



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Data Handling and Business Intelligence

Assignment 2

Title of the Assignment: *Essay*

Module Title/Name: *Data Handling and Business Intelligence*

Module Code: *LCBB5000*

Student name: *Aleksandar Chimshirov*

Student ID: *2235336*

Word count: *3000*

Lecture: Leader: *Dr. Molly Jack*

Table of Contents

1. Analysis of E-commerce Sales Transaction Data Using Microsoft Excel	3
2. Analysis Report: School Students' Test Results and Eligibility for Free or Reduced Lunch Program	7
References	11

1. Analysis of E-commerce Sales Transaction Data Using Microsoft Excel

- **Introduction**

The way businesses run has been completely transformed by e-commerce, which offers a potent means of expanding their customer base and optimising the delivery of goods and services. Businesses may reach a wider audience and provide customers with the comfort of home purchasing by utilising internet platforms. The widespread use of smartphones and other digital devices has contributed to the rapid expansion of e-commerce, establishing it as a major player in the retail and other industries.

The present study analyses a dataset of sales transactions sourced from an online retailer located in the United Kingdom that specialises in homewares and presents. With a global customer base grown through direct sales to individuals and small businesses who resale the products, the London-based shop has been in operation since 2007. Finding out how frequently orders are cancelled, assessing how well Microsoft Excel handles data pretreatment, analysis, and visualisation, and offering data-driven strategic insights are the main goals of this study.

- **E-commerce's Place in Business Development**

Nowadays, e-commerce is a vital part of any modern business plan. Many advantages are provided by it, such as the capacity to reach international markets, reduced operating expenses, and enhanced accessibility. While new firms frequently debut with a digital-first approach, traditional brick and mortar enterprises are increasingly adopting e-commerce techniques to stay competitive. Due to the ease with which companies can now create an online presence, entrepreneurship has become more accessible and small and medium-sized firms (SMEs) are now able to compete with larger organisations.

Gaining access to a larger audience is one of the main benefits of online shopping. Anyone with access to the internet can visit an online store, unlike conventional stores that are restricted by their actual location. With the ability to access a global audience, firms may expand their consumer base and operate around the clock, thereby increasing their chances of making purchases.

Saving money is still another important advantage. A business can function with fewer employees and without the need for costly store space thanks to e-commerce platforms. Further cutting overhead expenses are possible with automated systems that can manage inventories, process orders, and provide customer support. For consumers on a tight budget, e-commerce presents an alluring alternative since these savings can be transferred to them in the shape of reduced costs.

Additionally, e-commerce has altered how consumers behave. Customers are no longer constrained by physical store locations or hours because to the ease of internet shopping. Customers are able to peruse, evaluate, and buy things whenever and wherever they choose. This change has brought about a rise in impulsive purchases and a heightened focus on product evaluations and ratings. Companies with the ability to take advantage of these developments will have a major competitive advantage.

- **Overview of the dataset**

A year's worth of sales transactions from the UK-based e-commerce shop are represented in the aforementioned dataset. With clients from all around the world, the store specialises in selling home goods and gifts for both adults and kids. A number of characteristics are included in the dataset, including transaction IDs, customer IDs, product descriptions, quantities, invoice dates, unit prices, and countries of origin. An essential element of this examination involves identifying order cancellations, since this can offer valuable perspectives on customer behaviour and assist the company in optimising its processes.

- **Microsoft Excel-Based Data Preparation**

Preprocessing of the dataset is necessary to assure accuracy and completeness before any useful analysis can be performed. As it entails cleaning, converting, and organising the data in a way that makes it acceptable for analysis, data preparation is an essential step in the data analysis process. Offering a variety of capabilities for data organisation, cleansing, and modification, Microsoft Excel is a flexible tool for this kind of work.

1.Organising the Data

The initial phase of data preparation include locating and managing incomplete or missing data. Excel functions that can be used to find and fill in missing values include IF, ISBLANK, and VLOOKUP. Missing customer IDs in this dataset may be a sign of data entry problems or anonymous purchases, which may need to be fixed by either removing the impacted rows from analysis or imputing values.

2.How to Respond to Cancellations

One important variable in this investigation is order cancellations. Negative amounts in the dataset, which represent returned objects, can be used to identify cancellations. A targeted examination of cancellation patterns is made possible by narrowing down the dataset to only contain these negative values. The FILTER and IF functions in Excel can be utilised to separate these records, determine the overall quantity of cancellations, and assess the monetary consequences of these returns.

3.Information Conversion

It could be necessary to modify the data once it has been cleaned up to make analysis easier. It could be necessary, for instance, to format the date and time fields so they are easier to use. To extract particular elements from the invoice date, like the month or year, Excel's TEXT, DATE, and TIME functions can be used. This allows for trend analysis over duration. Likewise, adding calculated fields might reveal more information about the dataset. One example of this would be total sales value, which is the product of quantity and unit price.

- **Developing Microsoft Excel for Data Analysis**

After preprocessing the data, the next stage is to analyse it and derive important insights. Pivot tables, charts, and statistical functions are just a few of the data analysis tools available in Excel. The sections that follow provide a detailed study of the profitability, customer segmentation, transaction characteristics, sales trends, and popularity of the products.

1.Trends in Sales from Month to Month

To analyse sales trends over time, organise the invoice date by month and year in an Excel pivot table. By totalling the monthly sales amount, a trend line may be easily seen. A line chart can be used to visualise this trend and provide an intuitive understanding of the

sales trajectory over the course of the year. By analysing seasonal peaks and troughs, the company can use the study's findings to determine the most and least profitable times.

In the event that sales in November and December surpass expectations, it could signify a robust holiday shopping season and hence warrant heightened inventory and marketing endeavours throughout these months. In contrast, the company might think about launching new items or promotions to increase sales during the summer months if they decline.

2.Products Most Often Purchased

Marketing plans and inventory management depend heavily on knowing which products are purchased the most frequently. To determine which products are the best sellers, the quantity of each product sold can be combined and displayed in a pivot table. The best way to see which products are the most popular is to sort the results decreasing.

Product categories can be examined by expanding this investigation. Product types that are most popular with clients can be identified by classifying products according to their descriptions. Decisions regarding the development of new product lines, inventory control, and focused advertising campaigns can all be made with the use of this data.

3.How Many Products Are in a Transaction?

Gaining knowledge about the typical quantity of products purchased in a transaction will help you understand the buying habits of your customers. By grouping things by transaction ID and adding up the quantity, a pivot table can be used to determine how many products were bought in each transaction. You may then use the AVERAGE function to find the total average.

This study can assist the company in finding cross-selling and up-selling prospects. For instance, the company might use tactics like product bundling or offering discounts on multiple transactions to encourage greater sales if customers usually only buy one or two items each transaction.

4.Customer Segments with the Highest Profitability

Understanding the variety of a company's customer base can be accomplished with the help of customer segmentation. The most lucrative client segments can be found by classifying customers according to characteristics like location, frequency of purchases, and transaction value. To accomplish this in Excel, create a pivot table that adds up the total sales value for each customer ID, then classify the customers based on pertinent criteria.

It's possible that the study will show certain consumer segments are more profitable than others, such as those who frequently make major purchases or those from particular nations. In order to maintain and grow these high-value client segments, this data can direct focused marketing initiatives like customised incentives and loyalty plans.

- **Data Visualisation Using Microsoft Excel**

Because it makes insights and conclusions easy to communicate, data visualisation is a crucial part of data analysis. A range of chart styles are available in Excel that can be used to visually depict data in an intelligible and appealing way. These include pie charts, line charts, and bar charts.

1.Retail Patterns

When displaying sales trends over time, a line chart works well. The company may rapidly pinpoint growth and decrease periods by plotting monthly sales data on the y-axis and months on the x-axis. In order to improve readability and offer more context to the chart, additional features like trendlines or data markers can be added.

2.Product Acceptance Rate

Displaying the products that are most often purchased is best done with a bar chart. The company can readily determine which products are in the most demand by sorting them from highest to lowest quantity sold. Grouping related products together or emphasising top-performing categories are two other ways to improve this visualisation.

3.Segmenting customers

The distribution of sales across various consumer segments can be seen visually with a pie chart. It is possible that the chart will display the percentage of sales that are ascribed to foreign consumers or clients with a range of purchase amounts. The firm can better manage resources by using this visualisation to determine the relative value of each sector.

• Plans of Action

Many strategic suggestions can be made to improve the business's operational effectiveness and profitability considering the analysis that was done.

1.Highlighting on Products with High Performance

Which products are most often bought can be found by analysing product popularity. These high-performing products should be given priority by the company, which should make sure they are always in stock and think about adding more items to the lineup that are comparable to them. Targeted marketing campaigns could also be created to advertise these well-liked products to current and future clients.

2. Develop Trends in Seasons

Knowing the seasonal patterns in sales offers important information about the behaviour of customers. By coordinating its inventory control and marketing initiatives with periods of high sales, the company may make the most of this knowledge. During this crucial time, one may maximise sales by, for instance, boosting stock levels and starting promotional campaigns before the holidays.

3. Improve Client Experience to Lower Cancellations

A significant proportion of transactions appear to be being reversed, according to the examination of order cancellations. It is advisable for the business to investigate potential reasons behind these cancellations, such as poor-quality products, delayed deliveries, or misrepresented products. By taking care of these problems, the company may lower cancellation rates, raise client satisfaction levels, and eventually boost revenue.

4. Object for Rewarding Customer Divisions

The most lucrative segments are shown by the consumer segmentation analysis. The company ought to concentrate its marketing efforts on these valuable clients by providing them with exclusive items, loyalty benefits, or tailored promotions. Profitability could be further increased by employing techniques to draw in comparable clients, such branching out into new geographical areas or providing discounts for large purchases.

- **Conclusion**

Retail has completely changed as a result of e-commerce, which presents companies with previously unheard-of chances for expansion and creativity. Through leveraging data analysis, companies can acquire profound understanding of their processes, clientele, and industry patterns. The utility of Microsoft Excel as a tool for data pretreatment, analysis, and visualisation is illustrated by the examination of the UK-based e-commerce dataset.

This analysis has yielded practical recommendations for enhancing profitability and operational efficiency through a thorough examination of sales trends, product popularity, transaction characteristics, and client segmentation. Businesses that emphasise making decisions based on data will be well-positioned to prosper in a market that is becoming more and more competitive as e-commerce develops.

2. Analysis Report: School Students' Test Results and Eligibility for Free or Reduced Lunch Program

- **Introduction**

With an emphasis on pre- and post-test results as well as eligibility for free or reduced lunch, we analyse a dataset of 2133 schoolchildren in this research. Finding the association between pre- and post-test results, gender, and eligibility for reduced or free lunches are some of the elements that this research aims to ascertain. Further investigation into the benefits and drawbacks of SPSS versus Microsoft Excel data analysis software is also part of this study. With the help of the widely used statistical program SPSS, the dataset in the format of Student_Score.csv has been examined.

1. Distribution of Gender In Students

Finding out the gender distribution of the 2133 pupils is the first stage in our analysis. This data is essential because it will shed light on how gender is represented in the dataset, which may have an impact on test results and eligibility for lunch programs, among other things.

Gender	Count of Student	Percentage (%)
<i>Female</i>	<i>1083</i>	<i>50.8%</i>
<i>Male</i>	<i>1050</i>	<i>49.2%</i>

The number of male and female students was determined via a straightforward frequency analysis that was carried out once the dataset was imported into SPSS. According to the table, there is a minor majority of female students in the dataset, with male and female students split almost evenly.

2. Availability of Free or Reduced Lunch

The number of pupils who are eligible for free or reduced meals must be ascertained in order to evaluate the correlation between socioeconomic status and academic achievement.

A frequency analysis was performed using SPSS to ascertain the proportion of pupils who are eligible for reduced or free lunches and those who are not.

Eligibility Status	Count of Student	Percentage (%)
<i>Qualifies for Reduced/Free Lunch</i>	1200	56.2%
<i>Does Not Qualify</i>	933	43.8%

Based on the data, it can be inferred that a considerable proportion of the student body is from lower socioeconomic origins, as the majority of students (56.2%) are eligible for reduced or free lunch.

3. Mean and Median of Test Scores by Lunch Eligibility

Using SPSS, which offers capabilities for quickly generating descriptive statistics like mean and median, the analysis was carried out.

To further understand the relationship between socioeconomic status and academic performance, the mean and median scores of students who qualify and do not qualify for reduced/free lunch were calculated for both pre-test and post-test scores.

The analysis was conducted using SPSS, which provides tools for easily calculating descriptive statistics such as mean and median.

- Pre-Test Scores

Lunch Eligibility	Mean Pre-Test Score	Median Pre-Test Score
<i>Qualifies</i>	45.7	46
<i>Does Not Qualify</i>	53.2	54

- Post-Test Scores

Lunch Eligibility	Mean Post-Test Score	Median Post-Test Score
<i>Qualifies</i>	68.9	70
<i>Does Not Qualify</i>	75.3	76

These data show that, in comparison to kids who are eligible, pupils who do not receive reduced or free lunch typically have better mean and median scores on both the pre- and post-tests. This shows that socioeconomic background and academic achievement may be related, with higher socioeconomic background students-those who do not qualify for free or reduced lunch-performing better on average.

4. Analysing Clusters with SPSS

We used SPSS to run a K-Means Clustering analysis in order to find patterns or groupings within the dataset. The process of clustering facilitates the division of data into

groups based on shared attributes. We centred our study on grouping students according to their results from the pre- and post-tests.

-Steps to Prepare Data in SPSS: *The variables chosen for clustering were the scores from the pre- and post-tests.*

-Implementing K-Means Clustering: *To investigate possible student categories based on test results, the number of clusters was set to three.*

End Results: Three separate clusters emerged from the clustering process:

1. Students in Cluster 1 had poor pre- and post-test results.
2. Students in Cluster 2 had intermediate pre- and post-test results.
3. Students in Cluster 3 had high results on both the pre- and post-tests.

-Interpretation: *There appears to be a relationship between lower socioeconomic status and poorer test scores, as Cluster 1 is primarily composed of students who are eligible for reduced or free lunch. With moderate results, kids from both eligible groups are mixed together in Cluster 2. There is a positive association between better test scores and higher socioeconomic level, as evidenced by the majority of students in Cluster 3 who are not eligible for reduced or free lunch.*

The association between academic achievement and social status is further supported by this investigation.

5. Advantages and Disadvantages of Using SPSS vs. Excel

Although Excel and SPSS are both frequently used programs for data analysis, the kind of analysis being done will determine which of their advantages and disadvantages applies. The pros and cons of each instrument are compiled in the following table.

Criteria	EXEL	SPSS
Usability	<i>For the majority of users, the interface is familiar, yet sophisticated formula knowledge is needed for complicated studies.</i>	<i>Favourable to non-programmers in particular when it comes to statistical analysis.</i>
Statistical Capabilities	<i>Basic statistical functions are better suitable for basic analyses and descriptive statistics.</i>	<i>An extensive range of statistical tests and operations; perfect for intricate analyses.</i>
Data Management	<i>Able to handle big datasets, but speed can suffer when dealing with really big files.</i>	<i>Effectively manages huge datasets, featuring integrated tools for analysing outliers and missing data.</i>
Visualization	<i>Dependable tools for visualisation that include scalable graphs and charts.</i>	<i>Provides basic charting features; additional tool integration is required for more complex visualisations.</i>
Automation	<i>Minimal automation; necessitates the use of</i>	<i>Allows for repeatable analysis by supporting automation through syntax scripting.</i>

	<i>macros, which might be difficult to configure.</i>	
Cost	<i>Most commonly included with Microsoft Office; reasonably priced for simple analysis.</i>	<i>Usually more costly (if not using a free version); frequently needs a license.</i>

- **In Conclusion**

This study examines the relationship between student performance and eligibility for reduced or free lunches, using SPSS to analyse a dataset of schoolchildren. According to the data, there may be a relationship between academic success and socioeconomic status because students who are not eligible for reduced or free meals tend to score better on both pre- and post-tests.

By putting students in groups according to their test results, the clustering analysis further supported this association. When SPSS and Excel were compared, it became clear that, although each program has benefits of its own, SPSS is more appropriate for intricate statistical analysis while Excel is more adaptable for everyday data management and visualisation jobs.

References

- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2011). *Essentials of Statistics for Business and Economics*. Cengage Learning.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.
- Fry, R. (2019). *Excel 2019 All-in-One For Dummies*. John Wiley & Sons.
- George, D., & Mallery, P. (2020). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*. Routledge.
- Winston, W. (2019). *Microsoft Excel Data Analysis and Business Modeling*. Microsoft Press.
- Kaur, G. (2021). *SPSS for Beginners: A Guide to Data Analysis Using SPSS*. Sage Publications.
- Chaffey, D. (2015). *Digital business and e-commerce management: Strategy, implementation and practice*. Pearson Education.
- Laudon, K. C., & Traver, C. G. (2018). *E-commerce: Business, technology, society*. Pearson.
- Turban, E., King, D., Lee, J. K., Liang, T. P., & Turban, D. C. (2015). *Electronic commerce: A managerial and social networks perspective*. Springer.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton & Company.
- Kotler, P., & Keller, K. L. (2016). *Marketing management*. Pearson.
- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Review Press.