Dietary Restrictions Around the Globe

By Beth Hilbert, December 2018

Purpose

A recent study by Nielsen shows that consumers are making the decision of what they are going to eat based on what is not in their foods (fat, sugar, sodium). By examining these dietary restrictions around the globe, we can get a better view of which wellness habits are considered important overall and regionally.

Data

The Nielsen "Health and Ingredient Sentiment Survey" records self-reported dietary restrictions. The data was collected from 30,000 online respondents in 63 countries. The data was published by Nielsen in August 2016 and is available as an Excel spreadsheet from Makeover Monday Website titled "Restricted Dietary Requirements Around the Globe." http://www.makeovermonday.co.uk/data/data-sets-2017/

The initial data consist of fifty-five rows in one table with three columns: Diet, Region, Followers. There are eleven diets and five regions. The diets reflect wellness lifestyles (low carbs, low fat, low sodium, sugar conscious), allergies (lactose free, wheat free), and preferences/religion (vegan, vegetarian, Flexitarian, kosher, Halal). Respondents could indicate they were following multiple diets, so the diet restrictions add up to over 100% per region. Regions cannot really be compared with each other as there is no indication if the sizes of the regions are balanced.

The potential integrity issue in this data set is that diet names and region names are repeated in every row of table. Duplicate data can easily lead to data inconsistency when inserting, updating, and deleting records. Normalization is the process of reorganizing a database to decrease data redundancy and increase data consistency. Other than the repetition of diet names and region names, there aren't data quality issue. The column names reasonably reflect their values. There are not any missing values.

To fix this redundancy issue, I normalized the data (Fig1). I separated the one combined table into 2 tables with unique values (Diet, Region) which I then join by keys to form a composite table named Trend. To aid in analyzing the trends, I created an additional variable to capture a category label for diet reasons (allergy, meatless, religion, and wellness). To normalize this new information, I created a 4th table called Category which joins by key with the Diet table.

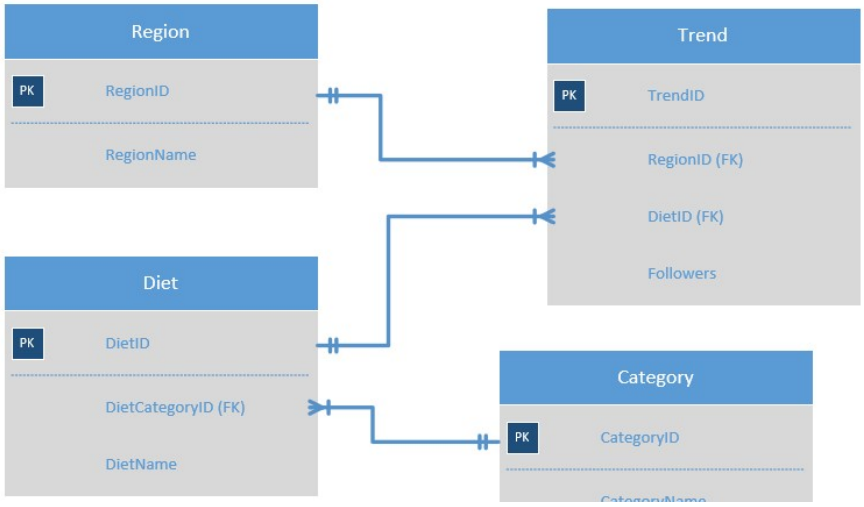

Figure 1: Food Trend Relational Diagram

Analysis

Global Insights:

First, I looked at the percent of followers globally, by diet and category (Fig2). The highest dietary restriction is “low fat” which is 10-fold higher (30.2%) than the lowest dietary restriction of “kosher” (3.3%). In fact, the top four dietary restrictions globally all reflect wellness choices.

	DietName	Avg_Followers
1	Low Fat	0.302
2	Sugar Conscious	0.262
3	Low Sodium	0.176
4	Low Carbohydrate	0.172
5	Halal	0.136
6	Vegetarian	0.108
7	Flexitarian	0.106
8	Lactose/Dairy Free	0.088
9	Wheat or Gluten Free	0.08
10	Vegan	0.046
11	Kosher	0.036

	CategoryName	Avg_Followers
1	Wellness	0.228
2	Meatless	0.087
3	Religion	0.086
4	Allergy	0.084

Figure 2: Dietary Restrictions Globally (by region, category, and diet)

Regional Dietary Category Insights:

I used a heatmap to drill down into the four categories within a region (Fig3). Darker colors represent more followers. By drilling down I see the Africa/Middle East region is heavily influenced by the 26% who follow religious diets and otherwise were actually a little less restrictive in their wellness diets (24%) than the next closest regions, Asia (27%) and Latin America (29%). I also see that Europe and North America are similar in the distribution of categories and much less restricted in their diets than the other regions across all four categories.

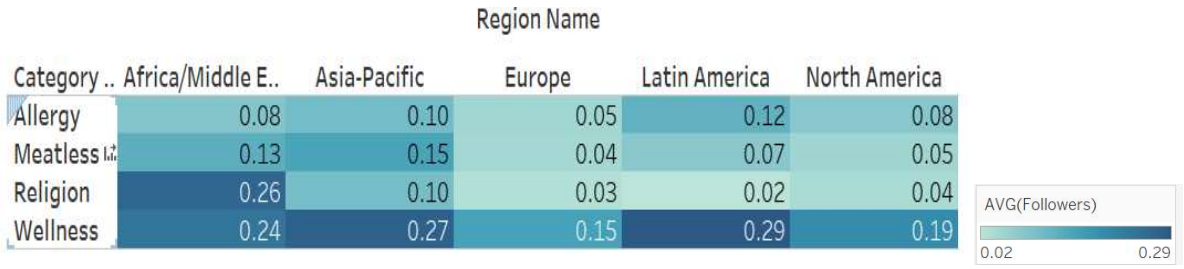

Figure 3: Percentage Followers With Dietary Restriction by Categories Regionally

Visually Fig4 shows the the consistency of focus on the wellness catetogires when it comes to dietary restrictions as all the bars in wellness category are similar heights. It also shows the Africa/Middle East emphasis on the religious dietary restrictions (26%) when compared with the other regional religious dietary restrictions (10%, 4%, 3% and 2%).

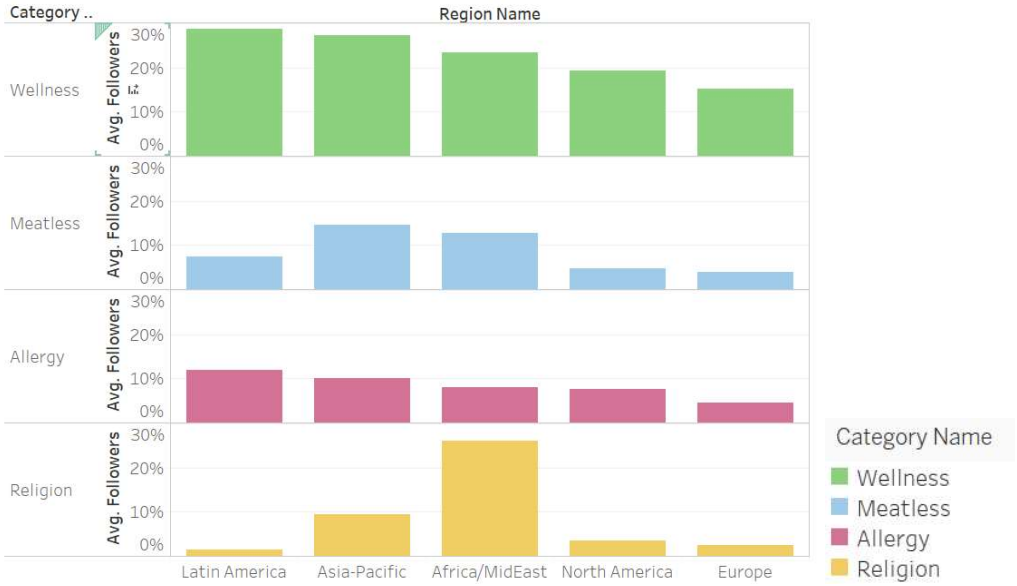

Figure 4: Dietary Restriction Categories Regionally

Regional Insights on Specific Dietary Restrictions:

Next I looked at the most popular diet restriction in each region (Fig5). It is interesting that half of the regions reported focusing on removing fat and the other half on removing sugar.

	RegionName	DietName	Max_Follower
1	Africa/Middle East	Halal	0.48
2	Latin America	Low Fat	0.39
3	Asia-Pacific	Low Fat	0.37
4	North America	Sugar Conscious	0.22
5	Europe	Sugar Conscious	0.22

Figure 5: Most Popular Diet Restriction In Each Region

I compared these most popular dietary restrictions with the other reported restrictions in each region (Fig6). I see that while the Middle East appears to have the highest percent of people within their region reporting dietary restrictions followers, they are very skewed by the 48% who follow the religious diet of Halal. Africa/Middle East is similar to Asia and Latin America when it comes to the wellness categories (24%, 27%, 29%).

Comparing Europe and North America (Fig6), I see that for most diets they are very similar differing by only a percent. However, the higher reported concerns for Low Carbs and Low Sodium (US with 15% and 21% and Europe with 11% and 8%) are where the largest differences show between these two regions. Both of these regions are much less concerned than the other three regions regarding Low Fat diets (19% and 20% versus 29%, 26% and 37% for other regions).

Fig6 also shows that Low Fat and Sugar Conscious Diets are the most prevalent reported dietary restrictions.

		Region Name				
Category ..	Diet Name	Africa/Middle E..	Asia-Pacific	Europe	Latin America	North America
Allergy	Lactose/Dairy Free	0.08	0.10	0.05	0.14	0.07
	Wheat or Gluten Free	0.08	0.10	0.04	0.10	0.08
Meatless	Flexitarian	0.16	0.16	0.05	0.10	0.06
	Vegan	0.06	0.09	0.02	0.04	0.02
	Vegetarian	0.16	0.19	0.05	0.08	0.06
Religion	Halal	0.48	0.12	0.04	0.01	0.03
	Kosher	0.04	0.07	0.01	0.02	0.04
Wellness	Low Carbohydrate	0.17	0.23	0.11	0.20	0.15
	Low Fat	0.36	0.37	0.20	0.39	0.19
	Low Sodium	0.13	0.22	0.08	0.24	0.21
	Sugar Conscious	0.28	0.27	0.22	0.32	0.22

AVG(Followers)
0.01 0.48

Figure 6: Dietary Restrictions Regionally

To continue drilling into specific diets, I created a plot to explore the distributions of diets between regions (Fig7). This time I exclude the two diets based on religious preferences. A few diets have overlapping percentages so don't show. But still I see that while the proportion of followers is pretty consistent for most diets, there are large gaps between regions in three categories: Vegetarian, Low Fat, and Low Sodium.

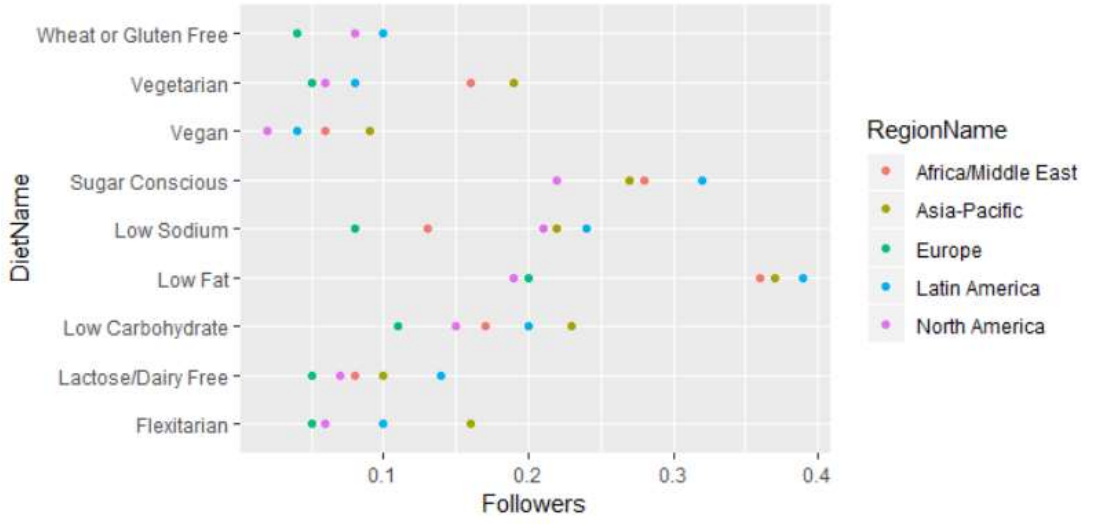

Figure 7: Dietary Restrictions Distributed Between Regions

I also look at the distribution of diets within a region (Fig8). The bar length shows how more people are reporting wellness dietary restrictions (colored green) over any other diet restriction. Looking at the data this way also reveals that Europe and North America are more sugar than fat conscious. Interestingly, Latin America rates allergy considerations higher and religious preference lower than other areas.

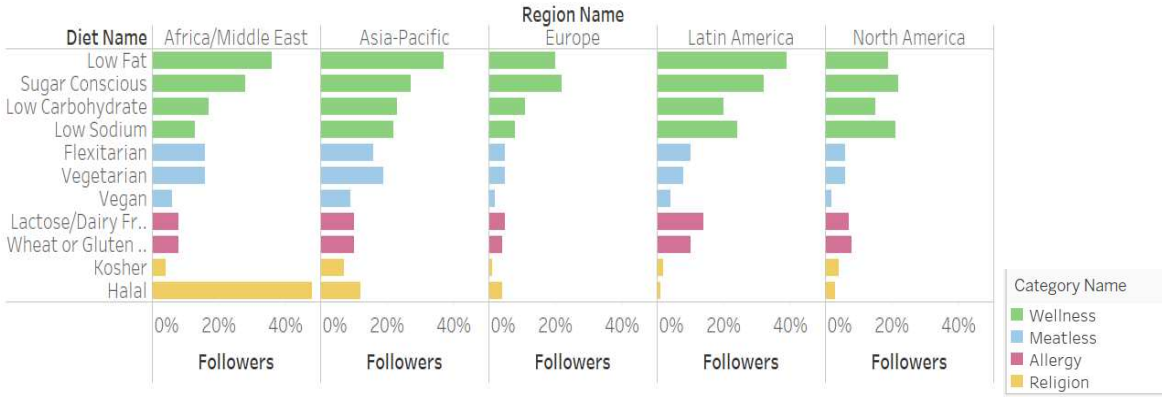

Figure 8: Dietary Restriction Within A Region

Statistical Relationship Insights:

To view the data from a statistical perspective, I created a correlation matrix to see if there are relationships between any of the variable pairs (Fig9). This matrix shows scatter plots on the left and the Pearson Correlation number on the right. Correlation measures the linear association between two variables. Values near 1 and -1 indicate strong linear correlation, values near 0 indicate little correlation. The variable Region is not correlated with anything, and the others were only very mildly correlated.

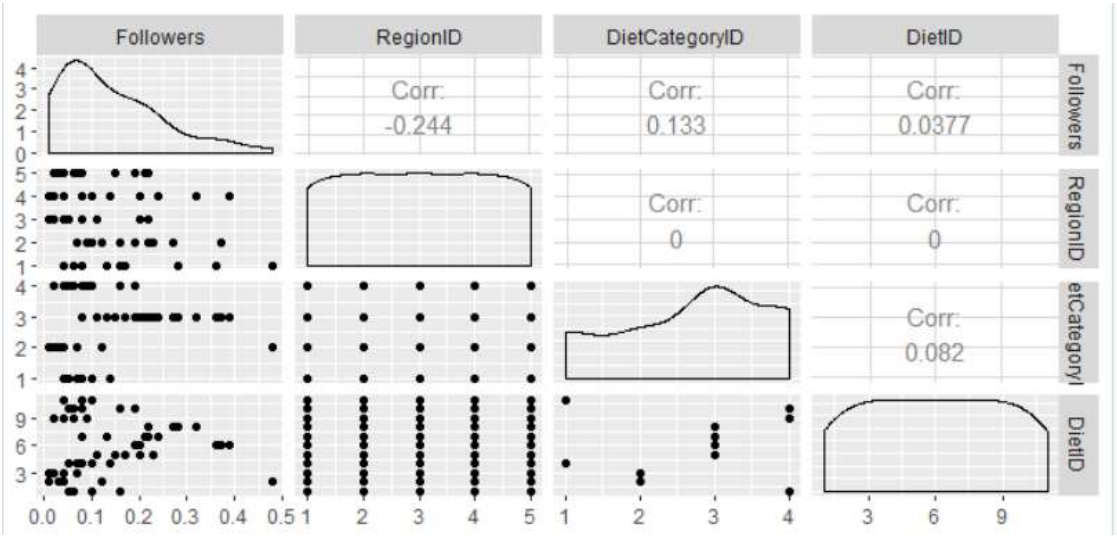

Figure 9: Correlation Matrix

I used linear regression to determine if it is possible to predict one of the variables (Followers, Region, Category, or Diet) based on the other 3 variables. A popular metric is R^2, which measures how much the model is explained by the predictor variables. A "good" score varies by industry, but generally you would prefer a score 80% or higher. Using this data, the highest R^2 is only 7.8% for the

Followers variable. This means that using the model to predict Followers, only 8% of the predicted value can be explained by Region, Category, and Diet variables and 92% is of the prediction is related to randomness (statistically referred to as error). Predictions for the other three targets are even lower: Category (2.5%), Diet (0.7%), and Region (0.6%). Clearly the current data can't be reliably used to generate a linear prediction model.

Conclusion

This data shows that across the globe, there is an interest in "clean" eating. While allergies and religious preferences do influence the things we choose to eat, overwhelming our choices are guided by choosing healthy foods that lack certain ingredients (sodium, sugar, fat, and carbohydrates). These choice preferences are true regardless of region.

In addition to healthy ingredients, vegetarian choices (including vegan and flexitarian) are important considerations for at least 1 out of 12 people globally, and 1 out of every 7 people in Asia. What goes on our plates and into our bodies are important considerations for many people.

There are a few limitations in this data. While there were 30,000 respondents, there is no indication if they are spread throughout the regions or concentrated in a few countries. A count of respondents per region would be helpful. Also, there is no way to tell if the meat-sensitive restrictions are a result of religion or preference. Finally, it would also be helpful to measure the daily impact of dietary restriction. For instance, is the restriction cyclical just around holidays or is it regular part of their daily lifestyle? What is the scale of the restriction, such does sugar conscious mean cutting back a little on sugar or completely eliminating sugar or from the diet? All these could be considered as an extension of this analysis.