IBM Training

## Demonstration 1

Working with Functions

At the end of this demonstration, you should be able to:

- Use some of the functions learned to work with and extract data in a sheet




*Demonstration 1: Working with Functions*

# Demonstration 1: Working with Functions

**Purpose:**
**You will work with the workbook that was created from the WordCount program. You will add new columns with specific functions to the sheet.**

User/Password: **biadmin/biadmin**
**Root/dalvm3**
Service Password: **ibm2blue**

## Task 1. Background.

The WordCount application generated a record for each unique character string found in the specified document. It then totaled the number of occurrences of each unique character string. Your goal is to create a sheet that has the number of occurrences for all character strings that occur the same number of times.

1. In **BigSheets** select the **Wordcount** workbook.

   Displayed for each row is a character string, followed by a tab character, followed by a number. The number represents the occurrences of that character string in the *last_of_the_mohicans.txt* document.

   Your goal is to count the number of character strings that occur the same number of times.

   In the previous exercise, you used Function sheets to apply functions to your data. This time, you are going to add new columns to a sheet and code the needed functions directly.

## Task 2. Create a new workbook.

1. From within the **Wordcount** workbook, click **Build new workbook**.
2. Click **Add Sheets** and then select **Formula**.

   Keep the default sheet name.

   You are to create a new sheet and specify a formula that references the *Header* column from the Wordcount workbook. The result is to have a new column called Occurrences. Then, use the *GETGROUPMATCH()* function to populate this new column. Here is the format of that function as shown in the BigInsights Knowledge Center. *GETGROUPMATCH (text,regex,group,number*)

You are provided the regular expression to use. You know that you want the second group. But what about the text parameter? You want this function to be applied to each row for the Header column. To indicate that, you specify **#Header**. As in:

GETGROUPMATCH(#Header,'(.+\t)([0-9]+)',2)

You are to reference the Wordcount sheet and, as stated, this new column is to have the name of **Occurrences**. The final parameter, 2, says to extract the second group from the regular expression.

3. Code the following in the **fx** field.

```
Wordcount!A1 : [Occurrences =
GETGROUPMATCH(#Header,'(.+\t)([0-9]+)',2) ]
```

4. Click **Apply settings**.

## Task 3. Count the occurrences.

Add another column to your sheet.

1. Click the drop-down for **Occurrences** and then select **Insert Right->New Column**.
2. Type **Num** for the name.

   You want to take the value for each row in the Occurrences column and select all rows that have that same value. Then count the number of those rows returned.
3. In the **fx** field for the **Num** column, type the following formula:

```
COUNT(SELECT(Occurrences,#Occurrences).Occurrences)
```

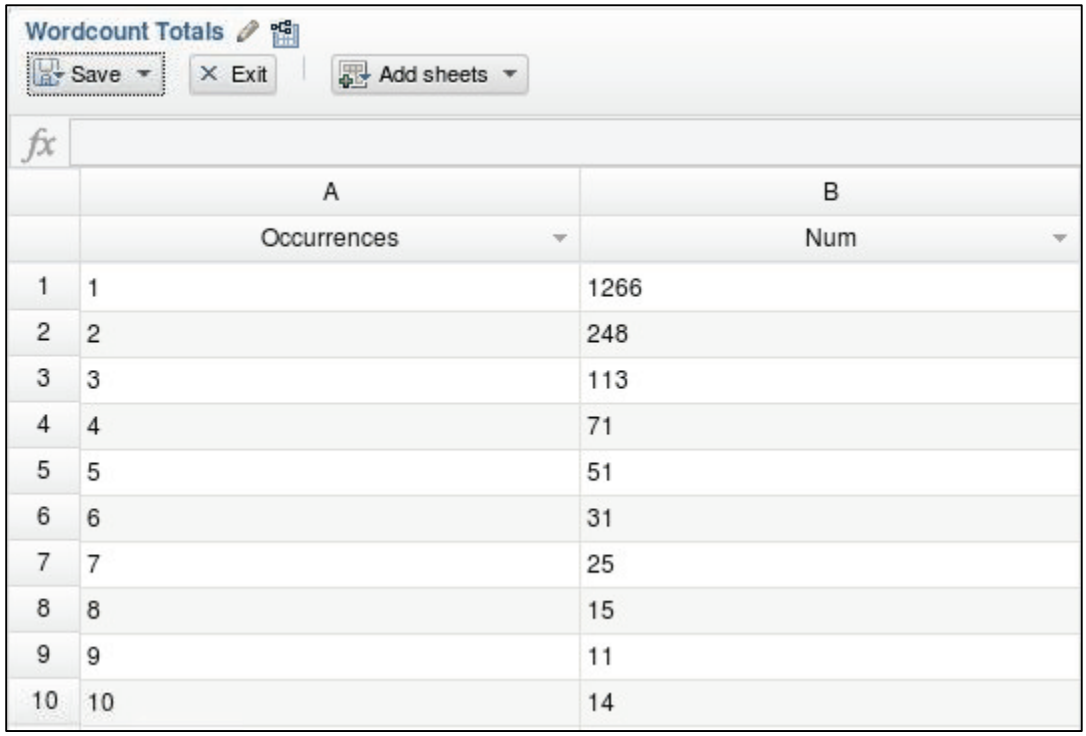
4. Click **Save formula**  .

   You can see that there were 1266 unique character strings that only occurred once in the document. But you do not want to see that 1266 times. You need to remove duplicate rows.
5. Make sure Sheet1 is selected, add a new sheet and then select **Distinct.**

6. Keep the default name for the new sheet and click **Apply settings**.

   Now you have your results. There are 1266 unique character strings that occurred once. Fourteen unique character strings that occurred 10 times, etc.

   The results appear as follows:




| | A | B |
|---|---|---|
| | Occurrences | Num |
| 1 | 1 | 1266 |
| 2 | 2 | 248 |
| 3 | 3 | 113 |
| 4 | 4 | 71 |
| 5 | 5 | 51 |
| 6 | 6 | 31 |
| 7 | 7 | 25 |
| 8 | 8 | 15 |
| 9 | 9 | 11 |
| 10 | 10 | 14 |

7. Click **Save**, **Save and Exit**, and name the workbook **WordCount Totals**.
8. Click **Save**.

**Results:**
**Using the workbook that was created from the WordCount program, you added new columns with specific functions to the sheet.**