

SCHOOL OF BUSINESS

LONDON CAMPUS

Component 2 – Data Analysis

Valentina Petkova 2232604

Data Handling and Business Intelligence

CODE / LCBB5000

TO / Molly Jack

DATE / 24 August 2024

WORD COUNT / 3000

Contents

Part 1 – E-Commerce	3
Introduction	3
Analysis	3
Order cancellation	3
1. Monthly sales trend and performance	4
2. Most purchased products	5
3. Overall purchases per transaction.....	5
4. Most profitable segment customers	5
5. Strategies for business growth and gaining more profit	6
Conclusion	8
Part 2 – Data Analysis / Score Report.....	8
1. Demographic distribution	8
2. Qualification rate.....	9
3. Mean and Median	9
4. K-means clustering.....	9
5. Advantages and disadvantages of SPSS vs Excel.....	12
References	14

Part 1 – E-Commerce

Introduction

Over the last few decades E-commerce, or also known as electronic commerce, has revolutionized the operation of business. This approach refers to the selling and buying of services and goods online. The difference between standard traditional commerce is the provision of more powerful channels that can reach global customers faster and more conveniently (Lim 2020).

Within the business world, e commerce has become a crucial component of the business strategy to many worlds' leading businesses due to the ability to remove geographical boundaries. This drastic shift from physical stores to digital platforms opens the door for wider range of products and services with incredible high efficiency and convenience. E-commerce not only broaden the access to the market but also reduces the cost of operation and minimizes the physical intermediaries (Lim, 2020).

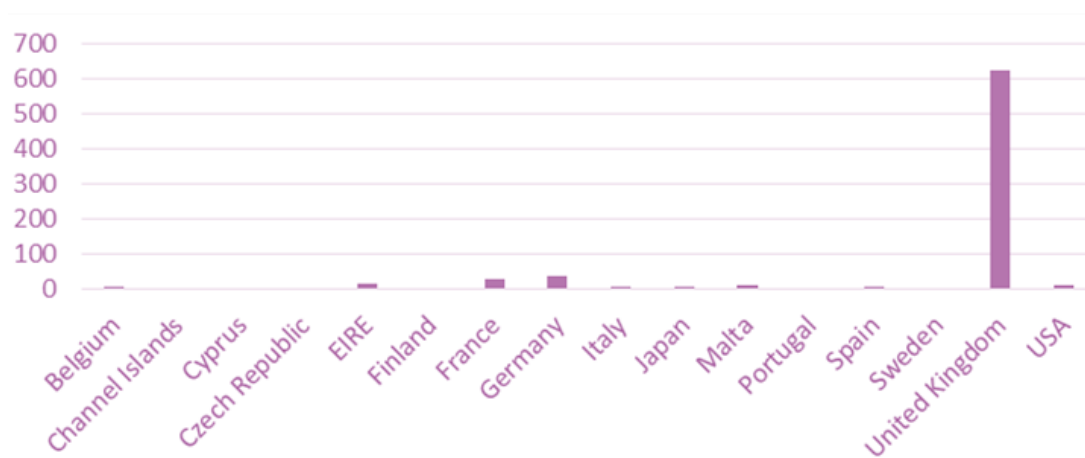
Analysis

Order cancellation

It is very common for a cancelation to appear in the orders in the moder establishments and the reasons for these are numerous. In the dataset that has been provided, a careful observation and analysis has been conducted and it has been noted that some transactions were of negative value whilst others could not be related to a customer. In total, 766 orders were cancelled during the period of transaction provided. The value of all cancelled transactions during the provided period is estimated to be £555,001.67 and majority of them

were recorded within the United Kingdom which had a total of 623 cancelled transactions.

Countries where least number of cancellations occurred (n=1) were Finland and Portugal.



1. Monthly sales trend and performance

The highest number of sales was recorded towards the end of November, which achieved a total of £359,034.12, an amount much higher than the total value of £74,291.61 recorded towards the end of the month. The monthly observations in the figure below were based on the accompanying transactions all of which varied in dynamic and specificity.

The consecutive month of December showed its highest sales within the second week of the month with a total of £650,126.36 and the lowest total sales within the previous week estimating a figure of £76,260.12. The nature of the transactions of December varied significantly than the ones in the previous months due to the frequency in observation. Within the beginning of December sales were totalling a sum of £214,210.63 which was a decline from £275,204.12 from three days before. However, transactions experienced a staggering growth that tripled its total sum in the following week, achieving £693,255.03 In sales.

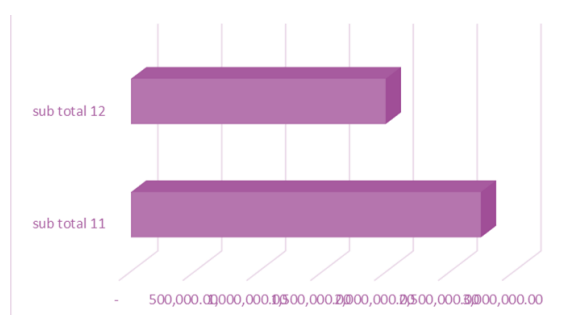


Figure 1 – Total Monthly Sales

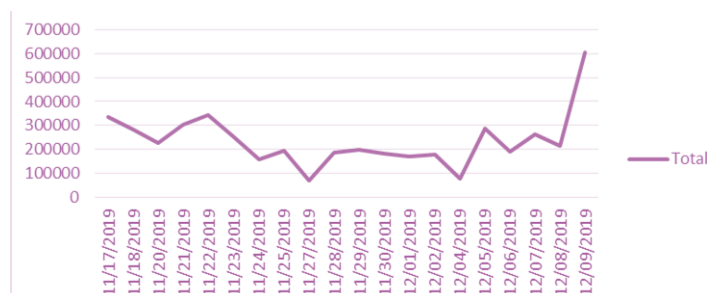
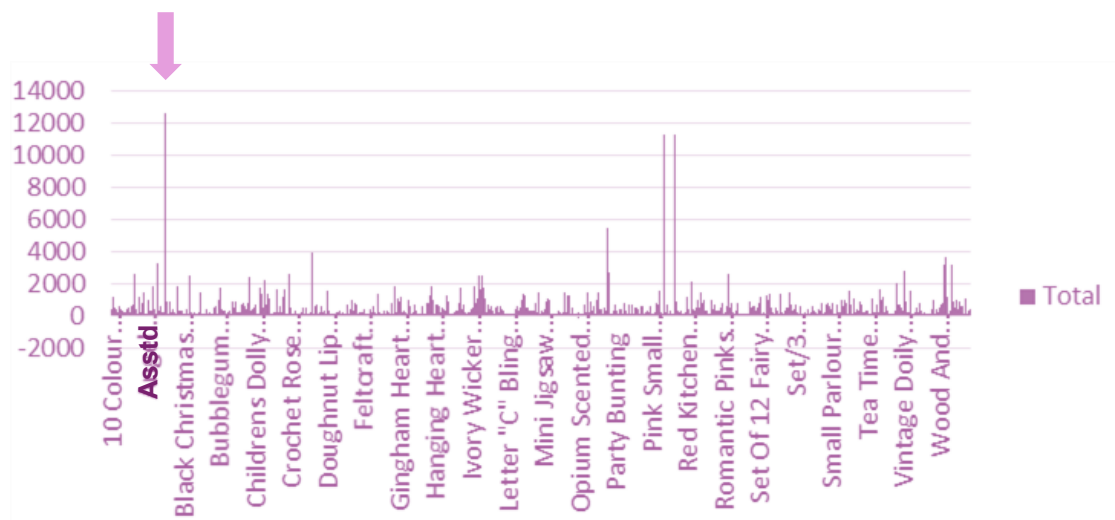


Figure 2 – Sum of total sales

2. Most purchased products

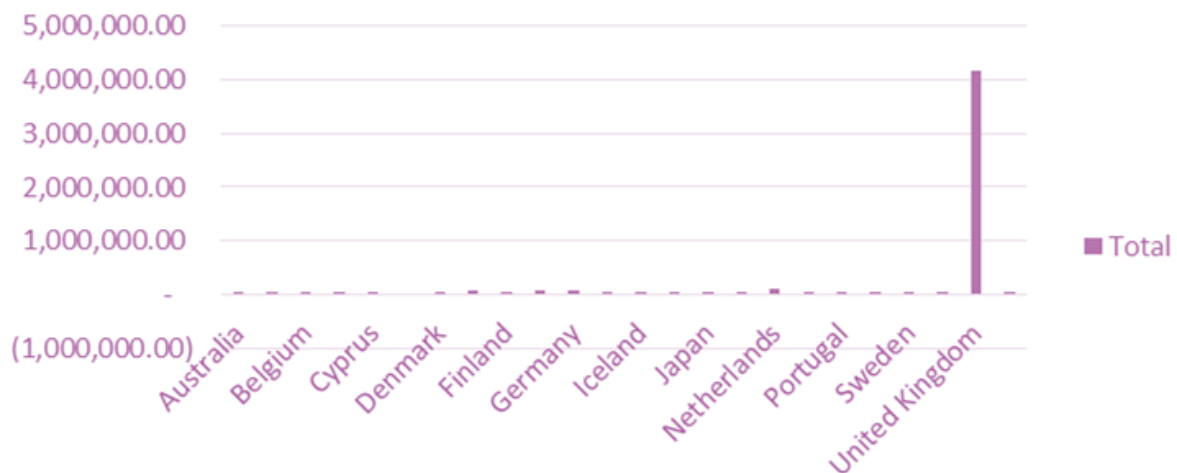


The most popular product that was sold was Asstd Design 3D Paper Stickers which sold more than 12,000 products. It was observed that more than 550,000 products were sold during the period of transaction which accounted for more than 2,500 items in total.

3. Overall purchases per transaction

The London based shop had nearly 2,500 transactions spread across 1,500 customers, which reveals that some customers initiated more than two transactions during the stated period. Some customers had single transactions on their records whilst others had two transactions on their records as well as more than three in some cases. Some of these transactions had multiple items in them whilst others just a single product. Overall the product-per-transaction ratio was calculated to be around

4. Most profitable segment customers



Sum of total sales per country

As shown on the graph above it is evident that almost 90% of all customers come from the United Kingdom and this accounts to more than 4 million in terms of value in sales.

5. Strategies for business growth and gaining more profit

Below are some recommendations to help the business growth and gain more profit based on the latest industry trends.

1. Optimize Inventory Management	Data-Driven Stocking Use sales data to identify high-demand products and ensure they are always in stock. This minimizes lost sales opportunities due to stockouts.
	Reduce Overstocking Analyse products with lower sales volume and adjust inventory levels accordingly to reduce holding costs and avoid tying up capital in slow-moving items.
2. Enhance Product Offering	Product Bundling Create attractive product bundles that offer a small discount compared to purchasing items individually. This encourages customers to buy more, increasing the average order value (Conolly, 2015).

	Introduce New Product Lines Based on sales trends, consider expanding the product range to include items that complement existing products or tap into emerging trends. For example, sustainable or eco-friendly products could attract a growing customer segment.
3. Improve Pricing Strategy	Targeted Discounts and Promotions Use customer purchase history data to offer personalized discounts or promotions. This encourages repeat purchases and builds customer loyalty without broadly discounting products.
4. Expand Marketing Efforts	Content Marketing Develop engaging content around home decor tips, gift ideas, or DIY projects to attract and retain customers. Content marketing can also improve SEO, driving organic traffic to the website.
5. Enhance Customer Experience	Improve Website Usability Ensure that the online shopping experience is seamless, with easy navigation, quick loading times, and mobile optimization. An improved user experience can reduce cart abandonment and increase conversion rates.
6. Expand Market Reach	International Shipping If not already offered, consider expanding shipping options to include more international destinations. This can open up new markets and increase sales.
	Wholesale Partnerships Strengthen relationships with small businesses that purchase in bulk and resell products. Offering them favourable terms can increase bulk orders and steady revenue streams (Conolly, 2015).
7. Leverage Customer Data	Customer Feedback Analysis Regularly gather and analyse customer feedback to identify areas for improvement. Addressing common pain points can lead to a better customer experience and higher sales.

8. Cost Management	Operational Efficiency Streamline operations, such as order processing and fulfilment, to reduce costs and improve margins. Automation tools can be leveraged to enhance efficiency in these areas.
9. Seasonal and Thematic Campaigns	Seasonal Promotions Run targeted campaigns during peak shopping seasons (e.g., Christmas, Black Friday) with themed products and limited-time offers. These can drive short-term sales spikes and increase overall revenue.
	Thematic Collections Develop collections based on popular themes (e.g., back-to-school, summer decor) that resonate with customers and encourage purchases.
10. Focus on Sustainability	Sustainable Products As sustainability becomes increasingly important to consumers, introducing more eco-friendly and sustainable products can attract a growing market segment willing to pay a premium.

Ensuring the implementation of the above strategies is vital in order to increase the profitability whilst simultaneously enhance the customer satisfaction and brand loyalty. Efficient operations and data-driven decisions are key to achieving a competitive position and long-term business success.

Conclusion

Part 2 – Data Analysis / Score Report

1. Demographic distribution

Gender					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	1056	49.5	49.5	49.5
	Male	1077	50.5	50.5	100.0
	Total	2133	100.0	100.0	

In the table above the data illustrates the demographic distribution within the school recognising almost equal split between the registering of males (50.5% / n=1077) and females (49.5% / n=1056).

2. Qualification rate

By applying an analysis to find out the frequency distribution, the data showed that of all students the ones that do not qualify for neither reduced nor free lunch are estimated to be 57% (n=1215) while the remaining 43% (n=918) qualified for free lunch.

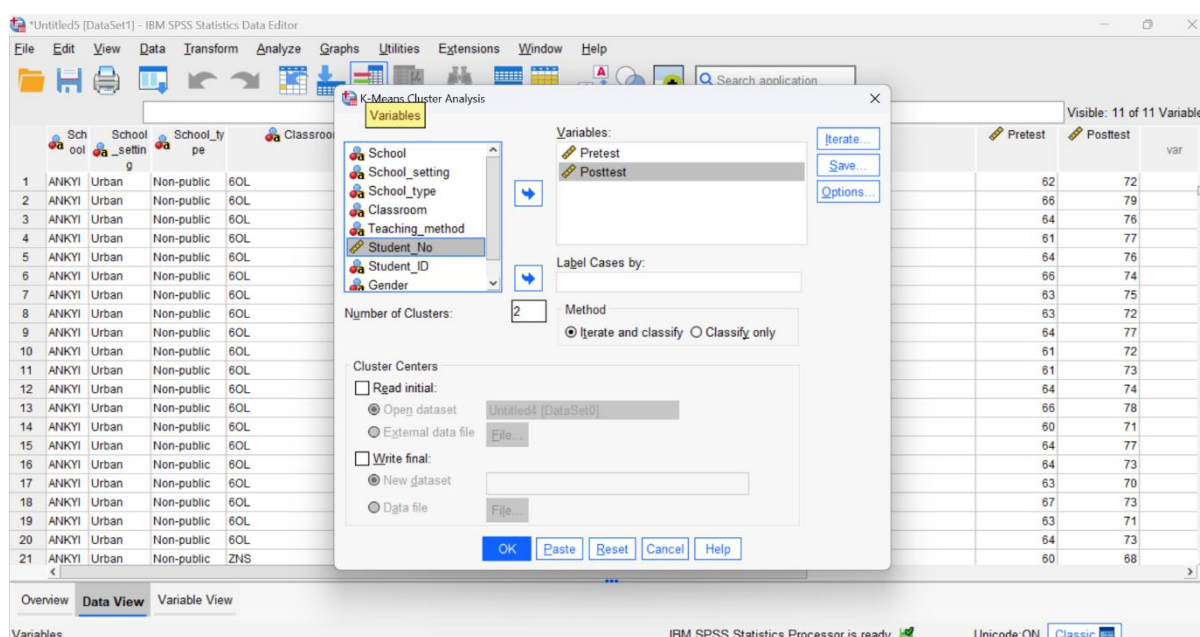
Lunch					
		Frequency	Percentage	Valid Percentage	Cumulative Percent
Valid	Does not qualify	1215	57.0	57.0	57.0
	Qualifies for reduced/free lunch	918	43.0	43.0	100.0
	Total	2133	100.0	100.0	

3. Mean and Median

Due to the lack of scale/ration mode of the variable measurement, no output was recorded that could display the students' mean and median. The dataset does not have a specific hierarchy in the way it measures.

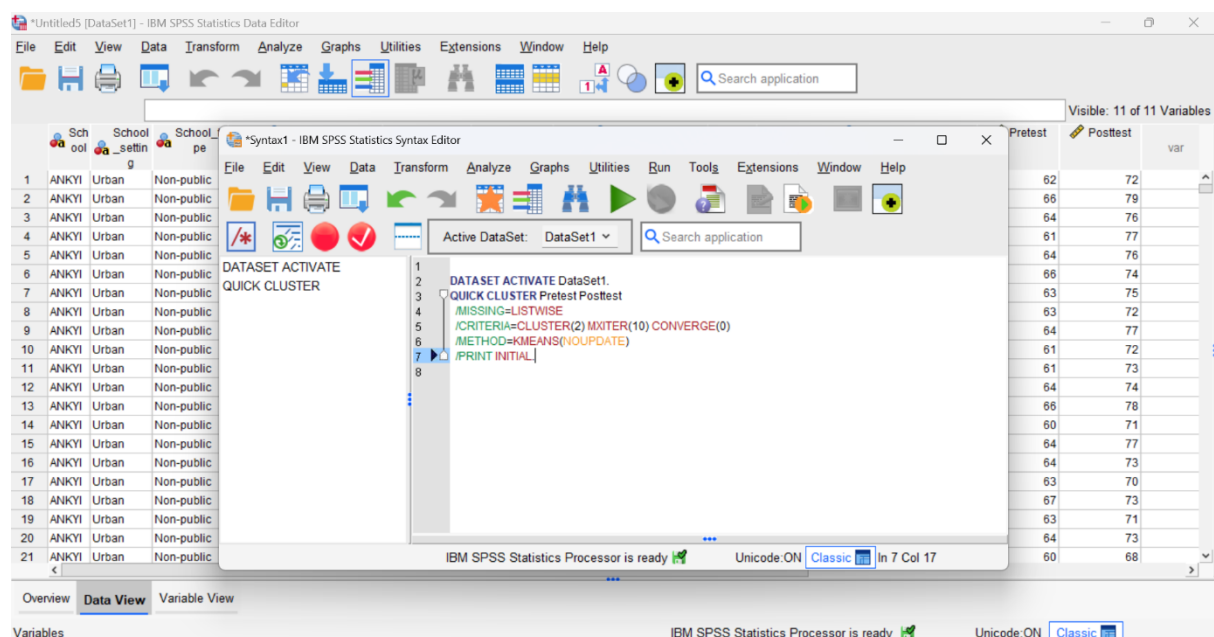
4. K-means clustering

K-means clustering is defined as a learning model that without supervision or algorithmic methodology is able to class unlabelled data and group them in clusters according to their unique



values (Makwana, 2013). This is different from other tools and methods for classification as the key focus remains on the common elements between the objectives. This clustering is especially helpful in SPSS as it offers the valuable analysis of selected numerical variables as it is illustrated further down in the pictures.

To achieve the code needed for the k-means clustering (shown below in the picture), after generating, selecting and defining each iteration the following step would be to press OK button after the code has been properly checked and satisfies the nature of the analysis. For more detailed information on the changes, selecting ANOVA table will bring the process to form an additional cluster centre options and this will generate variety in the dataset. To save the progress of the iterations it is recommended to use the save option once prompted.



Initial cluster centers		
	Cluster	
	1	2
Pretest	23	93
Posttest	35	100

The results on the left show us that the initial two clusters were variable in data groups, each of which has its unique cluster centre. The pretest showed a result of 23 (cluster 1) and 93 (cluster 2) and posttest was also of great difference with 35 in cluster 1 and 100 in cluster 2.

Upon defining the first cluster it is essential that iteration history is explored, in this scenario 7 different iterations were looked at and each of them had a change in the cluster distance whilst the analysis. The final cluster showed a variable significance from the initial cluster centre.

Final Cluster Centers		
	Cluster	
	1	2
Pretest	43	65
Posttest	55	78

Iteration History /a		
Iteration	Change in cluster Centers	
	1	2
1	30.191	33.606
2	.832	.325
3	.126	.532
4	.212	.181
5	.413	.300
6	.151	.128
7	.000	.000

The ANOVA model showed a value of 4351.122 for the F-statistic with a mean square error that was 60.505 and 2131 for the DF. From this it can be concluded that the difference between the cases under observation was quite strong. With DF being 2131 the difference between the cases was significant ($p=0.000$).

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Pretest	263262.668	1	60.505	2131	4351.122	.000
Posttest	285646.166	1	61.679	2131	4631.188	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximise the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

There is a significant variable between cluster 1 and 2 due to the number of cases as outlined in the table below.

Number of cases in each cluster		
Cluster	1	993.000
	2	1140.000
Valid		2133.000
Missing		.000

5. Advantages and disadvantages of SPSS vs Excel

Statistical Package for the Social Sciences, also known as SPSS, and Microsoft Excel are the key tools that are being largely used for data analysis. They do have distinctive advantages and disadvantages which are analysed below.

SPSS (Statistical Package for the Social Sciences)	
Advantages	Disadvantages
1. Specialized for Statistical Analysis SPSS is designed specifically for statistical analysis, offering a wide range of statistical tests, procedures, and models that are easy to access and apply.	1. Cost SPSS is a commercial software with licensing fees, which can be expensive for individuals or small businesses compared to free or less costly alternatives like Excel or R.
2. User-Friendly Interface SPSS features a point-and-click interface that allows users to perform complex statistical analyses without needing to write code. This is particularly beneficial for users who may not be proficient in programming.	2. Limited Flexibility for Non-Statistical Analysis SPSS is highly specialized for statistical analysis and might not be as versatile as Excel for tasks like data visualization, basic calculations, or handling non-statistical data.
3. Comprehensive Output SPSS provides detailed output for all analyses, including tables, charts, and descriptive statistics. The output is organized in a way that is easy to interpret and includes thorough documentation of the statistical methods used.	3. Learning Curve Although SPSS is user-friendly for basic tasks, mastering advanced features and understanding the statistical concepts behind them can require significant learning.
4. Customizability with Syntax While SPSS is user-friendly, it also allows advanced users to write syntax (scripts) for custom analyses, making it versatile for both beginners and experienced statisticians.	4. Less Suitable for Non-Statistical Data SPSS is not ideal for general data manipulation, organization, or for tasks like creating complex financial models, which Excel handles more effectively.

5. Compatibility with Other Statistical Software SPSS can import and export data in various formats, including Excel, CSV, and others, making it easy to collaborate and share data across different platforms.	
---	--

Microsoft Excel	
Advantages	Disadvantages
1. Widely Available and Accessible Excel is widely used across various industries and is often included as part of Microsoft Office, making it easily accessible for most users.	1. Potential for Errors Excel is prone to errors, especially in large or complex datasets. Manual data entry and formula creation can lead to mistakes that might not be immediately obvious.
2. Versatility Excel is highly versatile and can be used for a variety of tasks, including data entry, data analysis, financial modelling, reporting, and creating charts and graphs.	2. Handling of Large Datasets Excel can struggle with very large datasets, both in terms of performance and the maximum number of rows and columns it can handle. SPSS is better suited for large-scale data analysis.
3. Ease of Use Excel's interface is intuitive, and users can perform basic calculations, create charts, and manage data without needing advanced training.	3. Limited Data Management Excel lacks advanced data management capabilities such as those found in SPSS. Tasks like merging datasets, recoding variables, and conducting multiple regressions are less straightforward in Excel.
4. Integration with Other Tools Excel integrates seamlessly with other Microsoft products and can easily import/export data from various sources, making it a flexible tool in business environments.	4. Inadequate Documentation of Statistical Processes Excel does not provide as much detail in its output for statistical analyses, which can make it harder to review and verify the results compared to SPSS.
6. Low Cost Excel is generally less expensive than SPSS, and many users already have access to it as part of their Microsoft Office suite.	

--	--

Data analysis becomes easier with SPSS and Excel. However, there are some differences in between them. For example, SPSS is the preferable tool for more in depth statistical analysis that has built in features for detailing and outlining whose extensive approach is more attractive to researchers and scientists. On the other hand, Excel is more versatile, easy to use, and much more accessible which serves well for general data manipulation, modelling in finance and other basic analytical tasks. The choice between the two tools largely depends on the outcome goals and what is needed to be achieved as the complexity of the data analysis will differ, so it is important to be clear with aspects such as budget and deliverables.

References

- A.Azevedo & M.F. Santos (eds.) (2015) Integration of data mining in business intelligence systems . Hershey :, Business Science Reference.
- Azevedo, A. (2021) Integration challenges for analytics, business intelligence, and data mining . Hershey, PA :, IGI Global.
- A. Ishikawa & J. Nakagawa (eds.) (2013) An introduction to knowledge information strategy : from business intelligence to knowledge sciences . Hackensack, New Jersey :, World Scientific.

Connolly, T. (2015) Database systems : a practical approach to design, implementation, and management . Sixth edition, Global edition. Boston :, Pearson.

Howson, C. (2008) Successful business intelligence [electronic resource] : secrets to making BI a killer app . 1st edition. New York, NY :, McGraw-Hill.

Kodinariya, T. M., & Makwana, P. R. 2013. Review on determining the number of cluster in K-means Clustering. International Journal, l(6), 90-95

Lim, Y.Y. & Teoh, A.P. (2020) Realizing the strategic impact of business intelligence utilization. Strategic direction (Bradford, England). 36 (4), 7–9. doi:10.1108/SD-09-2019-0184.

Maheshwari, A.K. (2015) Business intelligence and data mining . New York :, Business Expert Press.