Probability, Statistics, and Excel – Store Data Lab

**Goal**

Understand the mechanisms driving sales for Basic Sun Preparations, a skincare product.

**Data description**

Monthly sales data for a skincare product.

The dataset contains sales data for multiple stores across the nation, and includes information about the region where the store is located, store characteristics, product shelf space use, and store sales performance.

**Instructions:**

* Read the lab all the way through
* Complete the lab
* Clearly label all worksheets and submit relevant Excel workbooks, Word documents and PowerPoint slides

**Part I: Explore the dataset and practice Excel**

1. **Basic Excel operations**
   1. ~~Read the dictionary to understand the variables in the dataset.~~
   2. ~~Use the data filter to find quick answers to:~~
      * ~~# of store formats included in the data~~
      * ~~# of regions~~
      * ~~# of years of sales~~
   3. ~~Use the status bar to find quick answers to:~~
      * Average sales transaction value (TXNS) across all stores for all time
      * Total tax-included sales (SALES\_TISP) of Store #19 for all time
   4. ~~Create a copy of the raw worksheet, rename it (“Store Copy”), and use the new copy for the rest of problems~~
   5. ~~Perform the following tasks: (~~**~~Challenge: use keystroke shortcuts~~**~~)~~
      * Traverse the cursor to the beginning/end of a row/column
      * Copy and paste a column to the end of the data panel: try different ways to do this
      * Find and Find/Replace a cell value: Ctrl+F / Ctrl+H
   6. **~~Challenge:~~** ~~Use subtotal, hide detail, and sort (sort is in the data filter menu) to determine:~~
      * # of stores that don’t have 12 months of sales information
   7. Save the file, and then use “Save as” to save the file as a CSV file. Note the prompt and message. Open the Store Sales YOUR NAME.csv file and check out the difference between this file and the saved .xlsx file. Use the .xlsx file for the rest of this lab.

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

1. **Enter and edit data and worksheets**
   1. ~~Go to the worksheet “Store Copy” and add new columns with the formulas below:~~
      * + **Avg\_Price** = SALES\_TISP/SALES\_UNITS
        + **Space\_Yield** = SALES\_TISP/SPACE
        + **NDSA\_Yield** = SALES\_TISP/NDSA
        + **LOG\_TISP** = LOG(SALES\_TISP)
        + **LOG\_TESP** = LOG(SALES\_TESP)
        + **LOG\_Space** = LOG(SPACE)
        + **LOG\_NDSA** = LOG(NDSA)
        + **LOG\_Price** = LOG(Avg\_Price)
        + **LOG\_Space\_Yield** = LOG(Space\_Yield)
        + **LOG\_NDSA\_Yield** = LOG(NDSA\_Yield)
        + **Date** = Date(Year,Month,1)
   2. ~~Use an empty cell and enter the following (try with keystroke shortcut):~~
      * + Current date
        + Current time
        + Current date and time
   3. ~~Practice selecting multiple rows/columns using Ctrl+🡪, Shift+🡪, Ctrl+🡨, Shift+🡨, etc.~~
   4. ~~Use find and replace (shortcut) and replace the values in “Region” with alias:~~
      * 1 🡪 “South”
      * 2 🡪 “Pacific”
      * 3 🡪 “Mid-Atlantic”
   5. Save the file

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

1. **Format data, use formulas, and use functions**
   1. Format all of the columns in the Store Copy sheet such that they have the following criteria:
      * ~~Numerical values have 2 decimal digits~~
      * ~~Integers such as SPACE do not have decimal digits~~
      * ~~Sales related columns use the accounting format: $XX,XXX.XX~~

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Avg\_Price | Space\_Yield | NDSA\_Yield | LOG\_TISP | LOG\_Space | LOG\_NDSA |
| $ 4.91 | $ 75.36 | $ 1.75 | $ 7.21 | 2.89 | 6.66 |
| $ 4.71 | $ 70.12 | $ 1.62 | $ 7.14 | 2.89 | 6.66 |

1. Go to a new worksheet and practice the following

* ~~Type in Monday and generate Tuesday – Sunday sequentially in the same row/column using Auto Fill~~
* ~~Type in numbers 1, 2 in two adjacent cells in the same row/column and generate the numbers 3-20 in the same row/column using Auto Fill~~

1. Use functions and formulas to build a report for the following metrics as in the table below. There descriptive statistics should be calculated for each store format.
   * + SALES\_TISP
     + NDSA

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Convenience** | **Chemist** | **Destination** |
| Total |  | Hint: Use SUMIF() and AVERAGEIF() |  |
| Average |  |  |  |
| Max |  |  |  |
| Min |  |  |  |
| Cutoff value for highest 1% | Hint: Use AGGREGATE() function |  |  |
| Cutoff value for lowest 1% |  |  |  |
| Count how many values are in the highest 1% |  |  |  |
| Count how many values are in the lowest 1% | Hint: Use COUNTIFS() function |  |  |

**PAUSE HERE, CHECK ANSWERS WITH CLASS, BRING FORWARD ANY QUESTIONS**

* 1. Remove the outliers identified from above; create a cross-tab to track the total sales (SALES\_TISP) by store format and month

|  |  |  |  |
| --- | --- | --- | --- |
| **Month** | **Convenience** | **Chemist** | **Destination** |
| **Jan** |  |  |  |
| **Feb** | Hint: Use PivotTable |  |  |
| **Mar** |  |  |  |
| **Apr** |  |  |  |
| **May** |  |  |  |
| **Jun** |  |  |  |
| **Jul** |  |  |  |
| **Aug** |  |  |  |
| **Sep** |  |  |  |
| **Oct** |  |  |  |
| **Nov** |  |  |  |
| **Dec** |  |  |  |

SUM SALES\_TISP

FORMAT

MONTH

* 1. Format the sales report to be attractive and easy to read.
  2. Use “Conditional Formatting” on data in “Store Copy” worksheet to highlight the outliers for the columns as shown above
  3. Save the file
  4. More reading for Excel filling: [http://www.techrepublic.com/bLOG/10-things/10-advanced-formatting-tricks-for-excel-users/](http://www.techrepublic.com/blog/10-things/10-advanced-formatting-tricks-for-excel-users/)