**Data Preparation**

1. Create a new column that has combined page descriptions for given pair under control and test category. Name this column as Pair\_page

For example :



Here we are combining 2 strings as they belong to pair1

1. SD:gp:mens-sweatshirts(category control)
2. SD:gp:mens-short-sleeve-shirts(category test)
3. We will need to write a formula using if conditions as follows –
   1. If values for next row of pair column is matching then concatenate values of page column for two rows else keep it blank
   2. For example, if cell number for pair is b3 and b4 then formula will be



IF(B3=B4,CONCATENATE(C3,"/",C4),"”)

1. Now do copy+ paste special and remove the formula present in Pair\_page column
2. Since pair 26 is present in both category and CLP group, pair 26 appears 4 times. And page descriptions of category test and CLP control gets combined



Ideally, row highlighted in red should have been blank. But since pair 26 is same for row 2 and 3 even though CLP group starts, we need to manually delete the highlighted row

1. Next step is to select the entire Page\_pair column. Then ctrl+g and go to special. Now select blanks. Refer to the immediate prior cell(for example, first blank is at D4 then refer to D3) and press ctrl+enter. Again do paste special and remove the formulas from Pair\_page column
2. Date columns are starting with number. In SAS, column names must start with alphabet. Hence rename the date column names starting with text. For example, “Dt\_24-03-2015”
3. Now save this file as .csv or .txt . This file is ready to be imported on SAS
4. Open the SAS script and import the data. Make sure import is successful and complete data is loaded on SAS
5. Separate the data for category and CLP. Transpose the data in order to get it in the format that can be used for t-test
6. Rename the columns and export the data. Repeat the same steps for CLP group and other metrics
7. This completes data preparation for t-test

**Running t-test**

1. Open R studio and import the file prepared in SAS
2. Run the t-test in r. T-test that we will run is “two sided”
3. Our null hypothesis and alternate hypothesis are as follows



1. Now compare the results. Interpretation of the results will be as follows –
   1. If p value is less than 0.05 then we reject the null hypothesis and conclude that there is a significant difference between control and test group
2. For final results, prepare the table as follows –



1. Groups where there is no significant difference are highlighted in red as they will not be considered in the analysis

**Projection**

1. Determine the unique number of values in Page\_pair column. Forecast control mean and test mean for additional number of pages using current number of pages as base
2. For example, for 11 unique values in Page\_pair columns, if we get the mean values as 1.83 and 3.27 for control and test respectively. Then forecast the mean expected for pages let’s say 100,500,1000,1500,2000 etc



1. Also plot the line chart for the same