

STUDY OF PESTICIDE RESIDUES IN FOOD

Pesticide Data Program (2015) by The United States Department of Agriculture

CIS 5270: Business Intelligence (Spring 2017)

By: Yashaswi Ananth





**Dataset:**

SUBMITTED TO – DR. SHILPA BALAN CALIFORNIA STATE UNIVERSITY, LOS ANGELES.

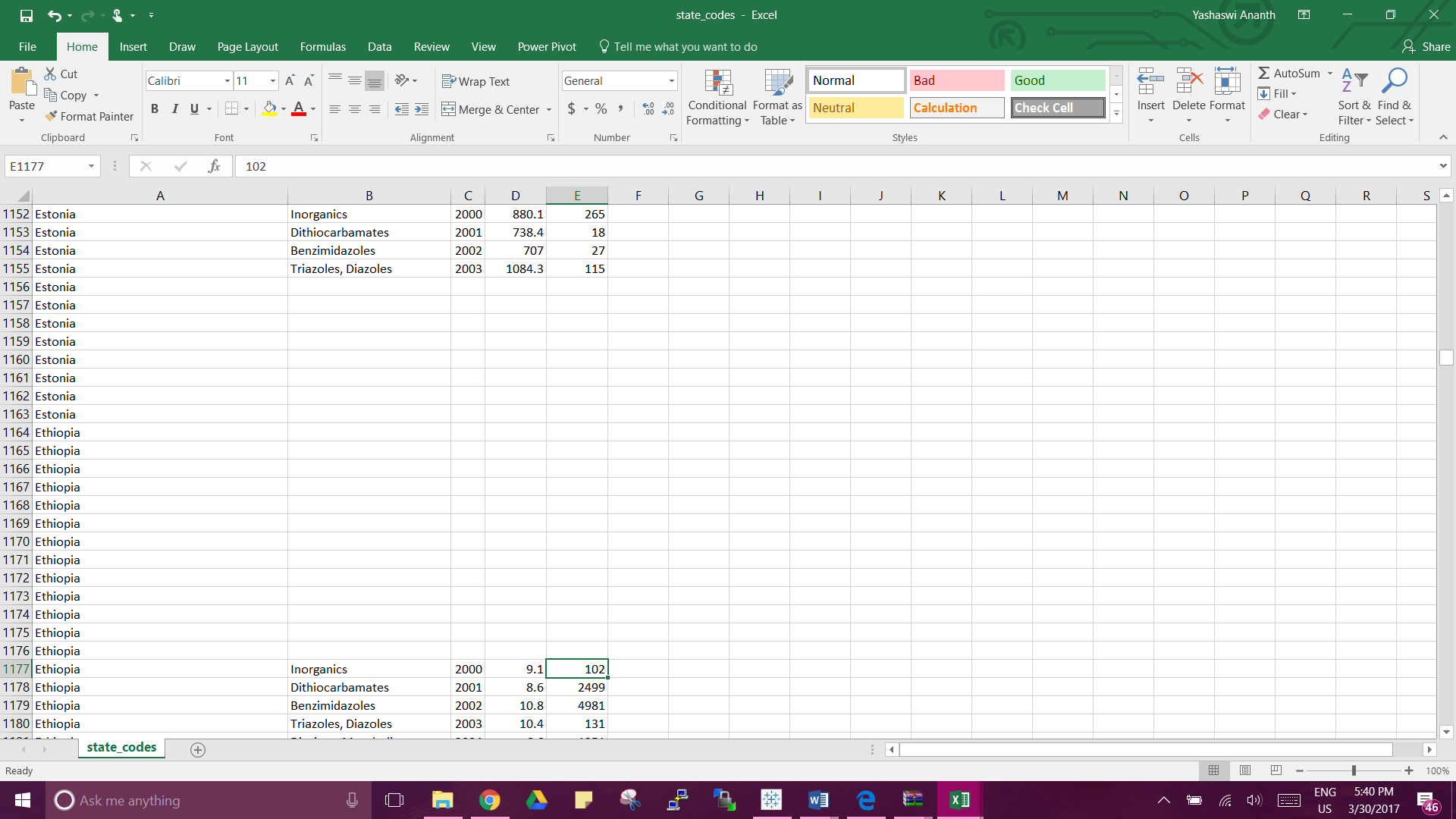
The dataset used for this project was extracted from the U.S. Department of Agriculture (USDA) Agricultural Marketing Service (AMS) that conducts the Pesticide Data Program (PDP) every year to help assure consumers that the food they feed themselves and their families is safe. Ultimately, if EPA determines a pesticide is not safe for human consumption, it is removed from the market. The dataset was abundant in information and contained 18 files with separately categorized tables for pesticide names and class, states and countries, tolerance values, commodities, laboratories, and extraction codes. These tables have been very helpful in deriving interesting insights detailed later in this report.

Dataset URL: <https://www.kaggle.com/usdeptofag/pesticide-data-program-2015>

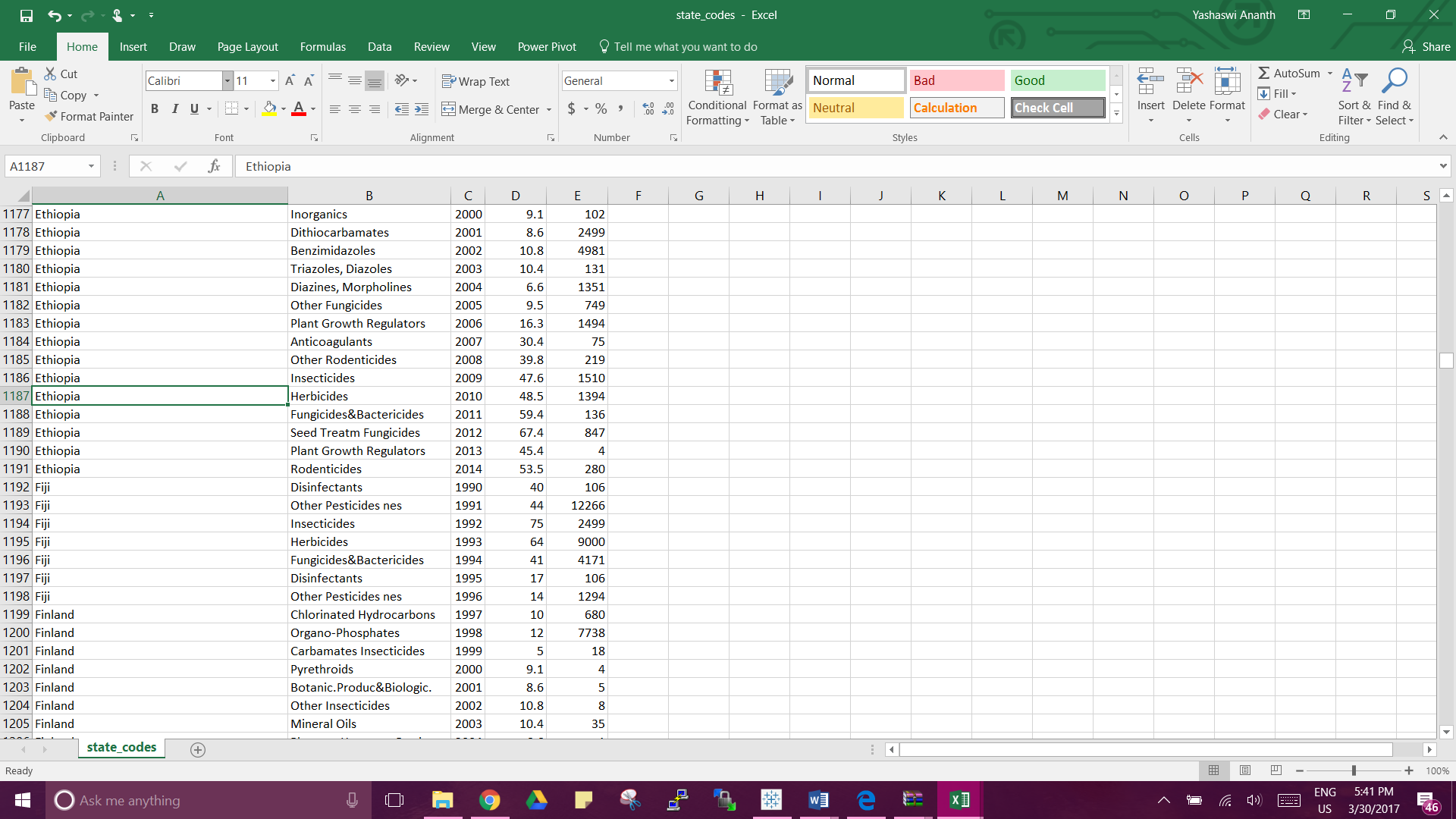
**Data Cleaning:**

Below were the modifications done to the data during the process of data cleaning.

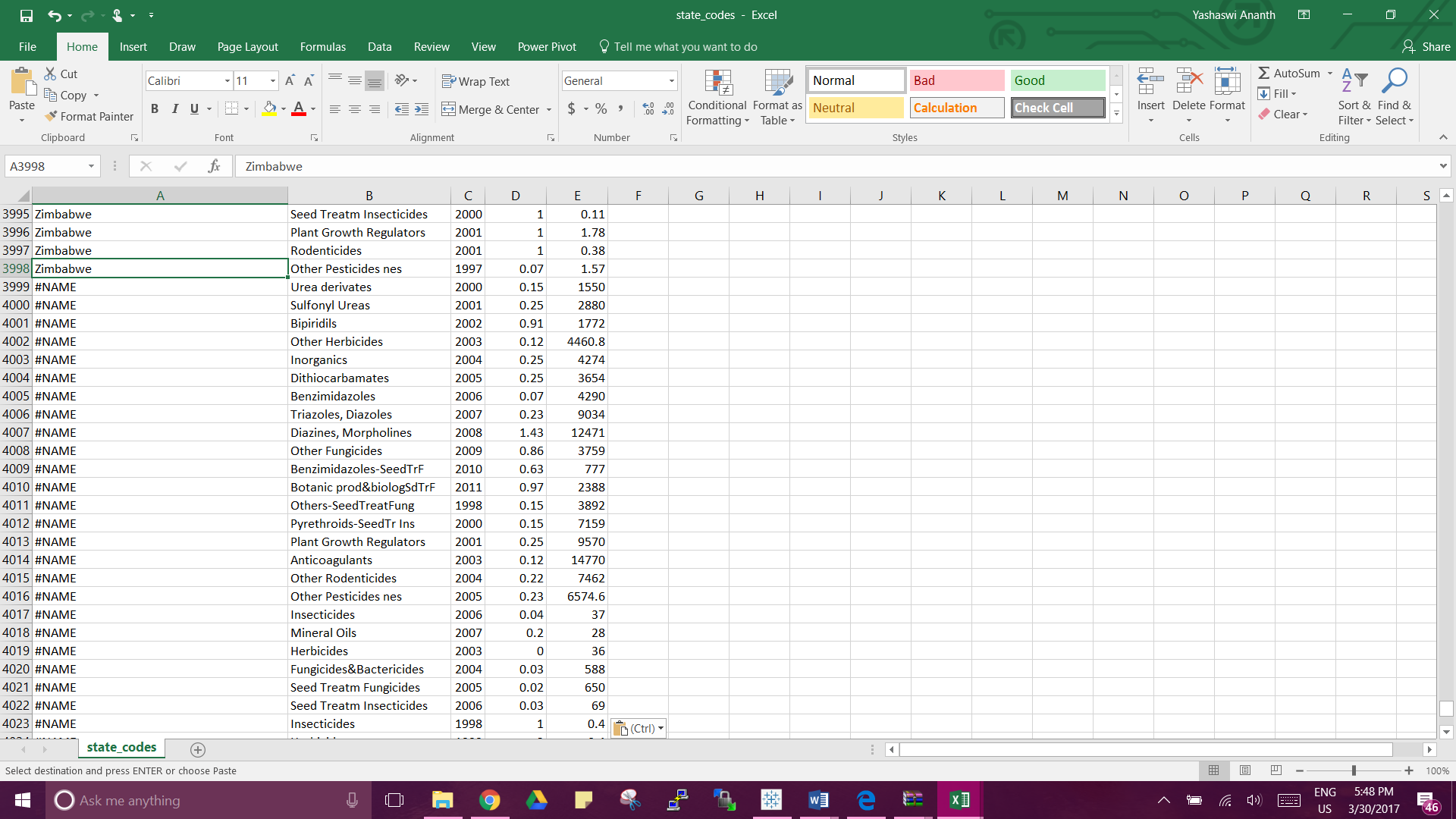
1. **Missing values**: Eliminating the NULL values



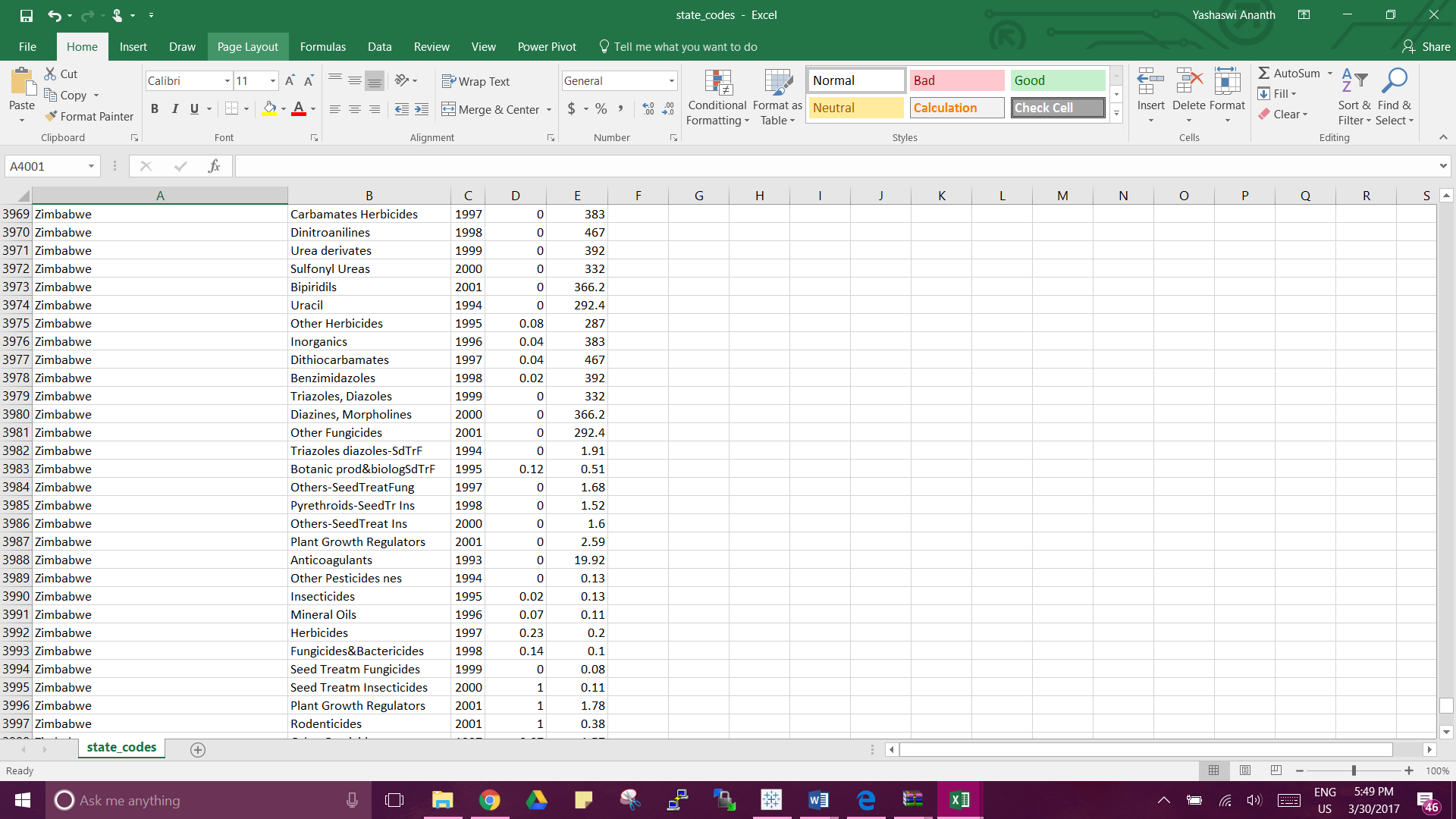
The dataset was thoroughly examined for NULL or missing information. All rows containing such data were deleted from the dataset to get accurate insights.



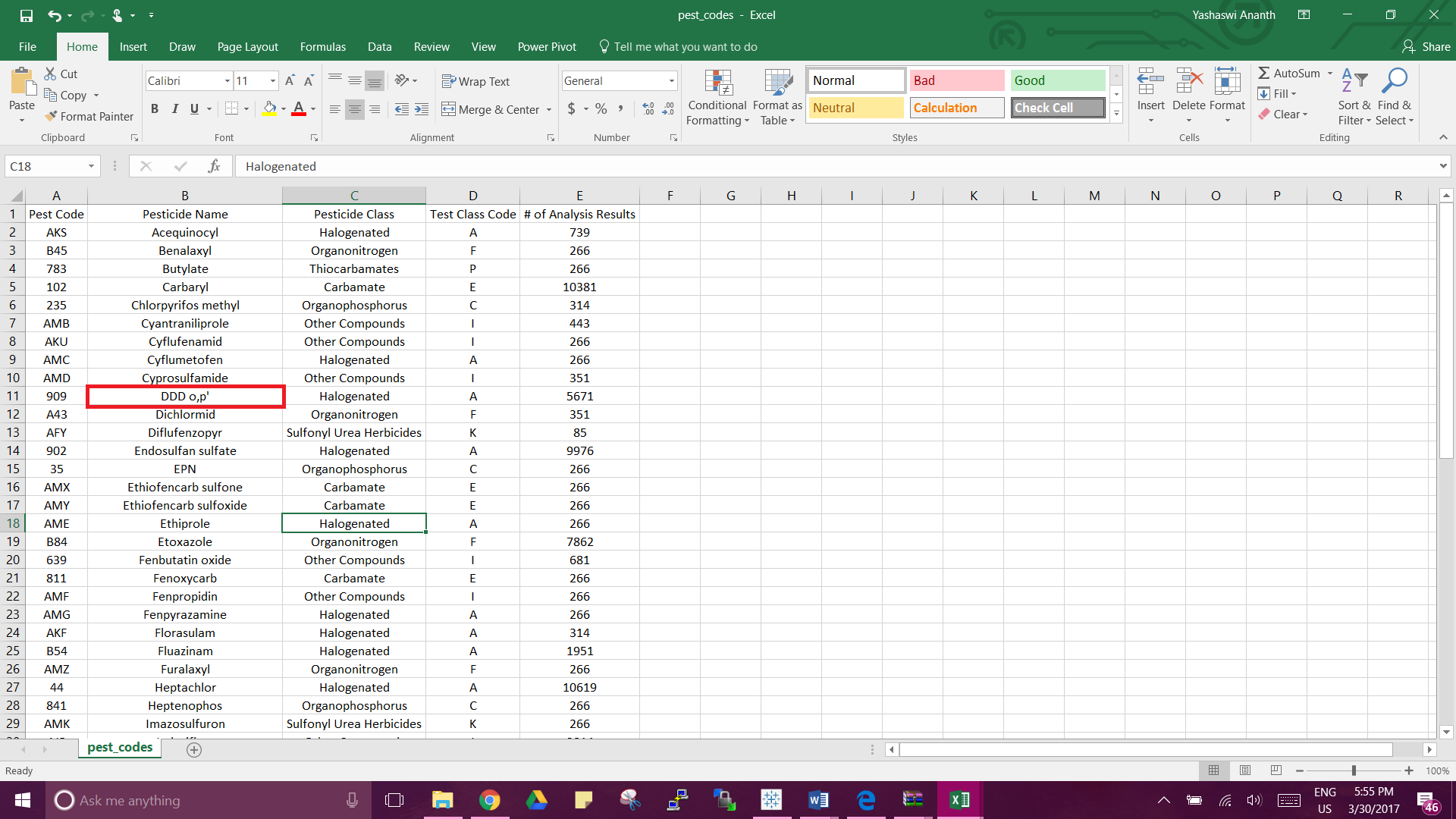
1. **Missing Values:** Eliminating the Default Values



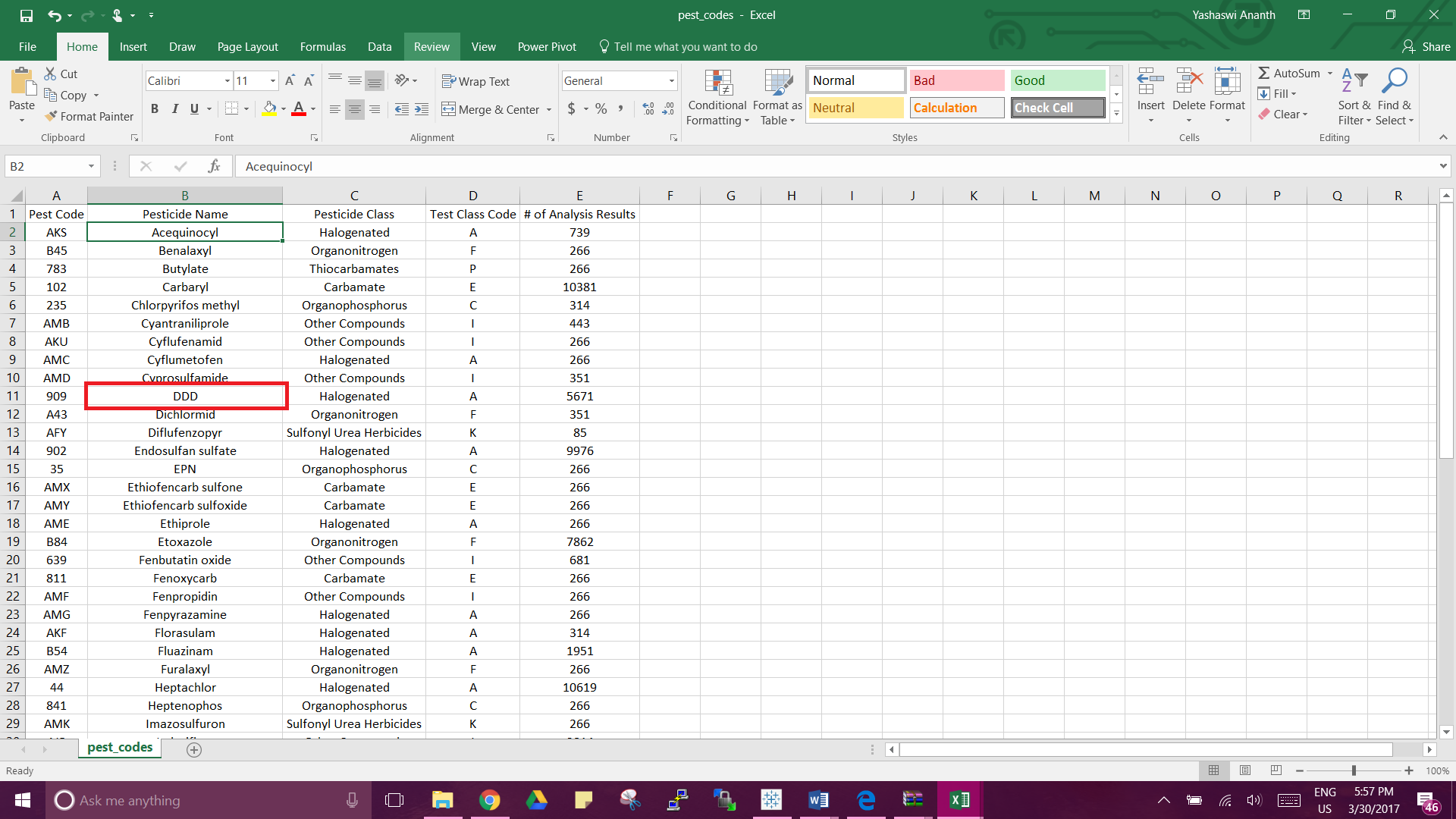
There were several rows with default values such as *#NAME* & *#VALUE.* Such values indicate that the authentic information is missing hence such rows were filtered out from the dataset.



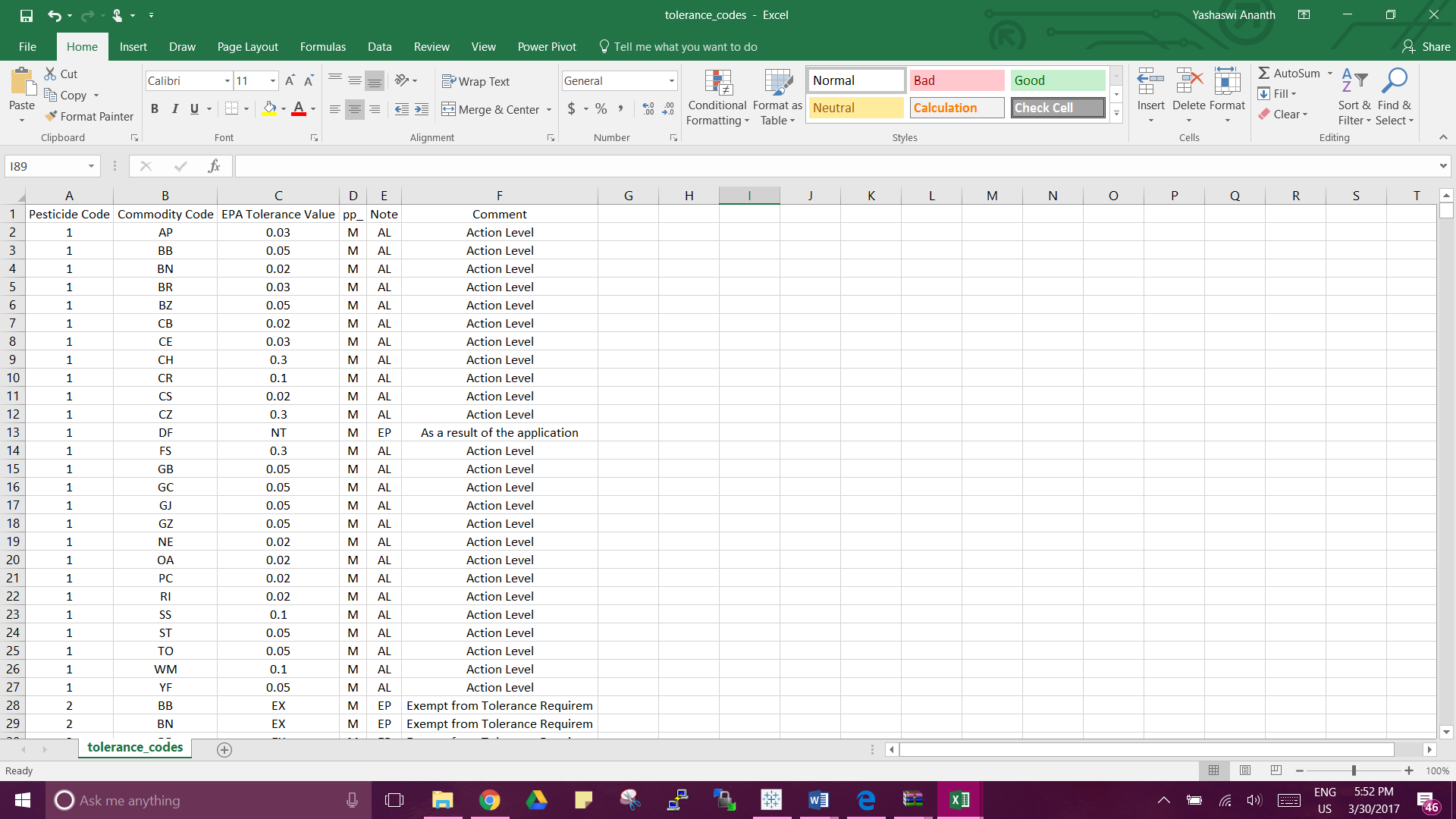
1. **Misspellings:** Rectifying the spelling errors



One of the pesticide name – DDD was misspelt as DDD o,p’. Since it was an invalid name, it was corrected as DDD.



1. **Illegal Values:** The dataset was checked for any possible values that do not validate or provide relevance to the column name. No such data was found.



In the above e.g. screenshot, data is checked to make sure that EPA Tolerance Value entries are valid. No negative or irrelevant values were found.

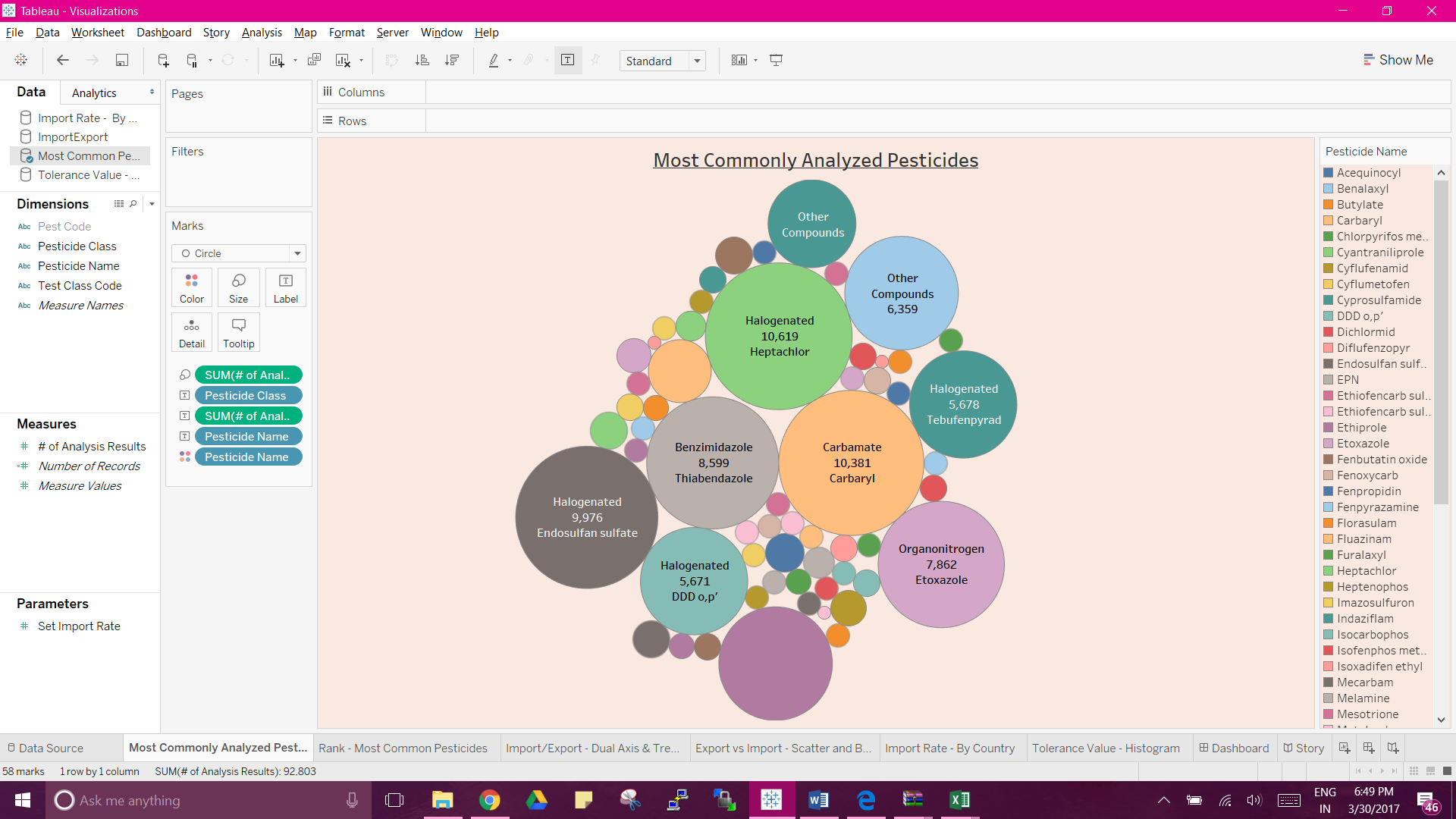
1. **Duplicates:** Elimination of duplicate entries

All the tables in the dataset were scanned for duplicate data. No redundant entries were found.

**Data Visualizations:**

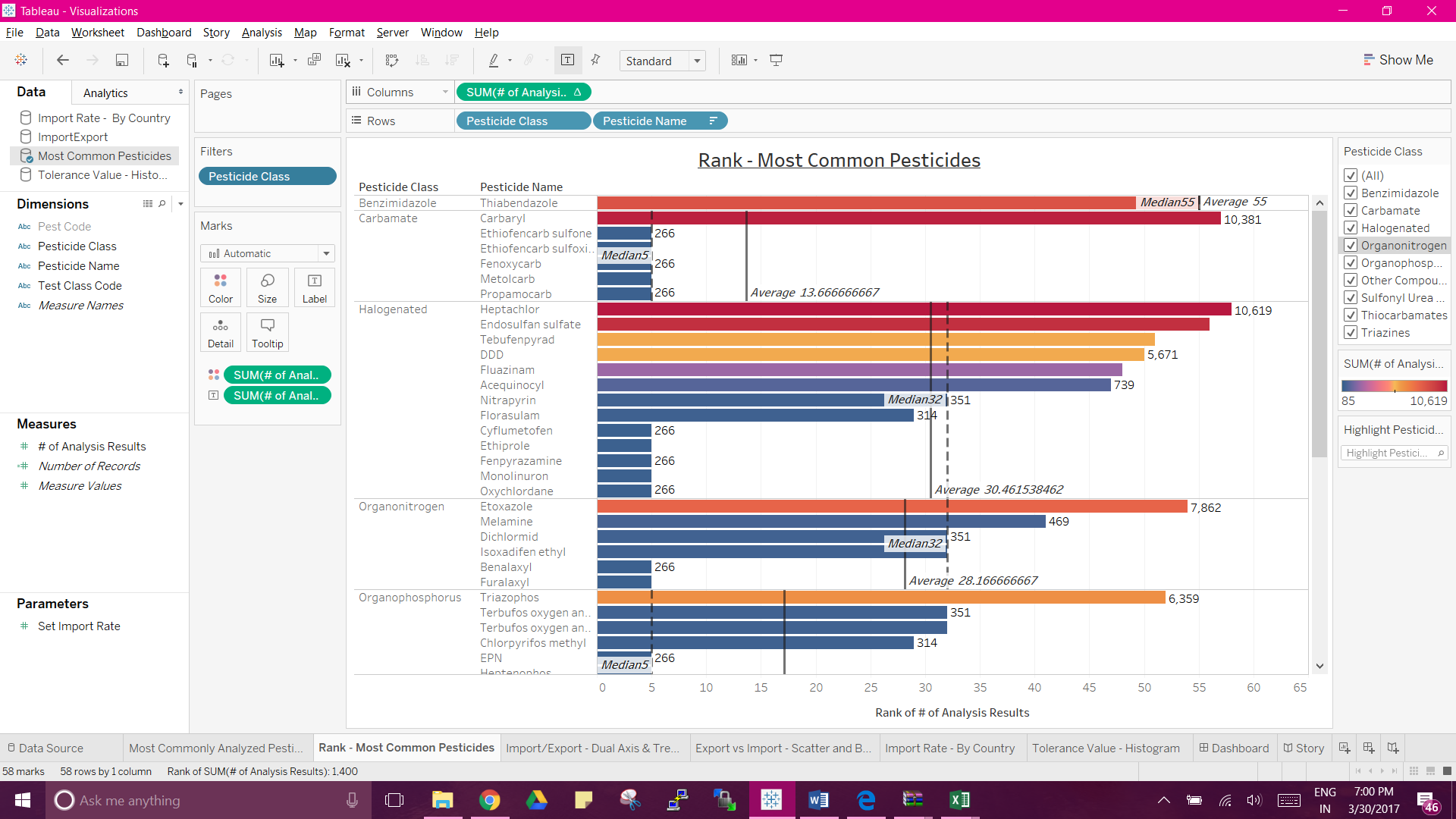
Following were the insights derived from the dataset.

1. **What are the most common pesticides that have been subjected to analysis?**



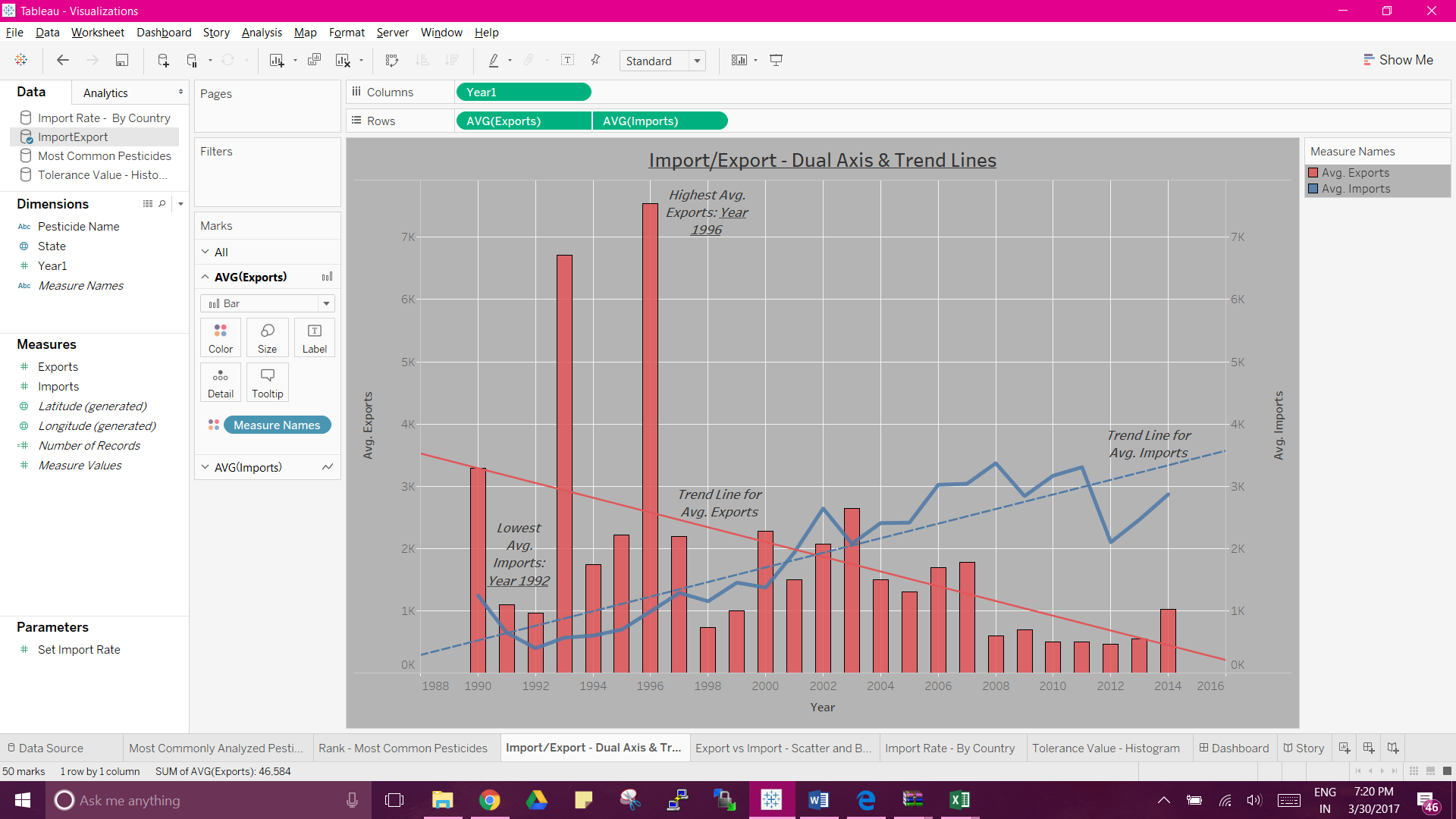
The above bubble chart gives an idea about the most prominently analyzed pesticides for the study. Inferring from the above visualization, we can conclude that the pesticide Heptachlor belonging to the Halogenated class has been analyzed most number of times (10,619).

1. **What is the ranking of the pesticides basis the number of analysis for each class?**



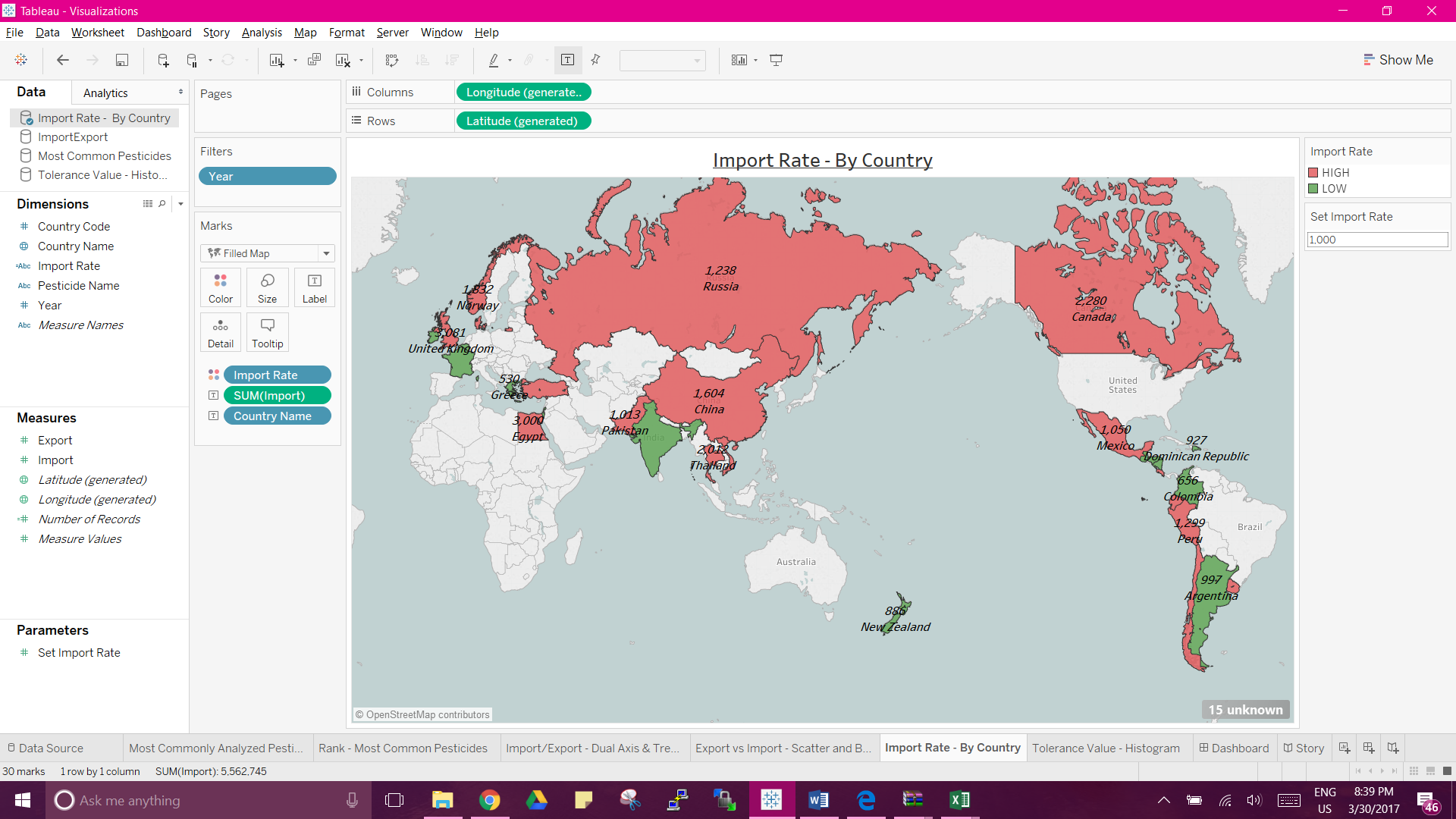
This visualization has been used to classify all the pesticides based on the class they belong to and then rank them in descending order of the number of times they have been tested for study. The ***Rank*** option has been used for each class to arrange the pesticides in the descending order of their ranking. A ***Reference Line*** has been inserted to better understand the average analysis number for each class. As an inference, it can be observed from the above graph that the pesticide Carbaryl belonging to class Carbamate ranks 57 while the other pesticides from the same class have been ranked 5 consistently. However, the reference line says that the average ranking for this class is 13.66.

1. **What are the average imports and exports of pesticides from the year 1990 – 2014?**



A ***Dual – Axis*** chart is a convenient way to show the variations of the imports and exports across a certain period if time. In this case the data ranges from the year 1990 – 2014. It can be easily analyzed that the years 1996 and 2008 were the time when the average exports and imports were highest respectively. To better understand the rate of rise and fall of the parameters, we can refer the ***Trend Lines*** in the chart. The patterns of exports & imports are volatile— making it difficult to see the overall pattern. The trend line for the exports implies that the amount of exports declined over the years whereas on the other hand, amount of imports increased. The P-value for the exports trend line being 0.01152 suggests that there was significant decrease in the exports across the world. The ***Dates*** in this visualization have been portrayed as continuous dimensions.

1. **What is the rate of the imports of the pesticides based on the region (countries)?**



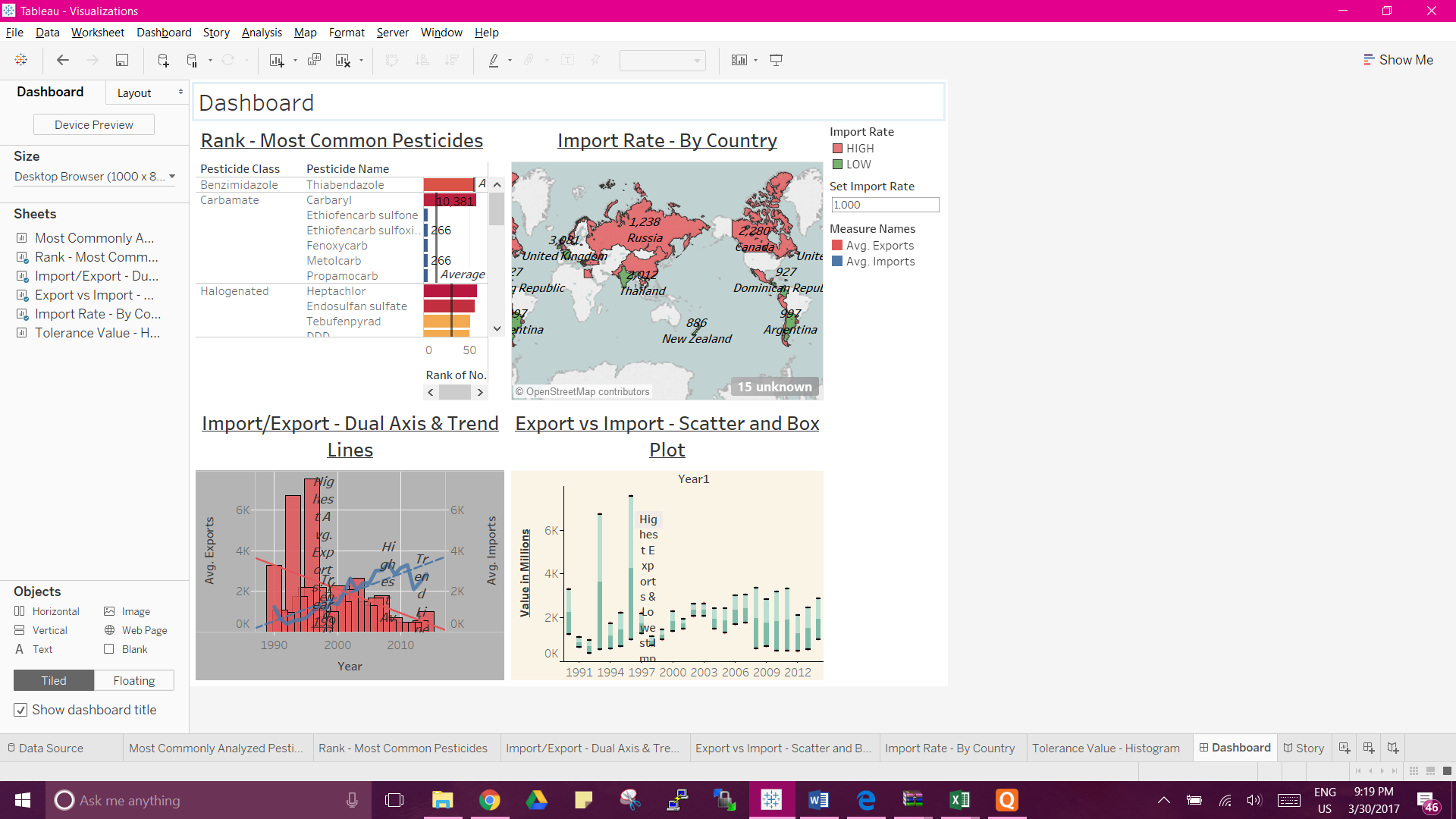
I have made use of the ***Geographic Maps*** (filled maps) for displaying region information. The countries in red denote high import rate; similarly, the countries in green denote comparatively low import rate. The define the margin line that differentiates the high and low import rates can be adjusted using the ***Parameter Control*** on the right. For the sake of the analysis, I have assumed that any country that has an import rate beyond 1,000 falls in the red category. From the analysis above, we can see that the countries Canada (2,280), Russia (1,238), Mexico (1,050), Peru (1,299), China (1,604) etc. have a higher annual import rate. Whereas, countries like India (596), Greece (530), Argentina (997), New Zealand (886) etc. have a lower annual import rate.

1. **What is the EPA Tolerance Value safely assigned by the U.S. Department of Agriculture (USDA) for different commodities of food products?**



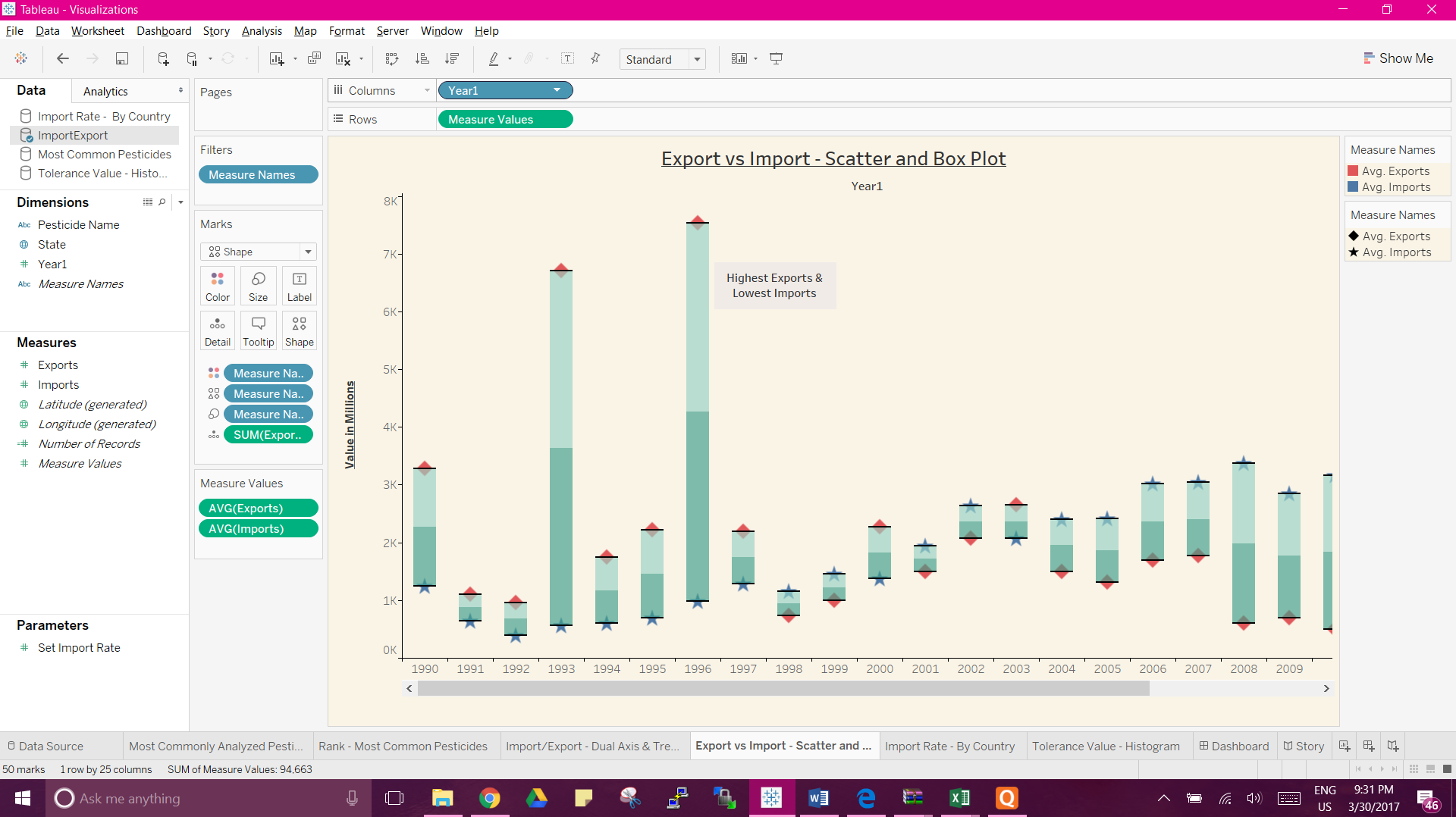
This visualization contains a ***Histogram*** to understand the different food commodities and their EPA Tolerance Value (maximum permissible amount of pesticide residue allowed). For the ease of analysis, I have assigned the bin size as 3 and also added a quick filter on the food commodities appearing on the right side of the visualization, to reduce the scope of the data that we are interested in looking at.

**Dashboard:**



The dashboard consists of the following visualizations:

1. **Rank – Most Common Pesticides:** This visualization best justifies the use of ***Ranks***. The pesticides are ranked basis the number of analysis and categorized by the class that they belong to.
2. **Import/Export –** **Dual Axis and Trend Lines:** In this chart, we can understand the rise and fall of the imports and the exports over a period of 15 years with the use of ***Trend Lines***.
3. **Import Rate – By Country:** The ***Parameter Control*** used in this geographical map can be used to specify the number figure that will define the high and low import rates.
4. **Export vs Import -**  **Scatter and Box Plot:**



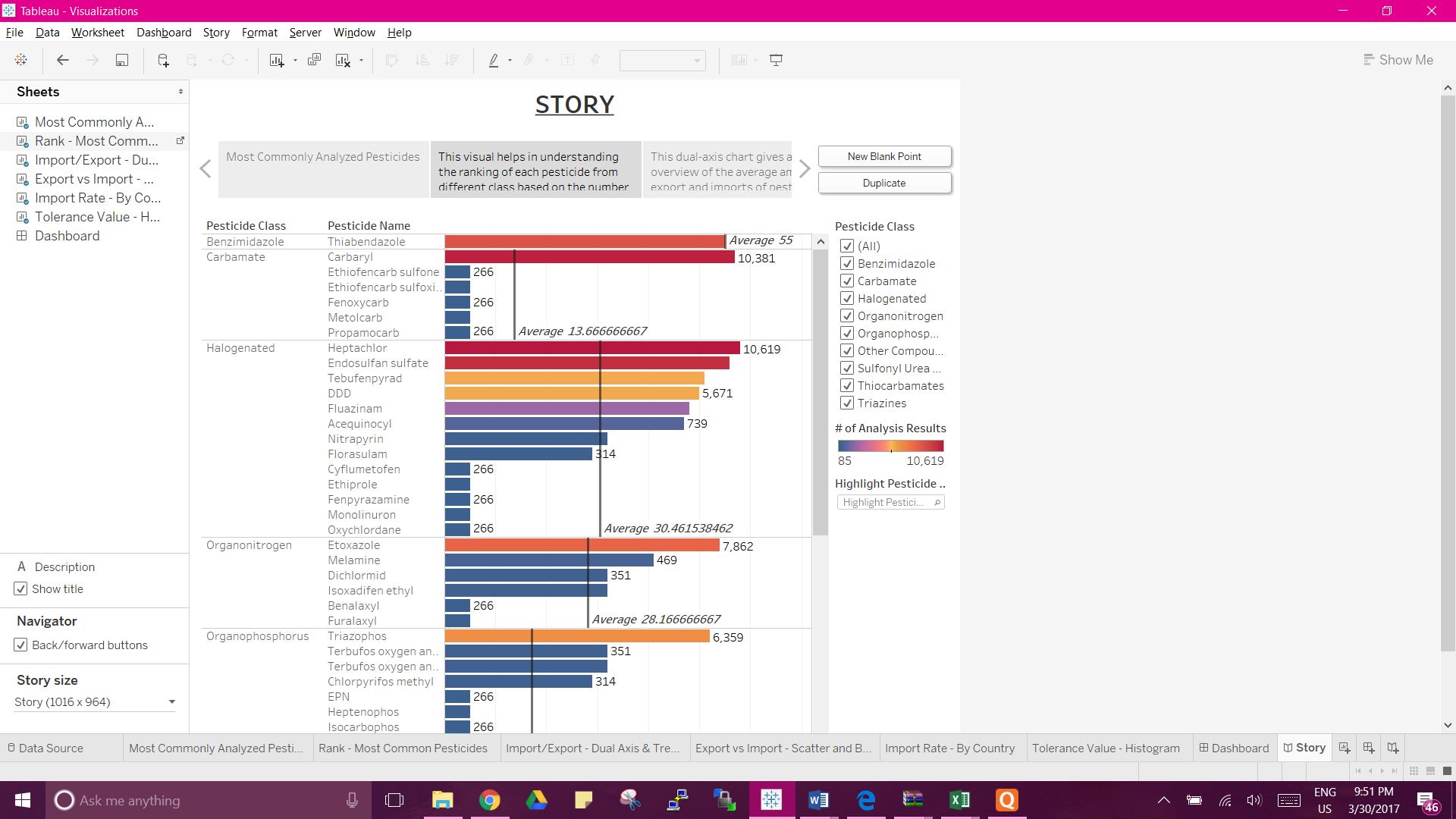
The above visualization uses the ***Scatter Plot*** to plot both exports and imports of pesticides for different years. To better showcase the difference in annual imports and exports, I have utilized ***Box Plot*** that highlights the values between the imports and the exports for a particular year. From the analysis, year 1996 was the time when there was a maximum difference in the exports and the imports. Per the statistics, exports were nearly 7.5k and imports were almost 1k.

**Story Telling:**

I first began by selecting a data source on Pesticide Residues in Food put together by the U.S. Department of Agriculture (USDA) Agricultural Marketing Service (AMS) that conducts the Pesticide Data Program (PDP) every year. This dataset includes worldwide information on the pesticides analyzed, the EPA tolerance value etc. In total, it has 18 tables and the size of the dataset is 18 MB.

While looking at the data, I sought to answer these questions;

* What are the most common pesticides that have been subjected to analysis?
* What is the ranking of the pesticides basis the number of analysis for each class?
* What are the average imports and exports of pesticides from the year 1990 – 2014?
* What is the rate of the imports of the pesticides based on the region (countries)?
* What is the EPA Tolerance Value safely assigned by the U.S. Department of Agriculture (USDA) for different commodities of food products?



To answer the 1st question, I used a packed bubble chart to represent different pesticides and the number of times they have been analyzed by the department for study. Every pesticide belongs to a certain class.

Answering the second question needed the help of the Rank option. This gave us the most popularly analyzed pesticide from every category. It is my assumption that the higher ranked pesticides are probably either the most used pesticides in agriculture or the most harmful ones if consumed. Hence, they are subjected to more analysis to ensure monitored use.

I then intended to find out the basic numbers for the imports and the exports of the pesticides across the world. As per the statistical data the exports have reduced whereas the imports have increased over a span of 15 years. Further and deeper analysis of the data can give a clear picture as to which are the most sought after pesticides and which countries tend to import more goods for food growth.

This was answered in the geographical visualization where we could clearly identify the countries having high average annual import rates. Canada being the country with the highest import rate probably is short of pesticides required for the commonly grown crops there like the rapeseed, barley, and corn. I assume this to be the reason for the high pesticide imports.

Different food commodities have different EPA tolerance values (probably based on the nature of pesticide being used). I built a Histogram to find out the tolerance values for various food items. It was found that apples and peaches had a higher tolerance value than the others. It’s a possibility that it is because these fruits have a slightly larger surface area than other food substances; for eg: Cherries.

**References:**

In order to do maximum justice in understanding the use of pesticides and other studies being undertaken across the world, below mentioned are few of the research papers that I went through during the course of carrying out this project: -

* *‘Impact of Pesticides and Other Chemicals on the Environment’* by Azhari Omer Abdelbagi, Adam Ali Mohamed Ahmed M. Elhindi and Ali Mohamed. Ali
* *‘Pesticides and health: A review of evidence on health effects, valuation of risks, and benefit‐cost analysis’* by Henrik Andersson, Damian Tago and Nicolas Treich, March 2014
* *‘FOOD SAFETY INFORMATION PAPERS’* by WHITE Technical Research Group