Cleaning, Wrangling and Analyzing Data using Excel Spreadsheets

-Sreya Ambrose

Objectives of the project

1.Cleaning and Wrangling of Data using spreadsheets

- Remove duplicate data, inaccurate data, and empty rows in Excel.
- Resolve inconsistencies in data.
- Manipulate and standardize data using the Flash Fill and Text to Columns features in Excel.

2. Analyzing Data using spreadsheets

- Describe the fundamentals of analyzing data using a spreadsheet
- Filter and sort data in a worksheet.
- Employ some of the most useful Excel functions for data analysis.
- Implement the VLOOKUP and HLOOKUP functions to reference data.
- Create pivot tables in Excel and utilizing its features

Software used: Excel for the web

Dataset used:
https://dataplatform.cloud.ibm.com/exchange/public/entry/view/f8ccaf607372882403a37d9019b3abf4

1.Cleaning and Wrangling of Data using spreadsheets

1.1 Removing Duplicated, Irrelevant or Inaccurate Data

Check Spelling

Select the desired column, go to **Review** Tab and select **Spelling,** click the correct suggestion to change the spelling

Remove empty rows

Select the whole datasheet and apply **Filter** to filter only the CUST_NAME column. Select all the blank columns and then delete them. Remove the filter.

Remove Duplicate rows

Select column ORDER_ID as it has the unique values, go to **conditional Formatting** and highlight the **Duplicate Values** using the **Highlight cell rules.** Now go to **Remove Duplicates**, select all columns and delete the duplicate values.

Find and Replace to correct misspelling

In order to replace jcb in the entire table to JCB, first select all jbc with **Find All** and replace it to JCB using **Replace All**

1.2 Dealing with inconsistencies in Data

Use the PROPER () function to change text from upper case to proper case

It is used for converting the headings from upper case to proper case

Use the UPPER () function to change text from proper case to upper case

It is used for converting the values in AG1 from proper case to upper case

Use the LOWER () function to change text from proper case to upper case

It is used for converting the values in AC from proper case to lower case

Change the date formatting

It is performed using the **Date** option in the **Number Format.** Date format with (*) symbol is selected as it displays the correct regional date format because of the chances of the data to get shared internationally.

Use Find and Replace to trim Whitespace

Find the areas with 2 spaces and replace them with single space using the **Replace** in **Find & Select**

1.3 Advanced Excel features for Cleaning Data

Use Flash Fill feature to clean Data:

The values *Mr.* In Column C and *Allen Perl* in Column B is combined as *Mr. Allen Per*l in Coulmn A using this feature

Use LEFT , RIGHT,LEN and SEARCH Functions to clean Data

Split the Customer Name into Customer First Name using **=LEFT(D2, SEARCH(" ",D2,1))** and Customer Last Name using **=RIGHT(D2,LEN(D2)-SEARCH(" ",D2,1))**

2. Applying useful functions for Data Analytics

Use IF function to apply one condition

Create column *Complete?* At AF and applied the function IF(AG2="complete","Yes","No") to have just 2 values for the column

Use of nested IF to apply multiple conditions

Create a column called *Order Size* base on the value in the column *Order values* using the formula *=IF(AF2>300,"Large",IF(AF2>100,"Medium",IF(AF2>0,"Small")))*

Use of IFS to apply multiple conditions(alternative of Nested IF)

IFS(AF2>300,”Large”, AF2>100,”Medium”, AF2>0,”Small”). This formula is applied to the cell *AE3*

Use of COUNTIF to count the number of cells that meet a specified criterion

Use the formula *=COUNTIF(O2:O195,"VISA")* to find the total number of VISA cards at column *BY2*

Use of SUMIF function to sum the values within w specified range that meet a specified criterion.

Use the formula *=SUMIF(AE2:AE195,"Large",AF2:AF195)* to find the total order value of *Large* orders

Use of SUMIFS function to sum thevalues within a specified range that meet multiple specified criteria

Use the formula *=SUMIFS(AF2:AF195,AE2:AE195,"Large",AL2:AL195,"*BABY_BOOMERS*")* to find the total order values of *Large* orders of *Baby_Boomer* Gens

Using VLOOKUP and HLOOKUP functions

VLOOKUP function is used inorder to obtain the customer contact information. Here a dropdown is created to select the cyustomer using the Data Validation. Address for the selected customer is generated using the function *=VLOOKUP(BV8,A2:BM195,6,FALSE)* similarly, city and phone number is also generated.

Pivot Table

Pivot Table has been created to analyze the order values and order size of Customers, filtered by the state