Google Sheets Tutorial Assignment

Due: April 2, 2025 @ 9 pm to D2L (Dropbox)

Knowledge (8)	Application Part 1 (10) Part 2 (6)	Thinking (8)	Communication Part 1(2) Part 2 (2)

This assignment is designed to introduce you to Google Sheets, an excellent tool for easily answering simple scientific questions, including those related to statistics and probability. It can also serve as a full-fledged scientific modeling platform when needed. The assignment will guide you through various features that Sheets offers to assist with statistical research and analysis.

For the Assignment…

Follow the instructions carefully. To obtain full marks for this assignment, you will require strict adherence to the instructions. All **BOLDED** characters inside the quotation marks are the commands you have to **TYPE** in(excluding the quotation marks). Please make sure you read carefully, as some of the instructions involve typing symbols and parentheses at the end or in between.
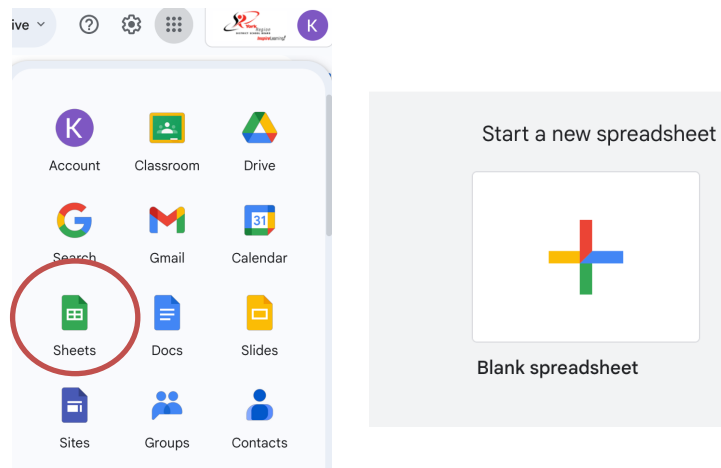
Are you ready?

.
.
.
.
.
.
.
.
.
.
.
.
.

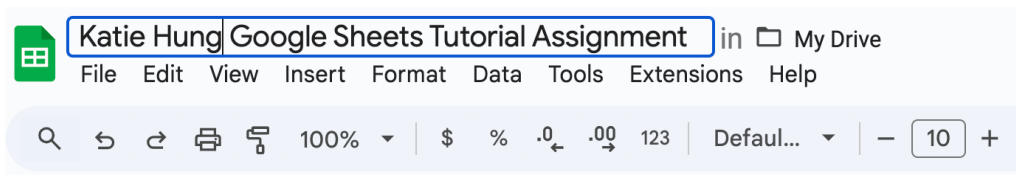Please begin your assignment on page 2.

Communication Part 1: [Total: 4 marks] (2 marks in this section)

1) Create a spreadsheet from Google Sheets

    a) Start a new spreadsheet from Google Sheets.

    b) Save your file name to "FirstName_LastName **Google Sheets Tutorial Assignment**" at the top left corner of the sheet.

    c) Click on the arrow from Sheet 1 tab at the bottom left and rename Sheet 1 to "**Knowledge**", Sheet 2 to "**Application Part 1**", Sheet 3 to "**Application Part 2**" and Sheet 4 to "**Thinking**".

Now you have completed Communication Part 1.
Part 2 of the **Communication** is at the end of this assignment.

Turn to the next page and begin Knowledge and Understanding section.

Now go to **Knowledge** Sheet and begin.

Knowledge & Understanding: [Total: 8 marks]

1) Make a **static list** of numbers. [2 marks]
   a) Click on cell A1 and type "**Number**".
   b) Click on cell A2 and type "**1**" and type "**2**" into cell A3.
   c) Select the range of cells from A2 to A3, move the cursor near the lower-right corner of cell A3. It should change from a hallow plus sign to a smaller, solid black plus "+" sign.
   d) While the cursor is the solid plus sign, click & hold the mouse and drag downward in column A to A31.
   e) You should get a list of numbers all the way to #30. (By selecting A2 and A3 cells, you are teaching Sheets to learn the pattern. When you select all the way to A31, Sheets is using the pattern it has learned from A2 and A3 to continue for the rest of the cells you have selected.
   f) Now go to column B and type **"Repeat"** in cell B1 and repeat steps a)-e) but change the pattern to "2" and "4" and the numbering pattern will show in an increment of 2 instead of 1. You should see "60" in cell B31.
   g) Next, you are going to learn how to number each cell in the column so that if you delete any one row, the number will automatically adjusted. In cell C1, type "**Row #**".
   h) In cell C2, type "**=ROW()-1**". You will see the number 1 appears in cell C2. This is because the Sheets picked up the command of the row number and use the row number and subtract one from it.
   i) Highlight cell C2 to C31, the press "Crtl + D" at the same time, it will copy the formula =row()-1 into all the cells highlighted. You should see cell C31 with the number 30 inside.
   j) Now, delete the entire row 24, you should be able to see column A and B do not have a continuation of the numbering system but column C is not being affected. Keep it the way it is right now and continue to instruction #2.

2) Make a list of **random** natural numbers. [2 marks]
   a) Click on cell D1 and type "**Random**".
   b) Click on cell D2 and type "**=RANDBETWEEN(1,10)**". This formula will instruct Sheets to make random numbers between 1 and 10 inclusively.
   c) Using one of the instructions of copying formulas from 1c) or 1i) to place the formula from 2b) to cells E2 to E30.
   d) Click on cell E1 and type "**Random 2**".
   e) Repeat the same steps from 2a) to 2c) but this time, generate a new set of random numbers **between 100 and 200** for Column D from E2 to E30. (You may find your column D numbers have changed which is normal because any formulas will trigger Sheets to re-generate a new set of random numbers, you will have to ignore the changes of generating random numbers throughout this assignment on these two columns)

3) Make a **calculated cell** and list. [4 marks]
   a) Click on cell G1 and type "**x**", H1 for "**y**", I1 for "**x-squared**", J1 for "**y-squared**", K1 for "**xy**" and then type the following numbers in the x and y columns.

G	H	I	J	K
x	y	x-squared	y-squared	xy
10	20			
4	12			
3	19			
9	10			
9	18			
2	13			
5	17			

b) Click on cell I2 and type "**=**" then press the left arrow on your keyboard. Note that you get a moving orange highlight, and the formula changes to a cell number as you move the highlight around. Move the highlight to cell **G2**, then type "*****", then highlight **G2** again, press Enter. The cell calculated value should be the product of x^2.

c) Use the black solid + sign to copy the formula for I3 to I8.

d) Repeat the same process for Column J to calculate y^2 and product of xy in Column K.

e) Click on cell G9 to generate the sum for all the x-values from G2 to G8 by typing "**=SUM(G2:G8)**".

f) Highlight cell G9 again and use the black solid + sign across to copy this formula to H9 to K9.

4) Use **IF** statements to choose values based on other values.

a) Think of a country (one word only) that you last visted: _____

b) Think of a course subject (one word only) you like the most in your high school years: _____

c) Click on cell F1 and type "**What IF**".

d) We are going to use the IF statement to choose values in column E that are smaller than 145 to be "your last visited country" and greater or equal to 145 will be "your favourite subject". On cell F2, type "**=IF(E2<145, "**type your last visited country**" , "**type your favourite subject**")**". Copy the formula for all the cells until it gets to cell F30.

5) Let's COUNT how many "your first name" and "your last name" in your list!

a) Click on cell E31 and type "**Count** your last visited country" and on E32, type "**Count** your favourite subject".

b) Click on cell F31 and type "**=COUNTIF(**" then highlight cell F2 to F30. Continue typing the rest of the formula by "**,** "your last visited country**")**" and press Enter. It will show you how many numbers are smaller than 150 on your list in Column E.

c) Click on cell F32 and count the number for "your favourite subject".

d) Now, click on cell F33 and use the sum formula to add up the number from F31 and F32 to confirm that you have a total of 29 counts.

Application Part 1: [Total: 10 marks]

In this section, you are going to generate a table to calculate several one-variable statistics.

The following data was recorded as the scores of a math contest:

105
105
104
103
99
98
98
97
95
95
89
88
88
86
84
84
83
77
76
76
75
73
73
72
71
68
67
67
64
63
54
53
42
23
22

6) Enter and Sort Data. [1 mark]

   a) Go to Sheet 2 tab, Application Part 1, at the bottom of the Sheets window.

   b) Copy and paste the table on the left on this page starting in cell A1 to A35.

   c) Select column A by clicking once at the label "A" at the top, you will see a down arrow box appears on the right-hand side of the "A". Click the down arrow box and choose Sort Sheet A →Z. This will help you to sort the data in column A from smallest to largest.

7) Calculate 1-Variable Statistics. (in this section, round all numbers to 3 decimal places using these two icons [.0←] [.00→] for column G so that whole numbers will be the values displaying as .000) **[4 marks]**

   a) In each cell from F1 to F11, type "**Mean**", "**Median**", "**Mode**", "**Sample Standard Deviation**", "**Q1**", "**Q3**", "**Max**", "**Min**", "**Variance**" "**90th Percentile**" and "**5th Percentile**" respectively. Go to the Capital letter F column and double click at the edge of the column width to adjust the width of the column to fit all your words inside.

   b) For Mean: in cell G1, type "**=average(**" select the data from column A1 to A35, type "**)**" then press Enter.

   c) For Median, in cell G2, type "**=median(**"select the data from column A "**)**" then press Enter.

   d) For Mode, in cell H3, type "**=mode.mult(**"select the data from column A "**)**", then press Enter. Since we have more than one mode, you need the space vertically down in the Sheets for it to display all the values. You should get 8 numbers to represent multi modes in this case.

   e) For Standard Deviation, in cell G4, type "**=stdev(**"select the data from column A "**)**" then press Enter.

   f) For Quartile 1: type in cell G5 "**=quartile.inc(**"select the data from column A "**, 1)**" then press Enter. Notice **.inc** presents numbers inclusively.

   g) For Quartile 3: type in cell G6 with the formula from Quartile 1 except, changing the number to "**3**".

   h) For Max: in cell G7, either use Quartile 4 or formula "**=max(**select the data from column A **)**".

   i) For Min: in cell G8, either use Quartile 0 or formula "**=min(**select the data from column A **)**".

   j) For Variance: in cell G9, use formula "**=var(**select the data from column A**)**".

   k) For the 90th Percentile: in cell G10, type "**=percentile.inc(**"select the data from column A "**, 0.90)**" then press Enter.

   l) For 5th Percentile: in cell G11, type "**=percentile.inc(**"select the data from column A "**, 0.05)**"

8) Create a Frequency Table. [3 marks]

    a) Create an interval chart using Column B (on cell B1 type "**Interval Start**"), C (on cell C1 type "**Interval End**"), D (on cell D1 type "**Midpoint**") and E (on cell E1 type "**Frequency**") to calculate the one-variable statistics.

    a) The two columns for the intervals (B and C) where one representing the value of the beginning of the interval and the next column is the value of the ending of the interval. It will be used to create the range of the interval. (i.e. Interval Start: 0 and Interval End: 10, means the interval 0-10)

    b) In order to show a frequency distribution for every 10 scores, you will need to define that for Sheets. The Frequency function refers to this limit as the "Bin Array". In cell B2 type the number "**20**". In the next cell below (B3), enter the formula: "**=B2+10**" where "B2" is the cell where you entered "20" and "+10" is the increment you want to set for each interval of your frequency distribution. After you type in the formula and press Enter, highlight cell B3 again, use the black solid + sign to copy the formula until the last number shows 110 scores.

    c) To calculate the ending value for each interval, in cell C2, type "**=B2+10**". highlight cell C2 again and copy the formula down to the same row you have stopped in Column B.

    d) Calculate the midpoint for each interval in column D. (Refer to Task 7 for how to calculate mean)

    e) Now, we are going to ask Sheets to prepare the frequency for each interval we have created. In the cell E2, type the formula "**=Frequency(**" then highlight the raw data cells in Column A to consider as the data array for the formula, then type a comma "**,**", then highlight values in Column C which is the Interval End to consider as the bin array for the formula, then type the closing bracket "**)**", then press Enter. Sheets will automatically complete the rest of the column for you. You will find an extra "0" in cell E12, ignore it.

    f) In cell E13, calculate the sum of the frequency.

9) To create a Histogram. [2 marks]

    a) Select/Highlight the data (A1 to A35), then go to Insert and select Chart and select Histogram.

    b) On the right side of the menu, choose Customize tab, select Chart & axis titles to label the Chart titles of the graph to "**Distribution of Math Contest Scores by** <u>type your first name as part of the title</u>" and then rename the two axes to **Scores** and **Frequency** by placing them to Horizontal axis title and Vertical axis title by selecting from the dropbox options.

    c) Remain on the Customize tab, go to Horizontal axis, adjust the minimum value to **20** and maximum value to **120**.

    d) Remain on the Customize tab, go to Histogram and manually change the bucket size to **10**.

    e) You may have noticed that your histogram is not showing the right value for the interval 20-30 as a frequency of 2. If that's the case in yours, double click the histogram to get into the setting of the graph on the right side of the screen. On the Set Up page, uncheck the box Use row 1 as header. It will correct the first bar of your histogram.

10)   Import Secondary Data from Statistics Canada. [1 mark]

a) Visit Statistics Canada website (http://www.statcan.gc.ca/eng/start)

b) On the top blue menu bar, go to Subject. This is the Statistics Canada's database for various subjects.

c) Select the subject Education, training and learning>Postsecondary education, you will see a list of tables available for you to select different statistics that have been collected by Statistics Canada. Search Canadian undergraduate tuition fees by field of study (current dollars): Table 37-10-0003-01. Click Add/Remove data to select specific statistics from the table before you download the file. For this assignment, you will have to do the following:

    a. For Geography tap: De-select "Canada", click the black "**+**" sign to expand, then select only "Ontario";

    b. For the tab of Field of Study: "Select all" except "Total field of study", "Optometry", "Agriculture, natural resources and conservation" and "Personal, protective and transportation services";

    c. For Reference period tap: choose from 2009/2010 to 2024/2025;

    d. In Customize Layout: select Field of Study as Row, then all the others as Column then click Apply.

d) The browser will reload to the table for things you have selected. Click "Download options", select CSV – Download as displayed (excluding accompanying symbols).

e) Once the .csv files (.csv file is a file format that is used to store database and allow majority of data management programs to recognize and import the data table) have been downloaded, add a new sheet after "Thinking" from the same spreadsheet of this assignment.

f) On this new sheet, no need to rename it.

g) Select File>Import, upload the csv file to the new sheet by selecting "Replace Current Sheet" from import location, Separator type as "Detect automatically", keep the box checked for "Convert text to numbers, dates, and formulas", then click "Import data". You should be able to cross-reference the table Cell A10 said "Field of Study" and Cell Q27 said "7,009".

h) Copy the table from the csv file sheet from cell A10 to Q27 and paste it to the Application Part 2 tab starting at cell A1.

i) Delete the entire row 2 where it said "Current dollars" not just the text in the cell.

j) Since the Reference Period has been imported as "text" 2009/2010, which cannot be identified as "year", we are going to change the value of our Reference Period now.

k) Select cell B1 and type "**2009**", then, in cell C1, type "**=**", then highlight cell B1, type "**+1**".

l) Move your cursor to the corner of cell C1 until you see the + sign, drag and copy the formula to cell Q1. You should be able to see cell Q1 with "2024".

11)	Create a bar graph [2 marks]]

a) Highlight the data from column Q including the Reference Period, 2024.
b) Insert Chart>Column Chart>Stacking: Standard>x-axis: column A1:A17.
c) Make sure check the box for "Use row 1 as headers" and "Use column A as labels".
d) Create a new title for the bar graph
e) Label the axes (Field of Study and Tuition $) of the graph appropriately using the method learned from Application Part 1, change the font size of the labels 36 pt.
f) Place the bar graph starting from column B under the table of data.

12)	Create a stacked Box-and-Whisker Plot [3 marks]

a) This section is testing your knowledge without step-by-step instruction.
b) You will be creating a table generating minimum, Q1, Median, Q3, and max in a table in the worksheet tab, Application Part 2, for the preparation of a stacked Box-and-Whisker Plot
c) To make it easier to view the columns after year 2022, highlight Column A, go to View>Freeze>1 column. Now, you should be able to scroll to column Q with the view of the first column, Field of Study.
d) Copy the fields of study from cell A5:A10 to R5:R10. You will be creating six box plots for the fields of:

Social and behavioural sciences, and legal studies
Law
Business, management and public administration
Physical and life sciences and technologies
Mathematics, computer and information sciences
Engineering

e) In cell S4 to W4, type "**Min**", "**Q1**", "**Median**", "**Q3**", and "**Max**" respectively.
f) Use the data from year 2009 to year 2024 for the above six fields to determine the minimum, Q1, Median, Q3, and Max to complete the table in cells S5:W10 using formulas learned in Application Part 1.
g) Highlight cells R5:W10, select Insert>Chart>Candlesticks chart to create a graph which includes all six box plots vertically displayed.
h) Make sure the label of the x-axis is field of study and y-axis is Tuition Fees.
i) Change the title of the graph to "**Box-and-Whisker Plots for Tuition Fees (2009-2024)**", center the title and change the font colour of the title in Red.
j) Double click on the scale of the y-axis, on the right-hand side of the screen, you will see the setup of the graph. Go to Customize>Vertical axis, adjust the scale to Min: "**4,000**" and Max: "**22,000**".
k) You will notice that the box plots are not having the correct values for Q3 and Max. This is because Candle Stick graph does not consider Q2. We will need to change the setup to include just Min, Q1, Q3, and Max. Double-click the graph to view the Setup page on the right-hand side of your screen, change the value range for Close which is Q3to "**V5:V10**", High to "**W5:W10**".

Thinking: [Total: 8 marks]

In this section, you are going to prepare an analysis for 2-variable statistics. We will investigate and analyze the cost of hotel and dinner in some major world cities.

13)   Create a Scatter Plot and generate the equation for a linear regression [3 marks]

a) The following table lists some major world cities. With each are cost in Canadian dollars of one night for a single room in a good mid-range hotel and the cost of dinner for one person including wine and a tip in the hotel restaurant.

Hotel Cost (in CAD$) - x	Dinner Cost (in CAD$) - y	City
237.3	107.9	Athens,
248.2	109.5	Caracas,
341.8	127	New Delhi,
335.9	106	Frankfurt,
458.6	191.1	Hong Kong,
260.4	325.2	Johannesburg,
289	86.2	Lisbon,
673.9	199.7	London,
308.1	95.6	Madrid,
818	272.7	Manila,
468.2	133.8	Mexico City,
222.2	35.9	Nairobi,
600	200	New York,
741.8	309.1	Paris,
464.1	149.7	Rio de Janeiro,
436.7	174.7	Rome,
506.9	190.1	Stockholm,
541.1	177.5	Sydney,
482.4	163.5	Tokyo,
502.6	137.8	Toronto,
397.7	106	Vienna,
458.9	137.7	Zurich,

b) Use the following table to copy and paste the data of the above table to worksheet tab: Thinking, starting at cell A1.

Hotel Cost (in CAD$)
237.3 248.2 341.8 335.9 458.6 260.4 289 673.9 308.1 818 468.2 222.2 600 741.8 464.1 436.7 506.9 541.1 482.4 502.6 397.7 458.9
Dinner Cost (in CAD$)
107.9 109.5 127 106 191.1 325.2 86.2 199.7 95.6 272.7 133.8 35.9 200 309.1 149.7 174.7 190.1 177.5 163.5 137.8 106 137.7
City
Athens, Caracas, New Delhi, Frankfurt, Hong Kong, Johannesburg, Lisbon, London, Madrid, Manila, Mexico City, Nairobi, New York, Paris,
Rio de Janeiro, Rome, Stockholm, Sydney, Tokyo, Toronto, Vienna, Zurich

*When you copy and paste, you will find that the data goes across in one cell. To create the table, you will have to do the following:

   i.   Highlight the cell that has the entire set of data for hotel cost in cell A3
   ii.   Go to menu Data>Split text to column
  iii.   The separator menu will show up with the arrow box for selection
  iv.   Choose Space to separate them into different cells
   v.   Then highlight all the data for hotel and Copy
  vi.   Go to cell L20 that is away from all your important data with enough space to prepare the data to be pasted in column
 vii.   Go to menu Edit>Paste special>Transposed and your data will be now organized in column
viii.   Repeat the steps for Dinner and paste them starting in cell M20.
  ix.   Repeat the steps for Cities except choosing Comma instead of Space at the separator and place them starting at cell N20
   x.   Label the titles of each column in cell A1: C1, A: Hotel Cost (in CAD$), B: Dinner Cost (in CAD$), C: Cities
  xi.   Delete any data in cells A2:V12
 xii.   Cut and paste the three columns in the order A: Hotel, B: Dinner, C: Cities starting cell A1 to C23

c)   Select the Hotel and Dinner columns, go to Insert and select the Chart>Scatter Plot.
d)   Go to Customize tab> Series menu, click the box Trendline to draw the linear regression line on the scatter plot.
e)   Click Label dropdown menu and choose Use Equation to display the equation of the linear regression on the graph.
f)   Check the box to Show R^2, note that you cannot change the value of R^2 on the graph to r so you need to manually calculate r-value on your own.
g)   Create an appropriate **Chart Title** and change the font size to 20pt, font type to Comic Sans MS, title format to Center, title text colour to Red and **BOLD** the title.
h)   Relocate the scatter plot to the right of the table.
i)   Now, you are going to calculate coefficient of correlation. In cell A24, type "**r-value**". In cell B24, type "**=CORREL(**", highlight the data in Dinner, then type a comma "**,**", then highlight the data in Hotel, then type the closing bracket "**)**", for the y-values and then x-values in the calculation, then press Enter.
j)   Round the r-value in cell B24 to 5 decimal places.

14)	Data Analysis [5 marks]

a) Given the set of data and the scatter plot in the thinking, answer the following questions in the cell(s) specified.

Questions:
1) Determine which city has the most expensive hotel cost.
   - i)   In cell A26, type "**Most Expensive Hotel**".
   - ii)   In cell B26, type "**=max(**" and highlight the data from the Hotel cost column, then type "**)**", press Enter.
   - iii) In order to find out which city has the most expensive hotel cost, we will use VLOOKUP and MATCH formulas together.
   - iv)   Copy and paste the title from the data set in cells A1: C1 to A27:C27.
   - v)   In cell A28, type "**=VLOOKUP($B$26, $A$1:$C$23, MATCH(A27,$A$1:$C$1,0),0)**", press Enter.

     *explanation:
     **$** - is to fix the location of the selected cell so that when you copy the formula to other cells, the selected cell with "$" will not be shifted
     $B$6 is to reference the value you need to look up
     $A$1:$C$23 is the range to look for the value
     MATCH(A27,$A$1:$C$1,0) is to look for the same title from the data set
     0 means exact match

   - vi)   Select cell A28, move the cursor to the right bottom corner of the cell until you see the + sign, drag to copy the formula to B28:C28.
   - vii) You should now see the most expensive hotel cost, along with dinner cost and the city name show up all together.

2) In the cells A30:C32, repeat the same process to determine which has the least expensive hotel cost?

3) In which city is the cost of dinner relatively expensive compared to the cost of hotel?
   - i)   In cell D1, type "**Dinner is more expensive than Hotel**".
   - ii)   In cell D2, type the "**=IF(**" formula that you have learned from Knowledge section to set the parameter to have dinner>hotel, then, the result needs to be using this symbol "✅" for true and the symbol "❌" for false.
   - iii) If you cannot copy and paste this symbol from this pdf file to Google Sheet, open Google Doc, insert a check from Insert>Special Character, search for check mark (✅) and cross (❌) to find the symbols there and copy to Sheet.
   - iv)   Repeat the steps for the rest of the cities.

4) If the cost of a hotel room in a particular city is $500, what would you expect the cost of a dinner to be?

    i)   Copy question #4 into cell A34.

    ii)   Highlight across from A34:D34 to merge the cells.

    iii)   In cell A35, use any methods of calculation involving Google Sheet formula to determine the answer for question 4.

    iv)   Then use cell A36 to type your concluding statement in full sentences.

    v)   Merge cell A36:B40, use Format>Text Wrapping>Wrap to put your conclusion in this merged cell. Center the conclusion in the middle of the box

***The length of your conclusion should not exceed the size of this box. It should just be a concluding statement that we normally have in a math question. i.e. Therefore, if the cost of the hotel room is…….. ***

i.e.

Your conclusion should be in the middle of the box, centered using the following icons from Google Sheet and look like this

d) On the menu bar, select File>Print. On the right-hand side of the menu, change the Print from Current Sheet to Workbook. That should make sure all five sheets in your assignment in one single pdf file.

e) Choose Landscape orientation, Letter size paper with Margin setting Normal.

f) Scale set Fit to Page. This will automatically adjust the scale of your sheets to fit in one page.

g) In Header & footer: check all boxes (Page numbers, Workbook title, Sheet name, Current date Current time)

h) Click EDIT CUSTOM FIELDS, type "your first name and last name" at the center box of the header, then type "**Ms. Hung**" in the center box at the footer.

i) Click Confirm.

j) Click Next.

k) Instead of printing it out in paper form, change the printer to save as PDF. This PDF file will be saved under your name "FirstName_LastName.pdf" as Original Work File.

l) Since D2L doesn't accept Google Sheet as a file for submission, you have to prepare a second pdf file for your assignment. Go to View>Show>Formulas, you should now see all your cells are showing the formulas instead of the final answers for all calculations.

m) Adjust the column width for all worksheets to make sure the formulas are clearly displayed in full.

n) Go to File>Download>PDF, and save this file under your name "FirstName_LastName View Formulas.pdf". Make sure you are saving all worksheets (which is the entire workbook) instead of just one worksheet. You will find that the symbols of ✅ and ❌ have gone missing in this pdf file. It is normal and you can ignore that and proceed to the next step.

o) Submit both of the PDF files (Original work file and Show formulas file) to D2L.

------ END OF THE ASSIGNMENT -----