

---

## Data Science and Marketing in E-Commerce Amid COVID-19 Pandemic

---

Submitted 10/03/21, 1st revision 28/04/21, 2nd revision 15/05/21, accepted 12/05/21

Olha Fedirko<sup>1</sup>, Tetiana Zatonatska<sup>2</sup>, Tomasz Wołowiec<sup>3</sup>, Stanisław Skowron<sup>4</sup>

### **Abstract:**

**Purpose:** The objective of this study involves the determination of data-driven solutions needed to increase the usability of e-commerce systems and its profitability.

**Design/Methodology/Approach:** In the research implementation process, logic generalization and induction to identify and analyze the most beneficial data science tools in e-commerce. The design of the study is to generalize existing approaches of data science usage in e-commerce, to develop practical recommendations to ensure the competitive advantages of e-commerce market participants and to estimate the cost of technical tools needed to launch the data science project in e-commerce.

**Findings:** The results clearly demonstrate that in 2020 businesses that have e-commerce system were financially successful and in next 3 years online sales will increase rapidly. The simple analytics will not cover the demand of online business and it is needed to implement advanced data-driven decisions now.

**Practical Implications:** The present research provides generalized knowledge on how to launch a data science project in e-commerce and how to choose the best programming and visualization app to ensure the profitability of a project. The scientific paper gives an instruction on the marketing contribution analysis, which is the tool of key importance for online marketplaces.

**Originality/Value:** The main research value drawn from the study is to launch the data-driven models in e-commerce company it is needed to observe the real business need and available data, find the best programming and visualization tools. It was defined that the most beneficial data science solutions are demand forecasting, estimation of the marketing contribution, customers clustering, recommendation system and customers' attitude analysis. The main business need for each e-commerce company is to estimate the contribution of all marketing channels and advertisement formats separately. This issue may be easily handled with a regression modelling, which helps to understand a set of factors influencing sales.

**Keywords:** Data science, marketing, e-commerce, online shopping.

**JEL codes:** L81, C8, M31.

**Paper type:** Research article.

---

<sup>1</sup>Corresponding Author, Faculty of Economics, Taras Shevchenko National University of Kiev, Ukraine, E-mail: [olya7fedirko@gmail.com](mailto:olya7fedirko@gmail.com)

<sup>2</sup>Same as in 1, E-mail: [tatonat@ukr.net](mailto:tatonat@ukr.net)

<sup>3</sup>Corresponding Author, Institute of Public Administration and Business, University of Economy and Innovation in Lublin, Poland, Email: [tomasz.wolowiec@wsei.lublin.pl](mailto:tomasz.wolowiec@wsei.lublin.pl)

<sup>4</sup>Corresponding Author, Department of Strategy and Business Planning, Faculty of Management, Lublin University of Technology, Poland, Email: [s.skowron@pollub.pl](mailto:s.skowron@pollub.pl)

## **1. Introduction**

Over recent years, the digitalization of life is progressing with an extremely high pace. Nowadays, even daily routine tasks such as taxi, shopping, paying bills are digitalized along with the digitalization of public services. Thus, digitalization becomes an integral part of life as it greatly simplifies it. One of the fastest growing sectors is e-commerce, which makes it possible to assert its superiority over offline services. Moreover, it is widely argued that e-commerce is going to replace offline sales in the next decade.

Undoubtedly, during the COVID-19 pandemic, the share of e-commerce sales increased dramatically. The e-commerce web traffic all over the world accounted for 21,96 billion visitors in June 2020, while in June 2019 the feature accounted for only 16,2 billion visitors (Statista, 2021). So, the e-commerce usage increased by almost 35% over the year and online turnover more than doubled during the quarantine. Thus, online services became so popular that it is unlikely to lose its advantage in the foreseeable future. It means that not all people are ready to come back to offline sales and even if the pandemic disappear completely people will constantly purchase online as it becomes a habit for them. As an outcome, businesses that do not have an e-commerce system will go bankrupt under competitors' that sale online.

Although, it should can't hidden that a huge number of companies that just started to sell online or did not have e-commerce at all lost a significant amount of money during the quarantine. For those businesses it is vital to launch e-commerce reinforcing with modern data science tools to establish the position of e-commerce as a convenient, fast, always available service. The truth is that the major number of shopping centers all over the world do not have an e-commerce system. So, the leadership will be gained by those stores that implement e-commerce with some kind of gamification before competitors. Furthermore, using data science and machine learning technologies is of key importance as it allows to manage sales constantly, predict the product's demand and retain customers by giving them personalized offers and recommendations.

The modification of consumer behavior and changes in business models in the aftermath of the covid-19 pandemic has influenced the search for new approaches to e-commers. The aim of the study is to generalize data science tools to define the main field to develop in e-commerce, the system analysis was used as a major method. To conduct the study a huge number of scientific papers and case studies were analyzed. Moreover, the main merits and drawbacks of technical methods in terms of suitability for e-commerce were defined. Due to the COVID-19 pandemic, existing behaviors and structures are being questioned. As crises can trigger fundamental economic and societal changes, companies need to understand consumer behavior at this particular time. The trends established during this crisis

may remain stable into the future, inflicting serious consequences on brick-and-mortar stores due to a rapid increase in e-commerce.

## **2. Previous Studies' Overview**

Despite the fact that e-commerce become popular in the last decade, there are many scientific articles regarding the efficiency of using data science tools to attract online sales. Data science for e-commerce implies retention and extension of the customer's life cycle. Moreover, in the time of COVID-19 pandemic, it is vital to use technological innovations to gain market leadership. Data science and machine learning technologies in the last six months become decisive in building trust and loyalty to companies using e-commerce.

Numerous scientific researches proved the efficiency of using data science and machine learning technologies in e-commerce. Regarding the outcomes of various studies data science may lead to the costs optimizing, increase in total sales, incline the brand awareness and improve the attitude to the brand. Moreover, it simplifies lots of operational work as the demand for products is forecasted and the consciousness of customer's preferences exists.

Previous researches confirmed the e-commerce efficiency in national economies growth (Fedirko and Zatonatska, 2020). Using ICT and e-commerce system may definitely improve the level of people's life and improve the economy. Nowadays, e-commerce is a key industry for strategic investment in Europe and it is needed to implement innovations into analyzing the ICT and e-commerce performance on national markets.

E-commerce as an online service is vital for businesses. Before the Covid-19 pandemic and quarantine researchers conducted scientific studies the main aim of those was to define the effectiveness of e-commerce systems on exporting performance (Gregory, Ngo, and Karavdic, 2019). Deeply interviewing 15 experienced exporters and testing the results of their answers on 340 other exporters, authors came to conclusion that e-commerce increase brand awareness and, as a result, make communication with other companies easier and more profitable. So, even before the pandemic it was vital for businesses to use the e-commerce system. Nowadays, it is essential not only to launch an e-commerce platform, but also analyze the data from it using data science.

It is not enough for e-commerce companies to make just simple data analytics to provide their businesses with predictions of a high accuracy, correct logistics and cost optimization. Now it is vital to use more advanced technologies such as data management, data engineering, data science and machine learning to work beneficially with data (Moorthi *et al.*, 2021).

To predict accurately the distribution, it is essential to pay close attention to the anticipation of demand on the products and, moreover, to the hidden data that may be easily collected and used for predictive models. It is vital to collect and analyze all possible business features and implement it into comprehensive data science solution to increase and control e-commerce sales (Hurtado, Dorneles, and Frazzon, 2019).

The special technique is able to predict the number of websites' visitors. Thus, it may be beneficial in terms of creating more accurate marketing strategy. Understanding all gaps in the prediction of the traffic to the website, it is possible to set right time bands and periods of the advertisement. Consequently, this technology leads to the budget optimization and in some cases to the cost reduction (Wang, Cai, and Zhao, 2020).

Among all data science tools the most frequently used are clustering and predictive analysis. The theoretical approach (Pliskunova and Klochko, 2020) define the best classification method among the random forest, k-means, linear discriminant analysis, support vector machine, classification and regression trees. The highest accuracy level was obtained using random forest method. That helps to identify 6 customer segments and, moreover, understand clearer those client's groups that are under the risk to be lost. More practical paper (Ballestar, Grau-Carles, and Sainz, 2018) observe the data of customer's experience duration on the site and profitability. Defined groups divided by customers behavior make possible to observe transferring processes between clusters.

Moreover, that empower marketing activity as for each cluster there was special strategy to keep clients. Particularly, those algorithms provide enterprises with better quality clients as it is understandable which specific group of people is more profitable or which group has a potential to bring more income. Clustering is also a benefit for those companies that need to track their performance frequently. Based on the visualization using k-means usability of such an instrument was proved (Kamthania, Pahwa, and Madhavan, 2018).

In addition, neural networks are widely used to deal with various problems in e-commerce or increase the profitability of services. Recommendation systems are one of those tools that show customization of the product and, thus, attract clients. Researchers widely use the neural network method reinforced by ant colony optimization to define the personalized recommendations for customers (Shao and Cheng, 2010). In recent year scientists paid attention even, at first glance, insignificant customer's activity such as reading the comments, asking questions about the product, spent time to order and other actions. In the study on those micro behaviors (Zhou *et al.*, 2018) the efficiency of usage that information was proved on a real data. Authors built their own frame work using neural networks to analyze micro behavior in the e-commerce sector. As an outcome, the system gives

customers better recommendations and owners explanation of chosen products that was recommended.

In addition, it is possible to prevent clients from fake and inferior goods with the help of machine learning algorithms (Chin *et al.*, 2021). The conducted model may predict with a high accuracy customers' behavior and prevent the infringing sales. Talking about more global issues, it is essential to mention the machine learning usability in logistic performance. The genetic algorithm is widely used for optimization the logistic system and increasing the logistic efficiency (He, Meng, and Liang, 2021).

It is also important for the e-commerce system to monitor continuously competitors' prices on the same products and services. As it is easy to navigate various web sites to find lower price, companies should ensure appropriate prices to attract customers. Neural networks may also forecast price using competitors' data from their sites and general data on operation costs. LSTM and SVR neural networks are useful to predict the price on the phone market for the next day (Bakir, Chniti, and Zaher, 2018). The optimization of e-commerce prices using neural network is also beneficial and widely implemented (Peng and Liu, 2007).

Moreover, with worldwide e-commerce popularization the importance of analyzing customers attitude to the brand and to the products they buy increased. The sentiment analysis is the key tool to detect clients' mood, indicate specific problems in products or services and prolong clients' life time value (Parveen *et al.*, 2021).

However, data science and machine learning technologies are not a magic pill for businesses as the efficiency of these tools directly affected by the implementation processes. It is noticeable that 85% of AI initiatives are not successfully implemented (Zhang, Pee, and Cui, 2021). The right management process in AI development and implementation is of key importance. All in all, the huge number of data science and AI tools exist and should be immediately implemented to the business processes. Though, it is vital to follow all general rules on how to work with innovation and efficiently use technologies in business to make this novation beneficial. In this work all popular data science and machine learning tools are generalized, the cost of innovation is calculated and the selection process of data science tools is defined.

### **3. Methodology**

As the aim of the study is to generalize data science tools to define the main field to develop in e-commerce, the system analysis was used as a major method. To conduct the study a huge number of scientific papers and case studies were analyzed. Moreover, the main merits and drawbacks of technical methods in terms of suitability for e-commerce were defined.

To achieve the objectives in this scientific paper general scientific methods were used. Logical generalization – to identify trends in the development of solutions for the e-commerce; comparison – to analyze commonalities and differences in the data science programming tools; induction – to define the main beneficial data science methods in e-commerce based on previous researches; analysis and synthesis – to analyze the theoretical and practical models used in e-commerce in terms of increasing its profitability and to build a holistic understanding of the necessary human and technical resources to ensure the launch of the data science project; methods of multicriteria analysis –rating technical tools for data science model building.

Overall, the study is conducted using system analysis, in particular generalization and systematization, classification, evaluation.

#### **4. E-Commerce Development during the COVID-19 Pandemic**

COVID-19 pandemic caused dramatic changes around the world as it affected all aspects of life from shopping to government policies. Impossible to imagine how the world would look like without those digital systems that help to communicate and connect every day, especially during quarantine. No doubt, that online purchases increased significantly over the past year. Particularly, the fact that not all industries have their e-commerce systems working is of key importance.

Regarding the retail turnover in 27 European Union countries, the feature of online sales increased significantly over first four months of global quarantine. The most revealing period is April in 2020 as the percentage of year-on-year change of online retail trade accounted for 30% increase, while offline retail trade decreased by approximately 18% to the April in the previous year. Moreover, the percentage of online retail's year-on-year growth reached a peak in May and accounted for 38%, when slightly decreased to 28% in June and to 18% in July (Eurostat, 2021). In 2020 the share of e-commerce in retail trade was accounted for 14% and it is forecasted the growth up to 22% in 2023 (Statista, 2021).

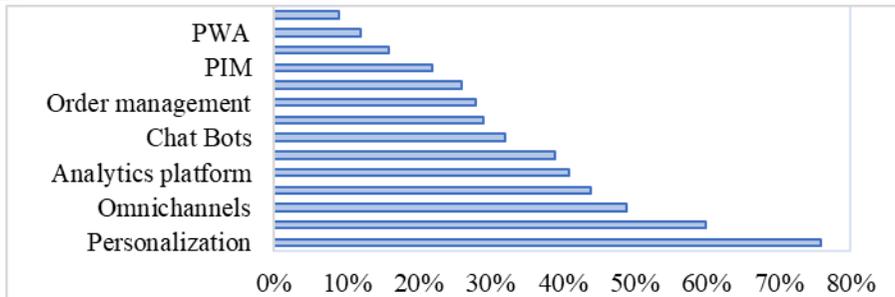
In fact, people all around the world were forced to change their shopping habits. For instance, 31% of people tried in-store pick up, 7% practiced pickup lockers and 26% experienced home delivery because of COVID-19 and quarantine. Moreover, the number of people who practiced whole e-commerce infrastructure is accounted for 19% (SearchNode, 2021).

Customers' behavior has changed dramatically due to the results of the poll (SearchNode, 2021) on whether the online revenue of the companies was affected positively or not. Respondents confirmed that for 90% of companies the revenue from online sales increased during the global lockdown. Moreover, for 50% of overall respondents number online revenue has increased by more than 100%. The

interesting fact is that after the global lockdown only 5% of responders mentioned the decline in online revenue and 32% experienced increase by 50-100%.

On the Figure 1 e-commerce companies' willingness to implement a specific technology is presented. Obviously, half of it may be realized with the help of data management, data science and machine learning. On the top of companies' purposes for 2020 was personalization. With the help of data science knowledge and data management techniques it is easy to analyze the data of each client and provide them with personalized advertisement, for instance, with the help of email marketing or chat bots. Also, there is a huge demand of analytics platforms (41%), which requires data management expertise. Speaking about fraud prevention, there are lots of machine learning algorithms that may be easily implemented into businesses.

**Figure 1.** Technologies that e-commerce companies were willing to implement in 2020



*Source:* SearchNode, 2020.

It is important to notice that companies still believe in the power of offline sales even amid COVID-19 pandemic. After the global quarantine businesses were asked whether they want to become a marketplace. Only 6% of respondent confirmed their willingness to build an e-commerce system, while 65% refuse this opportunity (SearchNode, 2021). Undoubtedly, the pandemic helps online marketplaces to promote faster, but there is still a great competition and the winners are those who started firstly. The share of online marketplaces that consider 2020 financially successful is accounted for 91% (SearchNode, 2021). It means that overall e-commerce experienced a rapid increase in popularity and usability. All in all, as lots of e-commerce companies gain high online revenues and are willing to implement data science solutions into their platforms it is needed to describe the possibilities of data-driven solutions and explain how to find the suitable data processing tool.

## 5. Results

Modern world offers dozens of programming tools for e-commerce development. There are numerous techniques for modeling and predicting the demand for the

product, prices optimization, creating recommendation systems, customers behavior analysis and even fraud prevention. In this work only data science projects are described and, consequently, cost estimation for technical tools as well as scope of technical possibilities for e-commerce are given only for data science projects.

In general, there are three main preparation stages of a data science project launch (Table 1). Speaking about the first stage of the data science project, it is essential to mention the tasks that each project manager or leader of a project should deal with to start the project and define correctly the subtasks and additions possibilities for the project. Next two stages required the solution on programming and visualization tools.

The process of selection of the data science programming tools is quite time-consuming, but utmost important in a long-term perspective. Those data science tools are expensive and differs a lot in their usability and value added. For this reason, the best programming tools were defined below to simplify the preparation process. Visualization tools are simpler and do not differs a lot. For data science projects in e-commerce were selected four main visualization platforms that may empower the perception of the modeling results.

**Table 1.** The required stages of preparation a data science project

№	Stage of project preparation	Tasks
1.	Briefing the team for project	<ol style="list-style-type: none"> <li>1) Deeply understand the client’s needs and client’s business processes</li> <li>2) Divide the client's global task into subtasks, determine their sequence and timings</li> <li>3) Form a team to solve each subtask or task as a whole</li> <li>4) Presentation to the team of the basic data about the client, global objectives and goals of the client, as well as the model of cooperation between team members</li> <li>5) Approval of project stages, subtasks and responsible employees for each type of work</li> </ol>
2.	Programming tools selection	<ol style="list-style-type: none"> <li>1) Define the model that suits client’s needs</li> <li>2) Think over possible options for additional solutions for the client and models for these tasks</li> <li>3) Choose the tool that allows to solve both issues – the main client’s task and additional options</li> </ol>
3.	Visualization tools selection	<ol style="list-style-type: none"> <li>1) Find out a convenient way for the client to obtain information</li> <li>2) Analyze whether there are options to connect chosen programming language to the visualization platform</li> <li>3) Find out the required visualization techniques for project</li> <li>4) Choose the visualization platform based on point 1)-3).</li> </ol>

**Source:** Own creation.

Obviously, before choosing technical tools it is needed to define the specific model due to client's needs. No doubt that there are some basic tools that each business should have to optimize costs and provide better services. First of all, it is demand forecasts, which give a benefit such as understanding the impact of season, marketing, competitors and other specific to each category factors. Predictive models could also help with defining categories of products that insure with the best value of money. In a long-term perspective those models ensure better service as all products are available any time people want to purchase.

Secondly, the most frequently used data science tool for business is customers clustering. In plain language it implies dividing all clients into categories based on their age, expected wage, frequency of purchases, average check, preferences. That will be a merit for business as it makes possible to introduce the practice of targeted advertising, which attracts people attention more and, as a result, not only involve more customers, but also make them long-term clients.

Nowadays it is also important to pay attention to comprehensive data-driven solution and gradually apply them to business processes. Based on clustering analysis and predictive analytics the recommendation system may be built. It also might be called market basket analysis for retail companies. For instance, we tracked that one person bought business outfit twice and that is a sign for us to belong that person to the category of those who bought the same type of clothes more than once. Next, the algorithm proceeds with analysis of the category and identifies key features of the category. Let's assume that it is a tendency to buy accessories such as bags and jewelry for those who favor stylish business clothes. Therefore, the outcome of such a model is to recommend person to buy not the next business suit, but specific accessories. Moreover, it means not only to provide clients with an article on those clothes, shoes or accessories that people often buy and also make customized compilation giving those attributes that suits the previous purchase.

In a perfect world it is better to add customer sentiment analysis to recommendation system. Based on the sentiment analysis it is easy to handle vast number of feedbacks and, moreover, respond to all of them. The first stage of algorithm is as follows: business launches the sentiment analysis and key words indicating functions to review all existing customer's feedbacks. The further step is to access the mood of the review and find all similar comments. After that the system should classify them by mood and category of product or brand as well as type of service (online order, delivery, consultation). It means that a company will have categorized types of problems and also may understand strong points.

So, there is worth using technologies such as predictive analytics, which will help to understand the demand, and therefore never have overstocks and out of stocks. To retain customers, it is better to deal with clustering by age, income, RFM indicators and market basket analytics. This will help to provide clients with

customized offers, as well as make recommendations for next purchases properly. It may also be useful to implement sentiment analysis, which will help to more determine the appropriate recommendations, as well as eliminate existing problems with the service.

Beyond doubt, technical improvements in the e-commerce system are made either to retain existing customers or to attract new ones. From this point of view, marketing and data science are connected and only their tandem is able to provide a continuous flow of customers to the online marketplace.

It's important to note that there is a two-way relationship between marketing and data science. On the one hand, the effectiveness of marketing investments is analyzed using data science. On the other hand, marketing is a tool for implementing data science solutions in e-commerce. There are two global tasks in marketing that are easily solved using data science methods – calculating the contribution of marketing communications to sales and personalizing marketing proposals.

First of all, the aim of each e-commerce business is to estimate the marketing contribution, which means to divide the effect of each specific marketing channel in terms of its added value. The best model for such a business need is regression. The regression model gives a possibility to calculate the share of contribution of each feature included into it. For example, if our indicators in the model are seasonality, special promos, google advertisement in the video format, google advertisement in the format of banner, YouTube advertisement, collaborations with influencers, then we may understand the number of sales given by each indicator separately.

Of course, there is a synergy effect of a group of indicators such as special promos and advertising, but the thing is that the regression model may estimate it separately with a high accuracy. So, if company needs to know the efficiency of each marketing channel and marketing format and to optimize the marketing budget it is highly recommended to use the regression model.

Speaking about more advanced decisions such as recommendation systems, it is also beneficial to emphasize the connection between marketing and data-driven solutions. Online companies as well as offline ones tend to plan their marketing activity in advance. The main problem here is that this kind of marketing calendar is based only on the expert's decisions and not on the real customers' behavior.

With the help of data science customer's behavior may be easily defined based on clustering. On the first stages only general clustering by gender, age, purchasing frequency and average check. Next, it is beneficial to divide customers by their special preferences, social status, having children, pets, family and other narrow groups. That will help to understand the main groups of the online marketplace.

After basic or advanced clustering, it is needed to test the efficiency of marketing activity. With the analytical tools data scientist may quickly identify the group of advertisement that works for each specific group of customers. After this analysis, that required Big Data expertise, the recommendation system may be built. It means that algorithm based on previous knowledge may define the best marketing channels and formats of advertisement for each group of customers. As a result, marketing can not work effectively without data-driven decisions. Data science solutions are designed to help marketers to find the best advertisement channel and format as well as to provide clients only with those promos to those they are sensitive.

Dealing with the data science project it is important to define a business need and suitable model for the business task. The fact that solutions such as budget optimization and recommendation system are comprehensive ones is of key importance. It means that if client needs to build a recommendation system, they also receive customers clustering as an additional option that may be used for another business need. It is obvious that the client may not always clearly know his need and, moreover, understand what process is needed to deal with their issues. That is why giving the client some knowledge about options and general modeling techniques is essential. Next, the technical tools for this scope of data-driven solutions are defined (Table 2). The important notion is that we do not mind that, for example, some issues can't be solved with another platforms, but we put in the table only those that are more beneficial, well-known and user-friendly for each specific task. In addition, the main features of the programming tools are observed in Table 3.

**Table 2.** Programming tools for the most useful data-driven solutions

Data science solution	Suitable programming tools
Demand forecasting	R, Python, MATLAB, BigML
Estimate marketing contribution	R, Python, BigML
Budget optimization	R, Python
Customers' attitude analysis	R, Python, BigML
Customers clustering	R, Python, MATLAB
Recommendation system	Python, BigML

**Source:** Own creation.

**Table 3.** The comparison of programming tools for data science project

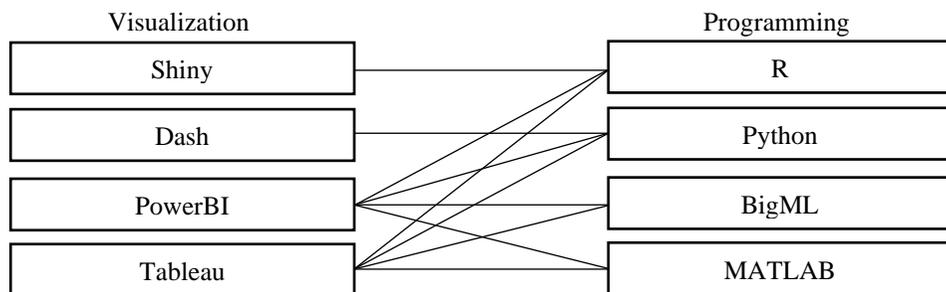
Feature	MATLAB	R	Python	BigML
Libraries	✓	✓	✓	✓
Speed			✓	✓
Ability to work with Big Data			✓	✓
Ease of use	✓	✓	✓	✓
Visualization		✓	✓	✓
Annual price	940\$	3749\$	1188\$	10000\$

**Source:** Own creation.

It should be mentioned that R and RStudio are user-friendly and frequently used at universities that makes the working process faster. Moreover, the special visualization tool Shiny is included to the annual price (RStudio, 2021). Python is more popular now and as well as R includes visualization tool Dash (PythonAnywhere, 2021). Furthermore, Python is faster and easily deals with Big Data, but it is more expensive to work with data scientists who code on Python. Speaking about MATLAB, it is essential to highlight the frequency of its usage at universities and for scientific purposes. The main drawbacks are in speed and ability to work with large flows of data. However, MATLAB’s data engineers confirmed that a tool is suitable for even machine learning solutions and may connect cloud storages to increase the capacity (MathWorks, 2021). The tool of a highest price BigML offers the most various number of possibilities. It is easy to build a community with this app and connect a team inside it. Also, it gives the company private cloud storage that is much bigger than other apps offer (BigML, 2021).

It is also quite essential to choose the appropriate visualization tool for the data science project to make it easier for data scientist to present the information and for clients to perceive it. As it was mentioned before, R has its own visualization tool Shiny and Python has an app Dash, that makes the process easier as no additional costs are needed. BigML also has an ability to visualize information, but it is easier to observe dashboards with Tableau and PowerBI. The compatibility of programming and visualization tools are given on the Figure 2.

**Figure 2.** The compatibility of programming and visualization tools



**Source:** Own creation.

Finally, all the stages of data science project preparation are described in details and the connection between data science, marketing and e-commerce is observed in this scientific paper. To conclude, the e-commerce field will develop rapidly in nearest three years and it is vital for the online business to implement data-driven decisions to collect and analyze data from the scratch. This work provides with generalized knowledge on how to start the data science project and how to find the best technical solution for each specific business purpose.

## 6. Conclusions

During the quarantine, online sales grew significantly by approximately 30%, since most often it was impossible to buy offline. Considering the fact that we live in an environment with high uncertainty level, it is not known how many times the quarantine will be repeated in the future. Based on this, the introduction of e-commerce into businesses is not a luxury, but almost a necessity. Moreover, given that there are not so many businesses selling online all over the world, those will gain leadership who will implement e-commerce systems first. To launch the data-driven models in e-commerce company it is needed to observe the real business need and available data, find the best programming and visualization tools. It was defined that the most beneficial data science solutions are demand forecasting, estimation of the marketing contribution, customers clustering, recommendation system and customers' attitude analysis. The four main programming tools that are user-friendly and useful for data science projects were chosen from various number of apps. In addition, the best visualization platforms were observed.

The main business need for each e-commerce company is to estimate the contribution of all marketing channels and advertisement formats separately. This issue may be easily handled with a regression modelling, which helps to understand a set of factors influencing sales. It helps businesses to manage sales efficiently and provide customers with personalized advertisements.

In general, the usage of data science technologies will help not only to increase sales, but also to incline loyalty and brand awareness, as well as redistribute sales between online and offline in equal shares.

## References:

- Bakir, H., Chniti, G., Zaher, H. 2018. E-Commerce Price Forecasting Using LSTM Neural Networks. *International Journal of Machine Learning and Computing*, 8, 169-174. <https://doi.org/10.18178/ijmlc.2018.8.2.682>.
- Ballestar, M.T., Grau-Carles, P., Sainz, J. 2018. Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research*, 88, 407-414. <https://doi.org/10.1016/j.jbusres.2017.11.047>.
- BigML. 2021. Machine learning for everyone. Available at: <https://bigml.com/pricing>.
- Chin, S.H., Lu, C., Ho, P.T., Shiao, Y.F., Wu, T.J. 2021. Commodity anti-counterfeiting decision in e-commerce trade based on machine learning and Internet of Things. *Computer Standards & Interfaces*, 76, 103504. <https://doi.org/10.1016/j.csi.2020.103504>.
- Fedirko, O., Zatonatska, T. 2020. The development strategies for national economies after Covid-19 pandemic. *Economica*, 99(2), 92-103. <https://doi.org/10.15388/Ekon.2020.2.6>.
- Gregory, G.D., Ngo, L.V., Karavdic, M. 2019. Developing e-commerce marketing capabilities and efficiencies for enhanced performance in business-to-business export ventures. *Industrial Marketing Management*, 78, 146-157. <https://doi.org/10.1016/j.indmarman.2017.03.002>.

- He, X., Meng, S., Liang, J. 2021. Analysis of cross-border E-Commerce logistics model based on embedded system and genetic algorithm. *Microprocessors and Microsystems*, 82, 103827. <https://doi.org/10.1016/j.micpro.2021.103827>.
- Hurtado, P.A., Dorneles, C., Frazzon, E. 2019. Big Data application for E-commerce's Logistics: A research assessment and conceptual model. *IFAC-PapersOnLine*, 52, 838-843. <https://doi.org/10.1016/j.ifacol.2019.11.234>.
- Kamthania, D., Pawa, A., Madhavan, S.S. 2018. Market Segmentation Analysis and Visualization Using K-Mode Clustering Algorithm for E-Commerce Business. *Journal of computing and information technology*, 26, 57-68. <https://doi.org/10.20532/cit.2018.1003863>.
- MathWorks. 2021. Pricing and licensing. Available at: <https://www.mathworks.com/pricing-licensing.html>.
- Moorthi, K., Dhiman, G., Arulprakash, P., Suresh, C., Srihari, K. 2021. A survey on impact of data analytics techniques in E-commerce. *Materials Today: Proceedings*, (in press). <https://doi.org/10.1016/j.matpr.2020.10.867>.
- Parveen, N., Santhi, M.V.B.T., Burra, L.R., Pellakuri, V., Pellakuri, H. 2021. Women's e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK. *Materials Today: Proceedings*, (in press). <https://doi.org/10.15388/Ekon.2020.2.6>.
- Peng, L., Liu, H. 2007. A Dynamic Pricing Method in E-Commerce Based on PSO-trained Neural Network. *Integration and Innovation Orient to E-Society*, 1, 323-329. [https://doi.org/10.1007/978-0-387-75466-6\\_36](https://doi.org/10.1007/978-0-387-75466-6_36).
- Pliskunova, O., Klochko, R. 2020. Classification of e-commerce customers based on Data Science techniques. *CEUR Workshop Proceedings*. Available at: <http://ceur-ws.org/Vol-2649/paper2.pdf>.
- PythonAnywhere. 2021. Plans and Pricing. Available at: <https://www.pythonanywhere.com/pricing/>.
- RStudio. 2021. RStudio Pricing. Available at: <https://rstudio.com/pricing/>.
- SearchNode. 2021. AI search solutions for ecommerce delivered by search experts. Available at: <https://searchnode.com>.
- Shao, K., Cheng, Y. 2010. E-Commerce Comparison-Shopping Model of Neural Network Based on Ant Colony Optimization. *Advances in Wireless Networks and Information Systems*, 72, 397-404. [https://doi.org/10.1007/978-3-642-14350-2\\_50](https://doi.org/10.1007/978-3-642-14350-2_50).
- Statista. 2021. Statistics and Market Data about E-commerce. Available at: <https://www.statista.com/markets/413/e-commerce/>.
- Wang, Q., Cai, R., Zhao, M. 2020. E-commerce brand marketing based on FPGA and machine learning. *Microprocessors and Microsystems*, 103446. <https://doi.org/10.1016/j.micpro.2020.103446>.
- Zhang, D., Pee, L.G., Cui, L. 2021. Artificial intelligence in E-commerce fulfillment: A case study of resource orchestration at Alibaba's Smart Warehouse. *International Journal of Information Management*, 57, 102304. <https://doi.org/10.1016/j.ijinfomgt.2020.102304>.
- Zhou, M., Ding, Z., Tang, J., Yin, D. 2018. Micro Behaviors: A New Perspective in E-commerce Recommender Systems. *Eleventh ACM International Conference on Web Search and Data Mining*, 727-735. <https://doi.org/10.1145/3159652.3159671>.