



SAS University Edition Challenge

Day 2 of 5 – Tuesday, 17th July 2018

Challenge Overview

You and your friend (let's call him) Kenny, have decided to start an online website that aggregates movie and TV show data, streamed by Netflix.

You have conducted some preliminary analysis on the Netflix data, but it doesn't look quite right. You notice the data only includes releases from 1940 to 2015. Titles with later releases are not in the data set.

Kenny explains there were some issues with the data and he sends you the data containing the 2016 releases in a csv file. You'll need to import the csv file into SAS as a data set.

Let's get the 2016 releases into SAS for further analysis.

Summary of Skills Demonstrated

- Importing a delimited file
- Recoding values
- Generating a Pie Chart

Submission Details

Submissions close **Monday, 23rd July 2018, 12:00pm.**

Submit your answers [here](#).

Be sure to use your SAS account when you submit your solution. If you don't use your SAS account, we can't put your entry in the draw!

If you haven't registered yet, it's not too late. Click [here](#) to register now.

Also, make sure to save your answers and tasks/code somewhere safe as a backup.

SAS University Edition

You'll need to download and install SAS University Edition to complete this challenge. Use the links below to install and set up SAS University Edition.

SAS University Edition download link:

https://www.sas.com/en_au/software/university-edition/download-software.html

How do I create a shared folder in VirtualBox?

https://support.sas.com/software/products/university-edition/faq/shared_folder_virtualbox.htm



Guided Exercises


Before we start

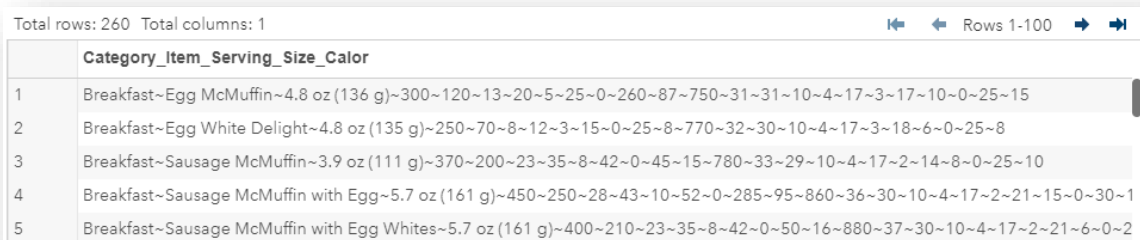
Make sure you've downloaded the resources for today's challenge from [here](#), placed the data in your shared folder repository and launched SAS University Edition.

Importing Delimited Files:

When working with data from various sources, you'll find that it can be structured differently and fields can be delimited using a wide array of special characters or symbols.

SAS is built on powerful algorithms that allow you to import data that has been formatted and structured in a wide variety of ways.

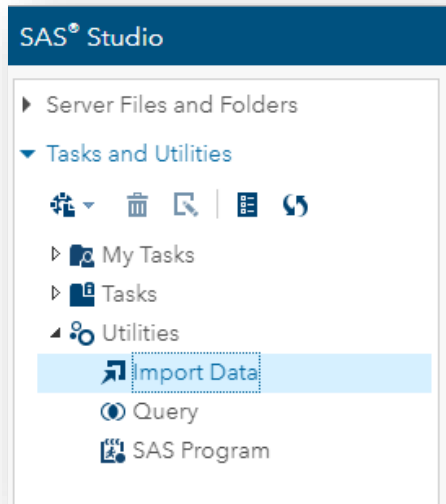
1. Open the **mcdonalds_menu.csv** in a text editor.
You'll notice that the fields or values are delimited by the hash symbol ("~"). For SAS to read the file in properly, we need to specify the character separating the values.
2. In **SAS Studio**, expand **Server Files and Folders** in the left pane.
3. Double-click **mcdonalds_menu.csv** to open the file as an **Import Data** task and click **Run**  .
Click the **Output Data** tab to view the result.



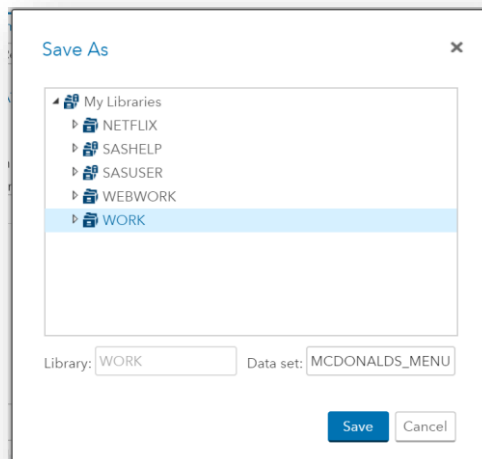
	Category_Item_Serving_Size_Calor
1	Breakfast~Egg McMuffin~4.8 oz (136 g)~300~120~13~20~5~25~0~260~87~750~31~31~10~4~17~3~17~10~0~25~15
2	Breakfast~Egg White Delight~4.8 oz (135 g)~250~70~8~12~3~15~0~25~8~770~32~30~10~4~17~3~18~6~0~25~8
3	Breakfast~Sausage McMuffin~3.9 oz (111 g)~370~200~23~35~8~42~0~45~15~780~33~29~10~4~17~2~14~8~0~25~10
4	Breakfast~Sausage McMuffin with Egg~5.7 oz (161 g)~450~250~28~43~10~52~0~285~95~860~36~30~10~4~17~2~21~15~0~30~1
5	Breakfast~Sausage McMuffin with Egg Whites~5.7 oz (161 g)~400~210~23~35~8~42~0~50~16~880~37~30~10~4~17~2~21~6~0~2

You'll notice there is only one variable where the values for each line have been squashed into.

4. Close the **mcdonalds_menu** tab without saving.
5. Expand **Tasks and Utilities**, then expand **Utilities** and double-click on **Import Data** task to open it.



6. Expand **Server Files and Folders**, drag `mcdonalds_menu.csv` onto the drop-zone of the **Import Data** task.
7. Under the **Output Data** heading, click the **Change** button.
8. Save the output data as `MCDONALDS_MENU` in the `WORK` library.



9. Under the **Options** heading, select **DLM (Delimited file)** from the drop-down menu.
10. Type '~' in the **Delimiter** input box and 260 (number of observations) in the **Guessing rows** input box.



▼ OPTIONS

File type:
DLM (Delimited file) ▼

Generate SAS variable names

Delimiter:
~


Quote delimiter value

Start reading data at row:
Default

Guessing rows:
260

11. In the **Code** tab you can see the SAS code that has been generated by the Import task which will be used to import the delimited file.

```
CODE LOG RESULTS OUTPUT DATA
Line #
5
6 %web_drop_table(WORK.MCDONALDS_MENU);
7
8
9 FILENAME REFFILE '/folders/myfolders/UE Challenge/mcdonalds_menu_2018.csv';
10
11 PROC IMPORT DATAFILE=REFFILE
12     DBMS=DLM
13     OUT=WORK.MCDONALDS_MENU;
14     DELIMITER="~";
15     GETNAMES=YES;
16     GUESSINGROWS=260;
17 RUN;
18
19 PROC CONTENTS DATA=WORK.MCDONALDS_MENU; RUN;
20
21
22 %web_open_table(WORK.MCDONALDS_MENU);
```

12. Click **Run**  and view the results in the **Output Data** tab.



CODE LOG RESULTS OUTPUT DATA

Table: WORK.MCDONALDS_MENU View: Column names Filter: (none)

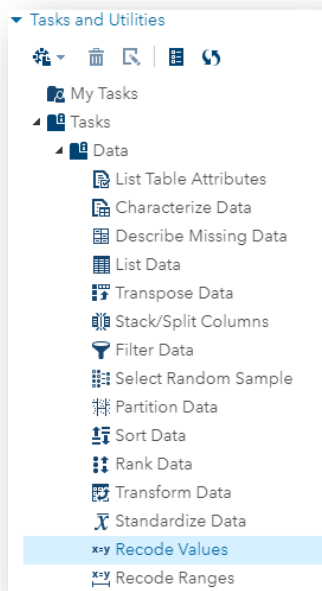
Total rows: 260 Total columns: 24 Rows 1-100

Category	Item	Serving_Size	Calories	Calories_from_Fat
1 Breakfast	Egg McMuffin	4.8 oz (136 g)	300	120
2 Breakfast	Egg White Delight	4.8 oz (135 g)	250	70
3 Breakfast	Sausage McMuffin	3.9 oz (111 g)	370	200
4 Breakfast	Sausage McMuffin with Egg	5.7 oz (161 g)	450	250
5 Breakfast	Sausage McMuffin with Egg Whites	5.7 oz (161 g)	400	210
6 Breakfast	Steak & Egg McMuffin	6.5 oz (185 g)	430	210
7 Breakfast	Bacon - Egg & Cheese Biscuit (Regular Biscuit)	5.3 oz (150 g)	460	230
8 Breakfast	Bacon - Egg & Cheese Biscuit (Large Biscuit)	5.8 oz (164 g)	520	270
9 Breakfast	Bacon - Egg & Cheese Biscuit with Egg Whites (Regular Biscuit)	5.4 oz (153 g)	410	180
10 Breakfast	Bacon - Egg & Cheese Biscuit with Egg Whites (Large Biscuit)	5.9 oz (167 g)	470	220
11 Breakfast	Sausage Biscuit (Regular Biscuit)	4.1 oz (117 g)	430	240
12 Breakfast	Sausage Biscuit (Large Biscuit)	4.6 oz (131 g)	480	280
13 Breakfast	Sausage Biscuit with Egg (Regular Biscuit)	5.7 oz (163 g)	510	290
14 Breakfast	Sausage Biscuit with Egg (Large Biscuit)	6.2 oz (177 g)	570	330
15 Breakfast	Sausage Biscuit with Egg Whites (Regular Biscuit)	5.9 oz (167 g)	460	250
16 Breakfast	Sausage Biscuit with Egg Whites (Large Biscuit)	6.4 oz (181 g)	520	280
17 Breakfast	Southern Style Chicken Biscuit (Regular Biscuit)	5 oz (143 g)	410	180
18 Breakfast	Southern Style Chicken Biscuit (Large Biscuit)	5.5 oz (157 g)	470	220

Recoding Variable Values:

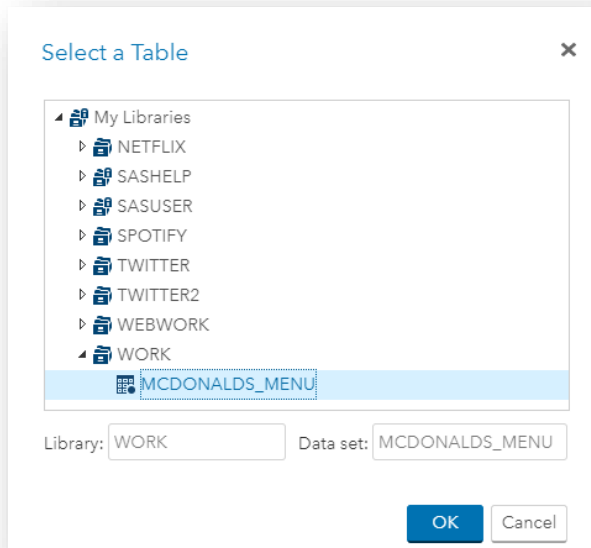
Often you may find that a variable has values that are not in the form you would like for your analysis. These values can easily be modified in SAS by using SAS code or by simply using the Recode Values task in SAS Studio.

1. Expand **Tasks and Utilities**, then expand **Data** and double-click on **Recode Values** task to open it.

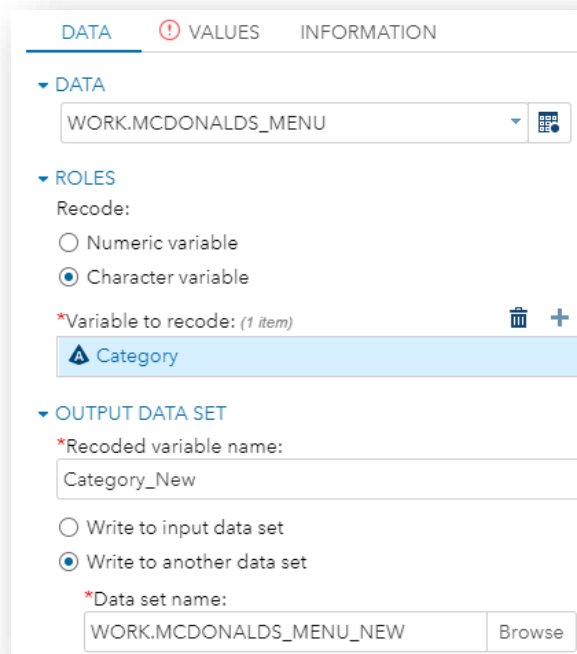




2. Choose the data set **MCDONALDS_MENU** in the **WORK** library as the Data source.



3. Change the **Recode** option to **Character Variable** and set the **Variable to recode** as **Category**.
4. Set the **Recoded variable name** to **Category_new** and write the data to a new data set called **MCDONALDS_MENU_NEW** in the **WORK** library.





5. Now click the **Values** tab to specify the values in the below table for the recoded variable **Category_New**. Look out for typos. Any existing values for **Category** which aren't specified here (such as **Breakfast**) will have their values passed on to the new variable.
Note: "Smoothies & Shakes" is the Old Value for value 5 in below screenshot.

Category	Category_New
Beef & Pork, Chicken & Fish	Main Food
Beverages, Coffee & Tea, Smoothies & Shakes	Drinks
Desserts, Salads, Snacks & Sides	Side Food

DATA **VALUES** INFORMATION

▼ VALUES


Recode values: (minimum 1 row)

Old value	New value
Beef & Pork	Main Food
Chicken & Fish	Main Food
Beverages	Drinks
Coffee & Tea	Drinks
Smoothies & Sha	Drinks
Desserts	Side Food
Salads	Side Food
Snacks & Sides	Side Food



- In the **Code** tab you can see the SAS code that has been generated by the Recode Values task which will be used to create the new recoded Category variable. SAS has used a **SELECT-WHEN** statement to execute this recoding.

```
CODE LOG RESULTS OUTPUT DATA
Line #
1 /*
2 *
3 * Task code generated by SAS Studio 3.71
4 *
5 * Generated on '27/06/2018 11:19'
6 * Generated by 'sasdemo'
7 * Generated on server 'LOCALHOST'
8 * Generated on SAS platform 'Linux LIN X64 2.6.32-696.20.1
9 * Generated on SAS version '9.04.01M5P09132017'
10 * Generated on browser 'Mozilla/5.0 (Windows NT 10.0; Win6
11 * Generated on web client 'http://localhost:10080/SASStudi
12 *
13 */
14
15 data WORK.MCDONALDS_MENU_NEW;
16 length Category_New $ 18;
17 set WORK.MCDONALDS_MENU;
18
19 select (Category);
20 when ('Beef & Pork') Category_New='Main Food';
21 when ('Chicken & Fish') Category_New='Main Food';
22 when ('Beverages') Category_New='Drinks';
23 when ('Coffee & Tea') Category_New='Drinks';
24 when ('Smoothies & Shakes') Category_New='Drinks';
25 when ('Desserts') Category_New='Side Food';
26 when ('Salads') Category_New='Side Food';
27 when ('Snacks & Sides') Category_New='Side Food';
28 otherwise Category_New=Category;
29 end;
30 run;
```

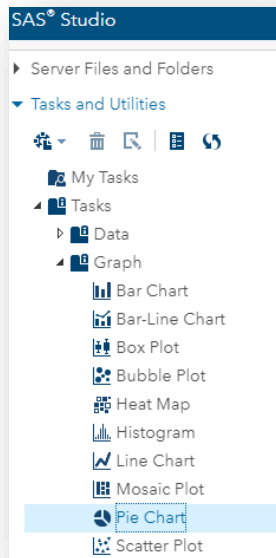
- Click **Run**  and view the results in the **Output Data** tab.



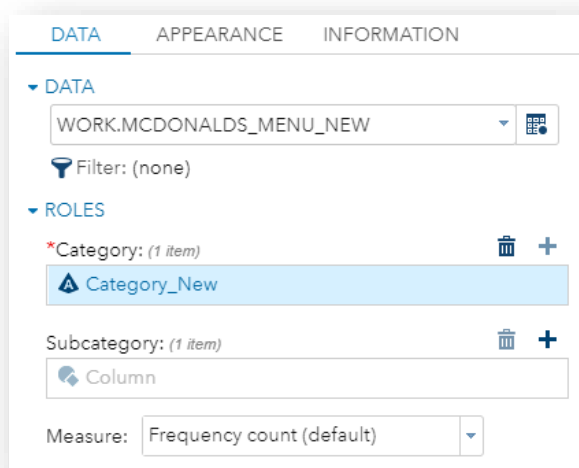
Graphing with Pie Chart:

Graphing can be easily achieved in SAS Studio using the provided Graph tasks. Below we will demonstrate how to create a Pie Chart using our new recoded Category variable created above.


1. Expand **Tasks and Utilities**, then expand **Graph** and double-click on the **Pie Chart** task to open it.

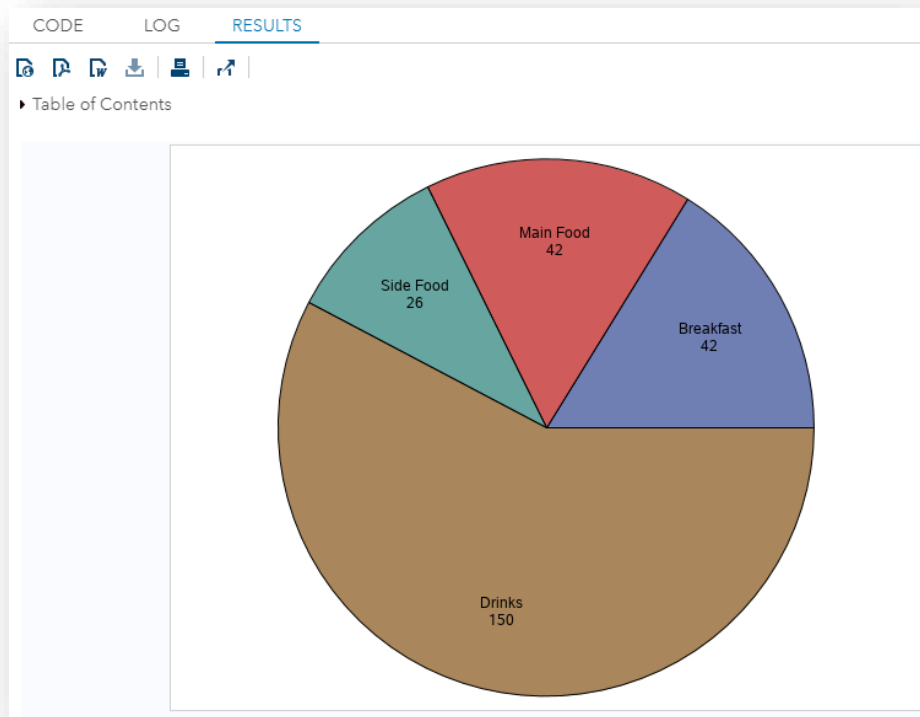


2. Choose the data set **MCDONALDS_MENU_NEW** in the **WORK** library as the Data source.
3. Set the **Category** variable to be our newly created variable **Category_New**.
You can change the **Measure** to from **Frequency Count** to **Frequency Percent**, or to a separate **numeric variable** depending on what you want your pie slice sizes to be determined by.





4. Click **Run**  and view the results in the **Output Data** tab.



End of Guided Exercises



Challenge Exercises

Submit your answers [here](#). Answer all 3 questions correctly to gain an entry into the prize draw.

Make sure you've downloaded the resources for today's challenge from [here](#) and placed it in **C:/SASUniversityEdition/myfolders**, the folder you created on Day 1.

It's a good idea to save your tasks in an accessible location once complete. If anything happens to your data, you can always rerun the tasks.

Import the delimited file **Netflix_2016.csv**. Save the data set to your **NETFLIX** library with the name **NETFLIX_2016**.

Question 1: *How many observations are in the NETFLIX_2016 data set?*

Use the **NETFLIX_2016** data set to create a new data set in the **NETFLIX** library called **NETFLIX_2016_RECODE** creates a new variable called **RatingGroups** that assigns each title a corresponding Rating Group base off its **rating**. Use the following to assign the groups:

Rating	RatingGroups
G, TV-Y, TV-G	Little Kids
PG, TV-Y7, TV-Y7-FV, TV-PG	Older Kids
TV-14	Teens
R, TV-MA, UR, NR	Adults

Question 2: *What is the RatingGroups given to the title Grey's Anatomy?*

Create a Pie Chart for the variable **RatingGroups** to calculate the frequency of each rating group in our 2016 Netflix data set.

Question 3: *How many titles in our 2016 Netflix data set have a rating group of Older Kids?*

Don't forget to visit our Facebook page tomorrow for the next challenge!



<https://www.facebook.com/SASAustNZ>