

Module 2: Exploratory Data Analysis on Assessment Records using Excel

By Associate Prof. Ts. Dr. Nurfadhlina Mohd Sharef Faculty of Computer Science and Information Technology Universiti Putra Malaysia <u>nurfadhlina@upm.edu.my</u>



Learning Outcomes and Outline

At the end of this module, participants will be able to:

- 1. Explain functions in Microsoft Excel for data analysis purpose
- 2. Use data exploration and visualization functions in Microsoft Excel functions to identify patterns and relationship in data

Outline								
I he outline of this presentation is as follows:								
1. Background								
2. Data Analysis Toolpak								
3. Questioning techniques								
4. Descriptive summary								
5. Conditional formatting								
6. Charting								
7. Recommended charts								
8. Analyze Data function								
9. Quick Analysis function								
11 COUNTIE								
12 Deducing Information								
IZ. Deducing information								

Background



Exploratory Data Analysis (EDA) refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

EDA is an iterative process comprising the following activities:

- 1. Generate questions about your data.
- 2. Identify patterns on the data to deduce information.
- 3. Search for answers by visualising, transforming, and modelling your data.
- 4. Use what you learn to refine your questions and/or generate new questions.

Various techniques can be used for EDA such as data preprocessing, descriptive statistics and visualization.

Tools such as Excel and PowerBI can be used to create dashboard visualizations. Programming languages such as Python and Java can be used to manipulate the data and spot trends.



Using Microsoft Excel for Data Analysis



Microsoft Excel is one of, if not the number 1 data analysis software in the world in terms of popularity mainly because its ease of use. It provides the ability to tabulate, analyze and visualize data to assist in data interpretation.

Chart, Formula, Analyze Data function, PivotTable, PivotCharts and Data Analysis Toolpak could support you to identify trends, patterns, and outliers in a data set, facilitating data analysis in seconds and empowering you to understand data through high-level summaries via Excel's Artificial Intelligence ability.

To visit tutorials in Excel, click the image below.



X Excel

Import and analyze data

Excel for Microsoft 365, Excel 2021, Excel 2019, Excel 2016, Excel 2013, Excel 2010

Import data	Tables	Sort and filter	Charts	PivotTables	Data models						
Available chart types in Office											
Create a chart											
Add a pie chart											
Present your data in a Gantt chart											
Create a waterfall	chart in Offi	ce 2016									
Create a histogra	m										
Change axis label	s in a chart										
Select data for a d	chart										
Add axis titles to	a chart										
Add a trend or m	oving averag	e line to a chart									

Using Microsoft Excel for Data Analysis (1)





Using Microsoft Excel for Data Analysis (2)





A pivot table is a **table of grouped** values that aggregates the individual items of a more extensive table within one or more discrete categories.

This summary might include sums, averages, or other statistics, which the pivot table groups together using a chosen aggregation function applied to the grouped values. Various charts can be created to **plot the trends** in the data. A sparkline is a tiny chart in a worksheet cell that provides a visual representation of data. Use sparklines to show trends in a series of values, such as seasonal increases or decreases, economic cycles, or to highlight maximum and minimum values.

Creating filters in the data allows analyst to focus on specific characteristics.

Using Microsoft Excel for Data Analysis (3)





Formulas are equations that can perform calculations, return information, manipulate the contents of other cells, test conditions, and more.

These **formulas** return a result, even when it is an error.

Among the formulas that are typically used by data analysts are:

- 1. SUM: to calculate total
- 2. COUNT: to identify frequency
- 3. AVERAGE: to get average value
- 4. MIN: to identify minimum value
- 5. MAX: to identify maximum value
- 6. LEN: to count the number of characters
- 7. IF: to check a condition
- 8. VLOOKUP: to refer a value in a vertical list
- 9. INDEX-MATCH: to lookup value dynamically
- 10. COUNTIF: to count how many matches the condition

Using Microsoft Excel for Data Analysis (4)





The sorting function allows data to be arranged according to increasing or decreasing order The filter function is to select some data based on certain characteristics The Text to Columns function **splits data according to occurrence** of certain characters

What-If Analysis is the process of changing the values in cells to see how those changes will affect the outcome of formulas on the worksheet. Three kinds of What-If Analysis tools come with Excel: Scenarios, Goal Seek, and Data Tables. Scenarios and Data tables take sets of input values and determine possible results.



Using Microsoft Excel for Data Analysis (5)



The Data Analysis Toolpak may be added to support for deeper data analysis exercises. To activate this, go to Home>Options>Add-ins.

cel Options			?
General	View and manage Microsoft Of	fice Add-ins.	
Formulas			
Data	Add-ins		
Proofing	Name 🗢	Location	Type
31/0	Active Application Add-ins		1.76-
ave	Analysis ToolPak	C:\Office16\Library\Analysis\ANALYS32.XLL	Excel Add-in
anguage	, analysis room an	estimornice to telefally that job that is bestee	
coossibility	Inactive Application Add-ins		
ccessionity	Analysis ToolPak - VBA	C:\ice16\Library\Analysis\ATPVBAEN.XLAM	Excel Add-in
dvanced	Date (XML)	C:\Microsoft Shared\Smart Tag\MOFL.DLL	Action
et a sella a Dilala a s	Euro Currency Tools	C:\root\Office16\Library\EUROTOOL.XLAM	Excel Add-in
ustomize Ribbon	Inquire	C:\soft Office\Office16\DCF\NativeShim.dll	COM Add-in
uick Access Toolbar	Microsoft Actions Pane 3		XML Expansion Pack
	Microsoft Data Streamer for Excel	C:\softDataStreamerforExcel.vsto vstolocal	COM Add-in
dd-ins	Microsoft Power Map for Excel	C:\ap Excel Add-in\EXCELPLUGINSHELL.DLL	COM Add-in
rust Center	Microsoft Power Pivot for Excel	C:\I Add-in\PowerPivotExcelClientAddIn.dll	COM Add-in
	Document Related Add-ins		
	No Document Related Add-ins		
			ſ
	Add-in: Analysis ToolPak		
	Rublicher Microsoft Office		
	Compatibility No compatibility infer	mation available	
	Compatibility: No compatibility infor	mation available	
	Location: C:\Program Files\Micro	osoft Office\root\Office16\Library\Analysis\ANALY532	XLL
	Description: Provides data analysis	tools for statistical and engineering analysis	
	Manage: Excel Add-ins	<u>G</u> o	



Data Analysis

 \times

?





Activity Time!

Now that you are familiar with the functions that can be used in Microsoft Excel for data analysis, let's do some activities using Microsoft Excel!

12 activities have been prepared for you, as basic learning analytics task. The remaining of this slide presentation contains step-by-step activities based on an example assessment records.



Example data 1: Assessment Record

	А	A B		D
1	StudentID	Test1-C3 (20%)	Asgmt1(10%)	Asgmt2(15%)
2	S1	10.00	6.17	13.80
3	S10	14.50	8.00	13.80
4	S11	14.50	9.50	13.20
5	S12	18.00	9.50	13.80
6	S13	10.00	9.50	15.00
7	S14	12.00	9.50	13.60
8	S15	17.50	9.50	14.20
9	S16	8.50	9.50	15.00
10	S17	13.50	9.50	11.40
11	S18	13.00	9.50	13.80
12	S19	13.50	9.50	12.00
13	S2	11.00	6.17	0.00
14	S20	4.00	9.50	13.20
15	S21	14.50	9.50	12.00
16	S22	4.50	9.50	13.80
17	S23	5.00	9.50	13.80
18	S24	9.00	9.50	14.00

Assessment record is one of the most common forms of data used for learning analytics because academic institutions usually keep the scores information. The scores are also used as the key performance indicator, not just for the learner, but also the instructor and the institution.

Therefore, typically some analytics mechanism is adopted such as using Excel template, developing a dedicated system or subscribing to a service.

The figure on the left shows a snippet of recorded scores by a batch of anonymised students in the PutraMOOC data analytics course. The columns are the assessment types and weight, while the rows are the scores for each student.

Download the **raw and processed assessment records using the activities in this file at** <u>http://putraoer.upm.edu.my/id/eprint/37</u>



Activity 1: Questioning techniques

Create questions as a driver of your exploration. Identify possible visualization and insights to be gained. Examples are as shown below.

Questions	Metrics	Visualization	Questions	Metrics	Visualization
What is the average, maximum, minimum, variance and stdev marks in each assessment?	Average, Max, Min, var, Stdev	Table	Is the performance of each student across assessment consistent?	Percentage	Table
How many students have scored more than the average score in each assessment?	Frequency	Bar chart	Which assessment has been easy or hard to be scored?	Percentage	Pie chart
How many students obtained more than 65% of each assessment's maximum score?	Frequency, Percentage	Bar chart	strength in terms of cognitive, psychomotor and affective skills?	Avelage	12



Activity 2: Descriptive summary

- Using the marks file, provide the descriptive summary: average, maximum (max), minimum (min), standard deviation (stdev) and variance (var). These values represent the distribution of the data.
 - a. Open the file
 - b. Go to Sheet 1
 - c. To get the average of values in column B, go to column B, row 45 and enter =AVERAGE(B2:B44)
 - d. To get the max among the values in column B, go to column B, row 45 and enter =MAX(B2:B44)
 - e. To get the min among the values in column B, go to column B, row 45 and enter =MIN(B2:B44)
 - f. To get the var among the values in column B, go to column B, row 45 and enter =VAR(B2:B44)
 - g. To get the stdev among the values in column B, go to column B, row 45 and enter =STDEV(B2:B44)



23.75

18.75

23.13

25.63

22.50

18.13

16.88

田 四 ---

24°C Rain ^ 0 10 6 41 ENG 11:45 PM

88.0000000 4.4000000

72.0000000 3.6000000

81,0000000 4,0500000

88.0000000 4.4000000

72.0000000 3.6000000

83.0000000 4.1500000

72.0000000

91 0000000

8.8000000 84.0000000 ######### ####### 18.16

7 2000000 87 0000000 ########## ####### 17 98

.0000000 ######### ####### 15.28

80.0000000 ######### ####### 17.80

87.0000000 ######### ####### 14.68

81.0000000 ######### ####### 8.65

INNANUAR RANNAR 15 69

7 2000000

8.1000000

8.8000000

Activity 3: Conditional formatting

Aut	Save 💽 🖸		2- 6- 6	8	anonymised marks - p	rocess.xlsx +	₽ Search (Alt)	+Q)	Nurfadhlina	Mohd Sharef NM	<i>lä</i> ee	- 0	×												
File	Home	e Inser	t Draw	Page Layou	ut Formulas D	Data Review	View Help	•			ස් s	hare 🛛 🖓 Comme	ents												
Pasta) X () ~	Calibri B I L	- 11 2 - ⊞ -	• A* A*	= = = ≫ . = = = = = =	한 Wrap Text 태 Merge & Cente	Number	3 6,%_%	Conditional F Imat as Cell Formatting * Hable * Styles *	Elinsert ∽ ∑ Delete ~ Format ~ ♦	Sort & Fin Filter * Sele	d & Analyze							Colo	or so	ales	s allo	ows	us to	
Clipb	oard 🕼		Font	ß	Aligr	nment	l⊊ N	umber 🛱		Cells	Editing	Analysis	^						see	data	a rep	ores	enta	tion	
A1		3 8	$\sqrt{-f_x}$	Studentl	D				Highlight Cells Rules				~								. '				
	A	в	с	D	E	F	G	н	Top/Bottom Rules	L		м	1						easi	ier v	isua	llγ, ι	usino	2	
St	ude Test1-	C3 (20 As	gmt1(10%) A	Asgmt2(15%)	carrying marks [/45]	project [100m] P	jk-CPS9 (5%) P	k-CPS3 (10%) Fi		Exercise 1-Conce	epts Exercise 2	2-Pre-processing												,	
2 S1	. 10.00	00000	6.1666667	13.8000000	29.9666667	91.0000000	4.5500000	9.1000000		1.0100	0000 18.13							-	traff	ic_lic	nht c	olor	' cod	ina	
3 51	.0 14.50	00000	8.0000000	13.8000000	36.300000	90.0000000	4.5000000	9.0000000 7	Data Bars	15.89	20.00								uan	IC-IIÇ	JIICO		COU	шg	
4 S1	1 14.50	00000	9.5000000	13.2000000	37.2000000	91.0000000	4.5500000	9.1000000			28.75							7		_	-			•	
5 S1	.2 18.00	00000	9.5000000	13.8000000	41.3000000	90.0000000	4.5000000	9.0000000	Color Scars		26.88														
6 S1	.3 10.00	00000	9.5000000	15.0000000	34.500000	81.0000000	4.0500000	8.1000000			26.88						•								
7 S1	4 12.00	00000	9.5000000	13.6000000	35.1000000	91.0000000	4.5500000	9.1000000 8	3		28.13														
8 S1	.5 17.50	00000	9.5000000	14.2000000	41.2000000	81.0000000	4.0500000	8.1000000	Ticon Sets		23.75		AudoSmin (a. 0. 6		anonymicarl marks - propose vis	w - 🔳	Q Search / Alt /			Martantillas	Mobel Sharef Day	/5 m	- 0 X
9 S1	.6 8.50	00000	9.5000000	15.0000000	33.0000000	81.0000000	4.0500000	8.1000000	7	\sim	18.13		Hallosere (anonymitted market protostation		yes search part	(G)		Huriourierio	WORD SHORE	<i>a</i> u	0 A
10 S1	7 13.50	00000	9.5000000	11.4000000	34.400000	83.0000000	4.1500000	8.3000000 8	New Rule	More Rules	25.00		File Ho	ome Insert	t Draw	Page Layou	it Formulas Data	Review	View Help					පි Share	Comments
11 S#	.8 13.00	00000	9.5000000	13.8000000	36.300000	90.0000000	4.5000000	9.0000000 8	Glear Bules	16.45	24.38		2 X	Calification	- la	Δ- Δ-	= = = 22		Munching		HER F		Insert ~ Σ	· A O	
12 St	9 13.50	00000	9.5000000	12.0000000	35.000000	74.0000000	3.7000000	7.4000000 8	3	18.61	26.25		Da -	, Caron	- 11		= com	rap sext	Number			34 HW	Delete -	- ZU P	
13 S7	11.00	00000	6.1666667	0.0000000	17.1666667	0.0000000	0.0000000	0.0000000	Manage <u>R</u> ules	19.55	27.50		v 🝼	BIU	! • ⊞ • .			erge & Center	r v 🜃 - %	· · · · · · · ·	Formatting * Tabl	le * Styles *	🖬 Format ~ 🛛 🔗	" Filter " Select "	Data
14 S7	4.00	00000	9.5000000	13.2000000	26.7000000	91.0000000	4.5500000	9.1000000	9.0000000 ######## #######	17.88	28.75		Tipboard	5	Font	5	Alignment		Ty Nu	mber 🕼	Styles		Cells	Editing	Anahsis /
15 S7	1 14.50	00000	9.5000000	12.0000000	36.000000	74.0000000	3.7000000	7.4000000 8	31.0000000 ######### ########	18.15	25.63														
16 S7	4.50	00000	9.5000000	13.8000000	27.800000	91.0000000	4.5500000	9.1000000 8	38.0000000 ######### ########	19.23	22.50	н	8		μ	8.1									
17 57	3 5.00	00000	9.5000000	13.8000000	28.300000	91.0000000	4.5500000	9.1000000 8	35.0000000 ######### ########	18.31	28.75	4	A	В	C	D	E	F	G	H	1 1	К	L	M	1
18 S7	4 9.00	00000	9.5000000	14.0000000	32.500000	72.0000000	3.6000000	7.2000000 8	38.0000000 ######### ########	18.75	25.63	1	Stude Te:	st1-C3 (20 Asg	mt1(10%) A	sgmt2(15%)	carrying marks [/45] project	t [100m] Pj	k-CPS9 (5%) Pjk	-CPS3 (10%) Fin	al [100m] Final [40%] Total	Exercise 1-Conce	pts Exercise 2-Pre	processing
19 S7	5 17.50	00000	9.5000000	14,4000000	41,400000	88.0000000	4,4000000	8.8000000 8	34.0000000 ######### ########	18.16	23.75	2	51 10	1 5000000	8.0000000	13.8000000	29,9666667 91.	0000000	4.5500000	9.1000000 78	8.0000000 #####		15.9100	50.00	
20 S?	6 13.00	00000	9.5000000	13.8000000	36.300000	72.0000000	3,6000000	7.2000000	74.0000000 ######### #######	15.20	18.75	4	S11 14	1.5000000	9.5000000	13,2000000	37.2000000 91.	.0000000	4.5500000	9.1000000 82	2.0000000 #####		19.75	28.75	
21 S	7 14.50	00000	9.5000000	13.8000000	37.800000	72.0000000	3.6000000	7.2000000	78.0000000 ######### ########	15.69	23.13	5	S12 18	3.0000000	9.5000000	13.8000000	41.3000000 90.	.0000000	4.5000000	9.0000000 83	3.0000000 #####		16.31	26.88	
22 ST	8 18.50	00000	9.5000000	13.6000000	41.6000000	91.0000000	4.5500000	9.1000000 8	37.0000000 ######### ########	15.28	25.63	6	513 10	0.0000000	9.5000000	15.0000000	34.5000000 81.	0000000	4.0500000	8.1000000 66	5.0000000 #####		17.93	26.88	
23 S?	9 15.50	00000	9.5000000	14,2000000	39,2000000	81.0000000	4.0500000	8.1000000 8	30.0000000 ######### ########	17.80	22.50	7	S14 12	2.0000000	9.5000000	13.6000000	35.1000000 91.	.0000000	4.5500000	9.1000000 82	2.0000000 #####	annu annuan	17.95	28.13	
24 S	20.00	00000	6.1666667	14,4000000	40.5666667	88.0000000	4,4000000	8.8000000 8	37.0000000 ######### ########	14.68	18.13	8	S15 1/	2.5000000	9.5000000	14.2000000	41.2000000 81.	.0000000	4.0500000	8.1000000 72	2.0000000 #####		18.24	23.75	
25 5	0 15.50	00000	9.5000000	13,8000000	38,800000	72.0000000	3,6000000	7.2000000 8	37.000000 #########	17.98	16.88	10	517 13	3.5000000	9.5000000	11.4000000	34,4000000 83.	.0000000	4.1500000	8.3000000 81	1.0000000 #####		18.95	25.00	
26 S	1 17.50	00000	9.5000000	11.4000000	38,400000	83.0000000	4.1500000	8.3000000 8	31.0000000 ##############################	8.65	25.63	11	518 13	3.0000000	9.5000000	13.8000000	36.3000000 90.	0000000	4.5000000	9.0000000 84	4.0000000 #####		16.45	24.38	
	Sh	eet1 St	neet2 (4	-)					- 4			12	\$19 13	8.5000000	9.5000000	12.0000000	35.0000000 74.	0000000	3.7000000	7.4000000 80	0.0000000 #####		18.61	26.25	
Bank										EEE (m	13	52 11	0000000	6.1666667	0.0000000	17.1666667 0.	0000000	0.0000000	0.0000000 0	0.0000000 0.000	00000 ========	19.55	27.50	
neady	-		-	100						ш .		14	S20 4	1.5000000	9.5000000	12.0000000	26.7000000 91.	0000000	4.5500000	7.4000000 81			18.15	25.63	
-	る型	-	C 0	K.					24°C	Rain ^ 📴 I		ENG 11:43 PM 16	522 4	1.5000000	9.5000000	13.8000000	27.8000000 91.	.0000000	4.5500000	9.1000000 88	8.0000000 #####		19.23	22.50	
												17	523 5	5.0000000	9.5000000	13.8000000	28.3000000 91.	0000000	4.5500000	9.1000000 85	5.0000000 #####		18.31	28.75	

7.5000000

23 529 15 5000000

24 53 20.0000000

26 531

14.400000

13.8000000

13,800000

13,6000000

14,4000000

9.5000000 14.2000000

17.5000000 9.5000000 11.4000000

25 \$30 15 5000000 9 5000000 13 8000000

Sheet1 Sheet2 (+)

비 오 비 🐂 😋 💁

41.4000000

36,3000000

37.8000000

41.6000000

39,2000000

38.8000000

38.4000000

40.5666667

Follow step 1 until 5 in the above figure. Click the corner between column A and row 1. And then, use the Color Scales option in the Conditional Formatting function.



Activity 4: Charting non-sorted and sorted values





Activity 5: Using Recommended Charts function

AutoSave 💽 🖪 🥠 🗸 🖓 🗸 🤿	anonymised marks - process.xlsx 👻	✓ Search (Alt+Q)	Nurfadhlina M	lohd Sharef N 🏑	ў в – о	×		
File Home Insert Draw Page Layou	Formulas Data Review	View Help			🖻 Share 🛛 🖓 Comm	nents		
PivotTable Recommended Table Tables	Add-ins v Recommended Add-ins v Add-ins	A A A A A A A A A A A A A A A A A A A	Line Ime Slicer Ime Column Ime Win/Loss Sparklines Filters	Link Comment	nt Text Symbols	^		
$x_1 \rightarrow i \times \sqrt{f_x}$ Studentic Insert Chart ? \times								
A B C D 1 Stude Test1-C3 (20 Asgmt1(10%) Asgmt2(15%) of 2 S1 4.000000 6.166667 13.800000 3 S10 4.500000 8.000000 13.800000 4 S11 5.000000 9.500000 13.800000 5 S12 8.500000 9.500000 13.800000 6 S13 9.000000 9.500000 13.600000 7 S14 10.000000 9.500000 14.200000 9 S16 10.000000 9.500000 14.200000 9 S16 10.000000 9.500000 13.800000 10 S17 10.500000 9.500000 13.000000 12 S19 11.000000 9.500000 13.200000 15 S21 12.000000 9.500000 13.800000 16 S22 12.000000 9.500000 13.800000 16 S22 12.000000 9.500000 13.800000 15 S21 12.000000		Clustered Column	Ci di vi gnt2(15%) Coss a few categories. Use it	L tercise 1-Concepts E 15.9100000 1 5.89 2 7.75 2 5.31 2 7.93 2 7.93 2 7.95 2 8.00 1 7.95 2 8.00 1 7.95 2 7.95 2 7.95 2 7.95 2 7.95 2 7.95 2 7.95 2 7.95 2 7.95 2 7.88 2 7.5 2 7.88 2 7.5 2 7.81 2 7.95 2 7.81 2 7.92 2 7.81 2 7.92 2 7.81 2 7.92 2 7.81 2 7.92 2 7.92 2 7.81 2 7.92 2 7.92 2 7.92 2 7.93 2 7.92 2 7.93 2 7.92 2 7.92 2 7.92 2 7.93 2 7.92 2 7.92 2 7.92 2 7.93 2 7.92 2 7.92 2 7.93 2 7.92 2 7.93 2 7.92 2 7.93 2 7.92 2 7.93 2 7.94 2 7.94 2 7.95 2 7.95 2 7.92 2 7	M xercise 2-Pre-processing 8.13 0.00 8.75 6.68 8.13 6.68 8.13 6.68 8.13 6.50 6.48 8.13 6.50 6.50 6.50 6.50 6.50 6.50 6.50 6.50			
Sheet1 Sheet2		[OK Cancel	.123 🖽 🗐	□ - <u> </u>	+ 100%		

Use the Recommended Charts function for some automatically generatable data plotting.

Activity 6: Using Recommended Charts function



Visualization choice matters!

Explore possible visualizations to identify the one that can best suit your need.

The clustered column shows the score in each assessment clearly, but to get information about the sum of some marks, stacked column is the best.

Meanwhile, the Line Chart makes it easy to compare each student's ranking compared to others in each assessment type.

We can also easily see which student has the highest, and the lowest mark.

Activity 7: Using Analyze Data function

AutoSa	ve 💓 日	9 • °	®,	anonymised r	marks - process (v	version 1).xlsb -	AutoRecovered 🝷	4	,O Nurfa	dhlina Mohd Sharef 🔞 🦨 🖬 – 🔿 🗙
File	Home In	sert Draw	Page Layo	ut Formulas Da	ta Review	View He	lp			남 Share 🖓 Comments
Ċì.	Calibri	~ 1	1 ~ A^ A*	≡≡∎ ≫~	ab c€ Wrap Text	Numb	per 🗸			$\begin{array}{c c} \blacksquare & \blacksquare $
Paste	B I	⊻ - ⊞ -	Ø - A -		🔁 Merge & Cen	ter - 🖙 -	% 9 50 30	Conditiona	I Format as Cell	Sort & Find & Analyze
Clichean	× ۲	Font		Aliana	oot	5	Number	Formatting	 Iable * Styles Styles 	Colle Editing Arabaia
Ciipboan		Fort	<u>د</u> ر در اید ا	-	en	14	Number 12		Jtyles	Cens County Analysis
Al	• :	×	Studentl	D						
A	В	С	D	E	F	G	н	1	J	K Daaluza Data
1 Stud	e Test1-C3 (20.	Asgmt1(10%)	Asgmt2(15%)	carrying marks [/45]	project [100m]	Pjk-CPS9 (5%)	Pjk-CPS3 (10%)	Final [100m]	Final [40%] Tota	Analyze Data
20 \$26	13.0000000	9.5000000	13.8000000	36.3000000	72.0000000	3.6000000	7.2000000	74.0000000	******	
21 S27	13.0000000	9.5000000	13.8000000	37.8000000	72.0000000	3.6000000	7.2000000	78.0000000	*******	🕂 🗠 Ask a question about your data
22 528	13.5000000	9.5000000	13.6000000	41.6000000	91.0000000	4.5500000	9.1000000	87.0000000	*****	
23 529	13.5000000	9.5000000	14.2000000	39.2000000	81.0000000	4.0500000	8.1000000	80.0000000		₩ Total 'Asgmt1(10%)' and total 'Asgmt2(15%)'
24 53	14.5000000	6.1666667	14.4000000	40.5666667	88.0000000	4.4000000	8.8000000	87.0000000		Q How many different 'StudentID' are there?
25 530	14.5000000	9.5000000	13.8000000	38.8000000	72.0000000	3.6000000	8.2000000	87.0000000		♀ Total 'Asgmt1(10%)' for 'StudentID' excluding 'S
20 331	14.5000000	9.5000000	12,000000	36,000000	74.0000000	2 700000	7 4000000	87.0000000		
28 533	14.5000000	9.5000000	15.0000000	37 5000000	81.0000000	4 0500000	8 1000000	83.0000000		Which fields interest you the most?
29 533	14.5000000	9,5000000	12 2000000	32,7000000	88,0000000	4.000000	8.8000000	84.0000000		
30 535	14.5000000	9,5000000	11.4000000	35,4000000	83.0000000	4.1500000	8.3000000	84.0000000		"# 'Test1-C3 (20%)', 'Asgmt1(10%)' by
31 \$36	14,5000000	9,5000000	13,6000000	39,1000000	91.0000000	4.5500000	9,1000000	75.0000000		" 'StudentID'
32 537	14,5000000	9.5000000	15.0000000	36.0000000	56.0000000	2.8000000	5.6000000	87.0000000		# 25.000000
33 \$38	15.0000000	9.5000000	15.0000000	39.0000000	81.0000000	4.0500000	8.1000000	84.0000000		# 20.000000
34 \$39	15.5000000	9.5000000	15.0000000	34.5000000	56.0000000	2.8000000	5.6000000	71.0000000		# 15,000000
35 S 4	15.5000000	6.1666667	13.2000000	31.3666667	91.0000000	4.5500000	9.1000000	78.0000000		10,000000
36 S40	15.5000000	9.5000000	15.0000000	39.5000000	81.0000000	4.0500000	8.1000000	82.0000000		
37 S41	16.0000000	10.0000000	13.0000000	34.0000000	87.0000000	4.3500000	8.7000000	71.0000000		#
38 S42	16.5000000	5.3333333	13.0000000	30.8333333	87.0000000	4.3500000	8.7000000	85.0000000		## StudentID
39 S43	17.5000000	5.3333333	13.0000000	30.8333333	87.0000000	4.3500000	8.7000000	76.0000000	*****	#
40 S5	17.5000000	6.1666667	14.4000000	36.0666667	88.0000000	4.4000000	8.8000000	82.0000000	##########	# + Insert Chart Is this helpful
41 S6	17.5000000	6.1666667	15.0000000	31.6666667	56.0000000	2.8000000	5.6000000	83.0000000	******	
42 S7	18.0000000	6.1666667	14.2000000	36.8666667	81.0000000	4.0500000	8.1000000	75.0000000	*****	#
43 S8	18.5000000	6.1666667	12.2000000	32.8666667	88.0000000	4.4000000	8.8000000	88.0000000	******	# 'Test1-C3 (20%)' and 'Asgmt1(10%)' by
44 \$9	20.0000000	6.1666667	12.2000000	32.8666667	88.0000000	4.4000000	8.8000000	87.000000	*****	₩ - 'StudentID'
	Sheet3	Sheet1 She	eet2 +			•				
Ready				Avera	ge: 11.6578811 C	ount: 176 Sum:	1503.8666667	We're starting th	e add-ins runtime, ju	st a moment 🌐 🔟 — — 🛨 — + 100
		2	o 🚺	R					💛 75°F N	Aostly cl ^ 🛱 🐿 焼 🕸 ENG 6:49 AM 💭



Simply select a cell in a data range, then select the Analyze Data button on the Home tab.

<u>Analyze Data</u> in Excel will analyze your data, and return interesting visuals about it in a task pane.

Choose any suggested ideas for data analysis listed or insert any readily made chart. You may also customize it using which field interests you the most. Analyze Data

Inc	lude fields	Summarize value by
	(Select All)	
	StudentID	Not a value
	Test1-C3 (20%)	Sum 🖣
	Asgmt1(10%)	Not a value
Asgmt2(15%)	Sum	
		Average

Update

18

Cancel



Quick Analysis function takes a range of data and helps you pick the perfect chart with just a few

Activity 8: Using Quick Analysis function

AutoSave 💽 🗂 🏷 🗸 🖓 👻 😓 anonymised marks - process (version 1).xlsb - AutoRecovered 🔹									
Fi	ile	Home In	sert Draw	Page La	yout Formulas	Data Review	/	View Help	
Piv	otTable	Recommende PivotTables	d Table Illu	ustrations	Get Add-ins	Recommended Charts	0 - ≪ - D -	■ · 『 · · @ [• · • • · · · · · · · · · · · · · · · ·	otChart
		Tables			Add-ins		Calib	rharte	TTTT 0/ •
F14	4	• :	× v	<i>fx</i> 91			B		
	А	В	с	D	E	F		G H	2 · .00 ->0 ·
1	Stude	Test1-C3 (20	Asgmt1(10%)	Asgmt2(159	%) carrying marks [/4	5] project [100r	X	Cut	ina
8	\$15	10.0000000	9.5000000	14.20000	00 41.20000	00 81.00000	De		72
9	S16	10.0000000	9.5000000	15.00000	33.00000	00 81.00000	LE	Сору	78.
10	S17	10.5000000	9.5000000	11.40000	34.40000	83.00000	Ĉ	Paste Options:	81.
11	S18	11.0000000	9.5000000	13.80000	36.30000	90.0000		r n	84.
12	S19	11.0000000	9.500000	12.00000	35.0000	00 74.00000			80.
13	S2	11.0000000	6.1666667	0.00000	17.16666	67 0.00000		Paste Special	0.
14	S20	11.5000000	9.5000000	13.20000	26.70000	00 91.00000	0	Constitution	79.
15	S21	12.0000000	9.5000000	12.00000	36.0000	00 74.00000	P	Smart Lookup	81.
16	S22	12.0000000	9.500000	13.80000	27.80000	00 91.00000		Insert	88.
1/	523	12.5000000	9.5000000	13.80000	28.30000	00 91.00000		Delete	85.
10	524	12.5000000	9.5000000	14.00000	32.50000	00 72.00000		Delete	88.
20	525	13.0000000	9.5000000	12 80000		00 72 00000		clear Contents	74
20	\$27	13.0000000	9.5000000	13,80000	0 37,8000	00 72.00000	23	Quick Analysis	78
22	528	13.5000000	9.5000000	13.60000	00 41.60000	00 91.0000		Sector Paralysis	87
23	\$29	13.5000000	9.5000000	14.20000	39.2000	00 81.00000		Filt <u>e</u> r	> 80.
24	\$3	14.5000000	6.1666667	14.40000	40.56666	67 88.00000		Sort	> 87.
25	S30	14.5000000	9.500000	13.80000	38.80000	00 72.00000	+7		87.
26	S31	14.5000000	9.5000000	11.40000	38.40000	00 83.00000	¢,	New Comment	81.
27	S32	14.5000000	9.5000000	12.00000	36.0000	00 74.00000	Ð	New Note	87.
28	\$33	14.5000000	9.5000000	15.00000	37.50000	00 81.00000		5	83.
29	\$34	14.5000000	9.5000000	12.20000	32.70000	00 88.00000	=	Eormat Cells	84.
30	\$35	14.5000000	9.500000	11.40000	35.40000	00 83.00000		Pick From Drop-down	List 84.
31	\$36	14.5000000	9.5000000	13.60000	39.10000	91.00000		Define Name	75.
1)	Sheet3	Sheet1 Sh	eet2 (+) 25.00000	55 19000		Denne Ngrne	07
Read	dy						ଡ	Link	> We'r

commands. Select a range of cells. Select the Quick Analysis button that appears at the bottom right corner of the selected data. Or, press Ctrl + Q .
<u>F</u> ormatting <u>Charts</u> Totals <u>Iables</u> <u>Sparklines</u>
Clustered Column
Recommended Charts help you visualize data.
14 11 11 11 11 11 11 11 11 11 11 11 11 1
<u>F</u> ormatting <u>Charts</u> <u>Totals</u> <u>Iables</u> <u>Sparklines</u>
Table PivotTable
Tables help you sort, filter, and summarize data.

19

Activity 9: Using Histogram to count frequency of marks of Test1 by a uniformed scale



Histogram is a visualization technique for supporting frequency analysis.

Add the Data Analysis Toolpak to perform Histogram analysis.

\frown						
1	Bin	Histogram		? ×		
Bronara tha	5.000000	Input	г			
Prepare trie	10.000000	Input Range:	\$B\$2:\$B\$44 1	OK		
scale for the	15.000000			Cancel		
bin	20.000000	<u>B</u> in Range:	\$B\$51:\$B\$55			
		✓ Labels		Help		
		Output options				
		Output Range:	\$H\$51			
		O New Worksheet <u>P</u> ly:				
		O New Workbook	Provide the v	ariahlas		
		Pareto (sorted histogram)	4 for creating t	ating the		
		Cu <u>m</u> ulative Percentage				
		✓ <u>C</u> hart Output	Histogram	1		





The table of answers will be displayed

		4.
Bin	Frequency	
5.000000	3	
10.000000	4	
15.000000	24	
20.000000	11	
More	0	
		Т

Activity 10: Using Histogram to find frequency of Test1 marks above average



			Bin	Histogram		?
Go to Data>Data Analysis		Prepare the	5.0000000 10.0000000	Input Input Range:	\$B\$2:\$B\$44	ОК
Choose Histogram		scale for the	13.0000000 15.0000000	<u>B</u> in Range:	\$B\$51:\$B\$56	Cancel
Data Analysis	? ×	placing the	20.000000	✓ <u>L</u> abels		<u>H</u> elp
Analysis Tools Anova: Single Factor Anova: Two-Factor With Replication Anova: Two-Factor Without Replication Correlation Covariance Descriptive Statistics Exponential Smoothing F-Test Two-Sample for Variances Fourier Analysis Histogram	OK Cancel Help	average as one of the bin variable		Output options	Provide the va for creating th Histogram	ariables e



2

3

The table of answers will be displayed. We can identify the number of students who obtained

marks above average

(average=13)

Bin	Frequency
5.000000	3
10.000000	4
13.0000000	12
15.0000000	12
20.000000	11
More	0

 \times

Activity 11: Using COUNTIF to find frequency of Final marks above average



1. The average value of final marks is **79.19**, which indicates that many student almost got grade=A (marks for grade=A is 80).

We can use this information to identify which students have gotten final marks below the average.

1. Enter this formula in cell N2:

=IF(K2>(AVERAGE(K2:K46)), TRUE, FALSE)

- 1. Then, copy this formula and paste into all cells in column K
- 2. Next, create a cell for counting the frequency of final marks that is larger than the average value of final marks. Enter this formula:

=COUNTIF(N2:N44,TRUE)

The result is 35, which indicates that 35 out of 42 students have gotten final marks above 35.

The instructor can use this **insights** to identify students who have gotten the high and low marks (can use average marks as a benchmark). The instructor can reflect the student's participation, and check whether there has been any comments by the students earlier on their learning problems. The instructor can also perform several additional analytics (using data analysis technique and visuals) to identify the gap among the students, and the trend of score by each student. These information can be used to strategize the next lessons and improve the curricula.

Activity 12: Deducing information

The figures on the right show the formula and the result of the descriptive summary. Based on this data, a few information may be deduced:

- a) In terms of average marks, Asgmt2>Test1>Asgmt1
- b) In terms of variance and Stdev, Test1>Asgmt2>Asgmt1

However, note that, THIS IS WRONG and we CANNOT compare across the assessments because their weight is different. Test1's weight is 20%, Asgmt1 is 10% and Asgmt2 is 15%.

This is why we need to STANDARDIZE the data so a fair comparison can be made.

In its current state, we can only report separately for each assessment. This limits a deep insight into understanding the student's performance.

Data standardization and normalization need to also be performed when we compare any two data sources. The next module will explain further on how to do this.

А		В		С	D			E	
StudentID	7 Test1-C3 (20%)) Asgm	t1(10%)	Asgmt2(15%)		carrying marks [/45]		
S4	12		6.166	66666666667	13.2	13.2		31.3666666666667	
S40	15		9.5		15		39.5		
S41	11		10		13	13		34	
S42	12.5 12.5		5.333	333333333333	13	13		33	
S43			5.333	333333333333	13		30.83333333333333		
S5	15.5		6.166	66666666667	14.4	14.4		36.0666666666667	
S6	10.5		6.166	66666666667	15	15		31.6666666666667	
S7	16.5		6.166	66666666667	14.2	14.2		36.866666666667	
S8	14.5		6.166	66666666667	12.2	12.2		67	
S9	14.5		6.166	66666666667	12.2	12.2		32.8666666666667	
Avg	=AVER	=AVERAGE(B2:B44)		RAGE(C2:C44)	=AVERAGE(D2	=AVERAGE(D2:D44)		14)	
Max	=MAX(B2:B44)	=MAX	((C2:C44)	=MAX(D2:D44)	=MAX(D2:D44)			
Min	=MIN(B2:B44)	=MIN	(C2:C44)	=MIN(D2:D44)		=MIN(E2:E44)		
Var	=VAR(I	32:B44)	=VAR	(C2:C44)	=VAR(D2:D44)		=VAR(E2:E44)		
Stdev	=STDEV(B2:B44)		4) =STD	EV(C2:C44)	=STDEV(D2:D4	=STDEV(D2:D44)			
		٨	D	C	D		E		
	1	A	D Tost1 C2 (20	C	D	corning			
	25	Stude	12 00	Asgmt1(10	Asgmt2(15%	carrying	21 27	_	
	30	54	12.00	0.17	13.20		31.37		
	36	540	15.00	9.50	15.00		39.50		
	37	S41	11.00	10.00	13.00		34.00		
	38	S42	12.50	5.33	13.00		30.83		
	39	S43	12.50	5.33	13.00		30.83		
	40	S 5	15.50	6.17	14.40		36.07		
	41	<mark>S6</mark>	10.50	6.17	15.00		31.67		
	42	S7	16.50	6.17	14.20		36.87		
	43	S 8	14.50	6.17	12.20		32.87		
	44	<u>s</u> 9	14.50	6.17	12.20		32.87		
	45	Avg	13.13	8.59	13.26		34.97		
	46	Max	20.00	10.00	15.00		41.60		
	17	Min	4.00	5 22	0.00		17 17		
	40	Vor	4.00	2.55	0.00 E.4E		21.92		
	48	var	12.68	2.41	5.45		21.82	23	
	49	Stdev	3.56	1.55	2.33		4.67		

41 42 43

44

45





Summary

Exploratory data analysis involves

- Understanding your variables
- Using visualization technique to dig into the data
- Cleaning your dataset
- Analyzing relationships between variables





Thank you!

By Assoc. Prof. Ts. Dr. Nurfadhlina Mohd Sharef Faculty of Computer Science and Information Technology Universiti Putra Malaysia <u>nurfadhlina@upm.edu.my</u>

