Data Mining and Modeling in Business Decision Making Amid COVID-19

This was done in collaboration

Contents

EXECUTIVE SUMMARY:

This study provides a comprehensive background of COVID-19, with its impact on various business sectors and steps taken by them to diversify to mitigate losses and stay afloat during these unprecedented times. Data has emerged as a new lifeline during this pandemic and has helped the world understand its potential strength. For instance, data has been used to predict the job losses due the lockdown, potential number of coronavirus cases & deaths with and without appropriate measures.

Various data mining techniques like association, clustering, classification and decision trees have been used to understand data and draw relevant insights from them. In addition, the study explains the data modeling techniques such as Agent-based modeling, Regression Analysis, System Dynamics Modeling, Neural Networks and Structural Equation Modeling. Data modeling techniques and methodologies are used to model data in a standard, consistent, predictable manner to manage it as a resource.

Business Processes such as operations, outsourcing, production, benchmarking, administration, business diversification and many more make use of the data and hence, can predict results using the data modelling techniques. The Analytical Hierarchy business model showcasing the application of different data modelling techniques in the different business processes has been drafted for more insight.

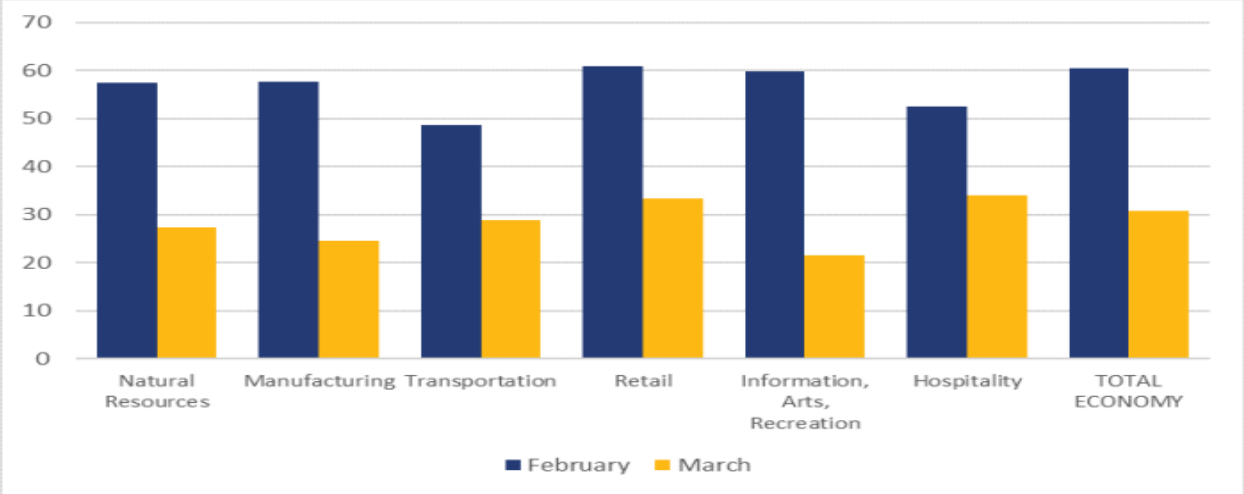
It is recommended that businesses use relevant data modeling techniques to support their businesses processes & make relevant changes to keep themselves afloat during these times. For instance, regression analysis or agent-based modeling can be used for demand planning and production to meet the evolving customer needs during this pandemic.

Significance of the study

The currents study therefore, aims at unearthing the roles of data mining and modeling during COVID-19 Pandemic to help businesses make informed business decisions. The paper also, contributes to existing literature and serves as a benchmark for future research in the field of Data Mining and Modeling. Furthermore, this study gives practical recommendation, to business decision and policy makers, to make informed decisions in the quest of surviving their business during these hard times.

INTRODUCTION:

COVID-19 is the infectious disease caused by the most recently discovered coronavirus. This new virus and disease were unknown before the outbreak began in Wuhan, China, in December 2019. COVID-19 is now a pandemic affecting many countries globally (Feng, H., Yu, D., & Weina, 2019). This pandemic is having a major impact on not only the Canadian economy but economies all around the world. Some industries have been particularly affected more than others (Salinas, E. M, Andres, E.F & Gil, 2007). The diagram below compares the how the various sectors of the economy suffered because of COVID-19 in a span of one month



Source: Figure prepared by the authors using data obtained from the Canadian Federation of Independent Business, *Business Barometer – March 2020*. Accessed 26 March 2020.

The federal government has left no stone unturned in trying to keep the economy afloat during these unprecedented times. To avoid layoffs, programs like Canadian Emergency Wage Subsidy (CEWS), Temporary 10% Wage Subsidy have been implemented. In addition, income tax & sales tax remittance deferrals have been allowed until June. More financial support has been offered through programs like Canada Emergency Business Account (CEBA) interest free loans, Regional Relief & Recovery Fund (RRRF)(Department of Finance Canada, 2020).

COVID-19 has depleted income sources for majority sectors such as aviation, tourism, retail, recreation, restaurants and many more as the country imposed a strict measure of lockdown to curtail the spread of COVID-19. More than a million workers were laid-off, the number highest since the Great Depression (Bachman, D., 2020).

Canadian Businesses and business leaders (Glaveski, 2020) are stepping up against COVID-19. For some, that means shifting their production lines to manufacture the personal protective equipment and medical supplies or launching philanthropic initiatives to help those in need such as buying zoom passes to support a local yoga studio in the community or distributing free meals. For other, it means ensuring Canadians have uninterrupted access to groceries, financial resources, medicines, communication.

The importance of data mining and modeling techniques has been realized like never. Several organizations have resorted to data on employment, consumer demand, preferred platform for communication, source of entertainment etc. to reap benefits from changing consumer habits due to COVID-19.

The report aims to provide a background into current COVID-19 situation and the various challenges and opportunities that have emerged in the face of this pandemic. The following sections will delve into the types of data needed along with the data mining techniques that can help businesses in their data modeling stage. In addition, this study also tries to explain the various business processes supported by various data modeling techniques. This report concludes with practical recommendations for businesses regarding incorporation of data modeling techniques into various processes to mitigate the negative impact of this pandemic.

STUDY BACKGROUND:

The Canadian Economy (Statistics Canada, 2020) changed dramatically in March because of COVID-19 and the situation has had a profound impact on the ability of businesses in Canada to operate. Across the country, over half the businesses in (Statistics Canada, 2020) Alberta (57.7%), Ontario (56.3%), British Columbia (54.8%), Newfoundland and Labrador (53.5%) and Saskatchewan (52.8%) saw declines of 20% or more in revenue. In contrast, close to one-third of businesses in Prince Edward Island (33.1%), the territories (32.4%) and New Brunswick (30.3%) reported either no change or all increase in revenue. The economic impact of COVID-19 is different for all the countries in the world. While many countries like China, Italy, France have flattened the curve and are bouncing back whereas some other countries are still reopening their economy in phases.

During the coronavirus pandemic, data emerged as a vital thing in predicting future outcomes and making important decisions. There are different types of data that can be used such as record data, transaction data, data matrix, market basket data, temporal data, sequence data etc. that can be used based on the defined problem(Gupta, 2019).

Regardless of the field of business or preference for defining data (quantitative, qualitative), accurate data collection is essential to maintaining the integrity of research. (Kabir, 2016)There are two methods for data collection as mentioned below:

- *Primary Method* is the technique in which data can be collected through surveys, questionnaires, interviews, focus groups, observational studies.

- *Secondary Method* is the technique in which data can be collected through books, records, newspapers, census or statistical data, databases, research articles, journals.

After determining the type of data to be collected, businesses can collect data & process it. By running descriptive analysis, they can summarize historical data by calculating various measures of central tendency and draw insights from it to take relevant decisions. (Asakiewicz, C., 2017) Cognitive Analytics can be used to study human behaviour and find patterns based on it.

(Unnisabegum et al., 2019) Data Mining refers to the exercise of going through huge volumes of datasets to draw some relevant information. Some of the Data Mining techniques that can help businesses in their modelling stage are:

- *Association*: It helps in making a correlation between two or more items to identify a pattern. Businesses like restaurants that are open for take outs can find which two or more items are popular in a combo. For instance, Pizza joints can make attractive combos of Coke, Pizza and Wings. Tim Hortons can focus on breakfast combo deal at attractive prices to drive customer demand.

- *Classification*: This analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes. For instance, furniture stores can classify their items into categories of high demand for such as work desk and ergonomic chairs due to increase in work from home during COVID-19. They can focus on stocking them up to meet the demand. Whereas items such as bed sets, couches, coffee table can be categorized as items of medium demand. Decorative items such as figurines, rugs, lamps are low priority items. Managing the supplies as per the demand can help furniture stores stay afloat during the pandemic.

- *Clustering*: Cluster analysis enables identifying a given user group according to common features within a database. These features can include age, geographic location, education level and so on. It is a data mining technique that is useful in marketing to segment the database and, for example, send a promotion to the right target for that product or service (young people, mothers, pensioners, etc.). Clustering has been used for contact tracing during COVID-19.

DATA MODELING TECHNIQUES:

Considering the current scenario with regards to the pandemic, data modeling plays a crucial role in the growth of any business who understands that data-driven decisions are key to their success, stability, and survival. Having data in the right format makes sure that the answers to several business questions can be found more easily and quickly. (Milne, J., Pourshahid, A., Lawrence, S., & Gamache, P. (2019)

It is becoming increasingly important for organizations and enterprises to access and analyze the relevant data to predict outcomes and improve services. Few techniques which will help businesses make informed decisions and survive are,

Systems Dynamics Modeling:

System Dynamics is mostly used in long-term, strategic models and assumes high level of aggregation of objects being modeled: people, products, events, and other discrete items are represented in the model by their quantities. Therefore, they lose any individual properties, histories, or dynamics. If this level of abstraction is applicable for your problem, systems dynamics may be the right method to use. Important things to know about system dynamics modeling are:

- If the model works only with aggregates, the items in that same stock are indistinguishable, they do not have individuality.

- The modeler must think in terms of global structural dependencies and must provide accurate quantitative data for them.

Agent-based Modeling:

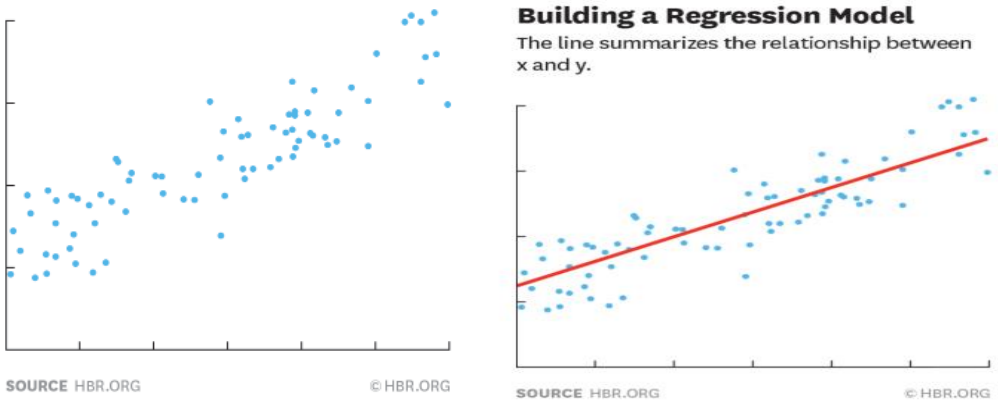
Agent-based modeling is a powerful simulation modeling technique. In this technique, a system is modeled as a collection of autonomous decision-making entities called agents. Each agent individually assesses its situation and makes decisions based on a set of rules. Agents may execute various behaviors appropriate for the system they represent—for example, producing, consuming, or selling. Repetitive competitive interactions between agents are a feature of agent-based modeling, which relies on the power of computers to explore dynamics out of the reach of pure mathematical methods. At the simplest level, an agent-based model consists of a system of agents and the relationships between them. Even a simple agent-based model can exhibit complex behavior patterns and provide valuable information about the dynamics of the real-world system that it emulates. (Bonabeau. E., 2002)

Regression Analysis:

Regression analysis is used to model the relationship between a response variable and one or more predictor variables. Regression analysis provides detailed insight that can be applied to further improve products and services. Regression analysis is a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other. it is essential to comprehend the following terms:

- *Dependent Variable*: This is the main factor that you are trying to understand or predict.

- *Independent Variable*: These are the factors that you hypothesize have an impact on your dependent variable.



Structural Equation Modeling:

Structural equation modeling is a multivariate statistical analysis technique that is used to analyze structural relationships. This technique is the combination of factor analysis and multiple regression analysis, and it is used to analyze the structural relationship between measured variables and latent constructs. This method is preferred by the researcher because it estimates the multiple and interrelated dependence in a single analysis. In this analysis, two types of variables are used

endogenous variables and exogenous variables. Endogenous variables are equivalent to dependent variables and are equal to the independent variable. There are two types of models:

- *Measurement model*: The measurement model represents the theory that specifies how measured variables come together to represent the theory.

- *Structural model*: Represents the theory that shows how constructs are related to other constructs.

Neural Networks:

Neural networks is one of the enhanced search techniques for processing information in the databases. In machine learning, it has been referred to as a pattern recognition methodology and its result model a neural network. It is a non-linear statistical data modelling tool to find patterns or relationships between input and output. Neural network is used for classification, clustering, feature mining, prediction and pattern recognition. There are three types in the neural network model:

- *Feed-forward networks*: the perception backpropagation model and the function network as representatives. It mainly used in prediction and pattern recognition.

- *Feedback network*: Hopfield discrete model and continuous model as representatives. It mainly used in associative memory and optimization calculation.

- *Self-organization networks*: it regards the adaptive resonance theory (ART) model and Korhonen model as representatives. It used for cluster analysis.

In the neural network, we have the model, the learning algorithm, and the activation functions. It will be trained to store, recognize, and retrieve patterns to solve combinatorial optimization problems to filter the noise from data and estimate sampled functions. And that is the reason it has been using in many projects with other techniques such as detecting fraudulent activities (Kirkos, E., Spathis, C., & Manolopoulos, 2007) and forecast the seasonal trends at the store-item level to predict demand for SKUs (Chris Rygielski, Jyun-Cheng Wang, 2002).

BUSINESS PROCESSES SUPPORTED BY DATA MINING AND MODELING:

Demand planning:

The demand planning process involves forecasting, lifecycle planning, promotion planning and consensus demand planning. It is very hard to do well in the demand planning process in a product-specific level. Because of the sale of specialty products are volatile each month. For predicting high volatile products based on historical sales product, a regression model against other data can be used. All types of data including industry trends, economic, weather forecasts, social media feeds can be used to improve the predictive power of demand planning. All types of data including industry trends, economic, weather forecasts, social media feeds can be used to improve the predictive power of demand planning.

For example, the news about the spread of COVID-19 coming from physical contact in close distance, and it led to the increase in demand in face masks and alcohol-based cleaning products. Many countries or stores do not predict this situation because no one knows the cause and effect of COVID-19 in early phase. Maybe we did not prepare for this time, but we can collect different

data for future scenario. However, if we use these data separately, the forecast results may not accurate. When combining all these data we can improve the forecast accuracy tremendously. And when we have accurate forecast information, we can restock inventory quickly and arrive when and where it is needed in smaller lot sizes or just-in-time inventory. Moreover, (Ellinger, Alexander E., Patricia J. Daugherty, 2000) found a significant association between the distribution service performance and firm performance which included the measures for sales growth and customer satisfaction. Organizations need to find or develop a supply service that can deliver on time with the right quantities to improve the organization's performance.

However, neural networks also can be used for the same reason. From the research of (Nguyen, N., & Cripps, 2001) they found that if we have enough data training size and appropriate parameters, neural networks analysis performs better than multi regression analysis. Therefore, we can use both techniques for comparison, and we may have better insights.

Human Resources Management:

The human resources management process consists of many activities from employee recruitment, employee training to performance appraisal and compensations. Turn over is a problem that HR managers need to solve to reduce the cost and improve the organization's performance.

For HR analytics, the structural equations modeling (SEM) has been used to:

- Consider multiple independent and dependent measures concurrently. It means employee-related data does not exist in a vacuum and organizations measure numerous outcomes.

- Cause and effect relationships. Correlations do not tell which events happen first. SEM can uncover the cause and effect relationship.

- Correct for measurement errors. This is important as correlation analysis assumes that everything was measured without error that can make the process a significant flaw.

Regression analysis can be used to look at multiple data simultaneously and prioritize the impact of employee data on business outcomes.

Customer Relationship Management:

The customer process is to gain insights on customers to increase the organization's performance. Customer purchasing behaviour is one of the key concepts to understand customers so we can attract and engage potential customers as well as improve customer retention.

One of the techniques used in analyses customer behaviour is regression analysis. Businesses can also use neural networks for predicting customer's purchasing behaviour based on customers' data. Many factors can affect the purchasing decision of customers so correlations cannot give any accurate information than multi regression analysis, neural networks, or logistic regression analysis. Using techniques can help in forecasting can develop models that predict sales and revenue which is crucial for making a budget. Gain more insights about customers can improve customer satisfaction. All these factors can increase the firm's performance.

Customer behaviour of consumers can be simulated using agent-based models to records and gain insights into the complexity of markets.

During the crisis, customer relationship management is an important process to test the loyal of customers and the organization's ability to boost relationship with customers. While the crisis makes customers' behaviors unpredicted, and limited data about the situations make the predicting analysis, machine learning become important more than ever in gaining insights. It may not help the current situation but after the crisis, these data can help organization understand more about their customer.

Productivity:

Employee productivity is a measure of both efficiencies such as time or other resources required for completing a task and effectiveness which is the degree of results after completing the target task. Therefore, productivity loss is a serious problem that can affect the organization's performance. There are many factors can cause productivity loss and measure through absenteeism and presenteeism which means the time employee worked or not work compared with expectations. Factors can be lifestyle, physical and mental well-being.

For analyzing influences of factors on workplace productivity loss, highly correlated factors will be grouped to construct distinct latent variables to analyze and transformed to observed data space. There will be links between constructed factors and observed variables are specified in a structural model. After combining all individual models into a single SEM framework, we can reduce the omitted variable bias and account for additional interrelations for all variables.

Multiple regression analysis can also be used to find the relationship between dependent and independent variables. Factors can be age, gender, constant, availability of equipment, experience.

Business Diversification

This is a growth strategy that involves entering a new market or industry - one that your business does not currently operate in - while also creating a new product for that new market.

Horizontal diversification is when you acquire or develop new products or services that are complementary to your core business and appeal to your current customers.

Concentric diversification involves adding new products that have technological or marketing synergies with existing product lines or industries, but appeal to new customers.

Conglomerate diversification occurs when you add new products or services that are entirely different from and unrelated to your core business.

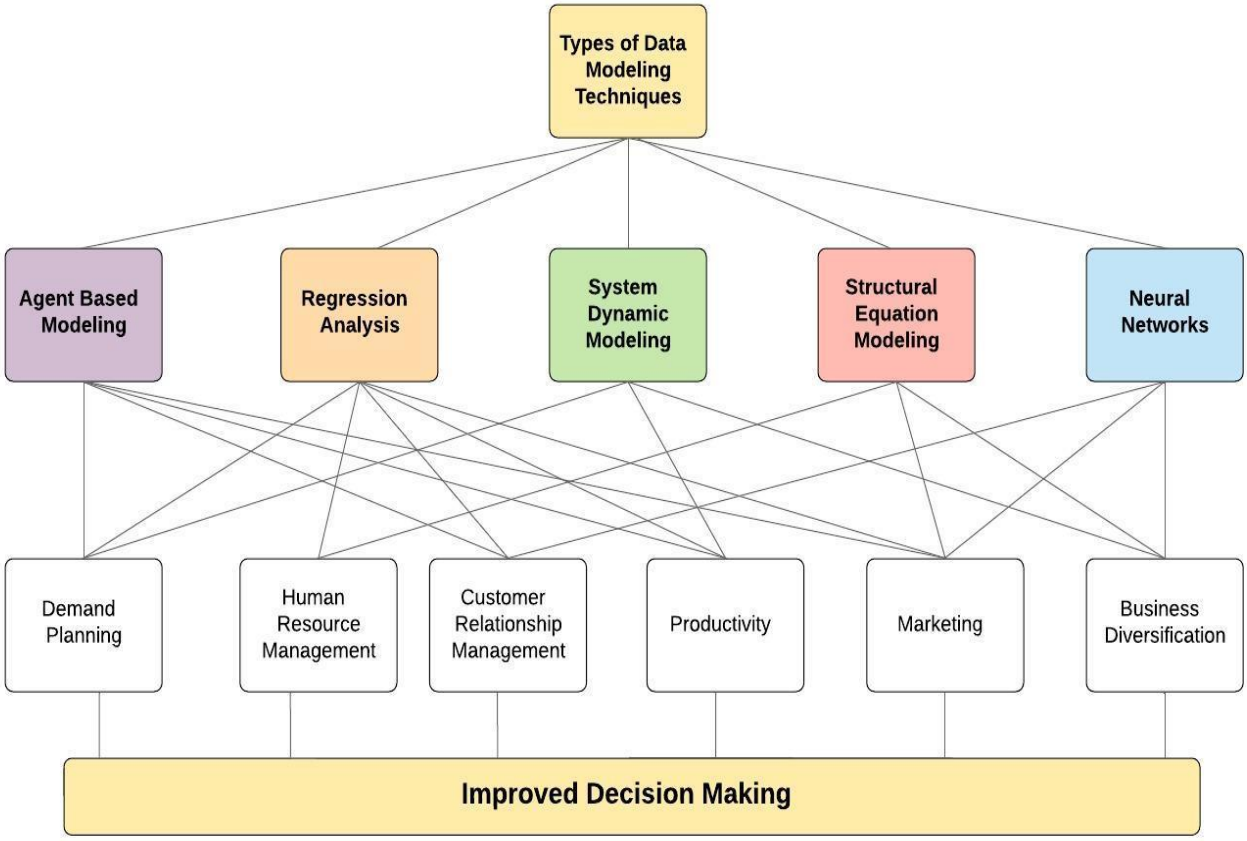
Vertical diversification or integration is when you expand in a backward or forward direction along the production chain of your product. In this approach, you may control more than one stage of the supply chain.

Marketing:

Marketing is very broad with many small processes. There are many forms of marketing to boost the sale of products or raise brand awareness for an organization such as viral marketing, World-Of-Mouth and market campaign. Because of the complexity of marketing, the conventional analytical or empirical approaches to be used. Agent-based modelling is a tool that can help analytics to understand and analyze the complex patterns which combined from many individual agents such as consumers, sellers, distributors and the choices made by these individuals can be affected by multiple factors such as brand, friend's advice, social network.

BUSINESS MODEL:

The Analytical Hierarchy model lists the five types of data modeling techniques mentioned in the study before – Agent based modeling, regression analysis, system dynamic modeling, structural equation modeling and neural networks. The business model depicts that more than one modeling technique can be deployed by a business processes to achieve desired outcome. For instance, Marketing (business process) can make use of modeling techniques such as regression, agent-based modeling, neural networks in their decision-making process.



RECOMMENDATIONS:

A crisis like COVID-19 can strengthen or weaken the relationships between organizations and customers due to unexpected events. Many things have been changed for the past few months and it will keep changing in the future. Over the years, businesses prioritized efficiency over other things by digital and automation. Analysts need to change the model to adapt to changes in customer behaviour as most people are going through something they never experienced before. Everything will not be the same and there is no past data so predicting analysis using data mining and machine learning can help businesses gain insights and give solutions. It is not only in demand and supply but other processes in business as well such as marketing, productivity, employee turn over and business diversification. Listed below are some practical recommendations for business processes:

Customer Relationship Management:

Data modeling techniques like Agent-based modeling, structural equation modeling, regression can be used by small businesses should look to their client base. They can apply clustering techniques for targeted servicing and reach-outs based on customer profiles and their behaviors. Companies should also collect data on the key product or service features that are proving most relevant for their customers during the pandemic and use customer relation dynamics to power upsells and potential sales.

Demand Planning:

With the appropriate modelling (discrete event simulation, neural networks) & analytics techniques, companies can model scenarios and project demand at the SKU level. Business leaders can make detailed plans for sourcing, production, etc., and then make informed spending decisions. Data analytics make it easier to understand downstream effects and optimize production capacity.

Productivity Analysis:

(Gattu, 2020) Data can help businesses streamline internal operations too. Organizations can conduct a productivity analysis through agent-based modeling, regression, system dynamics modeling, in which they measure the time invested individually and collectively in certain meetings and tasks and compare this data to pre-pandemic performance. Where are employees doing better or worse than before, and what factors might be able to improve productivity? For example, if the same amount of time is being spent in meetings as before, but output is lower, a team might be able to benefit from a collaboration software that includes features such as virtual whiteboards. Equally, if a company is finding that training on a certain skill is taking longer than usual, it might want to invest in digital classroom software.

Business Diversification:

Organizations from across industries have been forced to pivot their business strategies in ways that they likely would not have otherwise explored. Instead of seeing this as a cost of the pandemic, companies should treat the new stream of opportunities as a chance to generate growth. This is a great time for business owners to stand out as true winners in building a sustainable business.

Creative Marketing Strategy:

(McGillie, 2020) With the closure of non-essential brick and mortar businesses in many geographic areas, brands have become even more dependent on digital touchpoints as their primary way of communicating with consumers. Now is an excellent time for marketers to optimize strategies that will effectively grow their CRM database. One way to accomplish this is by creating engaging paid and organic content that is meant to attract and capture new audiences in a brand's digital ecosystem or to thoughtfully re-ignite relationships with customers who had previously churned. Understanding how audiences are consuming content and buying products or services differently than they did prior to COVID-19 should directly inform your media strategy by exposing areas of over or underinvestment.

REFERENCES:

Amadeo, K. (2020). *The 3 Types of Profit Margins and What They Tell You*. The Balance. https://www.thebalance.com/profit-margin-types-calculation-3305879

Asakiewicz, C., P. D. (2017). *Cognitive Analytics for Making Better Evidence-based Decisions*. https://ssrn.com/abstract=2965767

Bachman, D., D. (2020). *The economic impact of COVID-19 (novel coronavirus)*. Deloitte. https://www2.deloitte.com/us/en/insights/economy/covid-19/economic-impact-covid-19.html

Bonabeau, E. (2002). *Agent-based modeling: Methods and techniques for simulating human systems*. Pnas. https://www.pnas.org/content/99/suppl_3/7280

Chris Rygielski, Jyun-Cheng Wang, D. C. Y. (2002). Data mining techniques for customer relationship management. *Technology in Society*, *24*(4), 483–502.

Department of Finance Canada. (2020). *Economic Response plan*. Department of Finance Canada. https://www.canada.ca/en/department-finance/economic-response-plan.html

Elflein, J. (2020). *Coronavirus deaths worldwide by country*. Statistic Canada. https://www.statista.com/statistics/1093256/novel-coronavirus-2019ncov-deaths-worldwide-by-country/

Ellinger, Alexander E., Patricia J. Daugherty, and S. B. K. (2000). The Relationship Between Marketing/Logistics Interdepartmental Integration and Performance in U.S. Manufacturing Firms. *An Empirical Study Journal of Business Logistics*, *21*(1), 1–22.

Feng, H., Yu, D., & Weina, L. (2019). *Coronavirus disease 2019: What we know?* NCBI. https://www.ncbi.nlm.nih.gov/pubmed/32170865

Gallo, A., Davenport, T., & Kim, J. (2017). *A Refresher on Regression Analysis*. HBR. https://hbr.org/2015/11/a-refresher-on-regression-analysis

Gattu, N. (2020). *How Data Can Power Businesses Through the COVID-19 Crisis*. Worth. https://www.worth.com/how-data-can-power-businesses-through-the-covid-19-crisis/

Glaveski, S. (2020). *COVID19: Winners, Losers and New Business Opportunities*. Medium. https://medium.com/steveglaveski/covid19-winners-losers-and-new-business-opportunities-38a60badf579

Gupta, T. (2019). *Types of Data Sets in Data Science, Data Mining & Machine Learning*. https://towardsdatascience.com/types-of-data-sets-in-data-science-data-mining-machine-learning-eb47c80af7a

Kabir, S. M. (2016). *METHODS OF DATA COLLECTION* (pp. 201–275).

Kaul, A. (2020). *COVID-19: 5 Ways Analytics Can Help Businesses Prepare for Next*. Aithority. https://aithority.com/guest-authors/covid-19-5-ways-analytics-can-help-businesses-prepare-for-what-comes-next/

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003. https://doi-org.gbcprx01.georgebrown.ca/10.1016/j.eswa.2006.02.016

Léonard, A., & Bellerose, S. L. (2020). *Impacts of COVID-19 on Selected Sectors of Canada's Economy*. Hillnotes. https://hillnotes.ca/2020/04/08/impacts-of-covid-19-on-selected-sectors-of-canadas-economy/

McGillie, R. (2020). *Analytics During Coronavirus: 10 Helpful Strategies for Marketers*. Tinuiti. https://tinuiti.com/blog/marketing-news-covid-19/analytics-covid-10-helpful-strategies/

Milne, J., Pourshahid, A., Lawrence, S., & Gamache, P. (2019). *6 Data Modeling Techniques for Better Business Intelligence*. Klipfolio. https://www.klipfolio.com/blog/6-Data-Modeling-Techniques

Nguyen, N., & Cripps, A. (2001). Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. *Journal of Real Estate Research*, *22*(3), 313.

OpenStax. (n.d.). *3.1 Demand, Supply, and Equilibrium in Markets for Goods and Services*. Retrieved June 17, 2020, from https://opentextbc.ca/principlesofeconomics/chapter/3-1-demand-supply-and-equilibrium-in-markets-for-goods-and-services

Salinas, E. M, Andres, E.F & Gil, R. (2007). Family as a source of consumer-based brand equity. *Journal of Product & Brand Management*, *16*(2), 188–199.

Statistics Canada. (2020). *Canadian Survey on Business Conditions: Impact of COVID-19 on businesses in*. Statistic Canada. https://www150.statcan.gc.ca/n1/daily-quotidien/200429/dq200429a-eng.htm

Twin, A. (2020). *Why Companies Use Outsourcing*. Investopedia.

Unnisabegum, A., Hussain, M., & Shaik, M. (2019). *Data Mining Techniques for Big Data, Vol. 6, Special Issue,* https://doi.org/10.13140/RG.2.2.25408.07686