You are expected to provide a complete answer for each question that brings in material from the lectures and handouts, where needed. You must also provide all relevant output from Excel that is needed to answer each question. Submit both the Excel file and the written report on Bb by Friday 11 pm.

**Optional HW – HW1~HW5**

Name:

**Part I. Basic Coding**

For Part I, use the excel file Optional HW data file, “*Sales\_2016-2020”*

1. Use the “**Text to Columns**” function to separate the email domains from the *customer\_email list* into column D. (\****Hint****:* Use the “@” as delimiters to separate the domain name)
2. Use **VLOOKUP** function to identify the *Item Sizes* based on *Item Weight*, using the reference table in “*Shipping Cost Reference” tab*. (\****Hint:*** Use “$” value)
3. Use **INDEX** function to identify the shipping cost based on the *Zone* and *Item sizes.* The reference table is in “*Shipping Cost Reference” tab*.

*(****\*Hint:*** You should use “*Item sizes*” column you created rather than the “*Item weight*” column)

1. Identify the following descriptive values for *“Total Page Views”, “Duration”, “Basket Size (Net\_sales),* *“Item weight”,* and “*Shipping Cost”* to get a better understanding of the customers.

* Mean
* Mode
* Minimum (min())
* Lower Quartile (=quartile())
* Upper Quartile
* Maximum
* IQR
* Standard deviation (=stdev())

1. Create a **Frequency table** for: *Frequency by Email domain* and *Shipping by city*

* Use the advanced filter function to identify the “Unique records”
* Use the **COUNTIF** function to count the frequency for each unique records
* Calculate the relative frequency to understand what proportion   
  (\**Hint:* Use =sum() function and percentage formula (n/total x 100)
* Use conditional formatting to highlight the top 3 in the list

1. Identify the ***Net sales*** and ***Total order number*** by Year (2016, 2017, 2018, 2019)

***(\*Hint***: Use **SUMIF** for Net sales; and **=COUNTIF** for Total order number)

1. Calculate the **percent change** in revenue by year. Compare the percentage from the previous year. So, 2016 vs. 2017, 2017 vs. 2018, 2018 vs. 2019. (This means, X17 cell will be empty)  
   ***(\*Hint:*** Percent change formula: Increase (=Newer value – Older value) ÷ Older value × 100
2. Add “**Sparklines**” for both Net-sales (Line Sparklines) and Total order number (Column Sparklines) covering 2016 to 2020 data in cell W15 and AA15.
3. Create a scatterplot for ***Net sales by Year*** and add a “***Trend line.”***

* Display the **trendline equation** on chart and **R-squared value** on all the chart.
* Change the Trendline color to orange (can be different shades of oranges colors)
* Edit title of the chart and the axis titles
* Edit axis value to appropriately display the associations

1. Use the identified ***Net Sales by Year*** trendline equation to

* (a) Calculate the predicted net sales from 2016 to 2021.
* (b) Calculate the error estimates (i.e., difference between the predicted value and actual value) for year 2016, 2017, 2018, and 2019.

1. Explain what the scatter plot tells you about for following criteria

* *Direction*. Does it trend up or down? Positive or Negative? No relationships?
* *Curvature*. Is the pattern linear or curved?
* *Variation*. Are the points tightly clustered around the trend?
* *Outliers*. Is there something unexpected? Any visible outliers? How would you explain the outliers? Can it be ignored?

1. Identify the **correlations coefficient** for the scatter plot. Based on the correlations, DESCRIBE the relationships between variables. (NOTE: simply stating the value of a correlation does not describe the relationship. Is there a relationship? Is it positive or negative? Is it *strong, moderate, or weak*?)

(\*Hint: Use the =correl())

1. Create a Pivot Table with “*Device*” and “*Years*” under [Rows] and “*Net\_sale*s” under [Values]. Put “*Device*” first, then “*Years*”

Using the “**Shows Values As**” for the “Net\_sales” Values, create the following 7 Value columns. Make sure to format the cell appropriately.   
(***\*Hint***: you CAN put the same field under the [Values] multiple times)

* Revenue (Net\_sales – no calculation)
* Percent of Grand Total   
  (% of Total Column - Grand total. Everything combined becomes 100%)
* Percent of Location Sub-Total   
  (% of Parent – by shipping city. Each shipping city becomes 100%)
* Percentage Difference (% Difference – based on the previous year. NOT the base year)
* Running Total (Running Total – by year)
* Rank (Rank – Large 🡪 Small)
* Data Bars  
  (Use Conditional Formatting > Date Bars. Make sure to remove “numbers”, only show bars)

1. Based on the results from Q1 – Q13, suggest **3 different strategic plans** you can propose based on the identified values above. (NOTE: Be specific and creative as possible. Don’t simply say need more users. Use the insights you gathered from the data)
2. Use the Pivot Table from Q13 to create a Pivot Table Dashboard

* Create 3 Pivot Charts of your choice.
* ***Insert Slicer*** for your Pivot Chart. Connect the Slicer/Timeline to all the Pivot Table.
* Click the inserted Slicer. “Slicer Tools” will pop up on the top. Under the “Options”, select “Report Connections.” Select all the Pivot Tables that are on the list.
* Create a Readable Dashboard using the 3 Pivot Charts and Slicer/Timeline you created.
  + Hide the Pivot Table in the sheet (right-click > Hide)
  + Add a Title box (Insert > Illustrations > Shapes)
  + Reposition the Charts and Slicer/Timeline Easy-to-read manner (make sure to “Align” to make it look professional)
  + Use Page Layout to fit into 1-page print size (Page Layout > Orientation > Landscape)
  + Scale to Fit 1-page

