities in these industries.

The results revealed that the top three countries associated with lumber were Latvia, Poland, and Estonia. The importance of Latvia in relation to lumber is not surprising given the large percentage of land that is invested in forestry. The high regulation of this industry and the high level of wood exports could be seen as a hindrance to competing in this market or gaining a competitive advantage however, the vast resources and competitive labor force (low cost) available in Latvia could be a motivating factor for consideration in this market. Poland also provides potential opportunities for consideration since its forestry industry is growing, it is the 10th largest producer of lumber worldwide, and has 100% sustainable forestry practices. Poland is currently a primary source of lumber for IKEA but since it is an emerging economy and a low-wage country it presents opportunities for organizations to import this product at a more reasonable cost. Lastly, Estonia's involvement in high technology to digitalize forest management provides lower cost opportunities and could be an important consideration for competitors to enter the market.

Our findings related to the identification of Ukraine, New Zealand and Mexico as important markets for footwear apparel are quite interesting. All three countries have different market targets and their economies influence the quality of the products offered in footwear. For example, Mexico has become one of the top producers of footwear due to its low labor costs and high technological advances in footwear design. An organization wishing to compete in this industry would want to consider and assess the viability of benefiting from Mexico's low cost and technological advances. The close proximity to the United States also presents potential opportunities for footwear companies to sell their products to consumers in that region. Ukraine, is an untapped market, given the popularity of footwear among women and the lack of brand name athletic companies doing business in this part of the world. Lastly, New Zealand offers a much more sophisticated market since people are highly educated, and technologically advanced. One advantage of doing business in this market is the flexibility of starting a business there and catering to their sophisticated market. Overall, the exercise of using OGD data and LDA and pyLDAVIS to extract and visualize the data permitted us to identify markets (countries) with potential opportunities to compete in the lumber and footwear industries. Complementing this information with environmental scanning we were able to further examine the advantages and disadvantages of doing business in the three markets identified within each industry. Similarly, private organizations can find value in using OGD and apply these topic mining and visualization tools to help them assess the viability of doing business in various markets and gaining a competitive advantage.

This study makes important contributions to the open government data literature by examining the value that OGD can provide to private organizations; an area of research that is relatively unexplored. Prior studies have mainly focused on how government agencies share open data and its societal impact on citizens and other public entities [15,28,90].

Private organizations are distinct from citizens and public entities in that their primary focus is on generating profit, therefore the value and use of OGD differs because they are not typically concerned with societal issues. Rather, they view OGD from the perspective of how it can add value to their organization and assist them in achieving their financial and strategic objectives [14]. Our study extends research in this area by demonstrating the usefulness of OGD to private organizations for identifying market opportunities in various areas of business so that they may generate a competitive advantage and enhance profitability.

Jetzek et al. (2014) proposed an OGD value generation framework (see pg. 106) to highlight how OGD can be used to generate economic (monetary) and social value (improvement in the lives of citizens and society). The dimensions of the model included (1) transparency of government; (2) citizen participation/collaboration; (3) efficiency/effectiveness; and (4) innovation [8]. They contended that each of these four dimensions had the ability to generate a combination of social and economic value. Applying this framework to our study, we discuss the value of the use of OGD to generate business intelligence for the purpose of identifying competitive markets.

Our study reveals that the use of OGD can assist businesses in obtaining information on resources, technology, and industry competitors. This information is of economic value for businesses as it permits them to complement their in-house capabilities with additional BI as well as understand potential markets and build new data-driven products. The transparency of the information will be a key driver to generate value so governments will need to ensure that the information made available is transparent, accurate and that it reduces information asymmetry [8]. Improved transparency permits a more equitable allocation of resources which leads to the creation of economic and social value.

The use of OGD by organizations and citizens can also lead to efficiency and improvement in utilization of resources as well as improved benefits to citizens. Our study indicates that private companies can use the information obtained to delve into other market opportunities by taking advantage of resources that are lower in cost and maximize economic value. This efficiency also leads to social value since citizens are able to benefit from the improved products at more reasonable prices, thereby increasing their quality of life. The use of OGD also permits citizens to monitor government activities and public budget expenditures and how it impacts them, in addition to reducing the likelihood of corruption occurring at the government level.

OGD can also be a robust driver of economic growth through business innovation, business creation and efficiency. Organizations can gain more precise information on customer preferences, make export/import or plant location and expansion decisions, and identify areas of innovation and future opportunities. For example, businesses can build new innovative services and applications at minimal cost, similar to the two cases exemplified in Section 2 of the paper. The free use of open data can also be a driver for change and social technological advances. By gaining access to OGD, participants can provide opinion on public policy, and provide ideas and solutions to local and government issues. This permits individuals to voice their opinions and contribute to the generation of new ideas as well as create resources through the sharing of information. This leads to economic value for businesses as they are able to assist in solving societal issues while generating a profit.

Although the value of government data is difficult to measure, the ability to access OGD can also provide a competitive advantage for businesses since OGD can be used to supplement internal data and assist in making strategic decisions [7]. One caveat though is that the success of using OGD is dependent on organizations having in-house capabilities and resources such as knowledge, skills, and the ability to combine internal and external resources. Information Technology (IT) (i.e., internet connection, cloud computing, processing, linking and other tools); Information and Data (i.e., database with open data sets, company database, company products and services); and Human Resources (i.e., computer skills, finding and accessing open data, tool selection and use, data and result

interpretation, stakeholder network management) are all critical organizational resources that are necessary for making sense of government data. Without one of these three organizational resource categories it becomes more difficult to gain unique competitive benefits and gain a competitive advantage through OGD. The ability to use OGD and to integrate it into a firm's existing value proposition will therefore be vital to sustaining a competitive advantage [91].

This study has valuable implications for managerial practices, especially small and medium enterprises (SMEs), as the marketplace shifts towards a more open interconnected world and managers are required to rethink how business intelligence is generated and appropriated for their organization [8]. The amount of data created by the public sector is immense and is both freely available and of low cost. SMEs who are not able to invest in expensive business intelligence systems will find OGD a valuable source of data and information. Use of this data can serve as a starting point for SMEs to create added value, and identify new innovative business models. We recognize that government data is generated by the public sector for specific objectives [8], nevertheless as long as organizations take this into consideration when using the data, the open network and easy accessibility of this data source presents extensive opportunities for small businesses.

We also highlight the use of LDA and pyLDAVIS methodology as practical, affordable and easy to apply methods for generating business intelligence in various business contexts. By visualizing data, it is possible to discover trends and associations between concepts, which can provide a greater understanding of the ideas or concepts being explored [24]. Additionally, it can help organizations answer questions such as: how can this data be used to enter or expand into new markets? How can it assist in creating innovative products? Ultimately, the use of these methods can be used to bring about new innovative solutions of economic value, monitor their competitive environment, and assist strategic management in formulating decisions at a minimal cost to the organization.

We recognize our study is not without its limitations. First, we examined two specific industries within the consumer products market therefore the findings are not generalizable to other contexts. Additionally, we used only one genre of text-mining techniques, LDA and pyLDAVis, to extract terms from an open source government data site for the purpose of obtaining business intelligence. Future open source data gathering studies can further advance the field of open source intelligence by comparing different types of techniques and integrating multiple, dynamic data sources, including time-varying covariates and the combination of exploratory topic models with powerful predictive marketing models [18,48,92,93]. Scholars could also explore and compare how the use of OGD varies among different countries and affects their likelihood of adopting OGD. Lastly, triangulating the methodology, such as including a survey to assess the external validity of the text-mining and visualization outcomes, would allow the comparison of the results to data elicited from a survey-based approach [36].

Appendix A

Government-Generated Open Source Data Content

International Trade Administration (ITA) Market Research Library API can provide you with international trade market intelligence called the Market Intelligence API [75] (accessed 1 August 2018). It works similarly to the Market Research Library API and the

Market Intelligence API gives you additional control over your return set with the ability to also limit your search by:

- trade_regions
- world_regions
- first_published_date
- last_published_date

"The International Trade Administration's (ITA) Thesaurus of International Trade Terms is a controlled and structured list of words and phrases used to tag and index information found on the ITA's websites and databases. The thesaurus covers all subjects related to international trade and foreign investment with particular emphasis on exporting, trade promotion, market access and enforcement and compliance.

The thesaurus is structured into six domains or microthesauri:

- Trade Topics
- Industries
- Countries
- World Regions
- Trade Regions
- U.S. Trade Initiatives

The thesaurus is available to the software developer community as a JSON endpoint.

The thesaurus was developed by ITA's staff of international trade specialists, consulting several authoritative sources and vocabularies covering the language of international trade and investment" [75] https://developer.trade.gov/taxonomy.html (accessed 1 August September 2018).

Appendix B

Sample Metadata

https://api.trade.gov/market_intelligence/search?api_key=xxxxxxx&q=lumber

Key Word Search Term

N o t U s e d

"id": "ka0t0000000PYYIAA4","source": "COUNTRY_COMMERCIAL", "title": "Poland - Agricultural Sector","summary": "Sub-sector best prospects: Include the sub-sectors in which U.S. companies would have the best opportunity of exporting.","first_published_date": "2015-10-06","last_published_date": "2018-08-01","url": "https://www.export.gov/article2?id=Poland-Agricultural-Sector","references": "Prepared by our U.S. Embassies abroad. With its network of 108 offices across the United States and in more than 75 countries, the U.S. Commercial Service of the U.S. Department of Commerce utilizes its global presence and international marketing expertise to help U.S. companies sell their products and services worldwide. Locate the U.S. Commercial Service trade specialist in the U.S. nearest you by visiting http://export.gov/usoffices.","url_name": "Poland-Agricultural-Sector","industries": [

URL of Data

Example of one of the retrieved articles using the keyword "lumber" was downloaded using the URL of the data and included with the rest of the articles that used the keyword "lumber". These were located into a single file which became the "corpus". Dataset

extracted from International Trade Administration (ITA) Website, [75] Market Intelligence API (i.e., lumber, footwear).

Appendix C

LDA Technique Model—Steps Taken

Step 1: Locate: Locate key dimensions of search terms from an appropriate government-generated dataset extracted from International Trade Administration (ITA) Website, [75] Market Intelligence API (i.e., lumber, footwear)

Step 2: Downloading: Once identified, reports were downloaded from the specific government forum site, Market Intelligence Website, [75] Market Intelligence API, json format. The data set contains articles pulled from a specific week in time (i.e., week of 08/01/2018).

Step 3: Capture: Data json format is captured and re-formatted to Python List.

Step 4: Bag of Words: Used Python to produce a bag-of-words file

Step 5: Remove Stop Words: Used natural language tool kit (nltk) to remove stop words.

Step 4: Lemmatize Data: Lemmatization removes and/or consolidates words.

Step 5: Produce Corpus: Used gensim corpora dictionary module to produce the corpus.

Step 6: Create LDA Model and Populate: Used gensim LDA model with default parameters.

Step 7: Identification of Semantic Relationships: Using latent Dirichlet analysis (LDA) the exploratory data sets were converted into quantifiable perceptual maps based on key semantic terms.

We use a stand-alone HTML file which allows for easy sharing and users. https://pyldavis.readthedocs.io/en/latest/readme.html (accessed on 18 March 2021) [93].

References

1. Markets and Markets. Business Intelligence Market by Type (Platform, Software, Service), Data Type (Unstructured, Semi-structured, Structured), Business Application, Organization Size, Deployment Model, Industry Vertical, and Region-Global Forecast to 2021. March 2017. Available online: https://www.marketsandmarkets.com/PressReleases/social-business-intelligence-bi.asp (accessed on 27 July 2019).
2. Dishman, L.P.; Calof, J.L. Competitive intelligence: A multiphasic precedent to marketing strategy. *Eur. J. Mark.* **2008**, *47*, 766–785. [CrossRef]
3. Bergh, C.; Benghiat, G. Analytics at Amazon speed: The new normal. *Bus. Intell. J.* **2017**, *22*, 46–54.
4. Thomas, P. What Role does Business Intelligence Play in the Pharma Sector? 2019. Available online: www.asmag.com/showpost/30516.aspx (accessed on 1 June 2019).
5. Janssen, M.; Helbig, N. Innovating and changing the policy-cycle: Policy makers be prepared! *Gov. Inf. Q.* **2018**, *35*, S99–S105. [CrossRef]
6. Hagen, L.; Keller, T.E.; Yerden, X.; Luna-Reyes, L.F. Open data visualizations and analytics as tools for policy-making. *Gov. Inf. Q.* **2019**, *36*, 101387. [CrossRef]
7. Zuiderwijk, A.; Janssen, M.; Poulis, K.; van de Kaa, G. Open data for competitive advantage: Insights from open data use by companies. In Proceedings of the 16th Annual International Conference on Digital Government Research, Phoenix, AZ, USA, 27–30 May 2015.
8. Jetzek, T.; Avital, M.; Bjorn-Andersen, N. Data-driven innovation through open government data. *J. Theor. Appl. Electron.Commer. Res.* **2014**, *9*, 100–120. [CrossRef]
9. Safarov, I.; Meijer, A.; Grimmelikhuijsen, S. Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Inf. Polity* **2017**, *22*, 1–24. [CrossRef]
10. Huijboom, N.; Van den Brock, T. Open Data: An international comparison of strategies. *Eur. J. Epractice* **2011**, *12*, 1–13.
11. Jetzek, T.; Avital, M.; Bjorn-Andersen, N. The Value of Open Government Data: A Strategic Analysis Framework. In Proceedings of the Pre-ICIS Workshop, Orlando, FL, USA, 16 December 2012.
12. Jetzek, T.; Avital, M.; Bjorn-Andersen, N. The Generative Mechanisms of Open Government Data. In Proceedings of the ECIS 2013 Proceedings, Utrecht, The Netherlands, 5–8 June 2013.
13. Hughes-Cromwick, E.; Coronado, J. The value of US government data to US business decisions. *J. Econ. Perspect.* **2019**, *33*, 131–146. [CrossRef]
14. Kaasenbrood, M.; Zuiderwijk, A.; Janssen, M.; de Jong, M.; Bhrosa, N. Exploring the factors influencing the adoption of open government data by private organisations. *Int. J. Public Adm. Digit. Age* **2015**, *2*, 75–92.
15. Magalhães, G.; Roseira, C. Exploring the barriers in the commercial use of open government data. In Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance, Montevideo, Uruguay, 1–3 March 2016; pp. 211–214.

16. Hagen, L. Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Inf. Process. Manag.* **2018**, *54*, 1292–1307. [CrossRef]

17. Larson, D.; Chang, V. A review and future direction of agile, business intelligence, analytics and data science. *Int. J. Inf. Manag.* **2016**, *36*, 700–710. [CrossRef]

18. Reisenbichler, M.; Reutterer, T. Topic modeling in marketing: Recent advances and research opportunities. *J. Bus. Econ.* **2019**, *89*, 327–356. [CrossRef]

19. Park, J.; Fables, W.; Parker, K.R.; Nitse, P.S. The role of culture in business intelligence. *Int. J. Bus. Intell. Res.* **2010**, *1*, 1–14. [CrossRef]

20. Moore, S. News Release. 17 February 2017. Available online: https://www.gartner.com/en/newsroom/press-releases/2017-02-17-gartner-says-worldwide-business-intelligence-and-analytics-market-to-reach-18-billion-in-2017 (accessed on 5 January 2019).

21. Trieu, V.H. Getting value from Business Intelligence systems: A review and research agenda. *Decis. Support Syst.* **2017**, *93*, 111–124. [CrossRef]

22. Haupt, R.; Scholtz, B.; Calitz, A. Using business intelligence to support strategic sustainability information management. In Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologies, Stellenbosch, South Africa, 28–30 September 2015.

23. Open Government Partnership. Available online: https://www.opengovpartnership.org/about/ (accessed on 27 July 2019).

24. Graves, A.; Hendler, J. Visualization tools for open government data. In Proceedings of the 14th Annual International Conference on Digital Government Research, Quebec City, QC, Canada, 17–20 June 2013.

25. Janssen, K. The influence of the PSI directive on open government data: An overview of recent developments. *Gov. Inf. Q.* **2011**, *28*, 446–456. [CrossRef]

26. Janssen, M.; Charalabadis, Y.; Zuiderwijk, A. Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag.* **2012**, *29*, 258–268. [CrossRef]

27. Kucera, J.; Chlapek, D. Benefits and risks of open government data. *J. Syst. Integr.* **2014**, *5*, 30–41. [CrossRef]

28. Albano, C.S.; Reinhard, N. Open government data: Facilitating and motivating factors for coping with potential barriers in the Brazilian context. In *International Conference on Electronic Government*; Springer: Berlin/Heidelberg, Germany, 2014.

29. Wang, H.-J.; Lo, J. Factors influencing the adoption of open government data at the firm level. *IEEE Trans. Eng. Manag.* **2019**, *67*, 670–682. [CrossRef]

30. Altayar, M.S. Motivations for open data adoption: An institutional theory perspective. *Gov. Inf. Q.* **2018**, *35*, 633–643. [CrossRef]

31. Zeleti, F.A.; Ojo, A.; Curry, E. Exploring the economic value of open government data. *Gov. Inf. Q.* **2016**, *33*, 535–551. [CrossRef]

32. Kalampokis, E.; Tambouris, E.; Tarabanis, K. A classification scheme for open government data: Towards linking decentralised data. *Int. J. Web Eng. Technol.* **2011**, *6*, 266–285. [CrossRef]

33. Zimmerman, H.D.; Pucihar, A. Open Innovation, open data and new business models. In Proceedings of the IDIMT 2015—23rd Interdisciplinary Information and Management Talks, Poděbrady, Czech Republic, 9–11 September 2015; Doucek, P., Chroust, G., Oskrdal, V., Eds.; pp. 449–458.

34. Dörre, J.; Gerstl, P.; Seiffert, R. Text mining: Finding nuggets in mountains of textual data. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 15–18 August 1999.

35. Feldman, R.; Sanger, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*; Cambridge University Press: Cambridge, UK, 2007.

36. Netzer, O.; Feldman, R.; Goldenberg, J.; Fresko, M. Mine your own business: Market-structure surveillance through text mining. *Mark. Sci.* **2012**, *31*, 521–543. [CrossRef]

37. Feldman, R.; Fresko, M.; Kinar, Y.; Lindell, Y.; Liphstat, O.; Rajman, M.; Schler, Y.; Zamir, O. Text mining at the term level. In *European Symposium on Principles of Data Mining and Knowledge Discovery*; Springer: Berlin/Heidelberg, Germany, 1998.

38. Rosa, J.A.; Spanjol, J.; Porac, J.F.; Moorman, C.; Lehmann, D.R. Text-based approaches to marketing strategy research. In *Assessing Marketing Strategy Performance*; Marketing Science Institute: Cambridge, UK, 2004; pp. 185–211.

39. Sullivan, D. *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.

40. Linoff, G.S.; Berry, M.J.A. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*; John Wiley & Sons: Hoboken, NJ, USA, 2011.

41. Alamsyah, A.; Rahmah, W.; Irawan, H. Sentiment Analysis Based on Appraisal Theory for Marketing Intelligence in Indonesia's Mobile Phone Market. *J. Theor. Appl. Inf. Technol.* **2015**, *82*, 335.

42. Mäntylä, M.V.; Graziotin, D.; Kuutila, M. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Comput. Sci. Rev.* **2018**, *27*, 16–32. [CrossRef]

43. Kherwa, P.; Bansal, P. Topic modeling: A Comprehensive review. *EAI Endorsed Trans. Scalable Inf. Syst.* **2019**, *7*, 1–16. [CrossRef]

44. Biroscak, B.; Scott, J.E.; Lindenberger, J.H.; Bryant, C.A. Leximancer Software as a Research Tool for Social Marketers. Application to a Content Analysis. *Soc. Mark. Q.* **2017**, *23*, 223–231. [CrossRef]

45. Pascual, F. Introduction to Topic Modeling (MonkeyLearn Blog). 2019. Available online: https://monkeylearn.com/blog/introduction-to-topic-modeling/#:~{}:text=Topic%20modeling%20is%20an%20'unsupervised,able%20to%20automatically%20analyze%20texts (accessed on 8 February 2021).

46. Amado, A.; Cortez, P.; Rita, P.; Moro, S. Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *Eur. Res. Manag. Bus. Econ.* **2018**, *24*, 1–7. [CrossRef]
47. Calheiros, A.C.; Moro, S.; Rita, P. Sentiment classification of consumer-generated online reviews using topic modeling. *J. Hosp. Mark. Manag.* **2017**, *26*, 675–693. [CrossRef]
48. Lee, S.; Song, J.; Kim, Y. An empirical comparison of four text mining methods. *J. Comput. Inf. Syst.* **2010**, *5*, 1–10.
49. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
50. Krestel, R.; Fankhauser, P.; Nejdl, W. Latent dirichlet allocation for tag recommendation. In Proceedings of the Third ACM Conference on Recommender Systems, New York, NY, USA, 22–25 October 2009.
51. Sievert and Shirley. pyLDAvis. Python Library. 2015. Available online: https://pyldavis.readthedocs.io/en/latest/readme.html (accessed on 1 June 2019).
52. Sanchez-Monzon, J.; Putzke, J.; Fischbach, K. Automatic generation of product association networks using latent dirichlet allocation. *Procedia Soc. Behav. Sci.* **2011**, *26*, 63–75. [CrossRef]
53. Ma, B.; Zhang, D.; Yan, Z.; Kim, T. An LDA and synonym lexicon based approach to product feature extraction from online consumer product reviews. *J. Electron. Commer. Res.* **2013**, *14*, 304–314.
54. Tirunillai, S.; Tellis, G.J. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *J. Mark. Res.* **2014**, *51*, 463–479. [CrossRef]
55. Kim, Y.; Shim, K. TWILITE: A recommendation system for Twitter using a probabilistic model based on latent Dirichlet allocation. *Inf. Syst.* **2014**, *42*, 59–77. [CrossRef]
56. Hejazi Nia, M. A Decision Support System for Inbound Marketers: An Empirical Use of Latent Dirichlet Allocation Topic Model to Guide Infographic Designers. In Proceedings of the AMA Summer Educators, Chicago, IL, USA, 14–16 August 2015.
57. Moro, S.; Cortez, P.; Rita, P. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Syst. Appl.* **2015**, *42*, 1314–1324. [CrossRef]
58. Mahajan, A.; Dey, L.; Haque, S.M. Mining financial news for major events and their impacts on the market. In Proceedings of the Web Intelligence and Intelligent Agent Technology, Sydney, NSW, Australia, 9–12 December 2008.
59. Guo, Y.; Barnes, S.J.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [CrossRef]
60. Wang, W.; Feng, Y.; Dai, W. Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electron. Commer. Res. Appl.* **2018**, *29*, 142–156. [CrossRef]
61. Sievert, C.; Shirley, K. LDAvis: A method for visualizing and interpreting topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014; pp. 63–70.
62. April, K.; Bessa, J. A critique of the strategic competitive intelligence process within a global energy multinational. *Probl. Perspect. Manag.* **2006**, *4*, 86–99.
63. Liu, G.; Wei, Y.; Li, F. Understanding Consumer Preferences—Eliciting Topics from Online Q&A Community. In Proceedings of the 18th International Conference on Electronic Business, Guilin, China, 2–6 December 2018; pp. 690–698.
64. Li, F.; Du, T.C. Listen to me—Evaluating the influence of micro-blogs. *Decis. Support Syst.* **2014**, *62*, 119–130. [CrossRef]
65. Davey, L. The application of case study evaluations. *Pract. Assess. Res. Eval.* **1990**, *2*, 9.
66. Blichfeldt, B.S.; Andersen, J.P. Creating a wider audience for action research: Learning from case-study research. *J. Res. Pract.* **2006**, *2*, D2.
67. Yin, R. Validity and generalization in future case study evaluations. *Evaluation* **2013**, *19*, 321–332. [CrossRef]
68. Eisenhardt, K.M. Building theories from case study research. *Acad. Manag. Rev.* **1989**, *14*, 532–550. [CrossRef]
69. Corrales-Gaaray, D.; Mora-Valentin, E.M.; Ortiz-de-Urbina-Criado, M. Open data for open innovation: An analysis of literature characteristics. *Future Internet* **2019**, *11*, 77. [CrossRef]
70. Countants. 10 Leading Trends in Business Intelligence in the Year 2020. 18 January 2020. Available online: https://www.countants.com/blogs/10-leading-trends-in-business-intelligence-in-the-year-2020/ (accessed on 25 January 2021).
71. Manyika, J.; Chui, M.; Farrell, D.; Van Kuiken, S.; Groves, P.; Doshi, E.A. *Unlocking Innovation and Performance with Liquid Information*; McKinsey Global Institute: Missouri, St. Louis, USA, 2013. Available online: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information (accessed on 20 January 2021).
72. Camoinassociates Economic Development. Recent and Emerging Trends in Forestry and Lumber. 29 July 2019. Available online: https://www.camoinassociates.com/recent-and-emerging-trends-forestry-and-lumber (accessed on 29 August 2019).
73. Lahtinen, K. Linking resource-based view with business economics of woodworking industry: Earlier findings and future insights. *Silva Fenn.* **2007**, *41*, 149–165. [CrossRef]
74. Grand View Research. Footwear Market Size, Share, Global Industry Trends Report. 2025. August 2019. Available online: https://www.grandviewresearch.com/press-release/global-footwear-market (accessed on 25 June 2020).
75. Scott, A. The changing global geography of low-technology, labor-intensive industry: Clothing, footwear, and furniture. *Word Dev.* **2006**, *34*, 1517–1536. [CrossRef]
76. U.S. Department of Commerce. Trade.gov/Market-Intelligence/API/Enabling US Exports Through Open Data. 2018. Available online: https://developer.trade.gov/market-intelligence.html (accessed on 1 August 2018).

77. Chan, C. From open data to open innovation strategies: Creating e-services using open government data. In Proceedings of the 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, 7–10 January 2013.

78. Li, C.; Lu, Y.; Wu, J.; Zhang, Y.; Xia, Z.; Wang, T.; Dantian, Y.; Xurui, C.; Peidong, L.; Junyu, G. LDA meets Word2Vec: A novel model for academic abstract clustering. In Proceedings of the Web Conference, Lyon, France, 23–27 April 2018.

79. Kobuszynska, M. USDA Foreign Agricultural Service and Wood Sector in Latvia. 12 December 2016. Available online: https://gain.fas.usda.gov/Recent%20GAIN%20Publications/Wood%20Sector%20in%20Latvia_Warsaw_Latvia_12-12-2016.pdf (accessed on 1 June 2019).

80. Kobuszynska, M. USDA Foreign Agricultural Service. Forest and Wood Products in Poland. 2017. Available online: https://gain.fas.usda.gov/Recent%20GAIN%20Publications/The%20Forestry%20and%20Wood%20Products%20in%20Poland_Warsaw_Poland_3-23-2017.pdf (accessed on 1 June 2019).

81. Ikea. Wood—A Material with Many Qualities. 1999–2021. Available online: https://www.ikea.com/gb/en/this-is-ikea/people-planet/energy-resources/wood/ (accessed on 1 June 2019).

82. Kobuzynska, M. Foreign Agricultural Service and Wood Sector in Estonia. 20 December 2016. Available online: https://gain.fas.usda.gov/Recent%20GAIN%20Publications/Wood%20Sector%20in%20Estonia_Warsaw_Estonia_12-20-2016.pdf (accessed on 1 June 2019).

83. Estonian Timber. Digital Revolution in the Estonian Forestry and Wood Industry. 2018. Available online: https://estoniantimber.ee/best-practices/digital-revolution-in-the-estonian-forestry-and-wood-industry/ (accessed on 1 June 2019).

84. Government of Canada and Trade Commissioner Service. Canada-Ukraine Trade Deal Vastly Expands Opportunities for Exporters. 2020. Available online: https://www.tradecommissioner.gc.ca/canadexport/0004892.aspx?lang=eng (accessed on 19 August 2020).

85. Canada-Ukraine Trade & Investment Support Project (CUTIS). CUTIS Celebrates the First Anniversary of the CUFTA and Kicks off the CUTIS Investment Roadshow. 2018. Available online: https://www.globenewswire.com/news-release/2018/10/17/1622480/0/en/CUTIS-celebrates-the-first-anniversary-of-the-CUFTA-and-kicks-off-the-CUTIS-Investment-Roadshow.html (accessed on 1 June 2019).

86. Euromonitor International. Country Report. Footwear in Ukraine. 2020. Available online: https://www.euromonitor.com/footwear-in-ukraine/report (accessed on 29 January 2021).

87. Match Board. 10 Reasons You Should Consider Outsourcing to New Zealand. 2019. Available online: https://www.matchboard.com.au/10-reasons-you-should-consider-outsourcing-to-new-zealand/ (accessed on 1 June 2019).

88. Transparency Market Research. North America Footwear Market. 2019. Available online: https://www.transparencymarketresearch.com/north-america-footwear-market.htm (accessed on 1 June 2019).

89. Shoes from Mexico. Mexican Shoes and Globalization. 2019. Available online: https://shoesfrommexico.com/mexican-shoes-and-globalization/ (accessed on 29 January 2021).

90. Pereira, G.V.; Macadar, M.A.; Luciano, E.M.; Testa, M.G. Delivering public value through open government data initiatives in a Smart City context. *Inf. Syst. Front.* **2017**, *19*, 213–229. [CrossRef]

91. Barney, J. Firm resources and sustained competitive advantage. *J. Manag.* **1991**, *17*, 99–120. [CrossRef]

92. Nassirtoussi, A.K.; Aghabozorgi, S.; Wah, T.Y.; Ngo, D.C. Text mining for market prediction: A systematic review. *Expert Syst. Appl.* **2014**, *4*, 7653–7670. [CrossRef]

93. Short, J.S.; Palmer, T.B. The application of DICTION to content analysis research in strategic management. *Organ. Res. Methods* **2008**, *11*, 727–752. [CrossRef]