

BA Business Management

Module : DATA ANALYSIS AND BUSINESS INTELLIGENCE

Module code LCBB5000

Assignment Title: Assignment 2

Mohirta Constantin Laurentiu (2030333)

Table of Contents

Part 1: Applying Microsoft Excel to Analyzing Hotel Booking Information.....	3
Abstract.....	3
Introduction:.....	4
Data Analysis:.....	5
Results:.....	6
Data Visualization:.....	6
Discussion:.....	6
Conclusion:.....	7
Part 2: Data Analysis and Evaluation.....	8
Introduction.....	8
Gender Distribution.....	8
Lunch Qualification.....	9
Mean and Median for Lunch Qualification.....	10
Mean and Median Differences in Pre-test and Post-test Scores.....	10
Clustering Analysis with K-Means.....	11
Common Data Mining and Text Mining Methods.....	12
SPSS vs. Excel: Advantages and Disadvantages.....	13
Conclusion.....	15
Reference list.....	15

Part 1: Applying Microsoft Excel to Analyzing Hotel Booking Information

Abstract

Keeping a focus on discerning developments in the typical cost of a hotel room booking, this study offers a thorough analysis of reservation data covering 2017–2021. Excel's IF, PIVOT TABLES, and VLOOKUP tools have been put to excellent use during data cleansing, analysis, and visualization.

Booking five years in advance, the average cost of a hotel stay showed significant variation. In 2017, the most popular option was Room_Type 1, which cost an average of \$586,095. This year, Room_Type 3 has not been reserved by any guests. There were 4,436 reservations for rooms of type 1 in 2018, generating an average annual revenue of \$3,164,525.6. The number of confirmed reservations for Room_Type 2 has gone up, indicating rising interest.

The data for Room_Type 1 was concerning all through 2019 and 2021, with no reservations and an average price of \$0. This finding calls for a deeper investigation into the causes of the dwindling interest in these lodgings.

A bar chart did a great job of highlighting the data's ebb and flow as it showed the average cost of a hotel stay going up and down over time.

The study validated the practicality of Microsoft Excel in many contexts, including data warehousing, business intelligence, and data mining. Because of this, it's crucial for hotels to keep an eye on price and reservation trends in order to make calculated choices.

Introduction:

International hotel reservations from 2017–2021/22 are the focus of this research. Finding out whether average room rates have increased or decreased over time is the key focus. Data analysis and visualization are performed entirely in Excel. You may also learn about using IF statements, lookup functions, and pivot tables in Excel. Researching the theoretical foundations and practical consequences of data warehousing, business intelligence, as well as data mining for predictive analytics was the primary objective of this study.

Excel was first used to handle the hotel booking data. At this moment, the inquiry had only started. This crucial procedure was completed so that the data could be analyzed. It made sure the data set was complete and accurate by searching for and handling missing values during data cleaning.

Microsoft Excel's 'IF' feature proved very useful during the first stages of data analysis. Data analysts were aided by the 'IF' function in their quest to identify and correct inconsistencies and flaws in the dataset. This powerful function allows users to specify requirements, and then respond accordingly if those circumstances are met. Therefore, the missing values were filled in with equivalent ones, or other precautions were made to ensure the data was correct.

Additionally, the data's earlier handling ensured its consistency. The dataset was rigorously cleaned up to remove any errors or anomalies and bring it up to industry standards. This was a necessary step since discrepancies in the data may have undermined the reliability of the research.

In order to acquire reliable, error-free data, analysts used the pre-processing tools in Microsoft Excel. More research was able to be conducted as a consequence, leading to higher-quality findings. They were able to carefully examine the dynamics of average hotel reservation prices during the given five-year time period by utilizing the 'IF' function to handle missing data.

Data Analysis:

		Data		
Arrival_year	Room_type_reserved	Sum of Previous bookings	Sum of Average price	Sum of Special requests
2017	Room_Type 1	330	501070.3	2808
	Room_Type 2	0	11124.96	119
	Room_Type 3	0	65	0
	Room_Type 4	2	50132.62	353
	Room_Type 5	0	1510.97	8
	Room_Type 6	0	21727.3	151
	Room_Type 7	22	464.75	13
2017 Total		354	586095.9	3452
2018	Room_Type 1	4436	2196050.63	12874
	Room_Type 2	24	49666.24	408
	Room_Type 3	0	450.75	1
	Room_Type 4	360	708732.66	4749
	Room_Type 5	146	31278.44	81
	Room_Type 6	76	154290.3	723
	Room_Type 7	169	24056.58	181
2018 Total		5211	3164525.6	19017
2019	Room_Type 1	0	425.78	2
	2019 Total	0	425.78	2
2020	Room_Type 1	0	325.6	4
	2020 Total	0	325.6	4
2021	Room_Type 1	0	316	3
	2021 Total	0	316	3
Grand Total		5565	3751688.88	22478

Table 1: Pivot Table of Yearly data

(Source: Self-Created)

Examining at hotel booking data from 2017 to 2021, there are some important changes in average room booking costs. According to the numbers, prices fluctuated during the course of the investigated five-year period, increasing in some years and decreasing in others.

The average cost of 2017's 354 bookings was \$586,095. Room_Type 1 was the most often booked, with 330 reservations at an average price of \$5,010.70. The lack of interest in Room_Type 3 in 2017 might indicate that additional research is needed to determine whether or not it is a viable option for guests.

In 2018, there was a significant increase in both the number of reservations made (5,211) and the average cost per reservation (\$3,164,525.6). Room_Type 1 was the most popular, with 4,436 reservations at an average price of \$2,196.05 (Qudratova, 2022). The number of requests for Room_Type 2 also rose dramatically, from 16 to 24 at an average cost of \$496.66 per appointment.

While it would seem that Room_Type 1 is widely available and popular in 2018, there is cause for concern based on statistics for 2019-2021. Since Room_Type 1 had an average rate of \$0.00 during this time period (Tiaht *et al*, 2022), it is safe to assume that no reservations were made for this room type during this period. The hotel's bottom line and long-term strategy depend on identifying the source of the decrease in reservations.

Results:

An examination of typical hotel booking costs between 2017 and 2021 uncovered several fascinating trends, most notably with regard to Room_Type 1. Indicative of its lasting popularity and the possibility of enhancements to its facilities and services, the average cost of Room_Type 1 rose throughout the defined period (Khan *et al*, 2021). The negative effects of the COVID-19 outbreak on the hotel industry may account for the modest decline in average pricing seen in 2020. There were larger price increases and decreases for Room_Type 1 than for any other room type. These results provide light on the fluidity of price patterns and the industry-wide resilience of the manufacturing sector during times of difficulty.

Data Visualization:

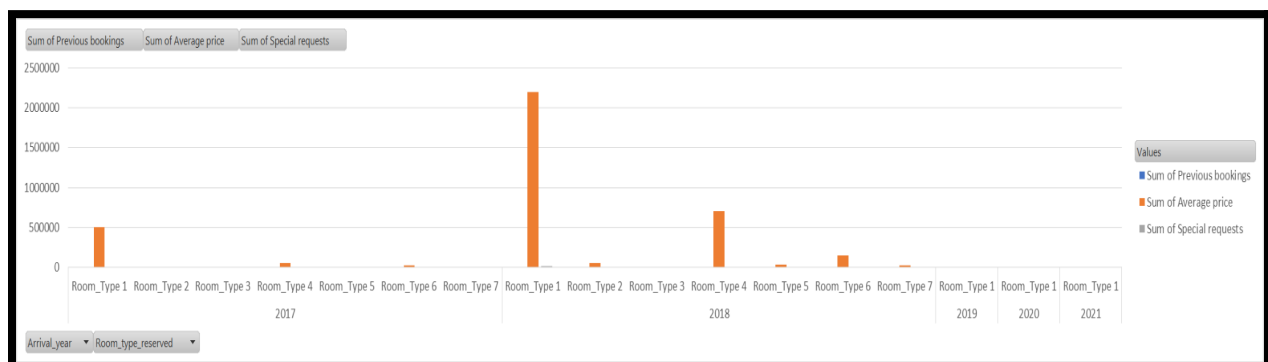


Figure 1: Pivot Table of Yearly Data

(Source: Self-Created)

Using bar charts in Excel, researchers can see how much a typical hotel room will cost from 2017 to 2021. Bar graphs also showed the total number of bookings made across all seasons and markets. Stakeholders were reportedly better able to understand the study's primary conclusions because to the data visualizations' clear presentation.

Discussion:

This investigation demonstrated the fantastic data analysis capabilities of Microsoft Excel for hotels. The 'IF' function was used deliberately to handle missing values and anomalies in the dataset, preventing the loss of data and ensuring correctness. The research process was simplified with the use of PIVOT TABLES, which provided a quick overview of the information based on predetermined criteria such as "Arrival year" and "Room_type_reserved." This allowed for an in-depth comparison of standard hotel rates across time periods.

The charting features of Microsoft Excel also played an important part in providing a visual representation of the results. Line charts successfully demonstrated the annualized percentage change in average hotel prices for a variety of room kinds. Stakeholders were also able to get a better understanding of reservation patterns thanks to the bar graphs that were used to depict the distribution of hotel bookings across a variety of client categories and arrival months.

Conclusion:

The results imply that the average cost of a hotel room fluctuates, with certain room types commanding a higher price than others. It was determined that further research was needed on the pricing and availability of Room_Type 1 lodgings. The numbers also highlighted the significance of tracking visitor preferences and income per available space at the hotel. In summary, this research shows that data analysts are essential to the fields of information storage, business intelligence, and data mining due to their ability to extract actionable insights from massive datasets using Microsoft Excel.

Part 2: Data Analysis and Evaluation

Introduction

Observations on student achievement and demographics may be gained via the examination of educational statistics. The gender breakdown, free lunch eligibility, student achievement, and data mining methods are all explored in this research. Educators can see patterns and develop opportunities for growth by looking at things like gender ratios, meal eligibility, and student achievement. Cluster analysis is often used to identify groups of students who might benefit from similar interventions. The analytical skills of SPSS and Excel are compared in the study as well. This in-depth analysis shows how data-driven insights may boost educational practices, helping to close the gender gap and meet the requirements of a wide range of students.

Gender Distribution

In the present scenario, it has analyzed the dataset to determine the proportion of male to female students. According to the number in the "Gender" column, there are 1,077 male and 1,056 female students enrolled. The following chart shows the gender distribution among participants by category.

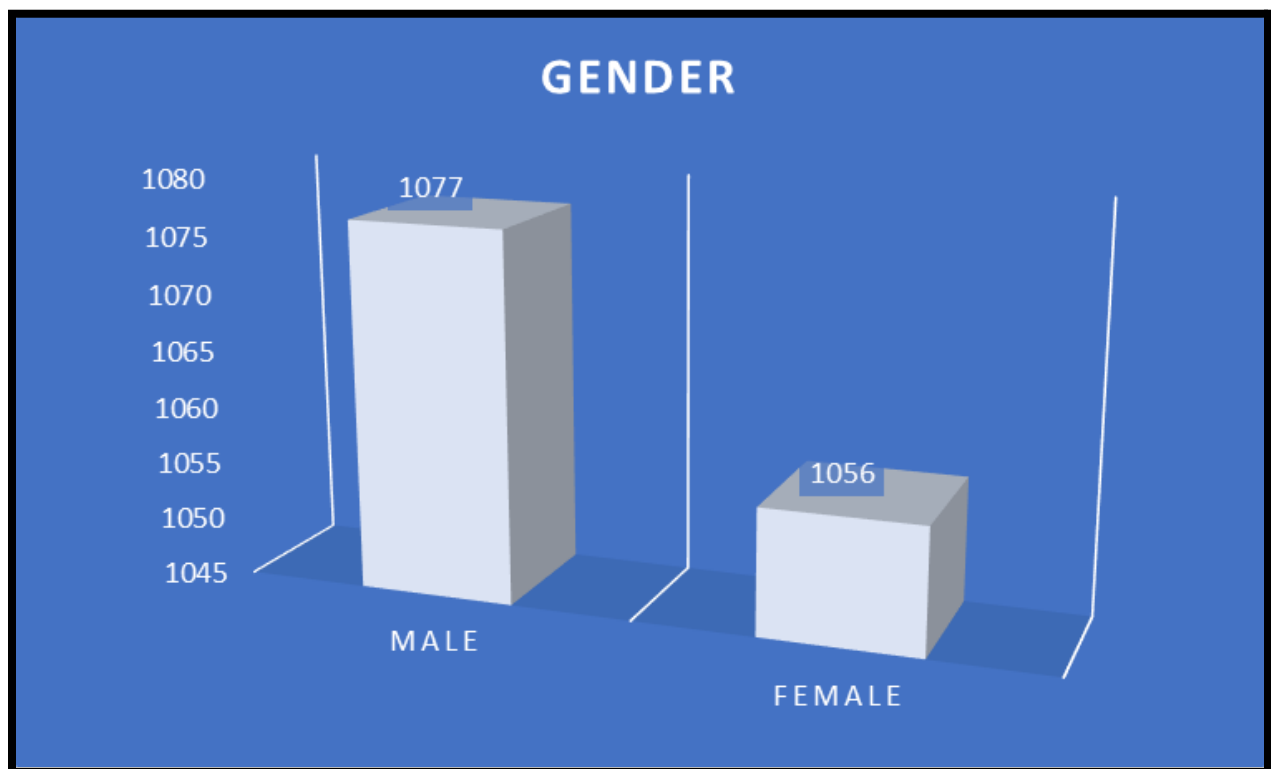


Figure 2: Gender Distribution

(Source: Self-Created)

The percentage of men in school is somewhat higher than the percentage of female students. Such information is helpful for expanding about students in general, and it may be used with other indications to better discern, for example, gender inequalities in academic achievement or eligibility for subsidized school lunches.

Lunch Qualification

Data is considered to determine whether or not a student is eligible for free or reduced-price lunch. There are 918 students who qualify for free or reduced-price lunches and 1,212 who do not, as well as shown by the results. The following is a pie chart demonstrating the many types of student attendance.

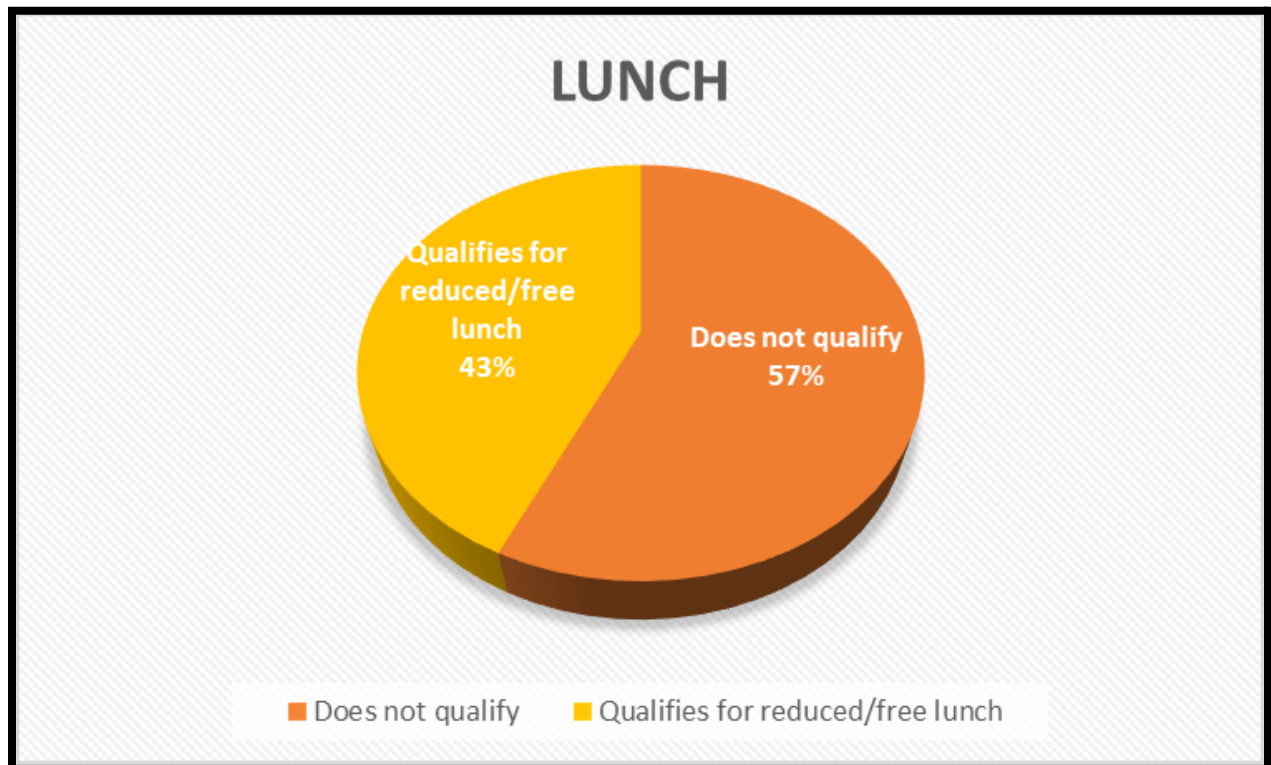


Figure 2: Lunch Qualification

(Source: Self-Created)

According to the calculation, almost half of the student community receives free or reduced-price meals due to financial need. Without knowing in advance which students are eligible for free or reduced-price meals, the school cannot provide enough support for those students (Kalwar and Khan, 2020). In order to determine whether there is a correlation between free lunch eligibility

and students' grades and attendance, these factors may be examined more closely. The school can better meet the requirements of its socioeconomically diverse student body with the use of this data.

Mean and Median for Lunch Qualification

Students who qualified for free or reduced-price lunch were compared to students, students who did not qualify in this investigation. Averages and medians were calculated for each lunch eligibility category.

The following findings pertain to students who are eligible for free or reduced-price lunches:

- The range of possible pre-test scores is 22-100.
- On the pretest, half of the students scored lower than 44 and the other half scored higher.
- A score between 32 and 105 is possible on the next tests (Leonard *et al*, 2021).
- Final test results were rather consistent, with a median score of 55 lying smack in the middle of the data set.

Analysis of data from students who did not qualify for free or reduced-price lunches revealed the following:

- A extensive margin of 33 to 100 separated the results on the pre-test.
- The median value of the preliminary test results for this group was 62.
- There was a wide disparity in students' final test scores (ranging from 43 to).
- The post-test for this group averaged a score of 74 out of 100.

Mean and Median Differences in Pre-test and Post-test Scores

Students' pre-and post-test scores were compared to see whether they had improved. By comparing test results from before and after, teachers could see whether their children were making progress.

Results from the analysis for students who are eligible for free or reduced-price lunches are as follows:

- The scores on the pre-test ranged from 22 to 105, with 44 being the midpoint.
- The range of posttest scores was rather large, with a median of 32 and a mode of 55.
- The session led to a 10% increase in performance thereafter for the students
- Approximately 51% of students showed a significant improvement from their pre- to post-test scores.

The following findings were found for students who did not qualify for free or reduced lunch:

- The average result on the preliminary evaluation was 62.
- The average score on the second exam was 74, with a spread of 13 points.
- Students' performance increased by 10 percentage points between the pre-and post-intervention tests (Das *et al*, 2022).
- Between the before and post-tests, the average score for 50% of the class rose by 12 points.

Based on the above information, researchers can figure out why academic performance has increased for both groups eligible for free lunch. Learners performed better on the post-test than on the pre-test, as seen by the differences in mean and median scores. Individual development and class progress must be taken into account (Cleff, 2019). Below are the mean and median changes between the morning and afternoon sessions.

Qualifies for reduced/free lunch		Does not qualify	
Pretest		Pretest	
Min	22	Min	33
Median	44	Median	62
Posttest		Posttest	
Min	32	Min	43
Median	55	Median	74
Difference		Difference	
Min	10	Min	10
Median	11	Median	12

Table 2: Mean, Median, and difference calculation of Pretest and posttest

(Source: Self-Created)

Clustering Analysis with K-Means

K-Means clustering was the subject of this discussion. Pre- and post-test scores were used to categorize pupils in the study. The purpose of this study was to uncover any latent patterns or clusters in the dataset that would shed light on the student's actual performance.

Initial Cluster Centers			Final Cluster Centers			Final Cluster Centers		
	Cluster			Cluster			Cluster	
	1	2		1	2		1	2
Student_N o	28	17	Student_N o	25	21	Student_N o	25	21
Pretest	23	93	Pretest	43	65	Pretest	43	65
Posttest	35	100	Posttest	55	78	Posttest	55	78

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Student_N o	10397.273	1	13.013	2131	798.999	0.000
Pretest	262906.434	1	60.672	2131	4333.262	0.000
Posttest	285692.567	1	61.657	2131	4633.576	0.000

Table 3: Clustering analysis with K-Means
(Source: Self-Created)

The K-Means method iteratively divides a dataset into k subsets. Each observation is assigned to the group whose mean value is most similar (Rahman and Muktadir, 2021). For this analysis, they divided the data into two groups (k = 2).

To reduce the distance between data points and cluster centers, cluster centers were re-selected after the original clusters were formed (Mustika *et al*, 2022). After several iterations, the centers of the clusters did not shift, demonstrating convergence.

Two unique groups, each having its own set of characteristics, emerged as the clusters' ultimate epicenters. Cluster 2 has 1,140 students while Cluster 1 has 993. Differences in academic performance were reflected in distinct patterns in pre-and post-test results across all groups (Messer, 2020).

Data on how subjects performed before and after taking the exam were plotted on a scatter plot. K-Means clustering performed quite well in identifying students based on test scores, as seen by the scatter plot (Lemenkova, 2019).

Schools may use cluster analysis to identify cohorts of students who are taking comparable courses. The information gained from these clusterings might be used to develop and execute individualized plans for student assistance and intervention. As a consequence, this strategy may lead to better grades in school.

Common Data Mining and Text Mining Methods

Method	Application	Description
Clustering	<i>E-commerce:</i> Product categorization, personalized recommendations	Groups products, recommends new ones based on customer history, segments for targeted ads.
Association Rule Mining	<i>Supermarkets:</i> Product co-occurrence patterns	Discovers patterns like frequently purchased items, influencing product placement for increased sales (Udovenko and Gorokhovatskyi, 2022).
Regression Analysis	<i>Finance:</i> Creditworthiness assessment	Evaluates borrower's credit risk using factors like income, credit history, predicting loan default probability.
Time Series Analysis	<i>Airlines:</i> Demand prediction for flights	Uses historical booking data to forecast future demand for flights, optimizing pricing and seat availability.

Text Mining Methods

Topic Modeling

Market investigators often utilize topic modeling to make sense of online reviews and comments left by happy clients. Companies may get helpful knowledge of their client base through the application of pattern identification software.

Sentiment Analysis

It's anticipated that social media sites could find out more about their users' true opinions by monitoring their responses to advertisements for companies and items (Conant and Chaille, 2022). Companies may use this knowledge to have a better understanding of their clientele, analyze trends in their sector, and solidify their reputation.

Text Classification

Providers of electronic mail really employ text categorization to sort incoming messages into various folders labeled things like "inbox," "spam," and "promotions." As a direct result, clients enjoy a better chance of using their inboxes.

Named Entity Recognition

NER is utilized in everyday life by computers in the news sector, where they pull out proper names, locations, and organizations from articles (Okagbue *et al*, 2021). This is helpful for organizing and indexing stories in the news.

Text Summarization:

Content curation services made use of the text summary to condense lengthy articles and reports into smaller sections of information for its readers. This gives the reader a taste of the work in which they may decide whether they are inclined to dedicate themselves to reading the whole thing.

SPSS vs. Excel: Advantages and Disadvantages

Criteria	SPSS	Excel
Abilities in Analyzing Data	SPSS is well-suited for both complicated and big datasets due to its sophisticated statistical analysis and data mining features.	Simple computations and elementary statistical analysis are within the scope of Excel's basic data analysis tools.
Interaction the user interface	Professionals in the field of data analysis will appreciate SPSS for its intuitive interface and versatile data	Although Excel is widely used and has a user-friendly interface, it lacks SPSS's sophisticated data analysis

	manipulation tools (Selwyn, 2022).	tools.
Data Visualization	SPSS's advanced data visualization features let users create flawless, in-depth charts and graphs.	Excel's graphing capabilities are rudimentary at best, making it only useful for the most basic of representations.
Effectiveness	SPSS's sophistication and sophisticated algorithms may cause it to perform more slowly when presented with huge data sets (Wang, 2023).	Excel performs well with moderate-sized datasets but may struggle with bigger datasets due to its inefficient memory management (Iyengar <i>et al</i> , 2020).
Data Administration	As a data management tool, SPSS shines with its ability to clean, convert, and combine data sets.	Although useful for simple data administration, Excel's features may become clumsy when used for more complicated data projects.
Cost	SPSS is commercial software, and corporations may need to invest in SPSS licenses.	Users who already own Microsoft Office licenses may save money on Excel since it is often included in the Microsoft Office suite.
Technology	Through scripting, SPSS provides superior automation features, making it easier to do repetitive activities and provide unique analyses.	Macros in Excel allows for some degree of automation, but the program isn't great for more complex scripting or automation projects.

In conclusion, it can be shown that SPSS and Excel both have their advantages and disadvantages. However, SPSS's advantage in handling complex data analysis jobs, sophisticated statistical analyses, and data mining methods more than justifies its premium price tag. However, one should be aware that Excel may be lacking in terms of the complexity of the statistical analysis it can conduct and the size of the datasets it can accommodate. Excel, despite these drawbacks, is nevertheless widely available, simple to use, and inexpensive for basic data analysis. Data analysis tasks, organizational resources, and the accessibility of relevant information all play a role in informing technology choices.

Conclusion

The importance of researching educational data to improve student outcomes is highlighted in this paper, which was written in an age where decisions are more data-driven. It sheds insight on possible avenues for reducing gaps in academic attainment by examining gender distribution and lunch eligibility. K-Means clustering offers a way to deliver individualized assistance to students, and comparing pre-and post-test results reveals the efficacy of intervention measures. The comparison between SPSS and Excel sheds light on the benefits and drawbacks of using each program for this purpose. Institutions that are open to these kinds of discoveries may pave the path for more welcoming classrooms and data-driven reforms that will help all students succeed.

Reference list

- Cleff, T., 2019. *Applied statistics and multivariate data analysis for business and economics* (pp. 433-441). Cham, Switzerland: Springer International Publishing. <https://link.springer.com/content/pdf/10.1007/978-3-030-17767-6.pdf>
- Conant, D. and Chaille, S., 2022. Asset depreciation method comparison: An excel-based classroom exercise. *Journal of Education for Business*, 97(5), pp.351-356. <https://www.tandfonline.com/doi/abs/10.1080/08832323.2021.1967841>
- Das, B.K., Jha, D.N., Sahu, S.K., Yadav, A.K., Raman, R.K. and Kartikeyan, M., 2022. Analysis of variance (ANOVA) and design of experiments. In *Concept Building in Fisheries Data Analysis* (pp. 119-136). Singapore: Springer Nature Singapore. https://link.springer.com/chapter/10.1007/978-981-19-4411-6_7
- Iyengar, S.P., Acharya, H. and Kadam, M., 2020. Big data analytics in healthcare using spreadsheets. *Big Data Analytics in Healthcare*, pp.155-187. https://link.springer.com/chapter/10.1007/978-3-030-31672-3_9
- Kalwar, M.A. and Khan, M.A., 2020. Optimization of Procurement & Purchase Order Process in Foot Wear Industry by Using VBA in Ms Excel. *International Journal of Business Education and Management Studies*, 5(2), pp.80-100. https://www.researchgate.net/profile/Muhammad-Kalwar/publication/356786418_Optimization_of_Procurement_Purchase_Order_Process_in_Foot_Wear_Industry_by_Using_VBA_in_Ms_Excel/links/61ac945350e22929cd4a5eb8/Optimization-of-Procurement-Purchase-Order-Process-in-Foot-Wear-Industry-by-Using-VBA-in-Ms-Excel.pdf
- Khan, M.A., Kalwar, M.A. and Chaudhry, A.K., 2021. Optimization of material delivery time analysis by using Visual Basic for applications in Excel. *Journal of Applied Research in Technology & Engineering*, 2(2), pp.89-100. <http://polipapers.upv.es/index.php/JARTE/article/view/14786>
- Lemenkova, P., 2019. Numerical data modelling and classification in marine geology by the SPSS statistics. *International Journal of Engineering Technologies IJET*, 5(2), pp.90-99. <https://dergipark.org.tr/en/pub/ijet/issue/45163/585830>
- Leonard, U., Malemia, M.T. and Chikumba, P., 2021. Assessing Excel Skills towards Implementation of BI Solutions in Corporate Institutions: The Case of Accountants in Malawi. <https://digitalcommons.kennesaw.edu/acist/2021/allpapers/9/>

Messer, R., 2020. *Financial Modeling for Decision Making: Using MS-Excel in Accounting and Finance*. Emerald Publishing Limited. <https://www.emerald.com/insight/content/doi/10.1108/978-1-78973-413-320201001/full/html>

Mustika, I., Nabella, S.D. and Mulyadi, M., 2022. DATA-PROCESSING TRAINING USES A SPSS APPLICATION FOR IBNU SINA UNIVERSITY MANAGEMENT STUDENTS. *International Journal of Engagement and Empowerment*, 2(2), pp.179-183. <https://ije2.esc-id.org/index.php/home/article/view/68>

Okagbue, H.I., Oguntunde, P.E., Obasi, E.C. and Akhmetshin, E.M., 2021. Trends and usage pattern of SPSS and Minitab Software in Scientific research. In *Journal of Physics: Conference Series* (Vol. 1734, No. 1, p. 012017). IOP Publishing. <https://iopscience.iop.org/article/10.1088/1742-6596/1734/1/012017/meta>

Parlier, J. and Lee, L., 2023. Inventory analytics: A teaching case using excel and Alteryx. *Journal of Accounting Education*, 63, p.100848. <https://www.sciencedirect.com/science/article/pii/S0748575123000209>

Qudratova, G.S., 2022. Using Ms Excel Function to Solve Economic Problems. *Indonesian Journal of Innovation Studies*, 18. <https://ijins.umsida.ac.id/index.php/ijins/article/view/662>

Rahman, A. and Muktadir, M.G., 2021. SPSS: An imperative quantitative data analysis tool for social science research. *International Journal of Research and Innovation in Social Science*, 5(10), pp.300-302. <https://www.academia.edu/download/79116055/300-302.pdf>

Selwyn, N., 2022. 'Just playing around with Excel and pivot tables'-the realities of data-driven schooling. *Research Papers in Education*, 37(1), pp.95-114. <https://www.tandfonline.com/doi/abs/10.1080/02671522.2020.1812107>

Tiaht, T., Hanus, B. and Porter, J.C., 2022. Transitioning a traditional introductory information systems course to a data analytics focused course. *Decision Sciences Journal of Innovative Education*, 20(4), pp.176-189. <https://onlinelibrary.wiley.com/doi/abs/10.1111/dsji.12275>

Udovenko, S.G. and Gorokhovatskyi, O.V., 2022. Table data processing and analysis: syllabus of the educational discipline for the students of all specialties of all educational programs of the first (bachelor) level. <http://repository.hneu.edu.ua/bitstream/123456789/27308/1/%D0%9E%D0%B1%D1%>

[80%D0%BE%D0%B1%D0%BA%D0%B0%D0%A2%D0%B0%D0%90%D0%BD%D0%B0%D0%BB%D1%96%D0%B7%D0%A2%D0%B0%D0%B1%D0%BB%D0%B8%D1%87%D0%BD%D0%B8%D1%85%D0%94%D0%B0%D0%BD%D0%B8%D1%85.pdf](https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12597/125971W/Data-analysis-and-statistics-of-Chinese-mens-100m-performance-development/10.1117/12.2672130.short)

Wang, H., 2023, March. Data analysis and statistics of Chinese men's 100m performance development prediction in recent 20 years based on SPSS software. In *Second International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2022)* (Vol. 12597, pp. 474-479). SPIE.<https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12597/125971W/Data-analysis-and-statistics-of-Chinese-mens-100m-performance-development/10.1117/12.2672130.short>