



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

BA BUSINESS MANAGEMENT

Assessment 2 - Essay

Module: Data Handling and Business Intelligence

Code: LCBB5000

Lecturer: Dr. Anupam Mazumbar

Name: Anderson Sanaguano Masache

Student ID: 2132325

SUBMITTED ON THE 2nd September 2023

Contents

Part 1: Analysis of Room Reservation Prices Over the Years (2017 - 2022)	3
1. Introduction.....	3
2. Methodology	4
2.1. Data Collection and Preparation	4
3. Analysis and Visualizations	4
3.1. Yearly, Quarterly and Monthly Analysis of Average Room Reservation Prices.....	4
Figure 1: Average Reservation Prices Over Time.....	5
Figure 2: Quarterly Room Reservation Prices Over the Years.	6
Figure 3: Monthly Room Reservation Prices Over the Years.....	6
Table 1: Yearly Occupancy For both Adults and Children	7
3.2. Average Prices Based on Reservation Characteristic, Context and History	8
Figure 4: Average Price by Meal Plan	8
Figure 5: Average Price by Parking Space Required.....	9
Figure 6: Average Price by Room Type	10
3.3. Booking Context and History	10
Figure 7: Average Price by Market Type	11
Figure 8: Number of Special Guests and Loyal Customers by Market Type and Booking Status	11
4. Recommendations	12
5. Conclusion	12
Part 2: School Free Lunch Eligibility	12
Introduction	12
Research Questions and Hypotheses	13
Results and Discussion	14
1. <i>Gender Distribution</i>	14
2. <i>Lunch Eligibility</i>	15
Table 1:.....	16
3. <i>Descriptive Statistics</i>	16
Table 2: Descriptive Statistics of Pre-test and Post-test Scores	16
4. <i>Cluster Analysis</i>	17
Five Most Common Data and Text Mining Methods	22
Advantages and Disadvantages SPSS and Excel for Data Analysis	24
References	25

Part 1: Analysis of Room Reservation Prices Over the Years (2017 - 2022)

1. Introduction

The hotel industry, and the whole hospitality sector, is a cornerstone within the larger domain of the economies. It has over the years dominated the to-do lists of many travellers and tourists seeking a haven of comfort and convenience during their journeys and vacations. This is because they offer various accommodation services in line with their customers' preferences to enhance the overall travel experience, be it for leisure or business purposes. This sector not only fulfils the basic requirement of shelter. In addition, its significance can hardly be overstated as hotels contribute to various communities' social and economic fabric. Thereby facilitating interactions, fostering cultural exchanges, and generating employment opportunities.

Families on vacation can relax and spend cherished moments in a hotel, which is their home away from home. Couples and lone visitors enjoy a sense of solitude and elegance, heightening the romance and the excitement of exploration. They also cater to the demands of corporate people through business packages, which promote networking and collaboration, spurred by the desire of the business traveller for effective services and facilities, such as conference rooms and dependable Wi-Fi.

Beyond the personal realm, tourism, job creation, and revenue generation are just but a few ways through which the hotel sector promotes economies. This is because they attract tourists who in turn move to different locations promoting the local enterprises, thus initiating economic advancements. The impact of their expenditures is finally felt across the wider community when they engage in gastronomic and shopping experiences.

2. Methodology

2.1. Data Collection and Preparation

The report was generated using a dataset which comprised of 36,275 rows and 21 columns. Each row corresponds to a distinct booking, identified by a unique Booking ID. The columns are the key booking characteristics such as the number of adults and children, reservation details, meal plan, market type, and average price. The absence of missing values in the dataset ensures the reliability of this report is attributed.

In order to determine to obtain deeper insights, new columns were created. These columns include "Total Visitors," "Total Nights," "Date," and "Quarter," achieved through the implementation of Excel's SUM and LOOKUP functions.

1. The "Total Visitors" column is derived by adding the count of adults and children for each booking.
2. The "Total Nights" column is established by aggregating the counts of both weekday and weekend nights.
3. The "Date" by combining the "Arrival Year," "Arrival Month," and "Arrival Date" columns.
4. The "Quarter" column is constructed using the LOOKUP function, mapping each month to its respective quarter.

3. Analysis and Visualizations

3.1. Yearly, Quarterly and Monthly Analysis of Average Room Reservation Prices

The average room reservation prices have fluctuated over the years from 2017 to 2022.

Figure 1: Average Reservation Prices Over Time

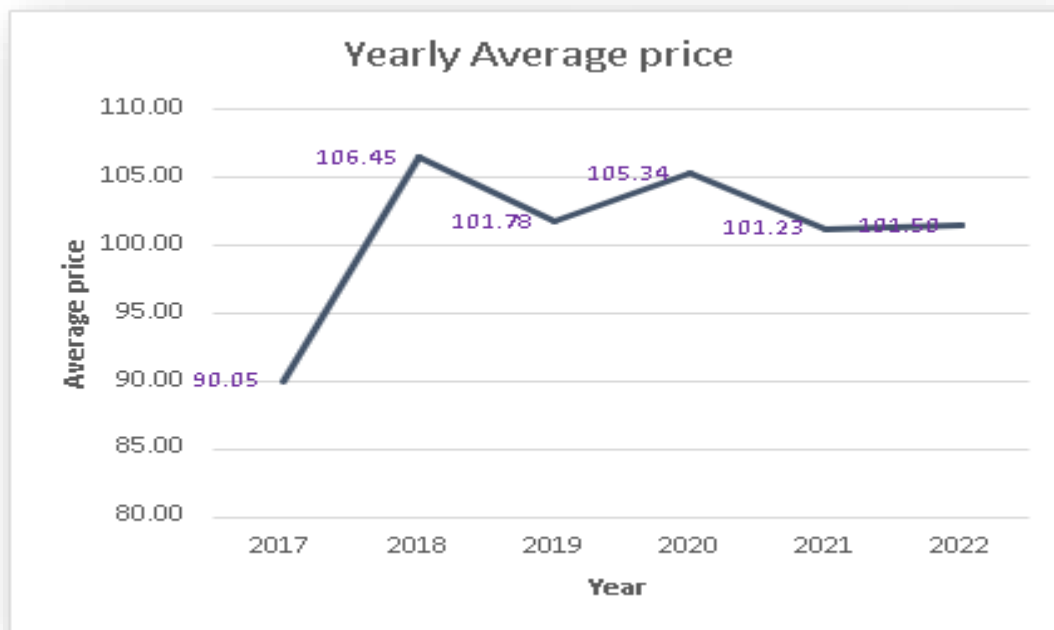


Figure 1 shows the average prices for each year and the corresponding increase or decrease compared to the previous year. The average price increased from 2017 to 2018 by \$16.40, indicating a huge jump. However, from 2018 to 2019, there was a decrease of \$4.67 in the average price. The average price exhibited fluctuations in the subsequent years, with a marginal increase in 2022.

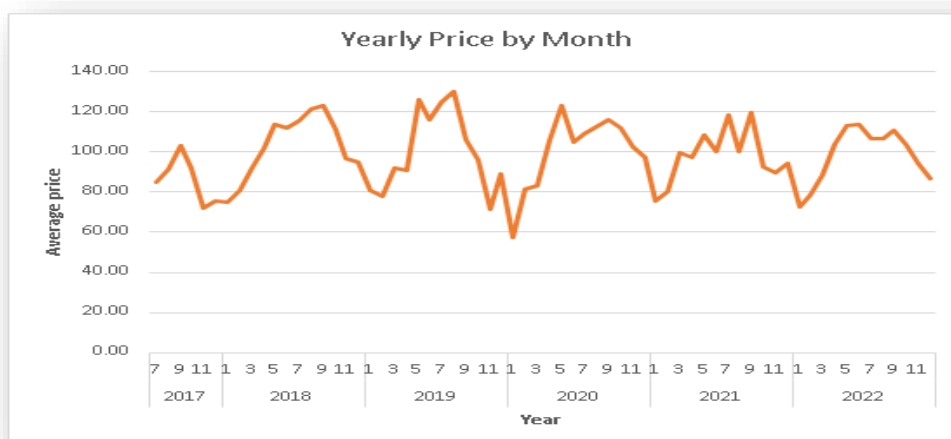
In order to understand how average room reservation prices have evolved over the years, we conducted a seasonally adjusted analysis. A quarterly analysis revealed that the second and third quarters (Q2 and Q3) have the highest average prices, indicating peak reservation periods while the first and fourth quarters (Q1 and Q4) experience relatively lower average prices.

Figure 2: Quarterly Room Reservation Prices Over the Years.



Furthermore, a monthly analysis provided a detailed view of how the average room reservation prices varied across months. From Figure 2, there is a significant variation in average room reservation prices across different months within each year.

Figure 3: Monthly Room Reservation Prices Over the Years.



Hotels in the hospitality industry experience distinct seasonal fluctuations in demand, which can be grouped into both peak and off-peak periods (Alrawabdeh, 2021). These

variations indicate the presence of seasonal pricing patterns where prices tend to fluctuate based on the time of the year. A closer look reveals that certain months consistently stand out as periods of higher prices across the years (June to August), while December, January and February exhibit the lowest room prices. This shows that the reservation prices are normally highest during the summer months and the lowest average prices are observed in the winter months.

Lardieri, 2021 attributed the failure to price recover to the outbreak of COVID-19 since the final months of 2019. The number of customers increased significantly from 2017 to 2018. This aligns with the increase and decrease in average prices during the same period. It was then followed by a drastic drop in 2019 which has since then struggled to recover despite the subsequent slight increases. The bookings averaged at 78.4% in 2018 before falling a historic low of 0.5% in 2019.

Table 1: Yearly Occupancy For both Adults and Children

Arrival_year	Children		Bookings	Percentage
	Adults No	No		
2017	10727	403	11130	15.7%
2018	52216	3215	55431	78.4%
2019	320	15	335	0.5%
2020	382	15	397	0.6%
2021	364	18	382	0.5%
2022	2917	153	3070	4.3%

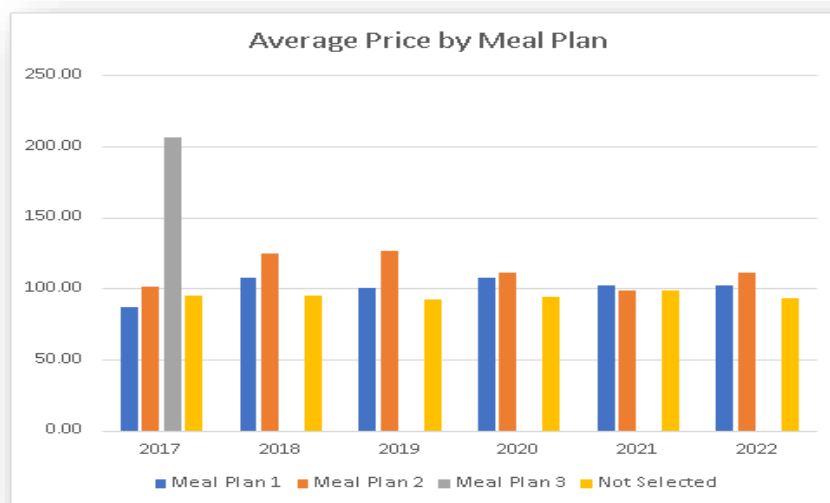
According to the American Hotel and Lodging Association, the effect of covid-19 is estimated to eliminate more than 10 years of job growth in the accommodations sector (Lardieri, 2021, pp1; AHLA'S STATE of the HOTEL INDUSTRY 2021). Hotels have

been trying to implement dynamic pricing in order to stabilize their revenue without compromising quality (Gibbs et al., 2017).

Certain reservation characteristics affect the average reservation prices differently. These include meal plans, parking space, and room type.

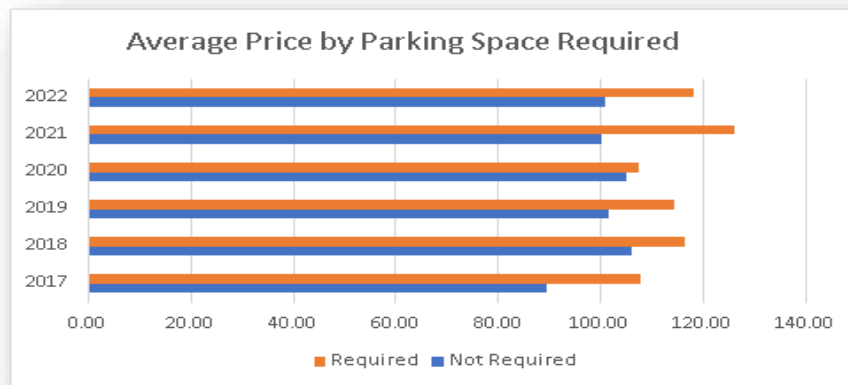
3.2. Average Prices Based on Reservation Characteristic, Context and History

Figure 4: Average Price by Meal Plan



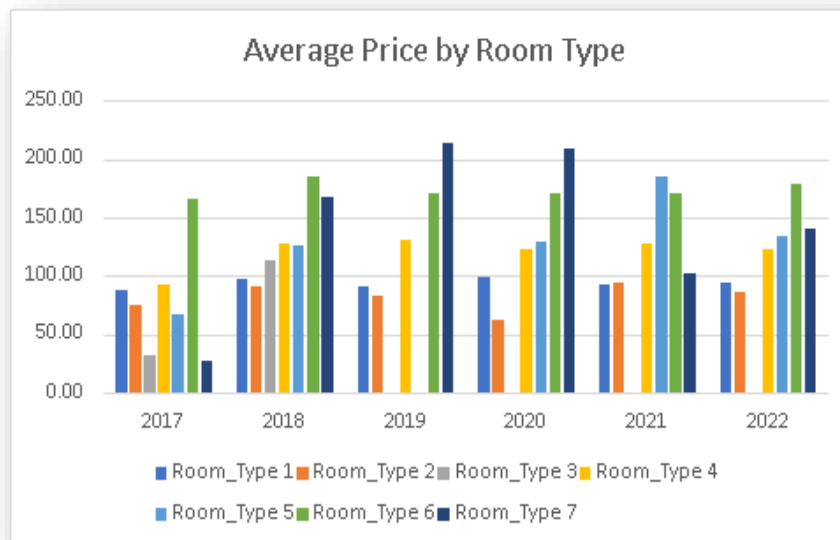
Customers who require car parking tend to pay higher average room reservation prices compared to those who do not need parking.

Figure 5: Average Price by Parking Space Required



The average reservation prices over the years are significantly associated with room type. These prices also tend to vary with popularity based on their features. nb"Room_Type 1" and "Room_Type 4" consistently show the highest popularity across all years, with substantial visitor counts. "Room_Type 4" consistently has higher average prices, indicating that it might offer premium features or additional amenities to room 1. Room_Type 6" shows an increasing trend in popularity over the years, with a notable increase in 2018. Additionally, "Room_Type 6" stands out with significantly higher average prices compared to other room types, suggesting it might be a luxury or specialty room.

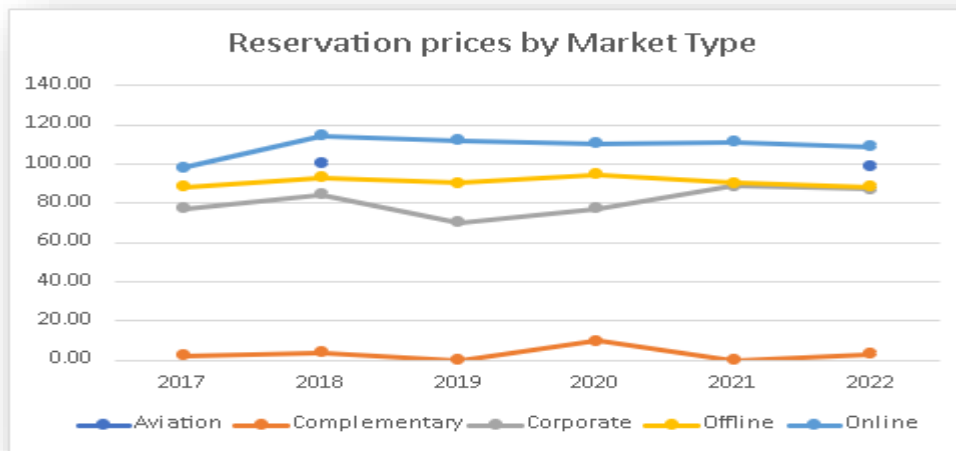
Figure 6: Average Price by Room Type



3.3. Booking Context and History

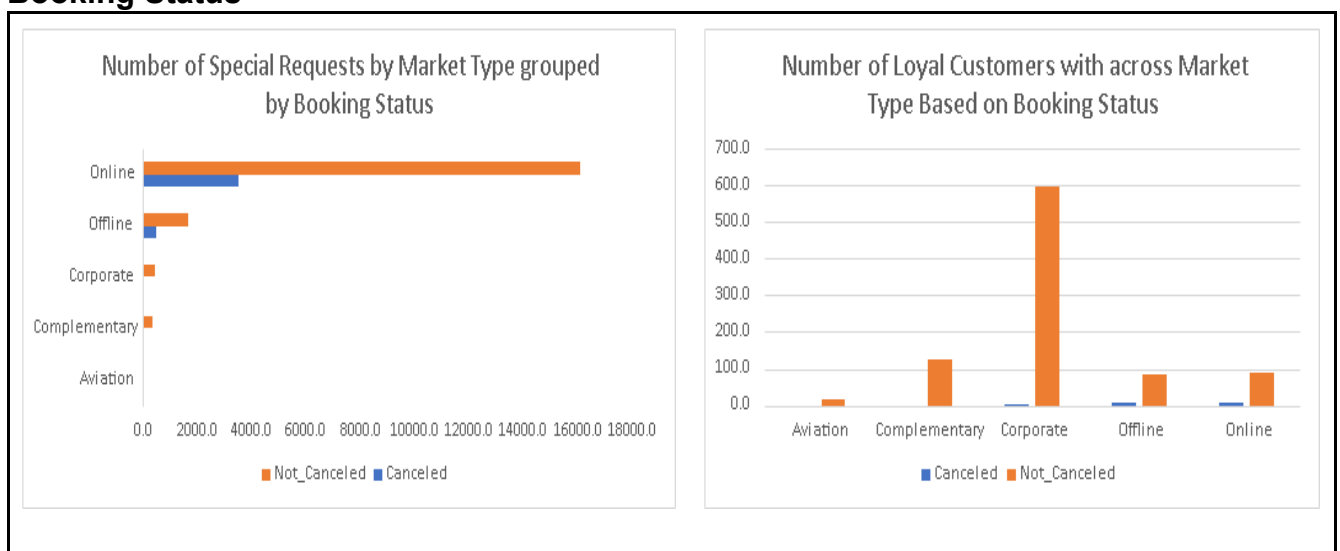
The premium pricing observed in the "Aviation" market in select years, the consistently lower prices in the "Complementary" market, and the generally moderate prices in the "Corporate" segment were observed. Both "Offline" and "Online" markets exhibit competitive pricing. Most customers choose Online as their market type for booking reservations. The choice of market type could be potentially attributed to factors like convenience, accessibility, and online-exclusive offers.

Figure 7: Average Price by Market Type



Most of the customers prefer Online method to making special requests and cancelling orders. In terms of loyalty, customers within the "Corporate" bookings segment show a substantial number of repeat guests compared to the other market types.

Figure 8: Number of Special Guests and Loyal Customers by Market Type and Booking Status



4. Recommendations

1. Implementing dynamic pricing strategies. For example, adjusting room rates based on demand fluctuations, seasonal patterns, and market trends.
2. Create marketing campaigns targeting peak and off-peak periods to attract customers during both high and low-demand periods.
3. Directed more focus on enhancing the special request experience, especially for online bookings, to provide personalized and convenient services.
4. Leverage the higher number of repeating guests within the "Corporate" segment to develop loyalty programs and exclusive offers for these customers.

5. Conclusion

The fluctuations in average room reservation prices are caused by various factors. These prices exhibit seasonal characteristics as shown by yearly, quarterly, and monthly trends. The seasonal patterns are caused by various factors which include social, economic, cultural, environmental and health factors. Other reservation characteristics and booking context also are determinants of price increase or decrease.



Part 2: School Free Lunch Eligibility

Introduction

Every decision and action made by educational institutions reverberates through the lives of students. To ensure that the correct decisions are made, schools should conduct thorough analyses of each and every situation at hand. To aid that, this report explores and elucidates the dataset collected by ABC school. The objective here is to illuminate the intricate connections between lunch eligibility, gender, and the very bedrock of academic performance (pre-test and post-test).

Amidst this exploration, a wealth of literature underscores the role of children's nutrition in shaping academic achievements (Gordanier et al., 2020). Stretching back to the inception of the National School Lunch Act in 1946, a succession of federal programs have been crafted to enhance childhood nutrition through school-based interventions (Pokorney et al., 2019). The Community Eligibility Provision (CEP) program, a more recent endeavour, emerged from the Healthy and Hunger-Free Kids Act of 2010, with the aim of widening access to school lunches. Under this program, schools receive subsidies to offer universal free lunch, irrespective of individual student eligibility for free or reduced lunch within the NSLP (Handbury & Moshary, 2021).

In this report, we classify students into eligible and ineligible categories for free lunch and how it might be associated to students demographics, pre-test and post-test scores, and lunch eligibility criteria. The significance of this analysis is the relief it extends to schools by alleviating administrative burdens (Tan et al., 2020). Furthermore, understanding the impact of lunch eligibility on academic performance can guide decisions about resource allocation.

Objective: The objective of this analysis is to explore the relationships that exist between lunch eligibility, gender, and academic performance among students.

Research Questions and Hypotheses

To achieve this objective, the analysis will address the following research questions:

1. Research Question 1: Does gender play a role in lunch eligibility?

Hypothesis 1: The lunch eligibility may vary based on students' gender.

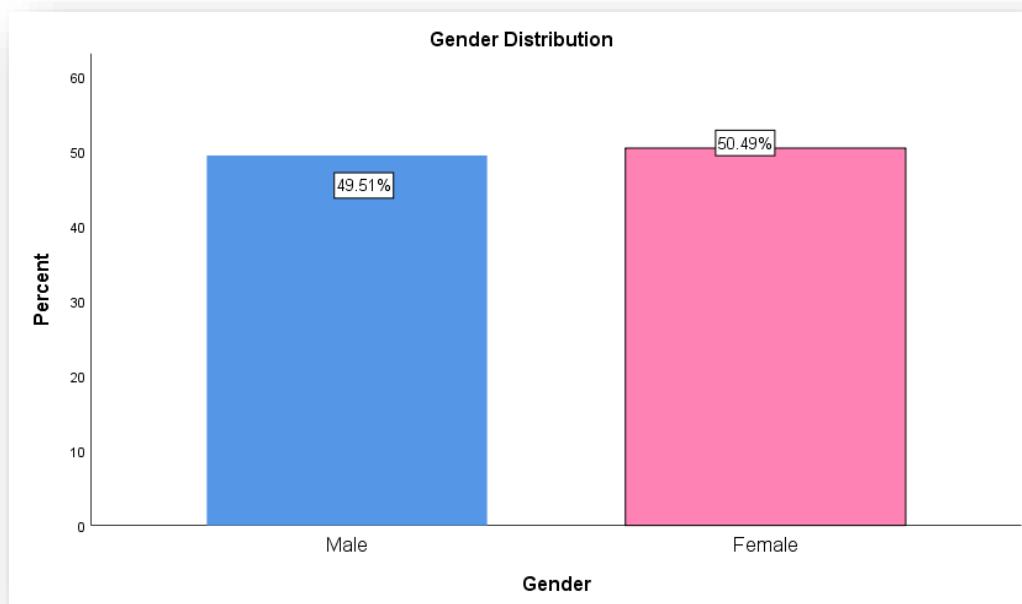
2. Research Question 2: Is there a significant relationship between lunch eligibility (qualifying for free/reduced lunch or not) and academic performance (pre-test and post-test scores) among students?

Hypothesis 2: Students who qualify for free/reduced lunch will exhibit different academic performance levels compared to those who do not qualify.

Results and Discussion

1. *Gender Distribution*

Among the 2133 students, the nearly equal representation of male (49.5%) and female (50.5%) students. The bar chart below visually represents this balanced gender distribution:

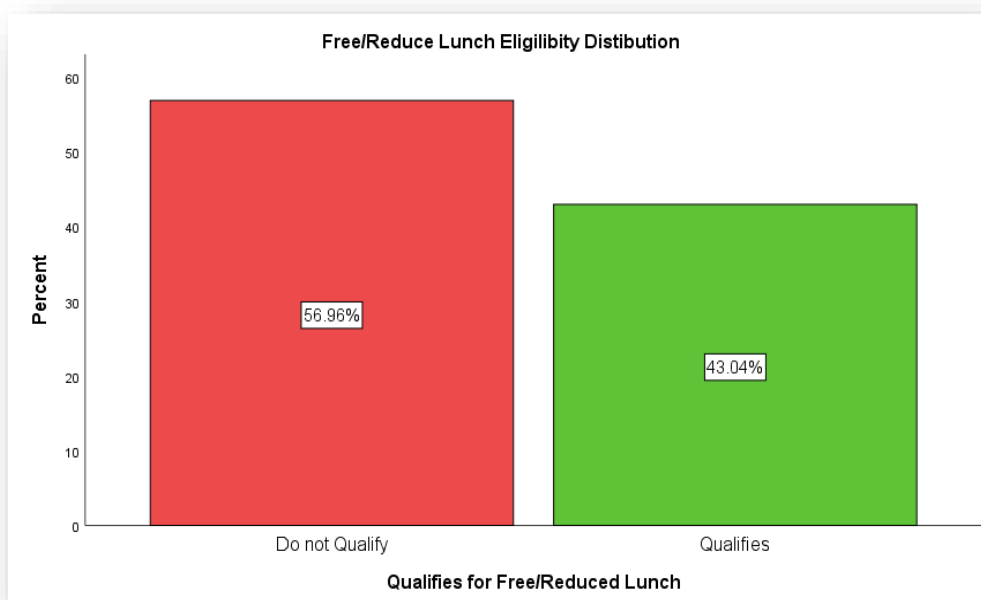


This reaffirms the School's commitment to fostering a balanced educational environment, addressing gender diversity, and providing equal educational opportunities.

2. Lunch Eligibility

The distribution of lunch eligibility unveiled that 57.0% of the students did not qualify for reduced/free lunch, while 43.0% qualified for it. Although smaller compared to its counterpart, the proportion of students who qualify for reduced/free lunch is still substantial in ensuring that the nutritional needs of these students are met.

The proportion of students (57.0%) who do not qualify for reduced/free lunch may come from families hovering just above the eligibility threshold. While this group is not eligible for the lunch program they could still face challenges related to food security. The bar chart below illustrates the distribution of lunch eligibility;



This distribution of lunch-eligible students varies slightly between genders. Specifically, among the 918 who qualify for reduced/free lunch, 463 are male students (43.84%) and 463 female students (43.09%). This shows that gender does not appear to play a significant role in lunch eligibility.

Table 1:

Qualification for free/reduced lunch by Gender				
Count		Lunch		Total
		Not Eligible	Eligible	
Gender	Male	593	463	1056
	Female	622	455	1077
Total		1215	918	2133

3. Descriptive Statistics

The mean and median pre-test scores (54.96, 56) and post-test scores (67.10, 68) indicate an average improvement in academic performance across the student population. Comparing pre-test and post-test scores between lunch-eligible and non-lunch-eligible students, lunch-eligible students initially exhibit lower mean and median pre-test scores (45.25 and 44) compared to their non-lunch-eligible peers (62.29 and 62). However, after participating in the educational program, lunch-eligible students demonstrated improvement, with mean and median post-test scores of 57.48 and 55, respectively. These underscore the positive impact of the free lunch program on students' learning outcomes.

Table 2: Descriptive Statistics of Pre-test and Post-test Scores

Qualification for free/reduced	Pre-test	Pre-test	Post-test	Pre-test / Post-test Difference
--------------------------------	----------	----------	-----------	---------------------------------

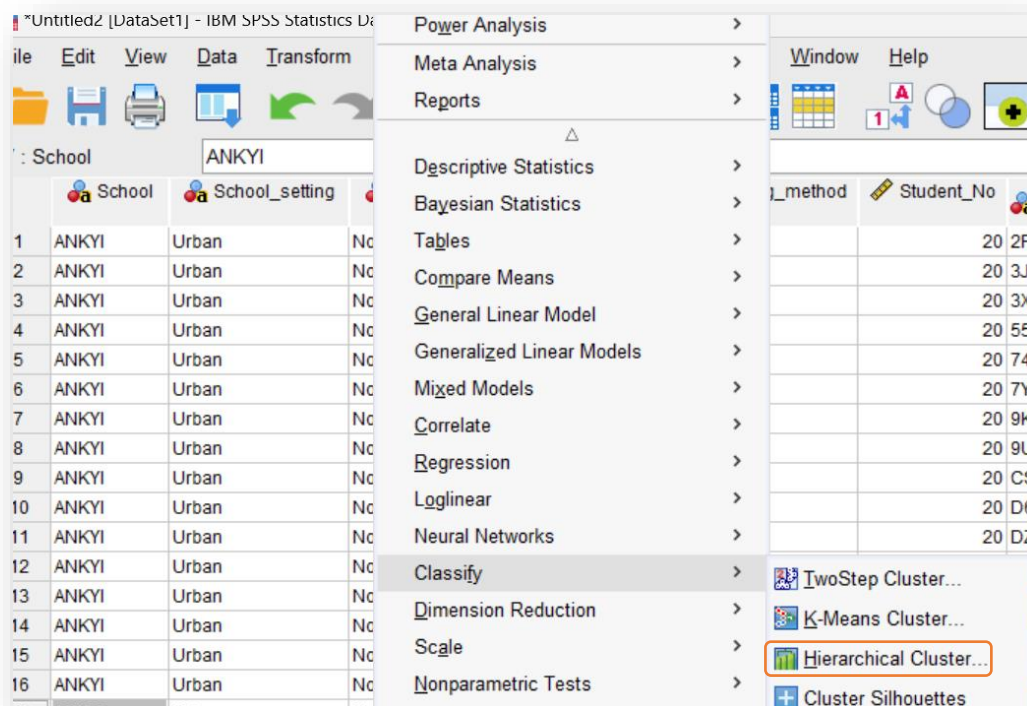
lunch				
Not Eligible	Mean	62.29	74.38	12.09
	Median	62	74	12
Eligible	Mean	45.25	57.48	12.23
	Median	44	55	11

4. Cluster Analysis

To determine the appropriate number of clusters, the hierarchical clustering method is used.

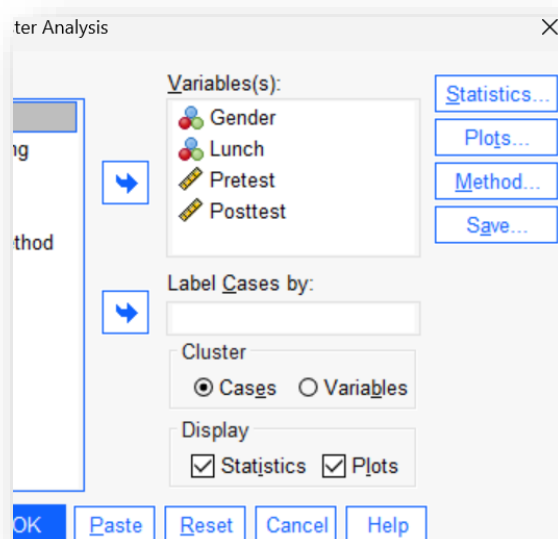
Step 1: Select the clustering method

From the "Analyze" menu, select "Classify" and then "Hierarchical Cluster..."



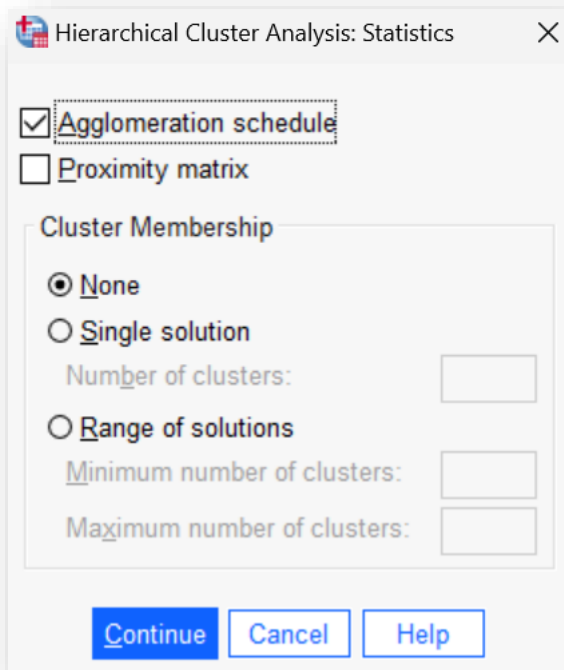
Step 2: In the "Hierarchical Cluster" dialog box, select the variables you want to base your clusters on (e.g., math, reading, and writing test scores).

Leave the tick marks for "Cases" in their default settings, as we want to cluster cases.



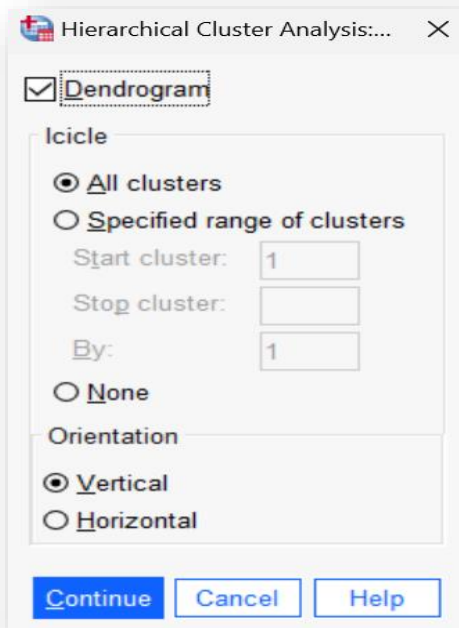
Step 3: Click on the "Statistics" button.

Leave the default settings.

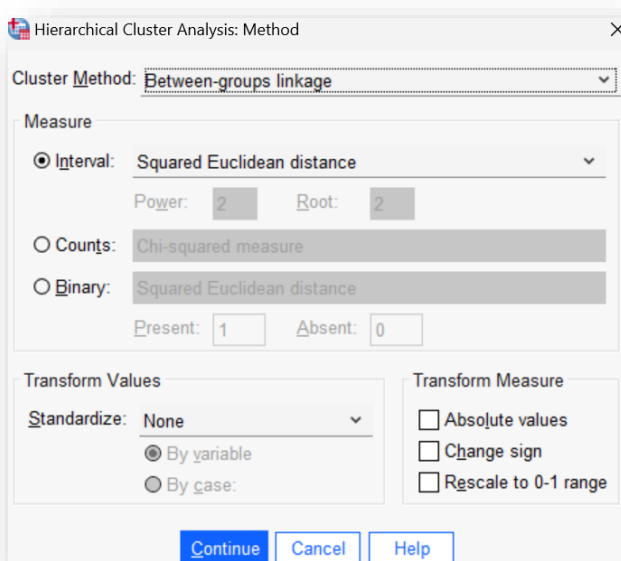


Step 4: Click on the "Plots" button.

Select "Dendrogram." The dendrogram graphically shows how clusters are merged and helps determine the appropriate number of clusters.

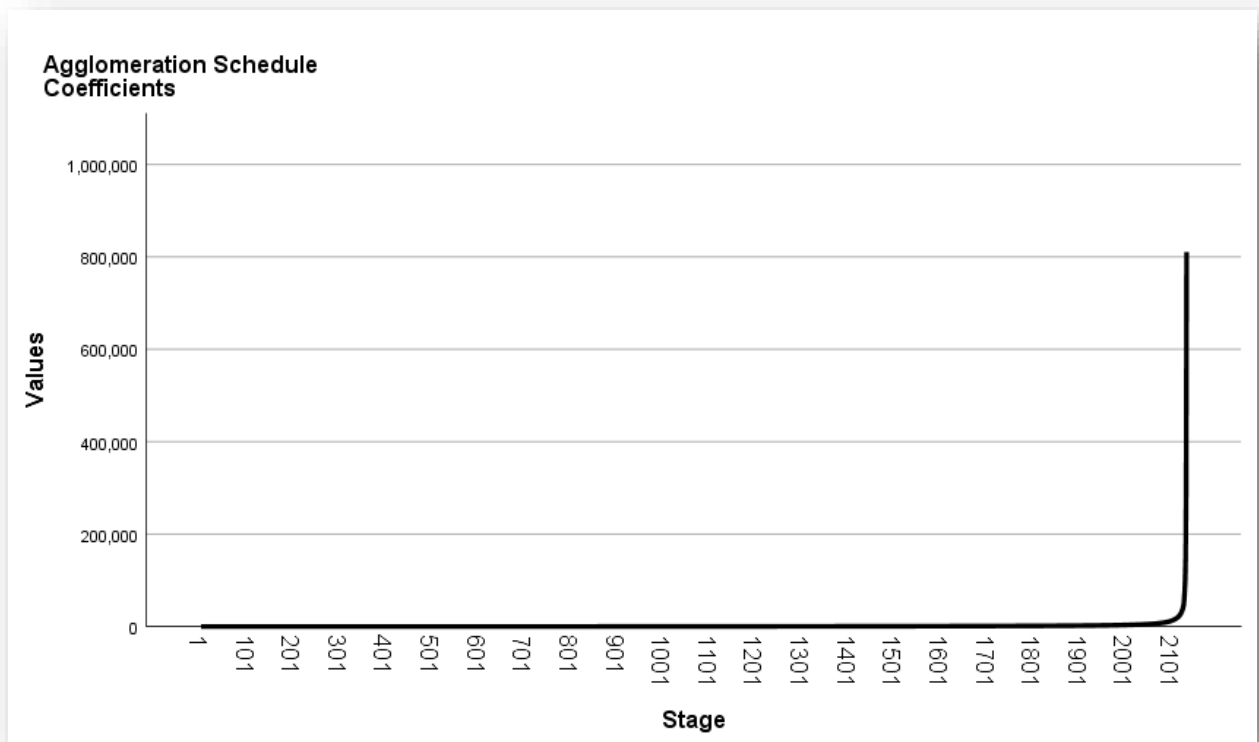


Step 5: Click on the "Method" button then choose "Interval" and select "Squared Euclidean Distance" as the distance measure and Ward method as the cluster method. Ward's method employs the F value, similar to that used in ANOVA, to enhance the significance of dissimilarities between clusters (Statistics Solutions, 2021).



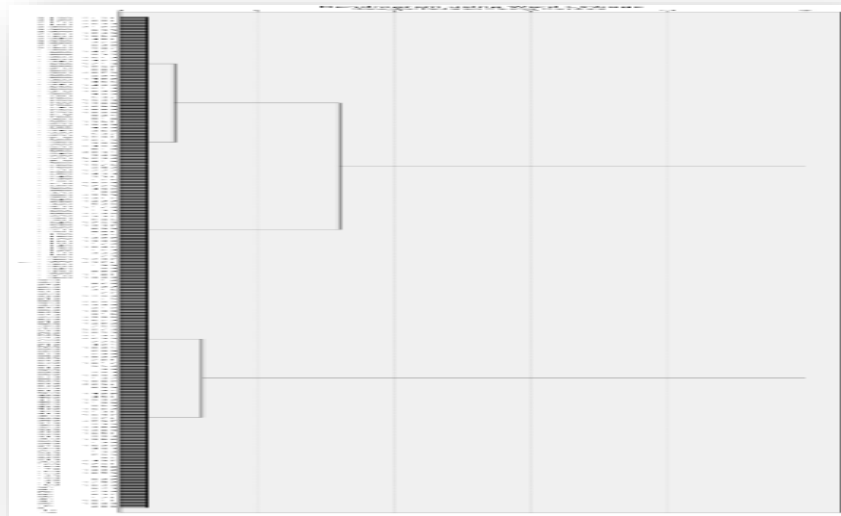
Step 6: Perform Analysis and Interpret Results

Click the "OK" button in the "Method" dialog to perform the analysis. Use the Ward linkage coefficients table to create an elbow chart, or observe the dendrogram to determine the clusters.



It appears from the elbow plot and dendrogram that there are 2 clusters.

Dendrogram



7. Repeat the whole process with k-means clustering, with 2 clusters

Five Most Common Data and Text Mining Methods

1. **Classification:** Classification involves categorizing data into predefined classes based on features. For instance, a bank might use classification to predict whether a loan applicant will default or not based on attributes like credit score, income, and employment status.
2. **Fraud Detection:** Anomaly detection is used in financial services to identify fraudulent transactions. Unusual spending patterns, large withdrawals, or transactions in unfamiliar locations can be flagged as anomalies, helping prevent fraud. For example, a credit card company can monitor customer transactions and detect unusual patterns such as multiple high-value purchases made in quick succession from different locations, triggering alerts for potential fraud.
3. **Clustering:** Clustering groups similar data points together based on similarity metrics. A retail chain can use clustering to segment customers into different

groups for targeted marketing strategies based on purchase history and demographics.

4. **Time Series Analysis:** Time series analysis deals with data points collected over time to identify trends, seasonality, and patterns. For instance, a utility company can use time series analysis to forecast electricity demand and plan energy production accordingly.
5. **Regression:** Regression analyses relationships between variables to predict an outcome. Regression can be linear regression or logistics regression. Regression can be used to predict sales based on the determining factors such as price, advertising spend, and time of year. For example, a car manufacturer, for instance, might use historical sales data, car features and model, and advertising spending to forecast future car sales.
6. **Sentiment Analysis:** Sentiment analysis extracts opinions and emotions from text, often used to gauge customer satisfaction and public perception. For instance, social media platforms use sentiment analysis to monitor user sentiments towards products and services.

Advantages and Disadvantages SPSS and Excel for Data Analysis

Criteria	SPSS	Excel
Advantages		
Data Handling	Designed for large datasets and complex analyses.	Suitable for small to medium-sized datasets.
Statistical Analysis	Provides a wide range of advanced statistical tests.	Offers basic statistical functions and analysis.
Data Visualization	Offers advanced graphing and visualization options.	Limited graphing features, primarily basic charts.
Ease of Use	Requires some learning but user-friendly interface.	Familiar interface, easy for basic tasks.
Disadvantages		
Cost	Requires a license, can be expensive for some users.	Generally comes bundled with Microsoft Office.
Learning Curve	Learning curve for advanced features and analyses.	Limited learning curve, suited for basic users.
Scalability	May slow down with extremely large datasets.	Can struggle with very large datasets.
Data Cleaning	Lacks Excel's flexibility for data cleaning tasks.	Requires manual steps for complex data cleaning.
Complex Calculations	Better equipped for complex statistical calculations.	Limited for complex statistical analyses.

References

PART 1

Alrawabdeh, W. (2021). Seasonal balancing of revenue and demand in hotel industry: the case of Dubai City. *Journal of Revenue and Pricing Management*. <https://doi.org/10.1057/s41272-021-00290-6>

Lardieri, A. (2021, January 27). Coronavirus Pandemic Sets Hotel Industry Back 10 Years, Report Finds. *US News & World Report*; U.S. News & World Report. <https://www.usnews.com/news/national-news/articles/2021-01-27/coronavirus-pandemic-sets-hotel-industry-back-10-years-report-finds>

AHLA'S STATE OF THE HOTEL INDUSTRY 2021 AHLA'S STATE OF THE HOTEL INDUSTRY 2021. (2021). https://www.ahla.com/sites/default/files/2021_state_of_the_industry_0.pdf

Economics, O. (2016). Economic impact of the US hotel industry. *American Hotel and*.

Gibbs, C., Guttentag, D., Gretzel, U., Yao, L., & Morton, J. (2018). Use of dynamic pricing strategies by Airbnb hosts. *International Journal of Contemporary Hospitality Management*, 30(1), 2-20.

PART 2

Pokorney, P. E., Chandran, A., & Long, M. W. (2019). Impact of the Community Eligibility Provision on meal counts and participation in Pennsylvania and Maryland National School Lunch Programs. *Public Health Nutrition*, 22(17), 3281–3287. <https://doi.org/10.1017/s1368980019002246>

Gordanier, J., Ozturk, O., Williams, B., & Zhan, C. (2020). Free lunch for all! The effect of the community eligibility provision on academic outcomes. *Economics of Education Review*, 77, 101999. <https://doi.org/10.1016/j.econedurev.2020.101999>

Handbury, J., & Moshary, S. (2021). *School food policy Affects everyone: Retail responses to the National School Lunch Program*. <https://doi.org/10.3386/w29384>

Tan, M. L., Laraia, B. A., Madsen, K. A., Johnson, R. C., & Ritchie, L. D. (2020). Community Eligibility Provision and School Meal Participation among Student Subgroups. *Journal of School Health*, 90(10), 802–811. <https://doi.org/10.1111/josh.12942>

Statistics Solutions. (2021, July 16). *Conduct and interpret a cluster analysis - Statistics solutions*. <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/cluster-analysis/>