



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

BA BUSINESS MANAGEMENT

Assignment 2

Module: Data Handling & Business Intelligence

Code: LCBB5000

Lecturer: Dr. Anupam Mazumdar

Name: Radu Maimescu

Student ID: 2131119

Table of Contents

Part 1:..... 3

Part 2:..... 13

References..... 24

Part 1:

Introduction

Hotel management has employed descriptive analytics to predict the future profits or catastrophes by analyzing trends, patterns and abnormalities. Hoteliers used to achieve exploration in diverse difficulties like reservations termination forecast, client subdivision, client satiation among others. Investigators can use the f sets of data to standardize reservations' forecast annulment replicas against outcomes now identified. Machine learning investigators can use the sets of data for benchmarking the presentation of diverse systems for resolving the identical kind of difficult (cataloguing, division, or other). Instructors can use the sets of data for device knowledge cataloguing or dissection problems. Instructors can use the sets of data to achieve either figures or facts excavating exercise. The research aims to work on five major items namely: enhanced pricing and request organization throughout ultimate plea periods; better communications between hotel control and with potential clients; market segmentation, overall discounting policies and strategies versus persons that are beleaguered to precise souk sections; impervious valuing; presentation quantification.

Data

In hotel booking management associated hotel businesses, the greatest of the study is on Income Supervision, request/bookings estimation and forecasting glitches by employing earlier database statistics available in journals and printouts. Nevertheless, the lasting travel and tourism

businesses like theme parks, cruising, and hospitality, etc., have diverse necessities and meticulousness that cannot be completely discovered without business's precise facts. Henceforth, the study collected various data from one hotel in which segmentation of data was carried on different items and periods of time. The sets of data now made accessible were gathered targeting at the expansion of forecast replicas to categories a guesthouse reservation's probability to be annulled. Nonetheless, owing to the features of the variables involved in these sets of data, their usage enthusiasms further than this annulment forecast difficult. Syed and Suroso (2018) stated that a unique significant property in data for forecast replicas is not to encourage seepage of forthcoming figures. To ensure this doesn't happen a time limit is set for the variable targeted. The extraction of these variables for the study are obtained from the table of bookings and reservation, and the arrival dates were defined in terms of the time limits from the day of booking. The general areas from which the data was extracted include the Booking Section, Segment of the market, Profile of the Customers, Meal's plan and the cash transactions section.

Experimental Design, Materials and Methods

The hotel's database from the bookings change log, market segment, customers profile, meals department and channels of distribution was compiled in Excel.

Table 1: Description of variables

Variable	Type	Description	Source
Required car parking space	Numerical	Provides the number of parking lots needed	-Bookings Charge log, Transactions, Customer Profile,

			Meals Section, Market Segments, Distribution channels
Room type reserved	Categorical	The rooms differ with sizes, number of beds, executive suite etc	Bookings Charge log, Transactions, Meals Section,
Lead time	Numerical	The time spent for the stay in the rooms	Transactions, Customer Profile,
Arrival year	Numerical	The year the guest clocked for the first time	Bookings Charge log, Transactions, Customer Profile, Market Segments,
Arrival month	Numerical	The month the guest clocked for the first time	Bookings Charge log, Transactions,
Arrival date	Numerical	The day the guest clocked for the first time	Bookings Charge log,
Market type	Categorical	This is either online, corporate, offline or supplementary	Bookings Charge log, Market Segments, Distribution channels

Repeated guest	Numerical	Number of times the guest has showed up for the same facility	Customer Profile,
Previous cancellation	Numerical	Cancellations done by a guest previously	Bookings Charge log, Transactions, Customer Profile, Meals Section, Market Segments, Distribution channels
Adults no	Numerical	The number of grownups who show as from the age of 21 years	Bookings Charge log, Transactions, Customer Profile,
Children no	Numerical	The number of children who show as from the age of 21 years	Bookings Charge log, Transactions, Customer Profile,
Weekend nights	Numerical	Number of weekend nights the guest has spent	Bookings Charge log, Transactions, Customer Profile,
Weekday nights	Numerical	Number of weekday nights the guest has spent	Bookings Charge log, Transactions, Customer Profile,

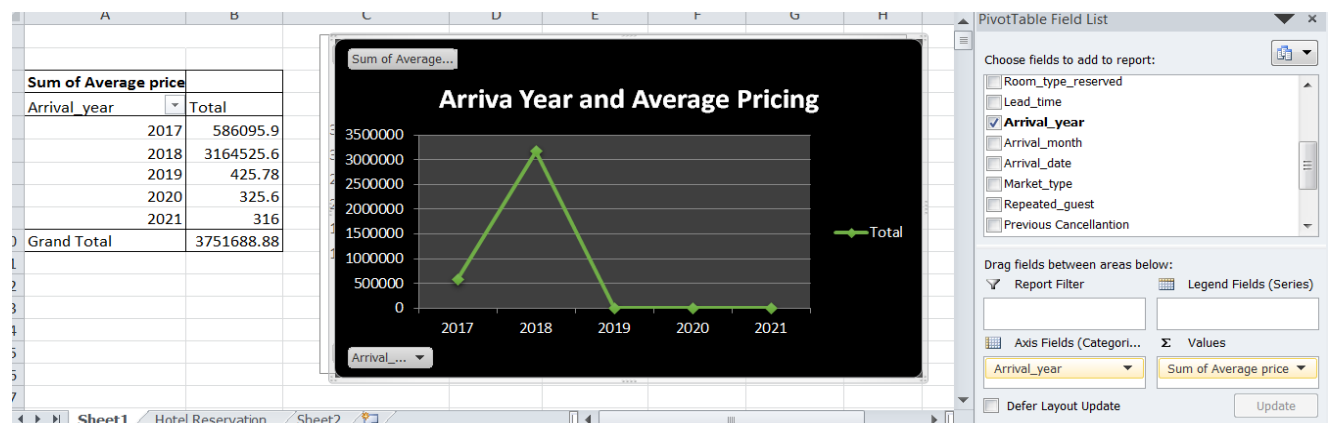
Meal plan	Categorical	The menu offered in combination or with association to a particular guest	Bookings Charge log, Transactions, Customer Profile, Meals Section,
Average price	Numerical	The charges for specified rooms	
Special requests	Numerical	Special guest bookings	Customer Profile,
Booking status	Categorical	Shows whether a reservation is cancelled or not cancelled	Bookings Charge log,

However, caution should be observed when making decisions on the operations of the hotels. The industry is often meeting challenges of customers altering their desire or behavior of booking in various ways such as persons booking a given type of room, the length of stay and time of the month of visit. It is imperative for the hotel to monitor the features or profiles of their customers in terms of nationality and reason for visiting (Theocharidis *et al.*,2020). Consequently, although the recording of statistics took a longer duration erstwhile to entrance time, the circulation of variables fluctuates amid non annulled and negated bookings. As a result, the utilization of the sets of data may need this change in spreading to be engaged into explanation. This change can be understood in the table (Jang *et al.*, p.51). Table plots are a dominant picturing technique and were created with Excel that permits for the examination and investigation of huge multivariate sets of

data. Analyzed domains were Assigned Meal type, Adults, Weeknights booking, children and others.

Results and Discussion

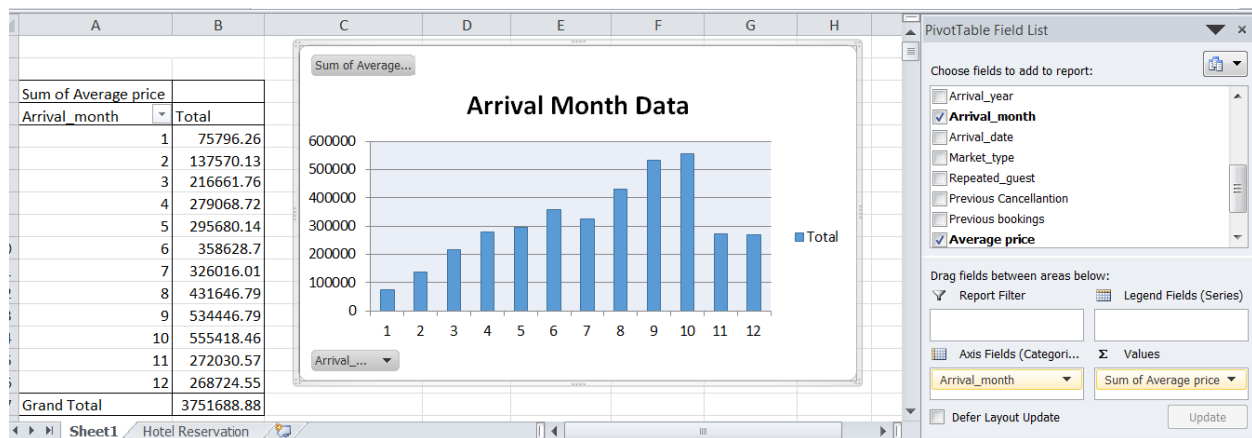
The Average pricing for the period 2017-2021 and over meal plans is depicted below using the pivot table and graphical representations.



Using predictions of forthcoming room demand and supply at substitute charge stages and then augmenting the number of bookings to admit at each dimension of visit, operative rating distinguishes which categories of request are the most lucrative for the state under consideration and founds charges to hearten or agree such plea while disheartening less money-making request. This is the objective to which guesthouses commonly seek; how near the originate to that vision evocatively influences the extent to which their incomes and revenues are capitalized on for instance, Chattopadhyay & Mitra (2019) noted that throughout highest periods a real valuing construction usually disheartens shorter vacations while instantaneously letting for lengthier breaks, even after poorer daily charges are obligatory to seize the lengthier vacations. The incremental income created from visitors who go for lengthier vacations, particularly if those

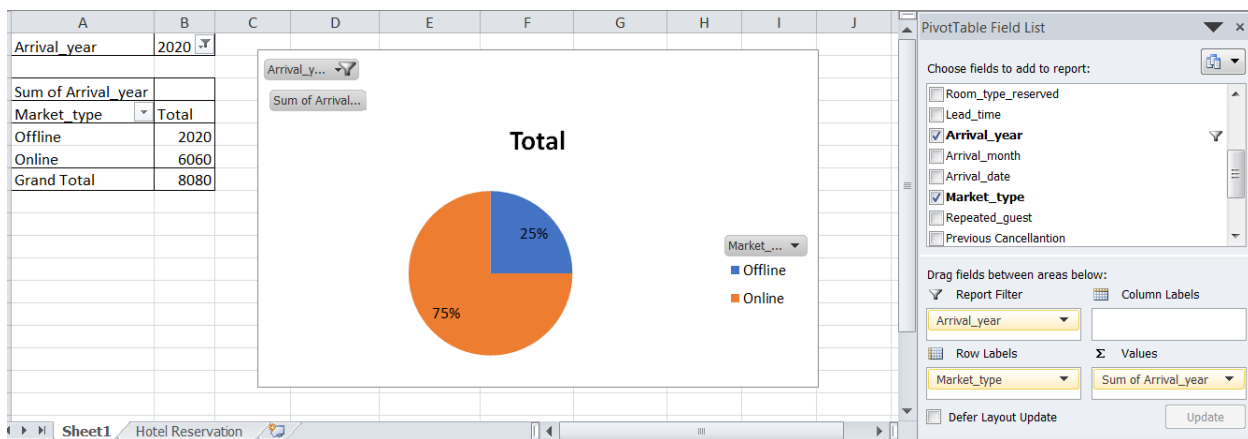
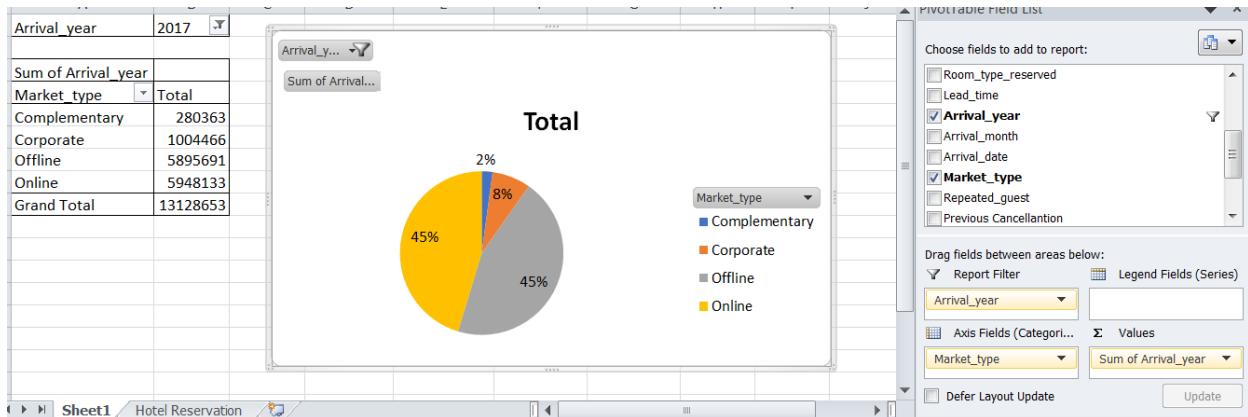
vacations distance days on which the guesthouse is not filled, is usually much better than the sum of income the guesthouse might have acknowledged from undersized vacations at advanced per diem charges.

The monthly average pricing for the years 2017 to 2021 was captured using the graphical Excel key functions as below.



An operative pricing platform also navigates requests to the correct days. Certain prospective visitors have plasticity in their travel-timetables. For those visitors, an operational pricing system empowers them to select days wherever they can obtain lesser charges, serving the guesthouse to shape tenancy where it is desirable (Deyá-Tortella *et al.*, p.103345). Lastly, advertising rates can be used to rouse additional bookings. When cut-rate rates are open, the incremental request stirred by these charges must create up for the decrease in income from visitors who recompense the lesser charge. As advanced tenancy levels normally suggest advanced variable prices.

The type of booking whether online, offline, corporate, or complementary is as below for the year 2017.



Customers initiated spending the Internet time to pursue for lesser fees and performing; nowadays it is correspondingly a suitable way to trade. Inconsistently, customers infrequently associate online charges. Rising statistics of Web links makes the discovery of a superior fee progressively tiresome. Clienteles give up on examining and finding the best prices before meeting the superior charges. Customers may treasure trove advanced charges online due to issues such as silent auction and value discrimination (Chang *et al.*, 2019, p.281). Customers spending time in a

co-operative setting ought to select an acquainted product over an unacquainted one owing to the identified trademark's understood assurance. Makes cuts fee pliability and upsurge the vendor's authority with one revision presentation online. To finish, customers have a duty to also discover superior charge spreading online than in off-line. The choice of values online is that clienteles can without difficulty equate facilities, products, and charges (Zhang *et al.*, 2019. 2745).

Finally, using the VLOOKUP Excel function to look for specific properties of a given ID was done as below. We examined the type of meal associated with a given ID No. of room book.

Function Library					
T3			$\sum$	=VLOOKUP(A2,A1:M19,6,FALSE)	
Q	R	S	T	U	
Average price	Special requests	Booking status			
65	0	Not_Canceled			
106.68	1	Not_Canceled	Meal Plan 1		0

The function is vital in locating particulars of a known value over the anticipated pricing or record.

Conclusion

Various programs and applications have been developed to extract, manage, and give data a meaning for forecasting. Excel and SPSS have come out to be among the best programs in statistical analysis and presentation. Through comparing various factors of time, space, and behavior one has a higher probability of forecasting the anticipated results after a given period of time. There has been a sharp decline in the number of rooms booked as from the year 2019

implying a low-income performance. We recommend the management to research mitigation ways to curb the trend through lowering prices and work on human resource hospitality.

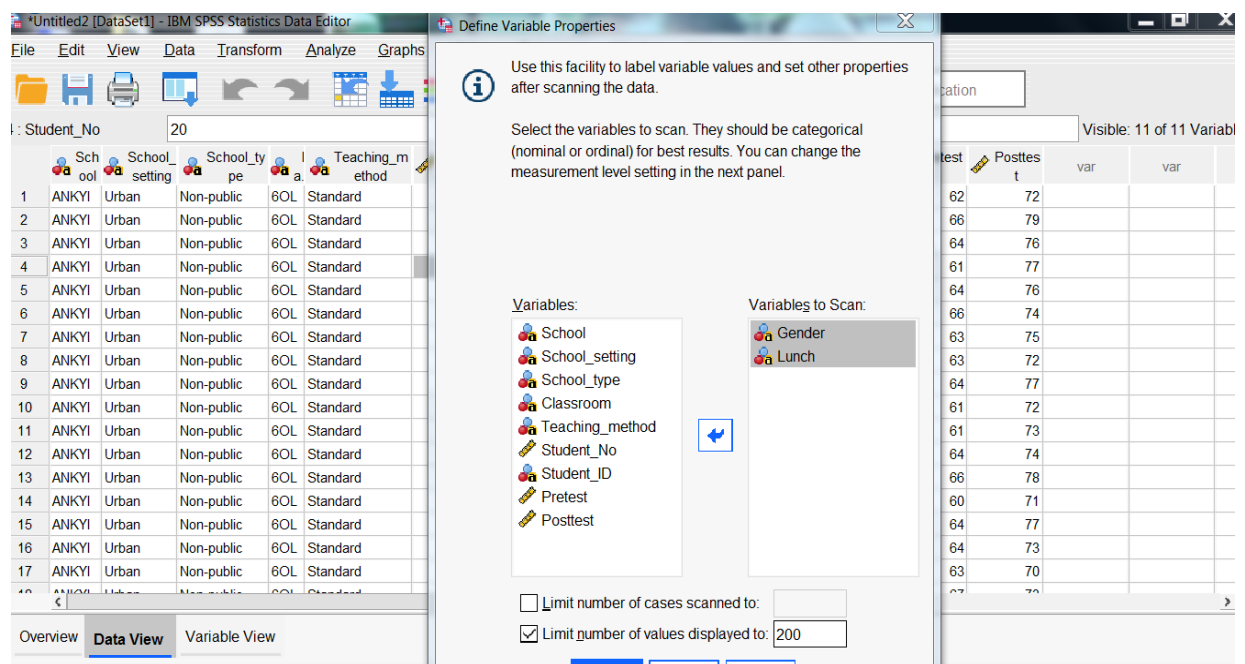
Part 2:

1. Male count is 1077 and female count is 1056.

After loading the data file Student_Score.csv into SPSS, under the go

Analyze>Classify>Discriminant

The display will be:



Add the “Gender” and “Lunch” data stacks to the Variable to Scan Section. The results will be portrayed as below.

Define Variable Properties

Scanned Variable List

Un...	M...	Role	Variable
☒	☒	Gender	Gender
☒	☒	Lunch	Lunch

Current Variable: Gender Label:

Measurement Level: Nomi... [Suggest](#) Type: String

Role: Input Width: 6 [Minimize Width](#)

Unlabeled values: 2 [Attributes...](#)

Value Label grid: [i](#) Enter or edit labels in the grid. You can enter additional values at the bottom.

	Changed	Missing	Count	Value	Label
1	☐	☐	1056	Female	
2	☐	☐	1077	Male	
3	☐	☐			

Cases scanned: 2133 Value list limit: 200

Copy Properties [From Another Variable...](#) [To Other Variables...](#) Unlabeled Values [Automatic Labels](#)

OK Paste Reset Cancel Help

Define Variable Properties

Scanned Variable List

Un...	M...	Role	Variable
☒	☒	Gender	Gender

Current Variable: Gender Label:

Measurement Level: Nomi... [Suggest](#) Type: String

Role: Input Width: 6 [Minimize Width](#)

Unlabeled values: 2 [Attributes...](#)

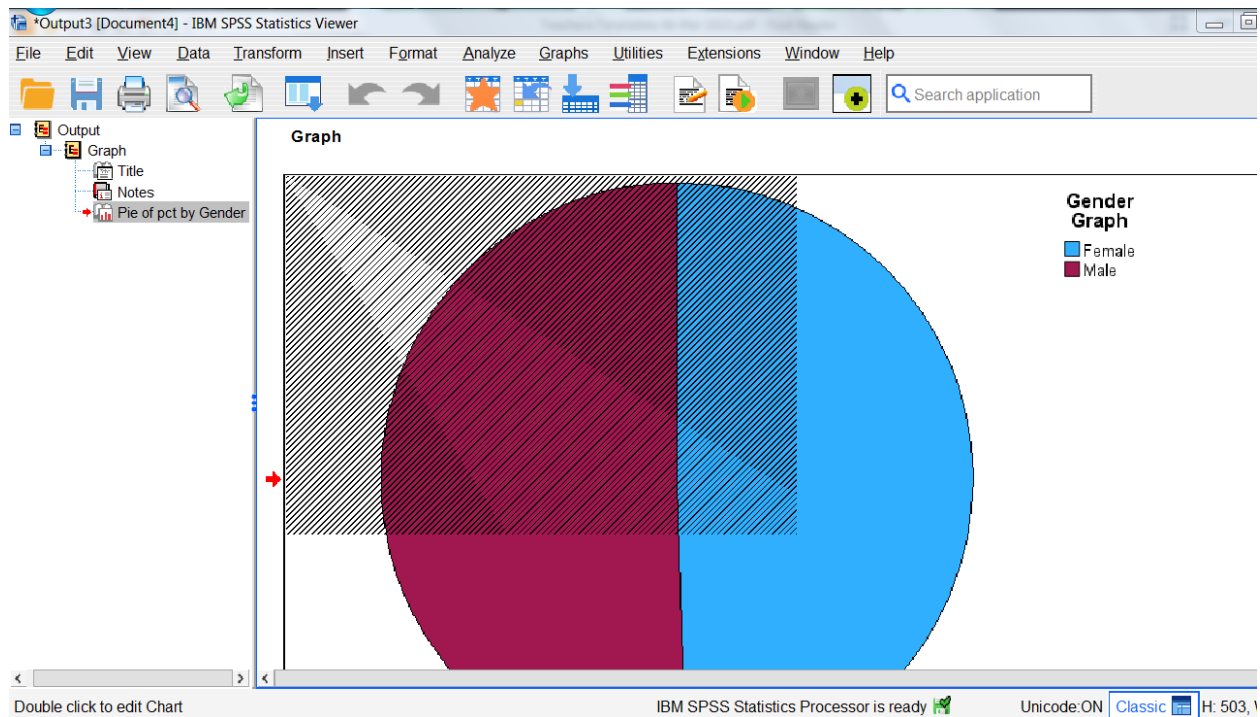
Value Label grid: [i](#) Enter or edit labels in the grid. You can enter additional values at the bottom.

	Changed	Missing	Count	Value	Label
1	☐	☐	1056	Female	
2	☐	☐	1077	Male	
3	☐	☐			

Cases scanned: 2133 Value list limit: 200

Copy Properties [From Another Variable...](#) [To Other Variables...](#) Unlabeled Values [Automatic Labels](#)

OK Paste Reset Cancel Help



- Students who qualify for reduced/ free meal are 918 and students who do not qualify for reduced/ free meal are 1215.

Student Score.sav [DataSet1] - IBM SPSS Statistics Data Editor

Define Variable Properties

Scanned Variable List

Un...	M...	Role	Variable
			Lunch

Current Variable: Lunch Label:

Measurement Level: Nomi... Suggest Type: String Width: 32 Minimize Width

Role: Input

Unlabeled values: 2 Attributes...

Value Label grid: Enter or edit labels in the grid. You can enter additional values at the bottom.

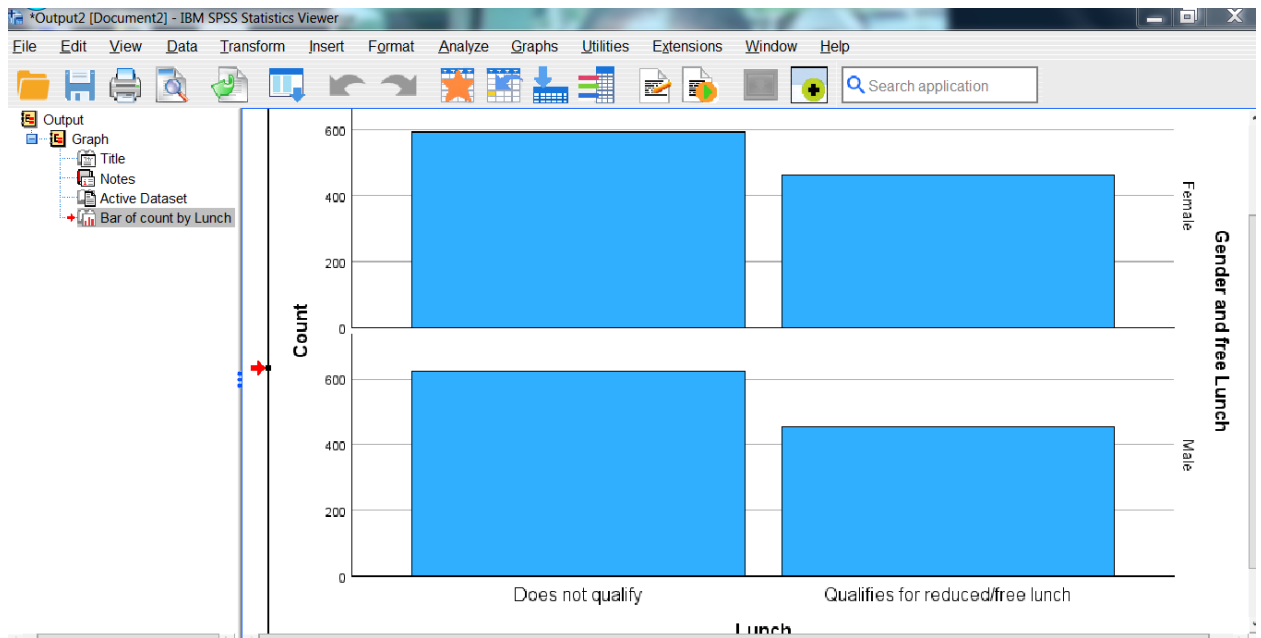
	Changed	Missing	Count	Value	Label
1			1215	Does not qualify	
2			918	Qualifies for reduced/	
3					

Cases scanned: 2133

Value list limit: 200

Copy Properties From Another Variable... To Other Variables... Unlabeled Values Automatic Labels

OK Paste Reset Cancel Help



3.

4. Mean for pretest and posttest results are 54.96 and 67.10.

Click the tabs in the procedure *Analyze>Descriptive Statistics> Descriptions*.

Descriptives

Variable(s):

- Pretest
- Posttest

Options... Style... Bootstrap...

Save standardized values as variables

OK Paste Reset Cancel Help

Pretest	Posttest
62	72
66	79
64	76
61	77
64	76
66	74
63	75
63	72
64	77
61	72
61	73
64	74
66	78
60	71
64	77
64	73
63	70
67	73

The output data is as below:

Case Processing Summary

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

Report

	Pretest	Posttest
Mean	54.96	67.10
N	2133	2133
Std. Deviation	13.563	13.987

Further analysis will yield.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Pretest	2133	22	93	54.96	13.563
Posttest	2133	32	100	67.10	13.987
Valid N (listwise)	2133				

- Clustering is basically used to quantitate data to show the spread of the scores. The K-means cluster for the pretest and posttest are as below.

*Output2 [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Extension

Output

- Quick Cluster
 - Title
 - Notes
 - Active Dataset
 - Initial Cluster Center
 - Iteration History
 - Final Cluster Center
 - Number of Cases in

Initial Cluster Centers

	Cluster 1	Cluster 2
Pretest	23	93
Posttest	35	100

Iteration History^a

Iteration	Change in Cluster Centers	
	1	2
1	30.191	33.606
2	.885	.791
3	.611	.532
4	.212	.181
5	.227	.200
6	.150	.128
7	.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 95.525.

*Output2 [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities

Output

- Quick Cluster
 - Title
 - Notes
 - Active Dataset
 - Initial Cluster Center
 - Iteration History
 - Final Cluster Center
 - Number of Cases in

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 95.525.

Final Cluster Centers

	Cluster 1	Cluster 2
Pretest	43	65
Posttest	55	78

Number of Cases in each Cluster

Cluster	1	2
Valid	993.000	1140.000
Missing		.000

6. Data and Text Mining methods that can be used in contemporary business.

Data Mining

Huge junks of data have been created following various research that have been carried out creating a need for specialized system software to analyze the data. Decision making has been reached quite often to meet the organizational needs through manipulation of the data on the software executions. The process of extracting these huge junks of data is called data mining. Techniques and algorithms have been developed to aid data mining and they include but not limited to regression, clustering, and neural networks.

Classification develops a model which groups the data at large by first pre-classifying them. Scam discovery and credit threat applications are principally fine matched to this sort of examination.

Goswami, *et al.* (2018) added that, this tactic normally hires neural network-based or decision tree cataloguing procedures. The statistics organization procedure comprises learning and grouping. In Learning the exercise statistics are examined by organization set of rules. In sorting trial statistics are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can best situate practical to the fresh statistics tuples. For a scam uncovering application, this would comprise whole registers of together false and lawful doings extracted on a record-by-record basis. The classifier-training procedure practices these pre-classified instances to regulate the usual limits compulsory for good judgement. The procedure then encrypts these strictures into a prototypical model named a classifier.

Clustering is generalized as identification of alike periods of items. By using clustering systems, we can supplementary recognize solid and scarce areas in entity interplanetary and can

realize general scattering outline and connections midst data characteristics. Organization method can also be cast off for operational way of differentiating clusters or programs of body but it turns out to be overpriced so clustering is utilized as pre-processing tactic for characteristic subgroup collection and cataloguing. For instance, to create a set of clienteles founded on buying patterns, to classes or genes with congruent purpose.

Regression method can be utilized for postulation. Regression analysis is basically utilized to type the association amongst one or more self-governing variables and reliant variables. In data mining self-governing variables are characteristics already recognized and response parameters to be predicted using the algorithm. Unluckily, many real-global glitches are not just forecast. For example, auctions dimensions, stock values, and product disaster rates are all challenging to forecast since they might rely on composite connections of numerous conjecturer variables. Consequently, more composite practices such as neural nets may be obligatory for prediction of upcoming standards. Connotation and association are typically to discover common article set conclusions amongst huge data arrays. This category of discovery benefits trades to create convinced verdicts, such as directory scheme, cross advertising, and client spending behavior examination. The rule of association ought to produce rubrics with sureness to classify standards.

A neural network is a customary grouping of linked input/output components, and every linking has a bulk existing with it. Throughout the knowledge stage, system studies by regulating masses to be capable to forecast the right class brands of the contribution tuples. Neural systems have the extraordinary aptitude to spring denotation from complex or vague statistics and can be castoff to excerpt designs and sense tendencies that are too composite to be observed by whichever persons or other processer performances. They are right for unceasing valued efforts and productions. For instance, handwritten info restructuring, for exercise a processer to articulate

English manuscript and numerous actual world professional glitches and have previously been effectively used in other companies.

Text Mining

Text mining has developed a significant investigation area. A huge amount of data kept in dissimilar places in formless assembly. This amorphous manuscript cannot be effortlessly used by processor for additional dispensation. Therefore, it is essential to employ the data mining process to have unstructured information to structured form. This involves: a) Gathering data from amorphous files, b) Translate this info acknowledged into organized statistics, c) Classify the design from organized data, d) Examine the arrangement, e) Excerpt the valued figures and storing in files as database. The techniques for text mining involve information retrieval (IR), information extraction (IE), categorization, clustering, and summarization (Amin, 2019).

Information retrieval uses the website to identify all the data linked to specified phrases or words. It is dignified as a leeway to article reclamation where the official papers that are reverted are utilized to excerpt the valuable material central for the consumer. Thus, file recovery is succeeded by a manuscript summarization phase that emphasizes the inquiry modelled by the handler. Retrieval of information in the larger logic entails the processing of text dealing with the whole. On the other hand, the objective of IE is the extraction of valuable info from manuscript. It classifies the drawing out of objects, proceedings, and relations from formless manuscript. Most valuable evidence such as designation of a person, setting and association are mined for correct consideration.

The categorization of text involves a keen supervisory in which the system allocates the development of each event or text in accordance with specific desired form. Categorization is the

consignment of standard dialectal leaflets to already established customary themes in relation to their scope. It is an assortment of manuscript pamphlets, the development of discovery of the correct subject or themes for every text. Tomasevic *et al.* (2020) stated that clustering aim is to discover inherent constructions in text and organize them into important subcategories for supplementary investigation. Clustering is valuable in data mining, image segmentation, retrieval of documents, and classification of designs to name a few. Lastly, summarization of texts produces a general overview or abstract of the files to be mined. Despite the techniques discussed herein, the major challenge poses and stands that all the texts to be retrieved must be stored or found in a readable format. “However, all of these methods have a common problem, that is, the quality of the development of classifiers is variable and highly dependent on the type of text being summarized.” (Souri, & Hosseini, 2018, p.12).

7. Advantages and Disadvantages of Excel and SPSS

Advantages	Disadvantages
1. Best for arranging and organizing a large set of information a into organized templates and tables. 2. Relaxed to digest and examine information. 3. Can generate data picturing like diagrams. 4. Can forward and deliberate data improved with C-Suite. 5. Restructures calculations using various functions. 6. SPSS delivers a stout set of tools that allows the institute to excerpt	1. Excel is not as precise in calculations involving huge or large sets of data. 2. Excel restricts perceptibility when creating composite models. 3. Excel is incompetent in managing models and data admission. 4. Limitations of Excel make keeping track of multiple spreadsheets challenging. 5. Microsoft Excel contains difficult to learn formulas. The IF,

actionable understandings from its statistics. 7. IBM SPSS tool empowers teachers to impart knowledge efficiently, helps scholars advance acute logical assistances, and it is a powerful tool in research presentation of data. 8. SPSS aids examination and improved comprehension of one's information and resolve composite occupational and study glitches through a comprehensible interface. 9. Assists In understanding huge and composite files sets rapidly with innovative statistical processes that aid safeguard high correctness and excellent making of decisions	VLOOKUP, and XLOOKUP function require manual entry which is at times complicated. 6. SPSS is a costly software the free student SPSS version can merely handle 1500 observations and 50 variables.
---	--

Excel may be regarded as the mother application of SPSS which is readily obtained from Microsoft Office Software. The ease of access to Excel gives it an upper hand in comparison to SPSS and any other statistical and financial analysis tools. The application is not complicated in the opening, sorting and entering of data; in fact, Excel is one of the major sources of format to export data to SPSS. Excel has shown great skills in optimization to maximize profits in an inbuilt tool called solver. This can be produced through the usage of a mathematical formulation of obtaining standard deviation which may at times produce inaccurate results.

References

- Amin, M.S., Chiam, Y.K. and Varathan, K.D., 2019. Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, 36, pp.82-93.
- Chang, C.M., Liu, L.W., Huang, H.C. and Hsieh, H.H., 2019. Factors influencing online hotel booking: Extending UTAUT2 with age, gender, and experience as moderators. *Information*, 10(9), p.281.
- Chattopadhyay, M. and Mitra, S.K., 2019. Determinants of revenue per available room: Influential roles of average daily rate, demand, seasonality and yearly trend. *International Journal of Hospitality Management*, 77, pp.573-582.
- Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A. and Chakraborty, B., 2018. A review on application of data mining techniques to combat natural disasters. *Ain Shams Engineering Journal*, 9(3), pp.365-378.
- Deyá-Tortella, B., Leoni, V. and Ramos, V., 2022. COVID-led consumption displacement: A longitudinal analysis of hotel booking patterns. *International Journal of Hospitality Management*, 107, p.103343.
- Jang, Y., Chen, C.C. and Miao, L., 2019. Last-minute hotel-booking behavior: The impact of time on decision-making. *Journal of Hospitality and Tourism Management*, 38, pp.49-57
- Mustika, I., Nabella, S.D. and Mulyadi, M., 2022. DATA-PROCESSING TRAINING USES A SPSS APPLICATION FOR IBNU SINA UNIVERSITY MANAGEMENT STUDENTS. *International Journal of Engagement and Empowerment*, 2(2), pp.179-183.
- Souri, A. and Hosseini, R., 2018. A state-of-the-art survey of malware detection approaches using data mining techniques. *Human-centric Computing and Information Sciences*, 8(1), pp.1-22.
- Syed, A.A. and Suroso, J., 2018. Factors affecting consumers' decision for E-Hotel booking. *CommIT (Communication and Information Technology) Journal*, 12(2), pp.111-123.
- Tomasevic, N., Gvozdenovic, N. and Vranes, S., 2020. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education*, 143, p.103676.
- .

- Theocharidis, A.I., Argyropoulou, M., Karavasilis, G., Vrana, V. and Kehris, E., 2020. An approach towards investigating factors affecting intention to book a hotel room through social media. *Sustainability*, 12(21), p.8973.
- Zhang, Z., Li, H., Meng, F. and Li, Y., 2019. The effect of management response similarity on online hotel booking: Field evidence from Expedia. *International Journal of Contemporary Hospitality Management*, 31(7), pp.2739-2758.