Data Handling and Business Intelligence

Student name: Mihaela Raducanu

Student ID: 2111701

BA (Hons) Business Management

Task 1

Spreadsheet software like Microsoft Excel is extensively used and has developed into a useful tool for data analysis and decision-making in a variety of commercial situations. It has an intuitive user interface, a vast array of functions, and the capacity to manage both small and big datasets. In this post, we'll look at how to utilise Microsoft Excel and how it may help with data analysis and decision-making (Sharda, et al., 2014).

Microsoft Excel's capacity to efficiently organise and analyse data is one of its main advantages. It is useful for storing data and doing various analysis since users may enter, modify, and process data in tabular form using rows and columns. Users may carry out mathematical, statistical, and logical operations on data using a variety of built-in formulae and functions supported by Excel. Excel gives users the tools they need to effectively analyse and summarise data, from simple mathematical operations to sophisticated statistical analysis. Bar charts, line charts, pie charts, scatter plots, and other chart styles that Excel provides make it simple for users to visualise data and spot patterns or trends.

Excel has capabilities for data validation to guarantee consistency and correctness of data. This capability reduces the possibility of inaccurate analysis by putting rules and restrictions on data entry. With the use of these tools, users may run what-if analyses by altering the input variables and evaluating the effects on the results. Goal Seek assists in identifying the essential input values required to attain particular desired results, while Scenario Manager enables the construction of several scenarios for comparison.

In order to make decisions, businesses frequently analyse data in order to spot trends, patterns, and prospective opportunities or hazards. Multiple techniques exist for using Microsoft Excel to help decision-making processes. By enabling users to gather and analyse data from numerous sources, including databases, spreadsheets, and external data connections, Excel supports business intelligence. These insights may be used by decision-makers to pinpoint problem areas, streamline processes, and create long-term strategies (Whitten, et al., 2017). Excel is a popular option for financial analysis and budgeting due to its financial operations and modelling features. Using Excel's built-in capabilities, users may develop financial models, carry out financial forecasts, and assess investment possibilities.

Excel's sensitivity analysis enables decision-makers to look at how altering factors affect results. Decision-makers can better comprehend risks and uncertainties related to particular decisions by identifying crucial elements. Making use of charts, graphs, and dashboards to effectively visualise data in Excel enables decision-makers to quickly comprehend complex data (Connolly & Begg, 2014). Visual representation supports informed decision-making and helps communicate findings. Excel is a helpful tool for decision support systems because it can handle hypothetical situations and create reports using PivotTables and graphics. Based on a study of the data, these technologies help choose the optimal course of action.

A flexible and strong tool, Microsoft Excel offers crucial capabilities for data analysis, corporate information, and decision-making. It is accessible to users with varied degrees of technical experience thanks to its user-friendly interface, data processing capabilities, and data visualisation features. Excel continues to play a significant role in assisting decision-making processes and aiding data analysis in contemporary commercial contexts. While Excel is a

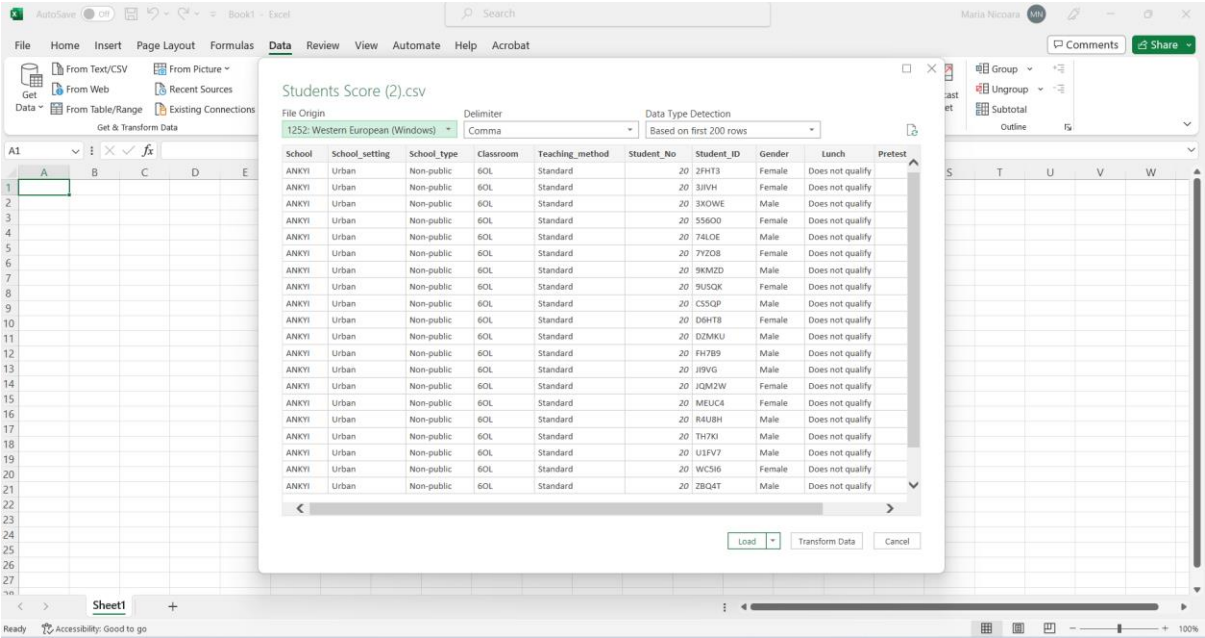
useful tool for many jobs, it's crucial to understand that more sophisticated analytics and complicated data processing may call for specialised tools and computer languages.

Evaluating the use of Microsoft Excel for pre-processing, analysing, and visualizing data is an important task to ensure that the tool is suitable and effective for the specific data analysis needs.

1. Pre-processing the data using Microsoft Excel:

Data Import: Excel provides various methods to import data from different sources, such as CSV files, databases, and web queries. Evaluate how easy it is to import the data and if it handles different data formats effectively.

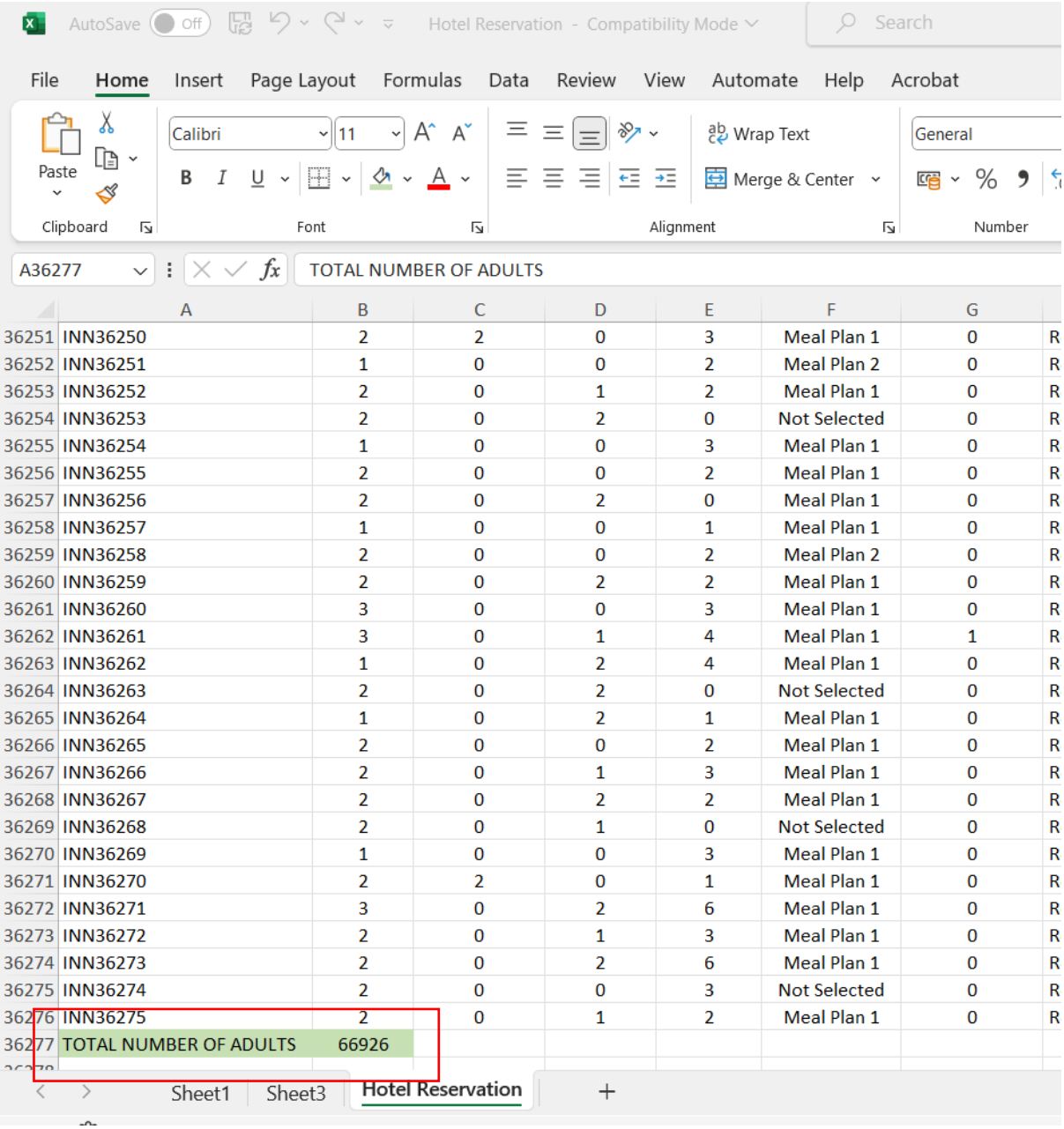
For example, you can easily import data from a CVS file.



Data Cleaning: Excel offers tools for basic data cleaning tasks like removing duplicates, handling missing values, and correcting data inconsistencies. Assess the efficiency of these data cleaning functionalities and if they are sufficient for the dataset at hand.

Data Transformation: Evaluate Excel's has an ability to transform data, such as splitting columns, merging data, and applying formulas for creating new variables.

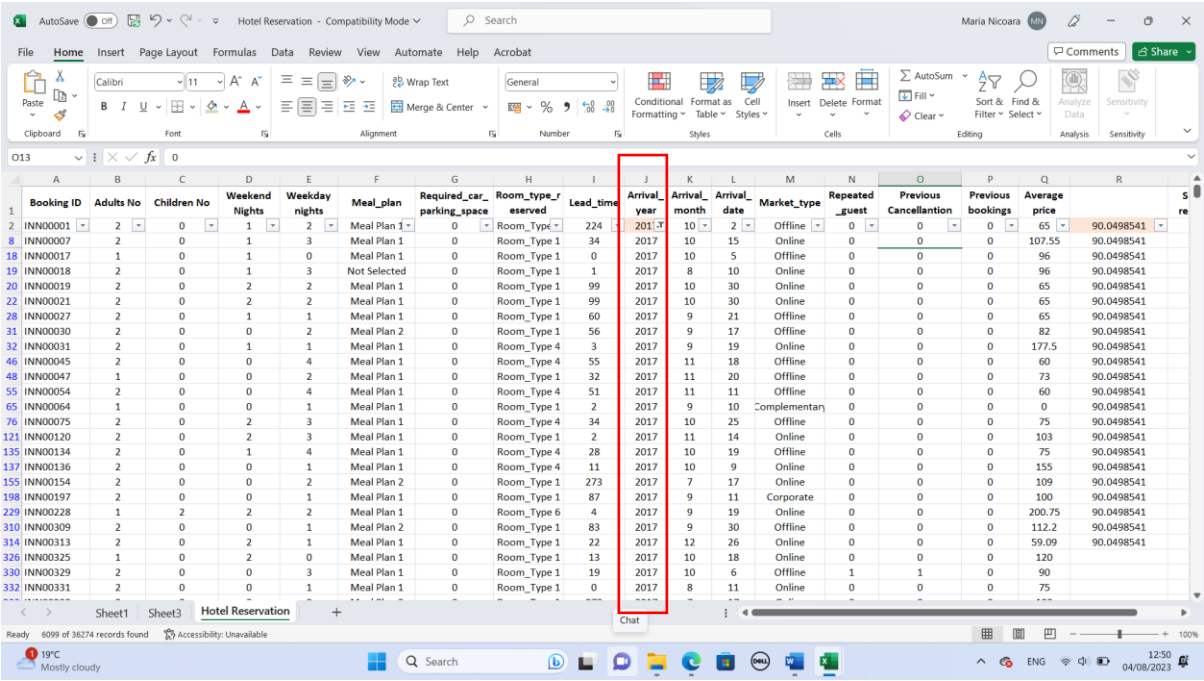
For example, we can easily find out how many adults have been in the hotel since opening, only by using the simple formula “SUM”.



	A	B	C	D	E	F	G	
36251	INN36250	2	2	0	3	Meal Plan 1	0	R
36252	INN36251	1	0	0	2	Meal Plan 2	0	R
36253	INN36252	2	0	1	2	Meal Plan 1	0	R
36254	INN36253	2	0	2	0	Not Selected	0	R
36255	INN36254	1	0	0	3	Meal Plan 1	0	R
36256	INN36255	2	0	0	2	Meal Plan 1	0	R
36257	INN36256	2	0	2	0	Meal Plan 1	0	R
36258	INN36257	1	0	0	1	Meal Plan 1	0	R
36259	INN36258	2	0	0	2	Meal Plan 2	0	R
36260	INN36259	2	0	2	2	Meal Plan 1	0	R
36261	INN36260	3	0	0	3	Meal Plan 1	0	R
36262	INN36261	3	0	1	4	Meal Plan 1	1	R
36263	INN36262	1	0	2	4	Meal Plan 1	0	R
36264	INN36263	2	0	2	0	Not Selected	0	R
36265	INN36264	1	0	2	1	Meal Plan 1	0	R
36266	INN36265	2	0	0	2	Meal Plan 1	0	R
36267	INN36266	2	0	1	3	Meal Plan 1	0	R
36268	INN36267	2	0	2	2	Meal Plan 1	0	R
36269	INN36268	2	0	1	0	Not Selected	0	R
36270	INN36269	1	0	0	3	Meal Plan 1	0	R
36271	INN36270	2	2	0	1	Meal Plan 1	0	R
36272	INN36271	3	0	2	6	Meal Plan 1	0	R
36273	INN36272	2	0	1	3	Meal Plan 1	0	R
36274	INN36273	2	0	2	6	Meal Plan 1	0	R
36275	INN36274	2	0	0	3	Not Selected	0	R
36276	INN36275	2	0	1	2	Meal Plan 1	0	R
36277	TOTAL NUMBER OF ADULTS	66926						

Data Filtering and Sorting: Excel allows users to filter and sort data based on specific criteria.

For example, the data can be filtered easily by the arrival year. In this way only the bookings from 2017 will be displayed.



Booking ID	Adults No	Children No	Weekend Nights	Weekday nights	Meal_plan	Required_car_parking_space	Room_type_reserved	Lead_time	Arrival_year	Arrival_month	Arrival_date	Market_type	Repeated_guest	Previous Cancellantion	Previous bookings	Average price	
INN00001	2	0	1	2	Meal Plan 1	0	Room_Type	224	201	10	2	Offline	0	0	0	65	90.0498541
INN00007	2	0	1	3	Meal Plan 1	0	Room_Type 1	34	2017	10	15	Online	0	0	0	107.55	90.0498541
INN00017	1	0	1	0	Meal Plan 1	0	Room_Type 1	0	2017	10	5	Offline	0	0	0	96	90.0498541
INN00018	2	0	1	3	Not Selected	0	Room_Type 1	1	2017	8	10	Online	0	0	0	96	90.0498541
INN00019	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	30	Online	0	0	0	65	90.0498541
INN00021	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	30	Online	0	0	0	65	90.0498541
INN00027	2	0	1	1	Meal Plan 1	0	Room_Type 1	60	2017	9	21	Offline	0	0	0	65	90.0498541
INN00030	2	0	0	2	Meal Plan 2	0	Room_Type 1	56	2017	9	17	Offline	0	0	0	82	90.0498541
INN00031	2	0	1	1	Meal Plan 1	0	Room_Type 4	3	2017	9	19	Online	0	0	0	177.5	90.0498541
INN00045	2	0	0	4	Meal Plan 1	0	Room_Type 4	55	2017	11	18	Offline	0	0	0	60	90.0498541
INN00047	1	0	0	2	Meal Plan 1	0	Room_Type 1	32	2017	11	20	Offline	0	0	0	73	90.0498541
INN00054	2	0	0	4	Meal Plan 1	0	Room_Type 4	51	2017	11	11	Offline	0	0	0	60	90.0498541
INN00064	1	0	0	1	Meal Plan 1	0	Room_Type 1	2	2017	9	10	Complementar	0	0	0	0	90.0498541
INN00075	2	0	2	3	Meal Plan 1	0	Room_Type 4	34	2017	10	25	Offline	0	0	0	75	90.0498541
INN00120	2	0	2	3	Meal Plan 1	0	Room_Type 1	2	2017	11	14	Online	0	0	0	103	90.0498541
INN00134	2	0	1	4	Meal Plan 1	0	Room_Type 4	28	2017	10	19	Offline	0	0	0	75	90.0498541
INN00136	2	0	0	1	Meal Plan 1	0	Room_Type 4	11	2017	10	9	Online	0	0	0	155	90.0498541
INN00154	2	0	0	2	Meal Plan 2	0	Room_Type 1	273	2017	7	17	Online	0	0	0	109	90.0498541
INN00197	2	0	0	1	Meal Plan 1	0	Room_Type 1	87	2017	9	11	Corporate	0	0	0	100	90.0498541
INN00228	1	2	2	2	Meal Plan 1	0	Room_Type 6	4	2017	9	19	Online	0	0	0	200.75	90.0498541
INN00309	2	0	0	1	Meal Plan 2	0	Room_Type 1	83	2017	9	30	Offline	0	0	0	112.2	90.0498541
INN00313	2	0	2	1	Meal Plan 1	0	Room_Type 1	22	2017	12	26	Online	0	0	0	59.09	90.0498541
INN00325	1	0	2	0	Meal Plan 1	0	Room_Type 1	13	2017	10	18	Online	0	0	0	120	
INN00329	2	0	0	3	Meal Plan 1	0	Room_Type 1	19	2017	10	6	Offline	1	1	0	90	
INN00331	2	0	0	1	Meal Plan 1	0	Room_Type 1	0	2017	8	11	Online	0	0	0	75	

2. Utilising Excel to analyse the data

Descriptive Statistics: Excel offers the fundamental descriptive statistics, such as mean, median, standard deviation, etc.

PivotTables and PivotCharts: PivotTables and PivotCharts in Excel are useful for summarising and displaying big datasets.

Data Analysis ToolPak: The Data Analysis ToolPak in Excel provides further data analysis capabilities. Inquire as to whether these resources are adequate for more complex studies like regression, correlation, and t-tests.

3. Microsoft Excel data visualisation:

Chart kinds: Consider the usefulness of the various chart kinds (such as bar charts, line charts, and scatter plots) for portraying various types of data.

Adaptability: the user can evaluate the degree of chart adaptability, including labelling, axis scaling, and colour possibilities.

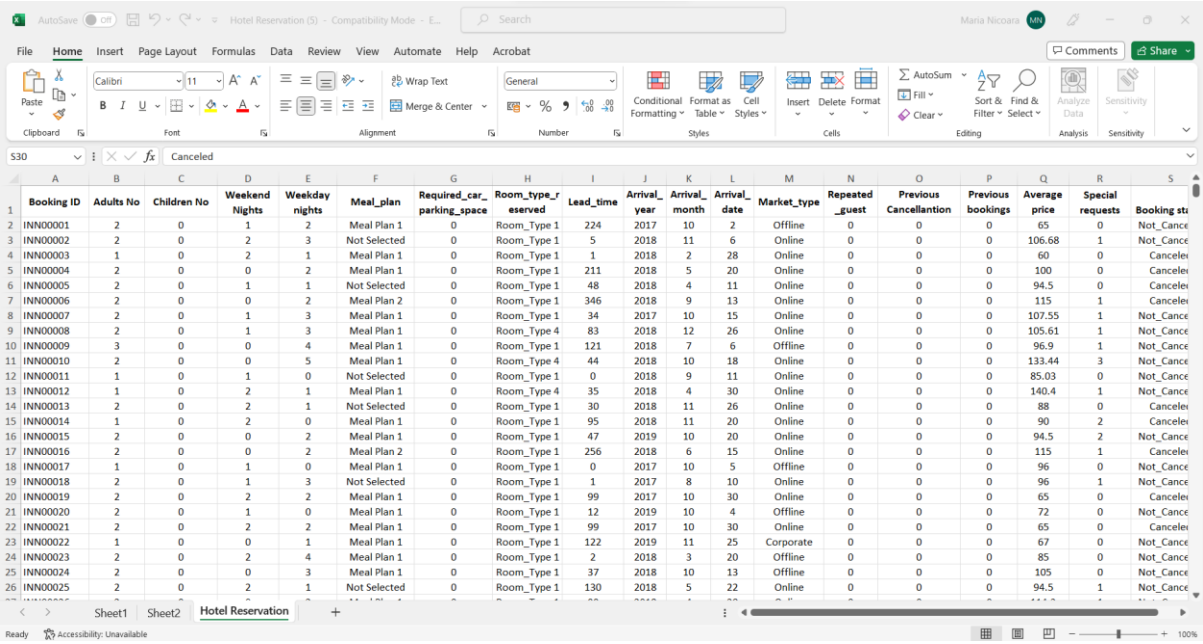
Interactive Features: the user can find out whether Excel has interactive visualisation features like data filtering or drill-down options.

Readability and Aesthetics: the user can assess the readability and aesthetic appeal of the Excel visualisations.

Excel is a strong tool for basic data analysis and visualisation, but it may be difficult to handle big datasets, sophisticated statistical methods, and complex data transformations. Data analysts may think about utilising specialised tools like Python, R, or specific data visualisation software when Excel falls short for more complex analysis and visualisations. The tools used

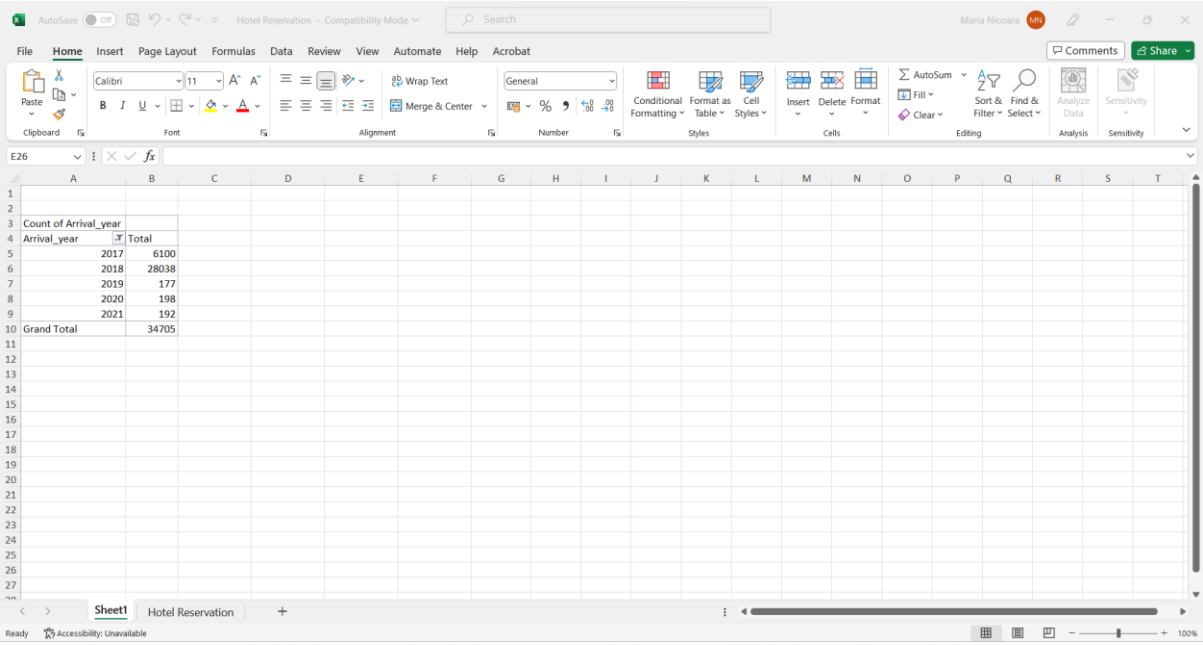
should always be in line with the project specifications and the degree of complexity required to extract valuable insights from the data.

In the data analysis, we are going to use the data from the “Hotel Reservation.xls” file .

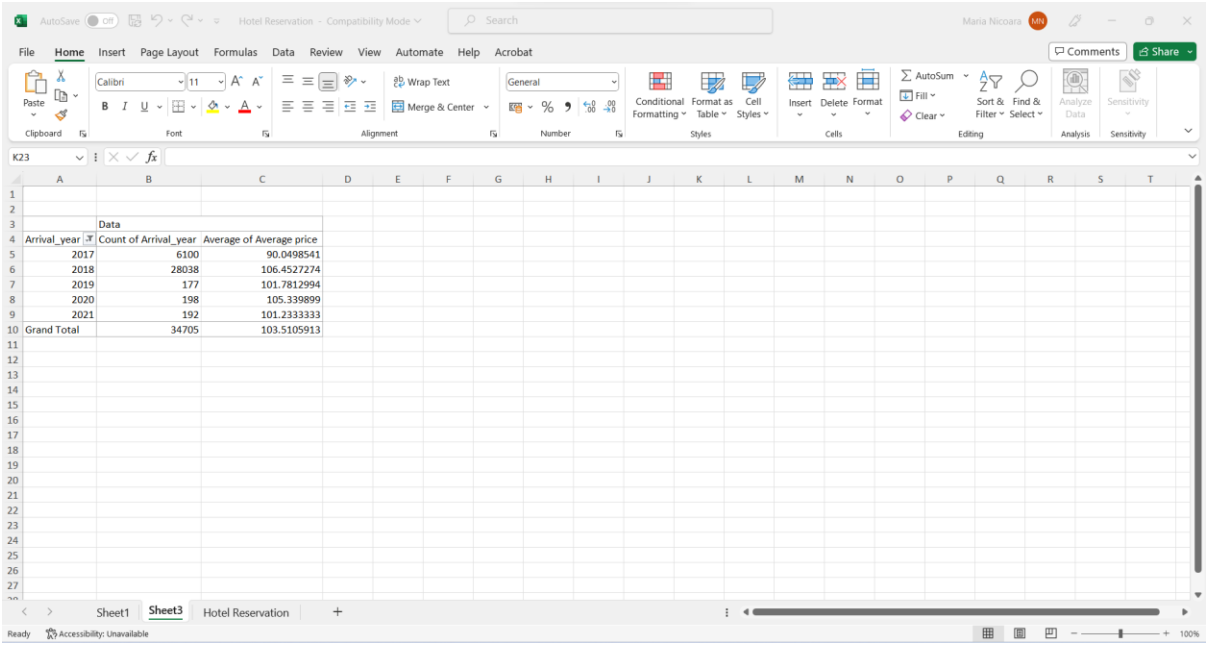


Booking ID	Adults No	Children No	Weekend Nights	Weekday nights	Meal_plan	Required_car_parking_space	Room_type_reserved	Lead_time	Arrival_year	Arrival_month	Arrival_date	Market_type	Repeated_guest	Previous Cancellantion	Previous bookings	Average price	Special requests	Booking st
INN00001	2	0	1	2	Meal Plan 1	0	Room_Type 1	224	2017	10	2	Offline	0	0	0	65	0	Not_Cance
INN00002	2	0	2	3	Not Selected	0	Room_Type 1	5	2018	11	6	Online	0	0	0	106.68	1	Not_Cance
INN00003	1	0	2	1	Meal Plan 1	0	Room_Type 1	1	2018	2	28	Online	0	0	0	60	0	Cancele
INN00004	2	0	0	2	Meal Plan 1	0	Room_Type 1	211	2018	5	20	Online	0	0	0	100	0	Cancele
INN00005	2	0	1	1	Not Selected	0	Room_Type 1	48	2018	4	11	Online	0	0	0	94.5	0	Cancele
INN00006	2	0	0	2	Meal Plan 2	0	Room_Type 1	346	2018	9	13	Online	0	0	0	115	1	Cancele
INN00007	2	0	1	3	Meal Plan 1	0	Room_Type 1	34	2017	10	15	Online	0	0	0	107.55	1	Not_Cance
INN00008	2	0	1	3	Meal Plan 1	0	Room_Type 4	83	2018	12	26	Online	0	0	0	105.61	1	Not_Cance
INN00009	3	0	0	4	Meal Plan 1	0	Room_Type 1	121	2018	7	6	Offline	0	0	0	96.9	1	Not_Cance
INN00010	2	0	0	5	Meal Plan 1	0	Room_Type 4	44	2018	10	18	Online	0	0	0	133.44	3	Not_Cance
INN00011	1	0	1	0	Not Selected	0	Room_Type 1	0	2018	9	11	Online	0	0	0	85.03	0	Not_Cance
INN00012	1	0	2	1	Meal Plan 1	0	Room_Type 4	35	2018	4	30	Online	0	0	0	140.4	1	Not_Cance
INN00013	2	0	2	1	Not Selected	0	Room_Type 1	30	2018	11	26	Online	0	0	0	88	0	Cancele
INN00014	1	0	2	0	Meal Plan 1	0	Room_Type 1	95	2018	11	20	Online	0	0	0	90	2	Cancele
INN00015	2	0	0	2	Meal Plan 1	0	Room_Type 1	47	2019	10	20	Online	0	0	0	94.5	2	Not_Cance
INN00016	2	0	0	2	Meal Plan 2	0	Room_Type 1	256	2018	6	15	Online	0	0	0	115	1	Cancele
INN00017	1	0	1	0	Meal Plan 1	0	Room_Type 1	0	2017	10	5	Offline	0	0	0	96	0	Not_Cance
INN00018	2	0	1	3	Not Selected	0	Room_Type 1	1	2017	8	10	Online	0	0	0	96	1	Not_Cance
INN00019	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	30	Online	0	0	0	65	0	Cancele
INN00020	2	0	1	0	Meal Plan 1	0	Room_Type 1	12	2019	10	4	Offline	0	0	0	72	0	Not_Cance
INN00021	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	30	Online	0	0	0	65	0	Cancele
INN00022	1	0	0	1	Meal Plan 1	0	Room_Type 1	122	2019	11	25	Corporate	0	0	0	67	0	Not_Cance
INN00023	2	0	2	4	Meal Plan 1	0	Room_Type 1	2	2018	3	20	Offline	0	0	0	85	0	Not_Cance
INN00024	2	0	0	3	Meal Plan 1	0	Room_Type 1	37	2018	10	13	Offline	0	0	0	105	0	Not_Cance
INN00025	2	0	2	1	Not Selected	0	Room_Type 1	130	2018	5	22	Online	0	0	0	94.5	1	Not_Cance

The Pivot Tables are used to easily analyse large amounts of data.

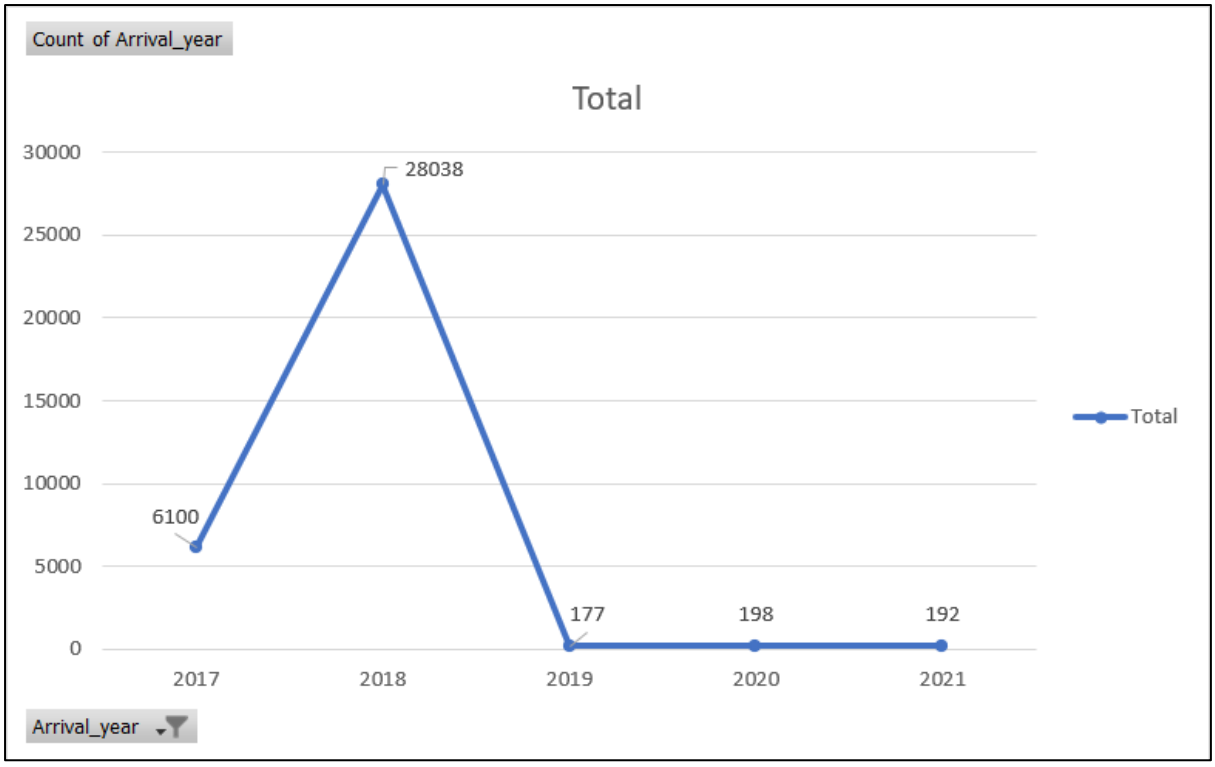


Count of Arrival_year	
Arrival_year	Total
2017	6100
2018	28038
2019	177
2020	198
2021	192
Grand Total	34705

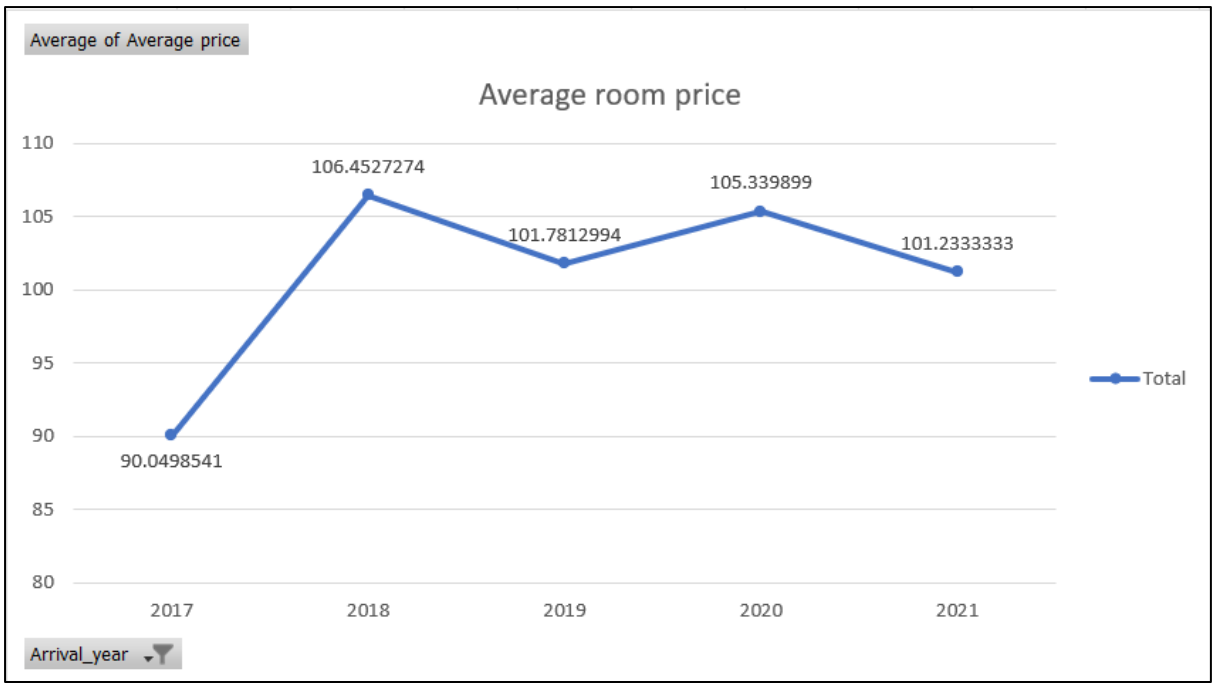


	Data	
Arrival_year	Count of Arrival_year	Average of Average price
2017	6100	90.0498541
2018	28038	106.4527274
2019	177	101.7812994
2020	198	105.339899
2021	192	101.2333333
Grand Total	34705	103.5105913

The Pivot Tables can be used to find the average price of the room for each year during the 2017-2021 period.



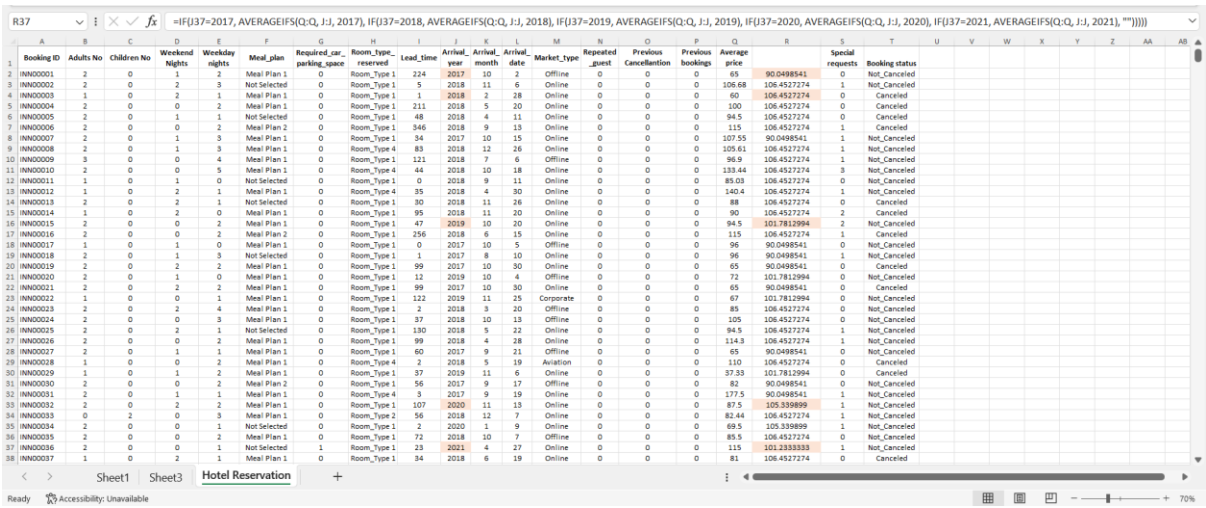
We can see from the above chart that in 2018 there were 28038 room reservations, a very good increase from the previous year. However, in the next three years the reservations decreased considerably.



The average price of the rooms has been increased considerably in 2018 and fluctuated in the period 2018-2021. The higher price of the rooms can be corelated with the lower number of reservations.

In Excel, there is also a more complicated way to find the average price of the rooms, using the following formula:

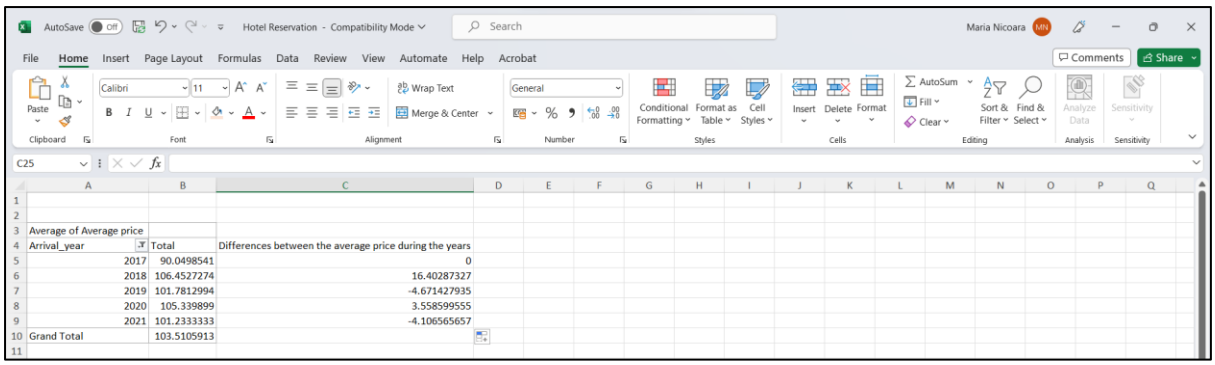
=IF(J2=2017, AVERAGEIFS(Q:Q, J:J, 2017), IF(J2=2018, AVERAGEIFS(Q:Q, J:J, 2018), IF(J2=2019, AVERAGEIFS(Q:Q, J:J, 2019), IF(J2=2020, AVERAGEIFS(Q:Q, J:J, 2020), IF(J2=2021, AVERAGEIFS(Q:Q, J:J, 2021), "")))))



R37 =IF(J37=2017, AVERAGEIFS(Q:Q, J:J, 2017), IF(J37=2018, AVERAGEIFS(Q:Q, J:J, 2018), IF(J37=2019, AVERAGEIFS(Q:Q, J:J, 2019), IF(J37=2020, AVERAGEIFS(Q:Q, J:J, 2020), IF(J37=2021, AVERAGEIFS(Q:Q, J:J, 2021), "")))))

Booking ID	Adults No	Children No	Weekend Nights	Weekday nights	Meal_plan	Required_car_parking_space	Room_type_reserved	Lead_time	Arrival_year	Arrival_month	Arrival_date	Market_type	Repeated_guest	Previous Cancellantion	Previous bookings	Average price		Special requests	Booking status
INN00001	2	0	1	2	Meal Plan 1	0	Room_Type 1	224	2017	10	2	Offline	0	0	0	65	90.0498541	0	Not_Canceled
INN00002	2	0	2	3	Not Selected	0	Room_Type 1	5	2018	11	6	Online	0	0	0	106.68	106.4527274	1	Not_Canceled
INN00003	1	0	2	1	Meal Plan 1	0	Room_Type 1	1	2018	2	28	Online	0	0	0	60	106.4527274	0	Canceled
INN00004	2	0	0	2	Meal Plan 1	0	Room_Type 1	211	2018	5	20	Online	0	0	0	100	106.4527274	0	Canceled
INN00005	2	0	1	1	Not Selected	0	Room_Type 1	48	2018	4	11	Online	0	0	0	94.5	106.4527274	0	Canceled
INN00006	2	0	0	2	Meal Plan 2	0	Room_Type 1	346	2018	9	13	Online	0	0	0	115	106.4527274	1	Canceled
INN00007	2	0	1	3	Meal Plan 1	0	Room_Type 1	34	2017	10	15	Online	0	0	0	107.55	90.0498541	1	Not_Canceled
INN00008	2	0	1	3	Meal Plan 1	0	Room_Type 4	83	2018	12	26	Online	0	0	0	105.61	106.4527274	1	Not_Canceled
INN00009	3	0	0	4	Meal Plan 1	0	Room_Type 1	121	2018	7	6	Offline	0	0	0	96.9	106.4527274	1	Not_Canceled
INN00010	2	0	0	5	Meal Plan 1	0	Room_Type 4	44	2018	10	18	Online	0	0	0	133.44	106.4527274	3	Not_Canceled
INN00011	1	0	1	0	Not Selected	0	Room_Type 1	0	2018	9	11	Online	0	0	0	85.03	106.4527274	0	Not_Canceled
INN00012	1	0	2	1	Meal Plan 1	0	Room_Type 4	35	2018	4	30	Online	0	0	0	140.4	106.4527274	1	Not_Canceled
INN00013	2	0	2	1	Not Selected	0	Room_Type 1	30	2018	11	26	Online	0	0	0	88	106.4527274	0	Canceled
INN00014	1	0	2	0	Meal Plan 1	0	Room_Type 1	95	2018	11	20	Online	0	0	0	90	106.4527274	2	Canceled
INN00015	2	0	0	2	Meal Plan 1	0	Room_Type 1	47	2019	10	20	Online	0	0	0	94.5	101.7812994	2	Not_Canceled
INN00016	2	0	0	2	Meal Plan 2	0	Room_Type 1	256	2018	6	15	Online	0	0	0	115	106.4527274	1	Canceled
INN00017	1	0	1	0	Meal Plan 1	0	Room_Type 1	0	2017	10	5	Offline	0	0	0	96	90.0498541	0	Not_Canceled
INN00018	2	0	1	3	Not Selected	0	Room_Type 1	1	2017	8	10	Online	0	0	0	96	90.0498541	1	Not_Canceled
INN00019	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	30	Online	0	0	0	65	90.0498541	0	Canceled
INN00020	2	0	1	0	Meal Plan 1	0	Room_Type 1	12	2019	10	4	Offline	0	0	0	72	101.7812994	0	Not_Canceled
INN00021	2	0	2	2	Meal Plan 1	0	Room_Type 1	99	2017	10	30	Online	0	0	0	65	90.0498541	0	Canceled
INN00022	1	0	0	1	Meal Plan 1	0	Room_Type 1	122	2019	11	25	Corporate	0	0	0	67	101.7812994	0	Not_Canceled
INN00023	2	0	2	4	Meal Plan 1	0	Room_Type 1	2	2018	3	20	Offline	0	0	0	85	106.4527274	0	Not_Canceled
INN00024	2	0	0	3	Meal Plan 1	0	Room_Type 1	37	2018	10	13	Offline	0	0	0	105	106.4527274	0	Not_Canceled
INN00025	2	0	2	1	Not Selected	0	Room_Type 1	130	2018	5	22	Online	0	0	0	94.5	106.4527274	1	Not_Canceled
INN00026	2	0	0	2	Meal Plan 1	0	Room_Type 1	99	2018	4	28	Online	0	0	0	114.3	106.4527274	1	Not_Canceled
INN00027	2	0	1	1	Meal Plan 1	0	Room_Type 1	60	2017	9	21	Offline	0	0	0	65	90.0498541	0	Not_Canceled
INN00028	1	0	0	2	Meal Plan 1	0	Room_Type 4	2	2018	5	19	Aviation	0	0	0	110	106.4527274	0	Canceled
INN00029	1	0	1	2	Meal Plan 1	0	Room_Type 1	37	2019	11	6	Online	0	0	0	37.33	101.7812994	0	Canceled
INN00030	2	0	0	2	Meal Plan 2	0	Room_Type 1	56	2017	9	17	Offline	0	0	0	82	90.0498541	0	Not_Canceled
INN00031	2	0	1	1	Meal Plan 1	0	Room_Type 4	3	2017	9	19	Online	0	0	0	177.5	90.0498541	1	Not_Canceled
INN00032	2	0	2	2	Meal Plan 1	0	Room_Type 1	107	2020	11	13	Online	0	0	0	87.5	105.339899	1	Not_Canceled
INN00033	0	2	0	3	Meal Plan 1	0	Room_Type 2	56	2018	12	7	Online	0	0	0	82.44	106.4527274	1	Not_Canceled
INN00034	2	0	0	1	Not Selected	0	Room_Type 1	2	2020	1	9	Online	0	0	0	69.5	105.339899	1	Not_Canceled
INN00035	2	0	0	2	Meal Plan 1	0	Room_Type 1	72	2018	10	7	Offline	0	0	0	85.5	106.4527274	0	Not_Canceled
INN00036	2	0	0	1	Not Selected	1	Room_Type 1	23	2021	4	27	Online	0	0	0	115	101.2333333	1	Not_Canceled
INN00037	1	0	2	1	Meal Plan 1	0	Room_Type 1	34	2018	6	19	Online	0	0	0	81	106.4527274	0	Canceled

Sheet1 Sheet3 Hotel Reservation



Average of Average price		
Arrival_year	Total	Differences between the average price during the years
2017	90.0498541	0
2018	106.4527274	16.40287327
2019	101.7812994	-4.671427935
2020	105.339899	3.558599555
2021	101.2333333	-4.106565657
Grand Total	103.5105913	

In the above Pivot Table, it can be seen the difference between the price of the rooms during the years.

To analyse the decline or increase in the average room reservation price over the years 2017 to 2021, we'll calculate the percentage change for each year compared to the previous year. This will help us understand the trend and identify whether there was an increase or decrease in the average room price over time.

Percentage Change from 2017 to 2018:

Percentage Change = ((New Value - Old Value) / Old Value) * 100

Percentage Change = ((106 - 90) / 90) * 100 ≈ 17.78%

Percentage Change from 2018 to 2019:

Percentage Change = ((New Value - Old Value) / Old Value) * 100

Percentage Change = ((101 - 106) / 106) * 100 ≈ -4.72% **(Negative value indicates a decrease)**

Percentage Change from 2019 to 2020:

Percentage Change = ((New Value - Old Value) / Old Value) * 100

Percentage Change = ((105 - 101) / 101) * 100 ≈ 3.96%

Percentage Change from 2020 to 2021:

Percentage Change = ((New Value - Old Value) / Old Value) * 100

Percentage Change = ((101 - 105) / 105) * 100 ≈ -3.81% **(Negative value indicates a decrease)**

From 2017 to 2018, there was an increase of approximately 17.78% in the average room price.

From 2018 to 2019, there was a decrease of approximately 4.72% in the average room price.

From 2019 to 2020, there was an increase of approximately 3.96% in the average room price.

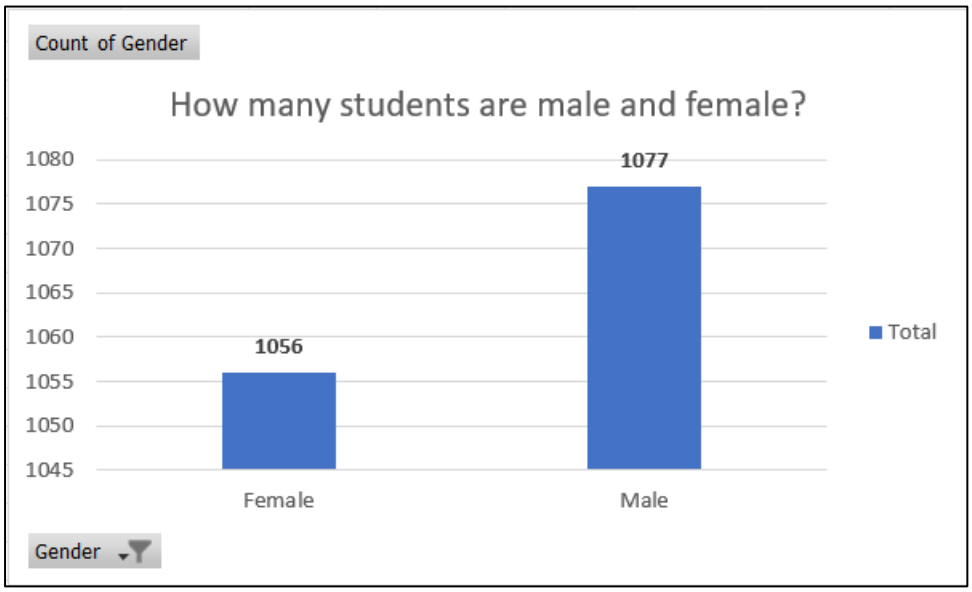
From 2020 to 2021, there was a decrease of approximately 3.81% in the average room price.

Overall, the analysis shows that the average room reservation price fluctuated over the years, with some years experiencing an increase and others experiencing a decrease. The fluctuations were not substantial, indicating relatively stable pricing trends over the five-year period.

Task 2

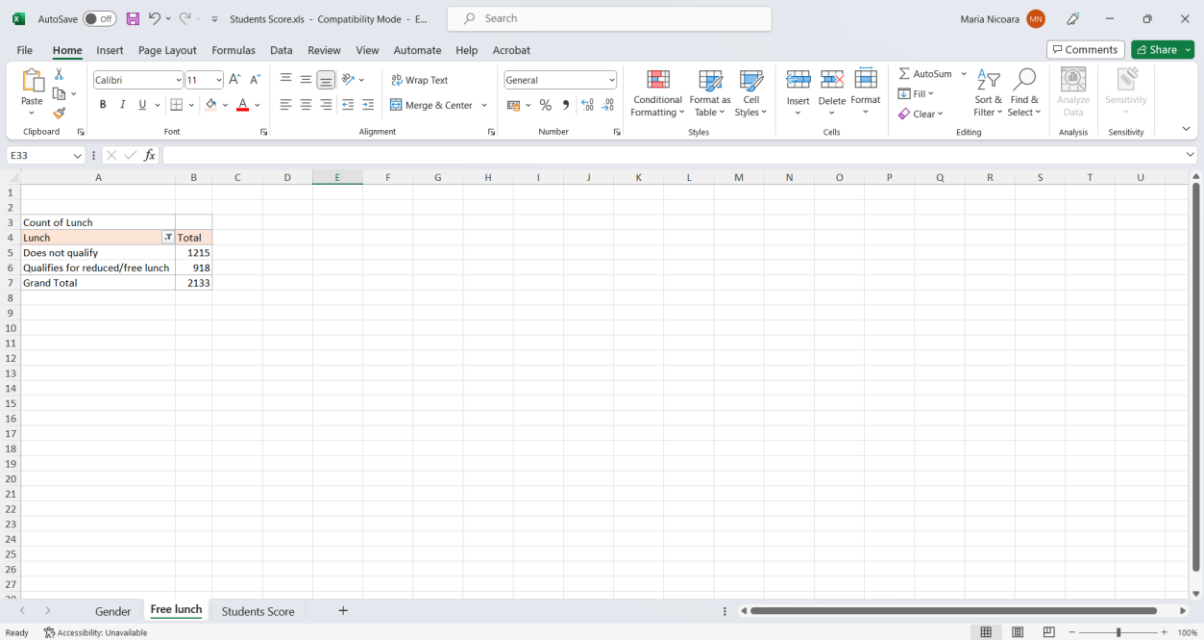
1) How many students are male and female?

Count of Gender	
Gender	Total
Female	1056
Male	1077
Grand Total	2133

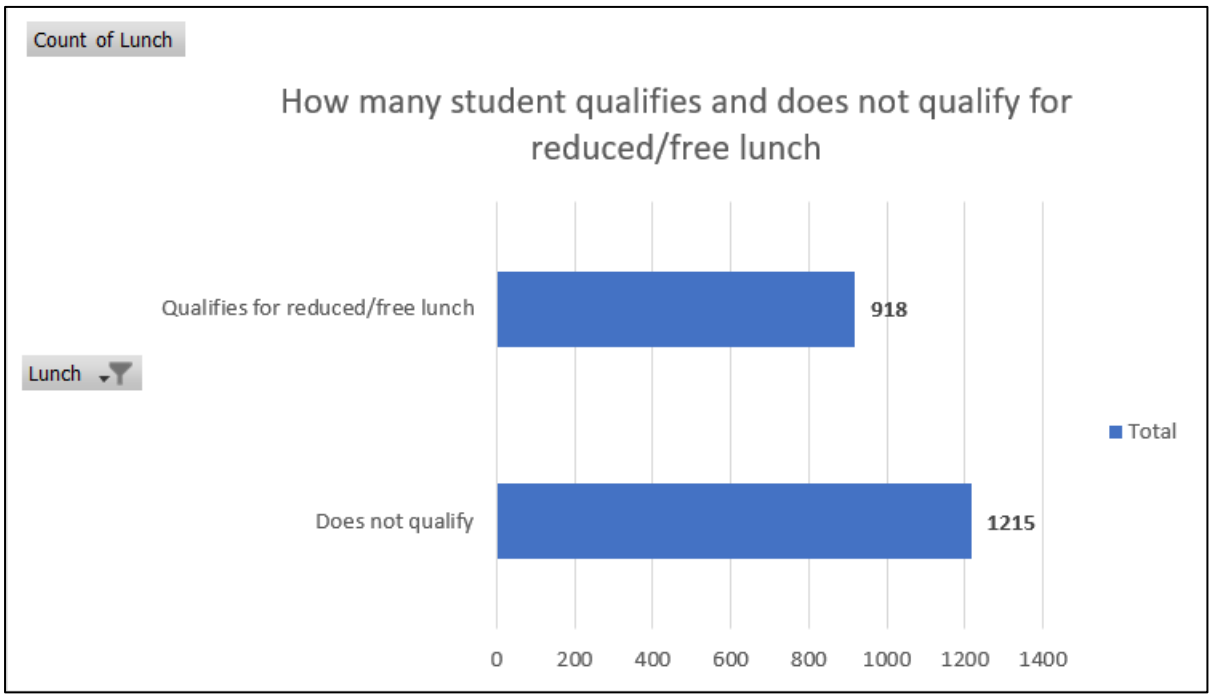


From this analysis, we can see that there are more little male students that female students. However, the difference is not significant. There are just 21 more male than females.

2) How many students qualifies and does not qualify for reduced/free lunch?

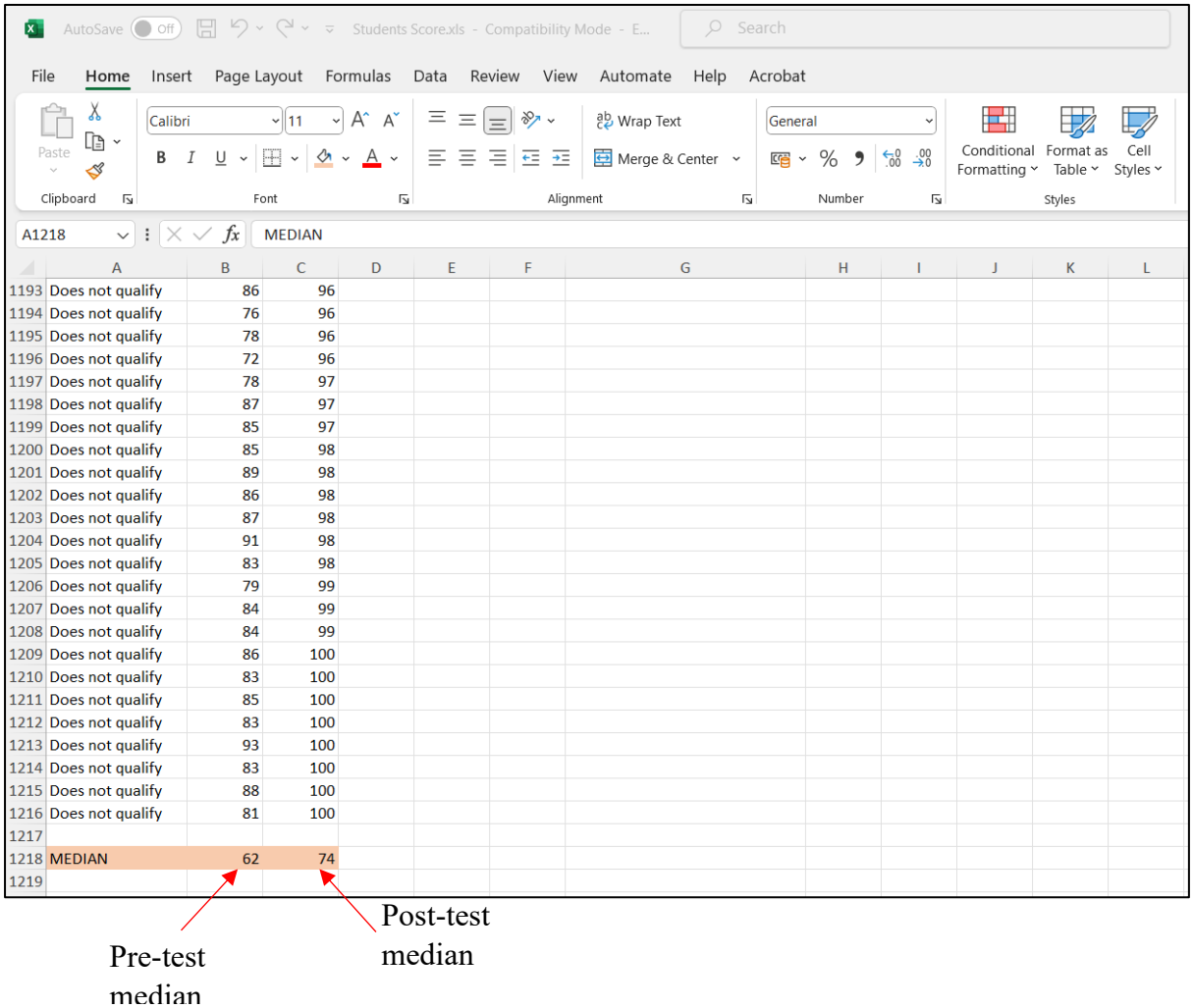


Count of Lunch	
Lunch	Total
Does not qualify	1215
Qualifies for reduced/free lunch	918
Grand Total	2133



Moreover, there are more students who does not qualify for free lunch. 918 students qualify for reduced/free lunch and 1215 does not qualify.

3) What are the mean and median of students who qualifies and does not qualify for reduced/free lunch?

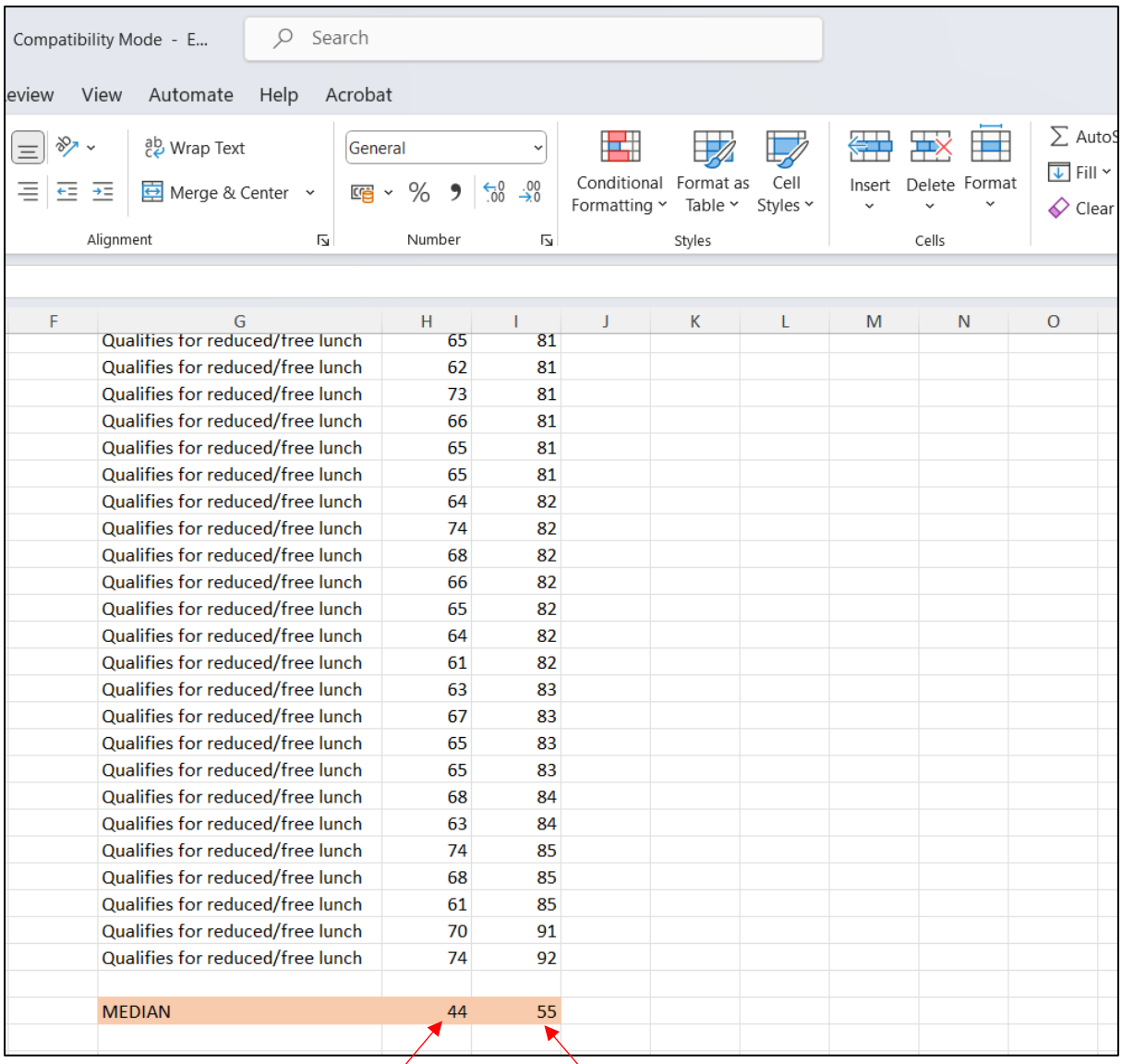


	A	B	C
1193	Does not qualify	86	96
1194	Does not qualify	76	96
1195	Does not qualify	78	96
1196	Does not qualify	72	96
1197	Does not qualify	78	97
1198	Does not qualify	87	97
1199	Does not qualify	85	97
1200	Does not qualify	85	98
1201	Does not qualify	89	98
1202	Does not qualify	86	98
1203	Does not qualify	87	98
1204	Does not qualify	91	98
1205	Does not qualify	83	98
1206	Does not qualify	79	99
1207	Does not qualify	84	99
1208	Does not qualify	84	99
1209	Does not qualify	86	100
1210	Does not qualify	83	100
1211	Does not qualify	85	100
1212	Does not qualify	83	100
1213	Does not qualify	93	100
1214	Does not qualify	83	100
1215	Does not qualify	88	100
1216	Does not qualify	81	100
1217			
1218	MEDIAN	62	74
1219			

Pre-test median

Post-test median

From the median analysis of the test results, we can see that students who does not qualify for free/reduced lunch have the average results of 62 pre-test and 74 post-tests. Thus, the post-test results are better.



Qualifies for reduced/free lunch	65	81
Qualifies for reduced/free lunch	62	81
Qualifies for reduced/free lunch	73	81
Qualifies for reduced/free lunch	66	81
Qualifies for reduced/free lunch	65	81
Qualifies for reduced/free lunch	65	81
Qualifies for reduced/free lunch	64	82
Qualifies for reduced/free lunch	74	82
Qualifies for reduced/free lunch	68	82
Qualifies for reduced/free lunch	66	82
Qualifies for reduced/free lunch	65	82
Qualifies for reduced/free lunch	64	82
Qualifies for reduced/free lunch	61	82
Qualifies for reduced/free lunch	63	83
Qualifies for reduced/free lunch	67	83
Qualifies for reduced/free lunch	65	83
Qualifies for reduced/free lunch	65	83
Qualifies for reduced/free lunch	68	84
Qualifies for reduced/free lunch	63	84
Qualifies for reduced/free lunch	74	85
Qualifies for reduced/free lunch	68	85
Qualifies for reduced/free lunch	61	85
Qualifies for reduced/free lunch	70	91
Qualifies for reduced/free lunch	74	92
MEDIAN	44	55

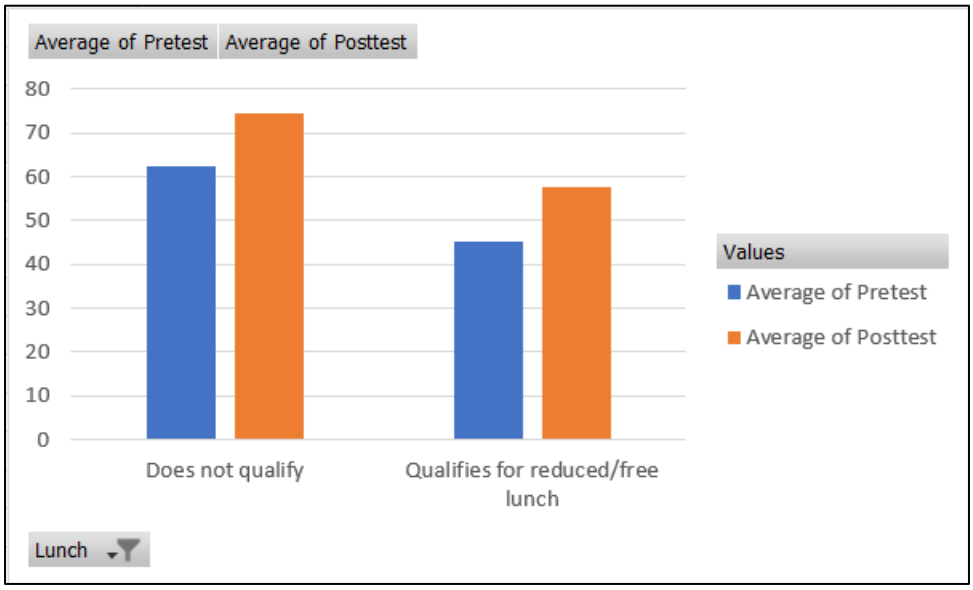
Pre-test median

Post-test median

For students who qualify for free/reduced lunch, the results are lower. 44 for pre-test and 55 for post-test. There is an increase in this case as well. However, the results are not as good as the ones from the other category.

Below, I have also calculated the mean for the two categories.

Lunch	Data	
	Average of Pretest	Average of Posttest
Does not qualify	62.29218107	74.37530864
Qualifies for reduced/free lunch	45.24618736	57.47603486
Grand Total	54.95593061	67.10220347



4) What are the mean and median differences of students' pre-test and post-test scores?

Based on these findings, it can be seen a correlation between the free/reduced lunch and the students' academic performance. The mean and median scores indicate that students who are eligible for reduced/free lunch have much lower results than the ones that does not qualify.

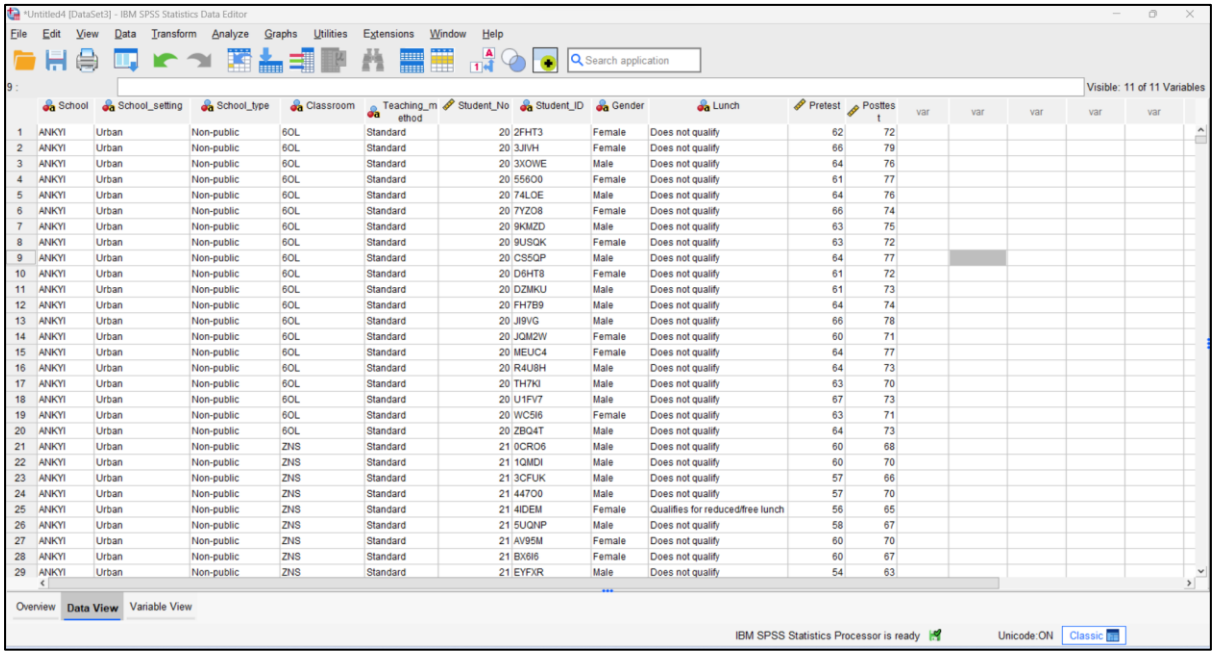
Students who qualify for free/reduced lunch experienced an average percentage increase of approximately 26.67% in their test scores from the pre-test to the post-test.

Students who do not qualify for free/reduced lunch experienced an average percentage increase of approximately 19.35% in their test scores from the pre-test to the post-test.

Based on these results, it appears that students who qualify for free/reduced lunch showed a higher percentage increase in their test scores compared to students who do not qualify. This could indicate that the lunch program might have had a positive impact on the academic performance of these students.

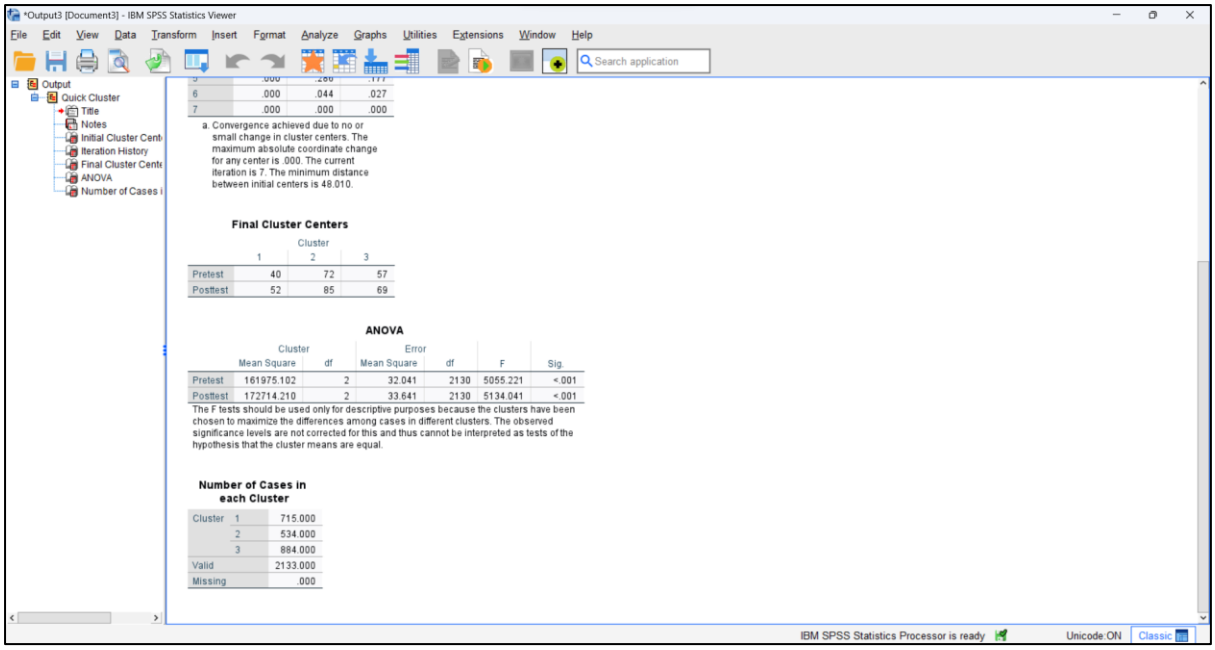
However, it's important to note that these findings are based on averages, and individual student outcomes may vary. Other factors, such as teaching methods, class size, and individual learning abilities, may also contribute to the observed differences.

5) Using the dataset Student_Score.csv provided in conjunction with SPSS, give a specific example of Clustering (ex. K-Means Clustering). Show your workings with screenshots and clearly explain your results.



	School	School_setting	School_type	Classroom	Teaching_method	Student_No	Student_ID	Gender	Lunch	Pretest	Posttest
1	ANKYI	Urban	Non-public	6OL	Standard	20	2FHT3	Female	Does not qualify	62	72
2	ANKYI	Urban	Non-public	6OL	Standard	20	3JIVH	Female	Does not qualify	66	79
3	ANKYI	Urban	Non-public	6OL	Standard	20	3XOWE	Male	Does not qualify	64	76
4	ANKYI	Urban	Non-public	6OL	Standard	20	556O0	Female	Does not qualify	61	77
5	ANKYI	Urban	Non-public	6OL	Standard	20	74LOE	Male	Does not qualify	64	76
6	ANKYI	Urban	Non-public	6OL	Standard	20	7YZO8	Female	Does not qualify	66	74
7	ANKYI	Urban	Non-public	6OL	Standard	20	9KMZD	Male	Does not qualify	63	75
8	ANKYI	Urban	Non-public	6OL	Standard	20	9USQK	Female	Does not qualify	63	72
9	ANKYI	Urban	Non-public	6OL	Standard	20	CS5QP	Male	Does not qualify	64	77
10	ANKYI	Urban	Non-public	6OL	Standard	20	D6HT8	Female	Does not qualify	61	72
11	ANKYI	Urban	Non-public	6OL	Standard	20	DZMKU	Male	Does not qualify	61	73
12	ANKYI	Urban	Non-public	6OL	Standard	20	FH7B9	Male	Does not qualify	64	74
13	ANKYI	Urban	Non-public	6OL	Standard	20	JI9VG	Male	Does not qualify	66	78
14	ANKYI	Urban	Non-public	6OL	Standard	20	JQM2W	Female	Does not qualify	60	71
15	ANKYI	Urban	Non-public	6OL	Standard	20	MEUC4	Female	Does not qualify	64	77
16	ANKYI	Urban	Non-public	6OL	Standard	20	R4U8H	Male	Does not qualify	64	73
17	ANKYI	Urban	Non-public	6OL	Standard	20	TH7KI	Male	Does not qualify	63	70
18	ANKYI	Urban	Non-public	6OL	Standard	20	U1FV7	Male	Does not qualify	67	73
19	ANKYI	Urban	Non-public	6OL	Standard	20	WC5I6	Female	Does not qualify	63	71
20	ANKYI	Urban	Non-public	6OL	Standard	20	ZBQ4T	Male	Does not qualify	64	73
21	ANKYI	Urban	Non-public	ZNS	Standard	21	0CRO6	Male	Does not qualify	60	68
22	ANKYI	Urban	Non-public	ZNS	Standard	21	1QMDI	Male	Does not qualify	60	70
23	ANKYI	Urban	Non-public	ZNS	Standard	21	3CFUK	Male	Does not qualify	57	66
24	ANKYI	Urban	Non-public	ZNS	Standard	21	447O0	Male	Does not qualify	57	70
25	ANKYI	Urban	Non-public	ZNS	Standard	21	4IDEM	Female	Qualifies for reduced/free lunch	56	65
26	ANKYI	Urban	Non-public	ZNS	Standard	21	5UQNP	Male	Does not qualify	58	67
27	ANKYI	Urban	Non-public	ZNS	Standard	21	AV95M	Female	Does not qualify	60	70
28	ANKYI	Urban	Non-public	ZNS	Standard	21	BX6I6	Female	Does not qualify	60	67
29	ANKYI	Urban	Non-public	ZNS	Standard	21	EYFXR	Male	Does not qualify	54	63

Cluster Means: We will have 3 clusters.



6	.000	.044	.027
7	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 7. The minimum distance between initial centers is 48.010.

Final Cluster Centers

	Cluster 1	Cluster 2	Cluster 3
Pretest	40	72	57
Posttest	52	85	69

ANOVA

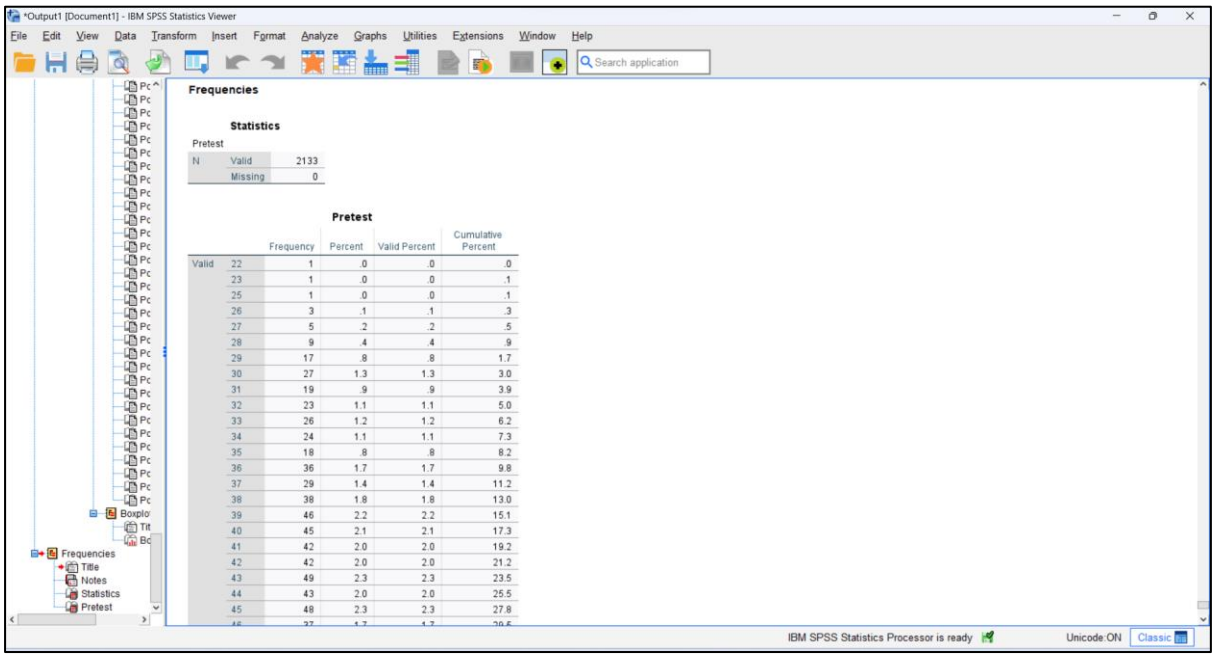
	Cluster Mean Square	df	Error Mean Square	df	F	Sig.
Pretest	161975.102	2	32.041	2130	5055.221	<.001
Posttest	172714.210	2	33.641	2130	5134.041	<.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize 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.

Number of Cases in each Cluster

Cluster	1	715.000
	2	534.000
	3	884.000
Valid		2133.000
Missing		.000

Cluster size:



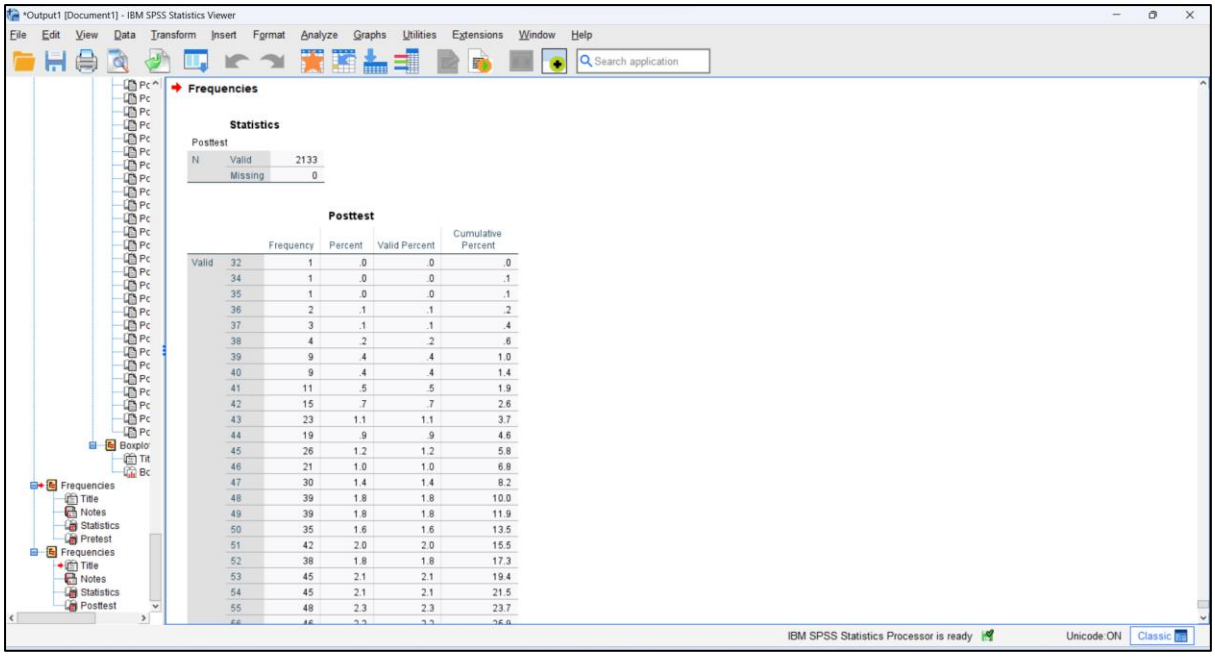
Frequencies

Statistics

Pretest		
N	Valid	2133
	Missing	0

Pretest

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	22	1	.0	.0	.0
	23	1	.0	.0	.1
	25	1	.0	.0	.1
	26	3	.1	.1	.3
	27	5	.2	.2	.5
	28	9	.4	.4	.9
	29	17	.8	.8	1.7
	30	27	1.3	1.3	3.0
	31	19	.9	.9	3.9
	32	23	1.1	1.1	5.0
	33	26	1.2	1.2	6.2
	34	24	1.1	1.1	7.3
	35	18	.8	.8	8.2
	36	36	1.7	1.7	9.8
	37	29	1.4	1.4	11.2
	38	38	1.8	1.8	13.0
	39	46	2.2	2.2	15.1
	40	45	2.1	2.1	17.3
	41	42	2.0	2.0	19.2
	42	42	2.0	2.0	21.2
	43	49	2.3	2.3	23.5
	44	43	2.0	2.0	25.5
	45	48	2.3	2.3	27.8



*Output1 [Document1] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help

Frequencies

Statistics

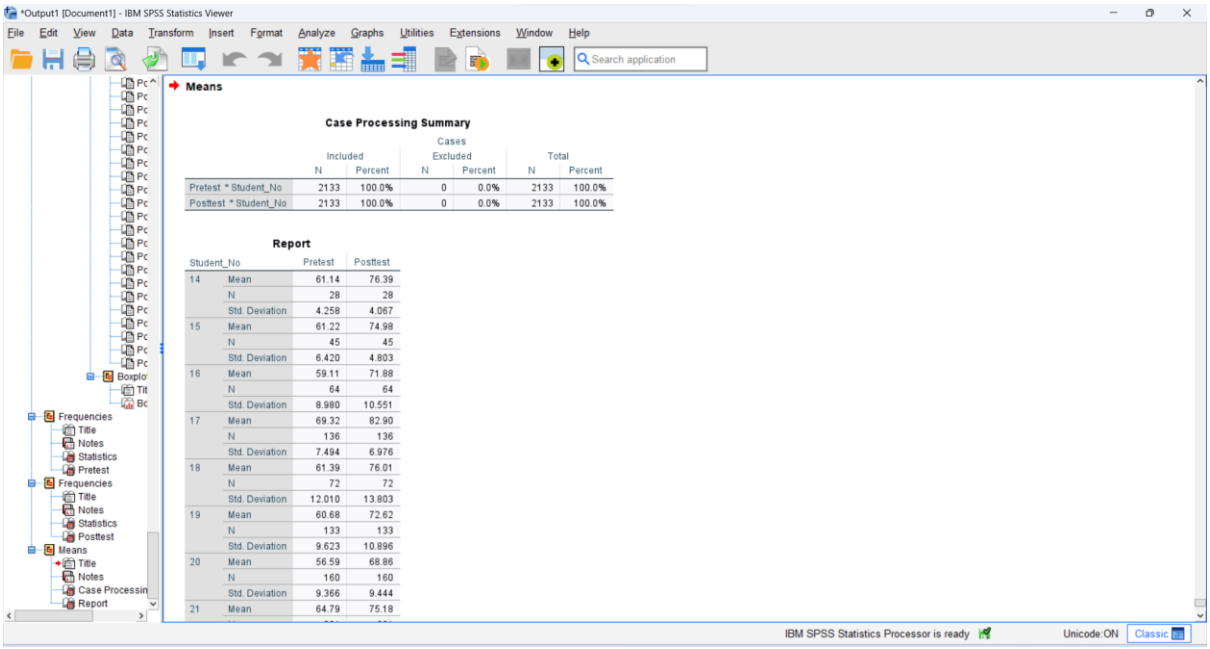
Posttest

N	Valid	2133
	Missing	0

Posttest

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	32	1	.0	.0	.0
	34	1	.0	.0	.1
	35	1	.0	.0	.1
	36	2	.1	.1	.2
	37	3	.1	.1	.4
	38	4	.2	.2	.6
	39	9	.4	.4	1.0
	40	9	.4	.4	1.4
	41	11	.5	.5	1.9
	42	15	.7	.7	2.6
	43	23	1.1	1.1	3.7
	44	19	.9	.9	4.6
	45	26	1.2	1.2	5.8
	46	21	1.0	1.0	6.8
	47	30	1.4	1.4	8.2
	48	39	1.8	1.8	10.0
	49	39	1.8	1.8	11.9
	50	35	1.6	1.6	13.5
	51	42	2.0	2.0	15.5
	52	38	1.8	1.8	17.3
	53	45	2.1	2.1	19.4
	54	45	2.1	2.1	21.5
	55	48	2.3	2.3	23.7

IBM SPSS Statistics Processor is ready Unicode:ON Classic



*Output1 [Document1] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Extensions Window Help

Means

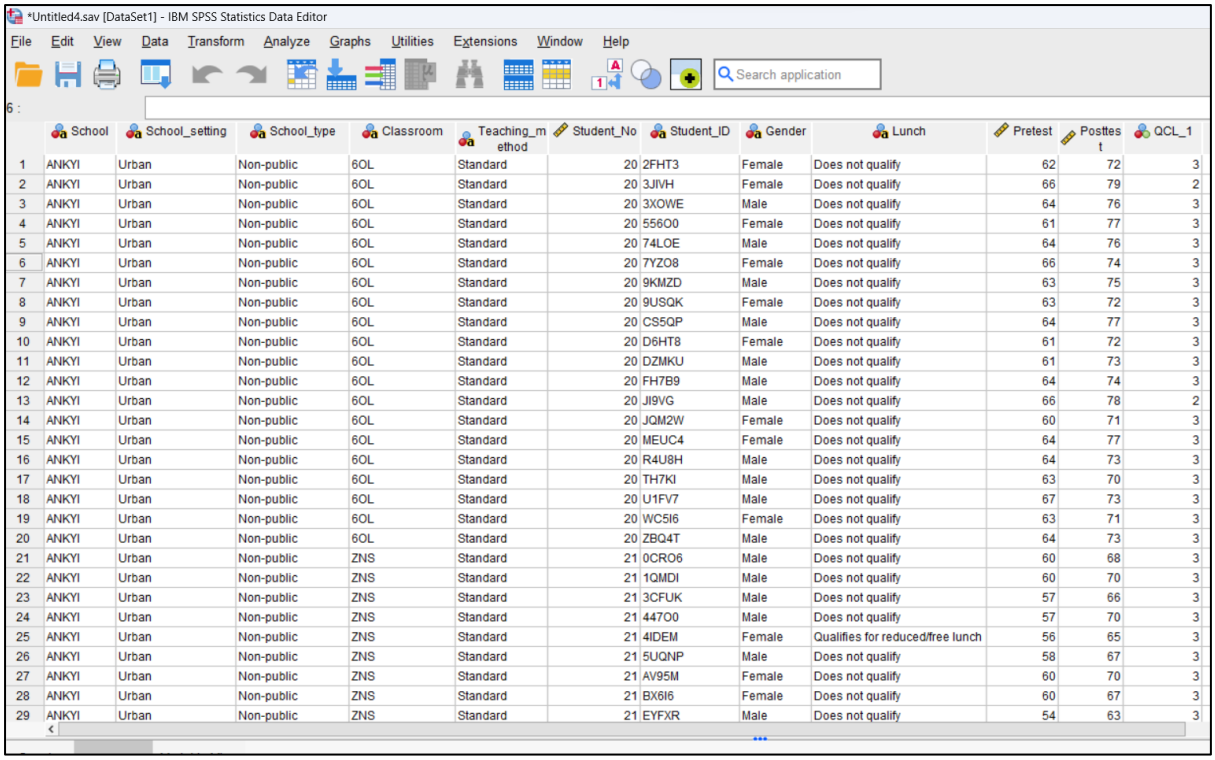
Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Pretest * Student_No	2133	100.0%	0	0.0%	2133	100.0%
Posttest * Student_No	2133	100.0%	0	0.0%	2133	100.0%

Report

Student_No		Pretest	Posttest
14	Mean	61.14	76.39
	N	28	28
	Std. Deviation	4.258	4.067
15	Mean	61.22	74.98
	N	45	45
	Std. Deviation	6.420	4.803
16	Mean	59.11	71.88
	N	64	64
	Std. Deviation	8.980	10.551
17	Mean	69.32	82.90
	N	136	136
	Std. Deviation	7.494	6.976
18	Mean	61.39	76.01
	N	72	72
	Std. Deviation	12.010	13.803
19	Mean	60.68	72.62
	N	133	133
	Std. Deviation	9.623	10.896
20	Mean	56.59	68.86
	N	160	160
	Std. Deviation	9.366	9.444
21	Mean	64.79	75.18

IBM SPSS Statistics Processor is ready Unicode:ON Classic



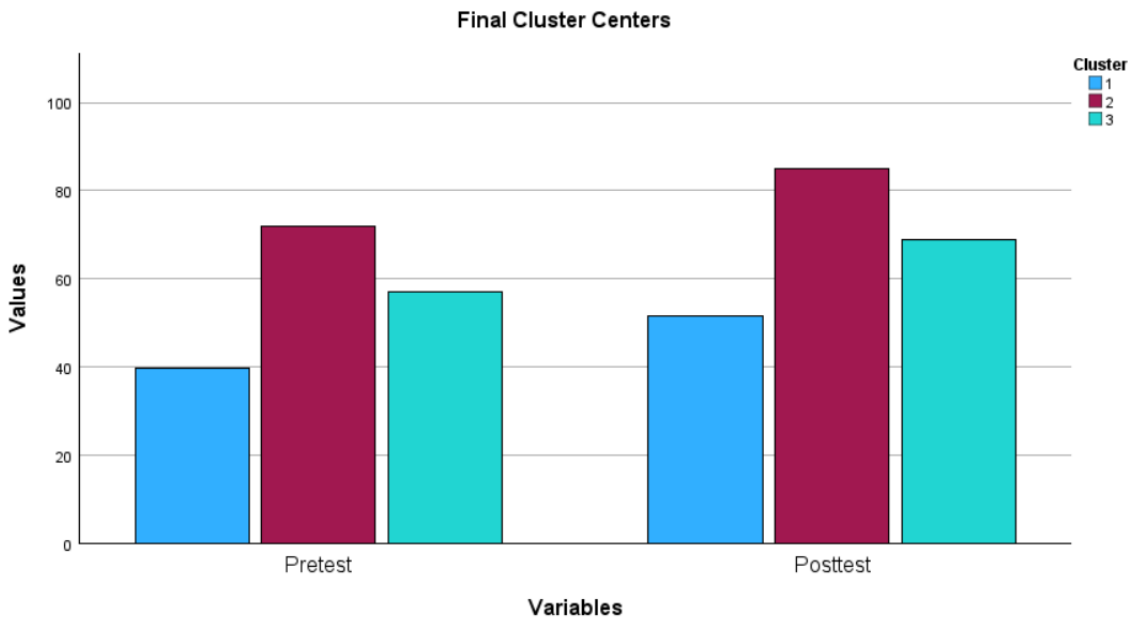
*Untitled4.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

Search application

6 :

	School	School_setting	School_type	Classroom	Teaching_method	Student_No	Student_ID	Gender	Lunch	Pretest	Posttest	QCL_1
1	ANKYI	Urban	Non-public	6OL	Standard	20	2FHT3	Female	Does not qualify	62	72	3
2	ANKYI	Urban	Non-public	6OL	Standard	20	3JIVH	Female	Does not qualify	66	79	2
3	ANKYI	Urban	Non-public	6OL	Standard	20	3XOWE	Male	Does not qualify	64	76	3
4	ANKYI	Urban	Non-public	6OL	Standard	20	556O0	Female	Does not qualify	61	77	3
5	ANKYI	Urban	Non-public	6OL	Standard	20	74LOE	Male	Does not qualify	64	76	3
6	ANKYI	Urban	Non-public	6OL	Standard	20	7YZO8	Female	Does not qualify	66	74	3
7	ANKYI	Urban	Non-public	6OL	Standard	20	9KMZD	Male	Does not qualify	63	75	3
8	ANKYI	Urban	Non-public	6OL	Standard	20	9USQK	Female	Does not qualify	63	72	3
9	ANKYI	Urban	Non-public	6OL	Standard	20	CS5QP	Male	Does not qualify	64	77	3
10	ANKYI	Urban	Non-public	6OL	Standard	20	D6HT8	Female	Does not qualify	61	72	3
11	ANKYI	Urban	Non-public	6OL	Standard	20	DZMKU	Male	Does not qualify	61	73	3
12	ANKYI	Urban	Non-public	6OL	Standard	20	FH7B9	Male	Does not qualify	64	74	3
13	ANKYI	Urban	Non-public	6OL	Standard	20	JI9VG	Male	Does not qualify	66	78	2
14	ANKYI	Urban	Non-public	6OL	Standard	20	JQM2W	Female	Does not qualify	60	71	3
15	ANKYI	Urban	Non-public	6OL	Standard	20	MEUC4	Female	Does not qualify	64	77	3
16	ANKYI	Urban	Non-public	6OL	Standard	20	R4U8H	Male	Does not qualify	64	73	3
17	ANKYI	Urban	Non-public	6OL	Standard	20	TH7KI	Male	Does not qualify	63	70	3
18	ANKYI	Urban	Non-public	6OL	Standard	20	U1FV7	Male	Does not qualify	67	73	3
19	ANKYI	Urban	Non-public	6OL	Standard	20	WC5I6	Female	Does not qualify	63	71	3
20	ANKYI	Urban	Non-public	6OL	Standard	20	ZBQ4T	Male	Does not qualify	64	73	3
21	ANKYI	Urban	Non-public	ZNS	Standard	21	0CRO6	Male	Does not qualify	60	68	3
22	ANKYI	Urban	Non-public	ZNS	Standard	21	1QMDI	Male	Does not qualify	60	70	3
23	ANKYI	Urban	Non-public	ZNS	Standard	21	3CFUK	Male	Does not qualify	57	66	3
24	ANKYI	Urban	Non-public	ZNS	Standard	21	447O0	Male	Does not qualify	57	70	3
25	ANKYI	Urban	Non-public	ZNS	Standard	21	4IDEM	Female	Qualifies for reduced/free lunch	56	65	3
26	ANKYI	Urban	Non-public	ZNS	Standard	21	5UQNP	Male	Does not qualify	58	67	3
27	ANKYI	Urban	Non-public	ZNS	Standard	21	AV95M	Female	Does not qualify	60	70	3
28	ANKYI	Urban	Non-public	ZNS	Standard	21	BX6I6	Female	Does not qualify	60	67	3
29	ANKYI	Urban	Non-public	ZNS	Standard	21	EYFXR	Male	Does not qualify	54	63	3



The "Low Improvers" (Cluster 1) started with a pre-test mean score of 40 and improved to a post-test mean of 52. Their improvement is moderate.

The "High Improvers" (Cluster 2) started with a high pre-test mean score of 72 and significantly improved to a post-test mean of 85. Their improvement is substantial.

The "Moderate Improvers" (Cluster 3) had a pre-test mean score of 57 and showed a moderate improvement to a post-test mean of 69.

Students in Cluster 2 (High Improvers) are performing well initially and making significant improvements. Schools might want to identify the factors contributing to their success and promote these practices among other students.

Students in Cluster 3 (Moderate Improvers) have a balanced performance with room for improvement. Targeted interventions could be designed to support their growth.

Students in Cluster 1 (Low Improvers) are struggling and need focused attention. Specialized assistance or interventions might be needed to help them catch up.

6) With real world (practical) examples, explain minimum 5 most common data mining and text mining methods for each that can be used in the contemporary business.

Data mining and text mining are effective tools used in modern industry to glean insightful information from massive datasets and textual material. Here are five of the most popular techniques for each, along with applications in the actual world:

Data mining techniques

Classification:

Data is categorised via classification into predetermined classes (Bharati, 2010). For instance, categorization algorithms may be employed in the e-commerce industry to forecast whether a client will make a purchase based on their browsing habits. This makes it possible for firms to personalise marketing to potential customers.

Regression:

Based on other relevant factors, regression aids in forecasting numerical values. Regression analysis may be used in the financial sector to predict stock prices by looking at past price trends, interest rates, and market indicators.

Clustering:

Without using specified classifications, clustering combines comparable data points. Retail businesses may utilise clustering to divide their clientele into several groups depending on their buying habits, enabling them to develop individualised marketing plans.

Association Rule Mining:

With the use of association rule mining, connections between variables in huge datasets may be found (Liao, et al., 2012). To discover which products are often bought together, grocery retailers frequently use association rule mining. For instance, people are more likely to purchase butter if they purchase bread.

Anomaly Detection:

This technique aids in finding rare or exceptional data points that deviate from predicted trends. Anomaly detection in cybersecurity can highlight questionable network behaviour that might be a sign of an impending cyberattack.

Text mining techniques

Sentiment analysis:

Sentiment analysis determines whether text conveys positive, negative, or neutral sentiments in order to evaluate public opinion (Dang & Ahmad, 2014). This method is applied by social media platforms to track brand reputation and analyse client comments to enhance goods and services.

Topic Modelling:

Using topic modelling, significant themes or subjects are gleaned from a group of texts. Topic modelling may be used by news organisations to automatically classify and label news stories, facilitating information organisation and retrieval.

Named Entity Recognition (NER):

Named Entity Recognition (NER) detects and categorises entities in a text, including names, dates, locations, and more. Legal companies may use NER to extract pertinent data from court records, saving time on case preparation and research.

Text summarization:

Text summary reduces lengthy papers into clearer, more concise versions while preserving key information (Talib, et al., 2016). Text summary can be used by news aggregators to give readers brief summaries of news pieces, keeping them informed without overwhelming them.

These techniques enable organisations to make data-driven decisions in modern business and achieve a competitive edge. They provide firms the ability to efficiently manage risks, improve product offerings, optimise processes, and analyse customer behaviour. These data mining and text mining approaches are crucial resources for contemporary businesses looking for meaningful insights from their data, whether it is for anticipating consumer preferences, monitoring market trends, or enhancing customer interaction.

7) Discuss the advantages and disadvantage of using SPSS and Excel using relevant practical examples and present your discussions in a table.

Criteria	SPSS	Excel
Advantages		
Data Analysis Capabilities	SPSS is specifically designed for complex statistical analysis. It supports a wide range of statistical tests, regression, ANOVA, factor analysis, and more.	Excel offers basic data analysis functions but lacks the advanced statistical tools available in SPSS.
User-Friendly	SPSS has a user-friendly interface with menus and dialogs, making it easier to perform advanced analyses without coding.	Excel is familiar to many users and offers a simple interface for basic calculations and graphs.
Data Visualization	SPSS provides advanced visualization tools such as scatter plots, histograms, and box plots for in-depth insights.	Excel offers basic charting options, suitable for simple data representation.
Disadvantages		
Learning Curve	SPSS requires a learning curve, especially for those new to statistical software. It may take time to become proficient.	Excel's basic functions are easy to learn, but more advanced features can be challenging for complex analyses.
Limited Spreadsheet Functionality	While SPSS has some spreadsheet functionality, it's not as versatile as Excel for general data management tasks.	Excel is primarily designed for spreadsheet tasks and may lack the depth needed for certain data analysis tasks.
Data Manipulation	SPSS focuses on analysis rather than data manipulation. Excel offers more powerful tools for data cleansing and transformation.	Excel provides limited data cleaning and transformation capabilities compared to dedicated data analysis tools like SPSS.
Integration with Office Suite	SPSS lacks seamless integration with Microsoft Office Suite, which can hinder workflow if using other Office tools.	Excel is tightly integrated with other Microsoft Office applications, making it easier to share and collaborate on data.

Practical example:

Consider that you are analysing customer preferences using a market research survey. Using clustering techniques and extensive segmentation analysis, SPSS enables you to pinpoint separate client groups based on their preferences. Instead of digging into complex segmentation, Excel may be used to construct simple frequency distributions or bar charts that display the general distribution of preferences.

When gathering sales information for several items across various areas. We might make sophisticated visualisations using SPSS, such as scatter plots that illustrate the link between sales and marketing expenditure. Simpler line charts may be made in Excel to show historical sales patterns.

In conclusion, Excel and SPSS are both useful tools, but they are best suited for particular jobs based on their individual strengths and weaknesses. Excel is fantastic for simple data administration and calculations, whereas SPSS shines in complex statistical analysis. You may pick the one that best meets your demands based on the intricacy of your analysis and your expertise with the tools.

References

Bharati, M., (2010). Data mining techniques and applications. *Indian Journal of Computer Science and Engineering,* 1(4).

Connolly, T. & Begg, C., (2014). *Database Systems: A Practical Approach to Design, Implementation, and Management.* 5th Ed. ed. Boston, MA: Pearson.

Dang, S. & Ahmad, P. H., (2014). Text Mining : Techniques and its Application. *IJETI International Journal of Engineering & Technology Innovations,* 1(4).

Liao, S.-H., Chu, P.-H. & Hsiao, P.-Y., (2012). Data mining techniques and applications. *Expert Systems with Applications,* Volume 39, p. 11303–11311.

Sharda, R., Dursun, D. & Turban, E., (2014). *Business Intelligence and Analytics: Systems for Decision Support.* 3rd Edition ed. Boston, MA: Pearson.

Talib, R., Kashif, M., Ayesha, S. & Fatima, F., (2016). Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications ,* 7(11).

Whitten, I., Frank, E., Hall, M. & Pal, C., (2017). *Data Mining: Practical Machine Learning Tools and Techniques.* 4th Edition ed. s.l.:Morgan Kaufman.