**Project**

**WORLD TOURISM DATASET**

**By**

**BOJJA VENKATA ALEKHYA RAIDU**

**CIN: 305829386**

**SUBMITTED TO**

**PROFESSOR SHILPA BALAN**

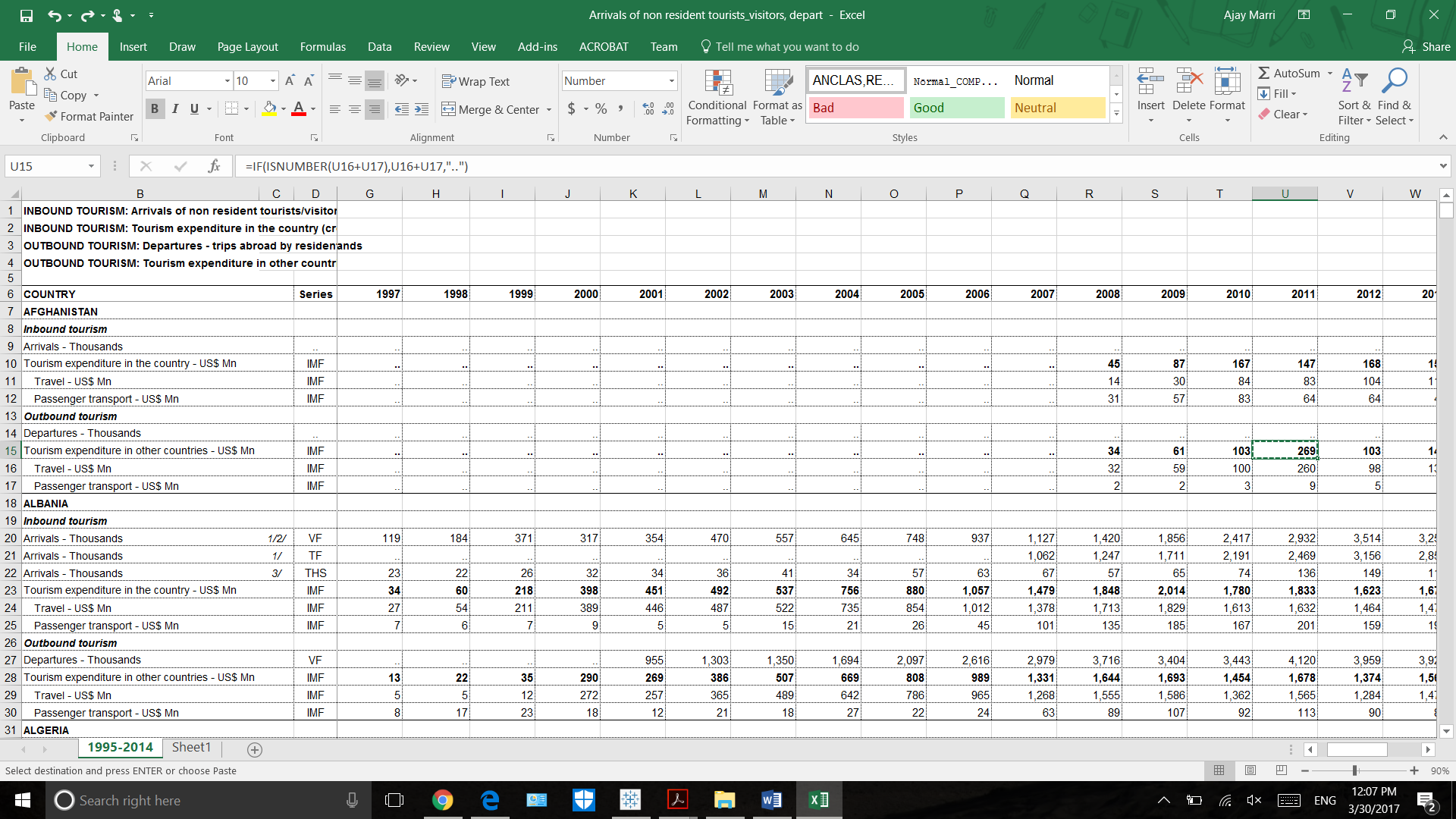
1. **URL’s of data set:**

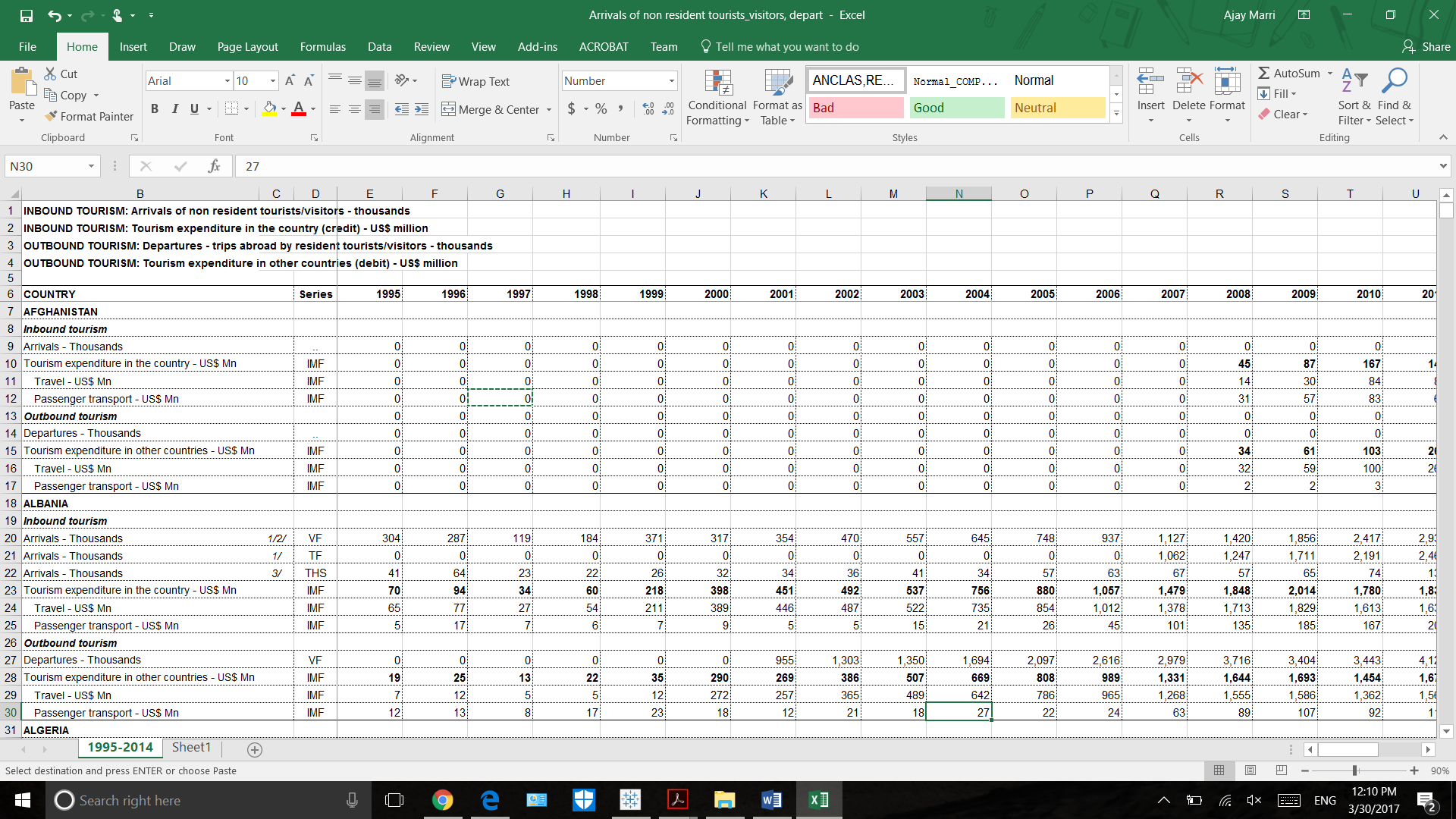
[**http://data.un.org/DocumentData.aspx?id=375**](http://data.un.org/DocumentData.aspx?id=375)

1. **Data Cleaning:**
2. Missing values:

Explanation: Data is given with missing values in the dataset. Like in hundreds of rows with n number of columns, data with missing values will create disturbance in analyzing the data. All the blank and null fields of the dataset is set to 0 for the mathematical operations to be used in tableau.

Before:

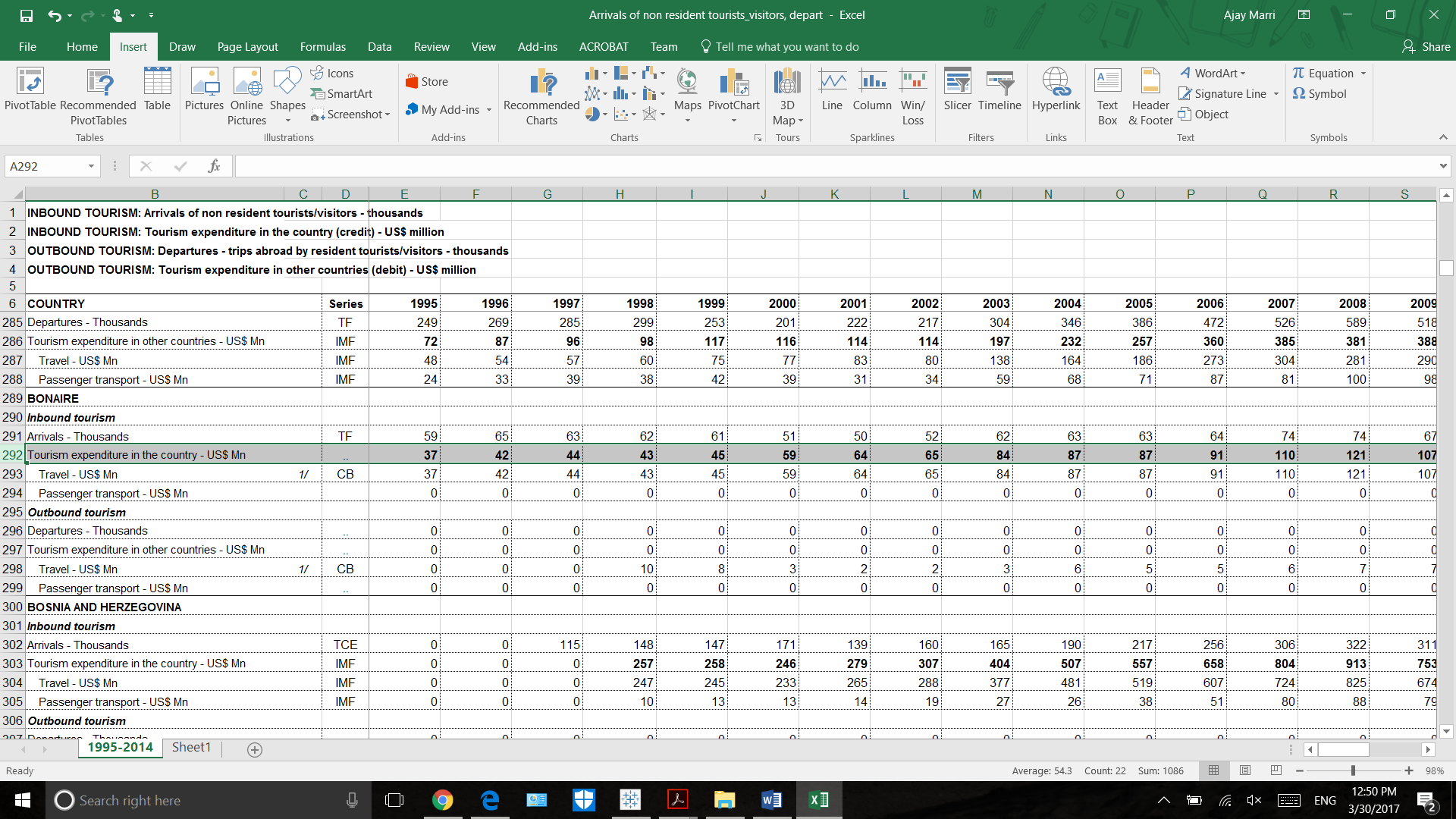


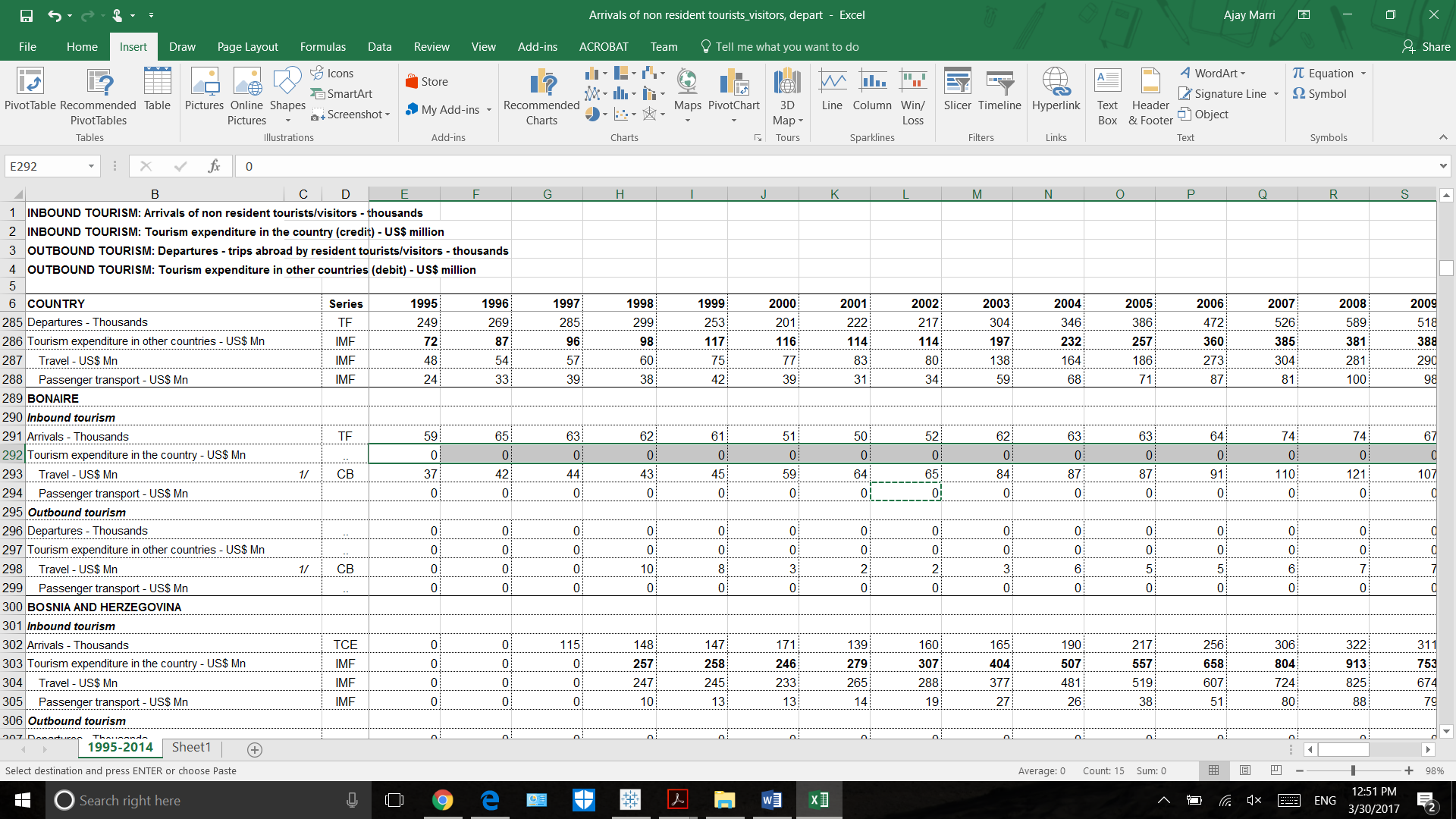
After:

1. Duplicate data:

Data with duplicate values will cause inconsistency in data. This will result in wrong predictions. Given below, for the tourism expenditure in the country row is duplicated with the values of the next row. So the values of the row is deleted to reduce inconsistency in data.

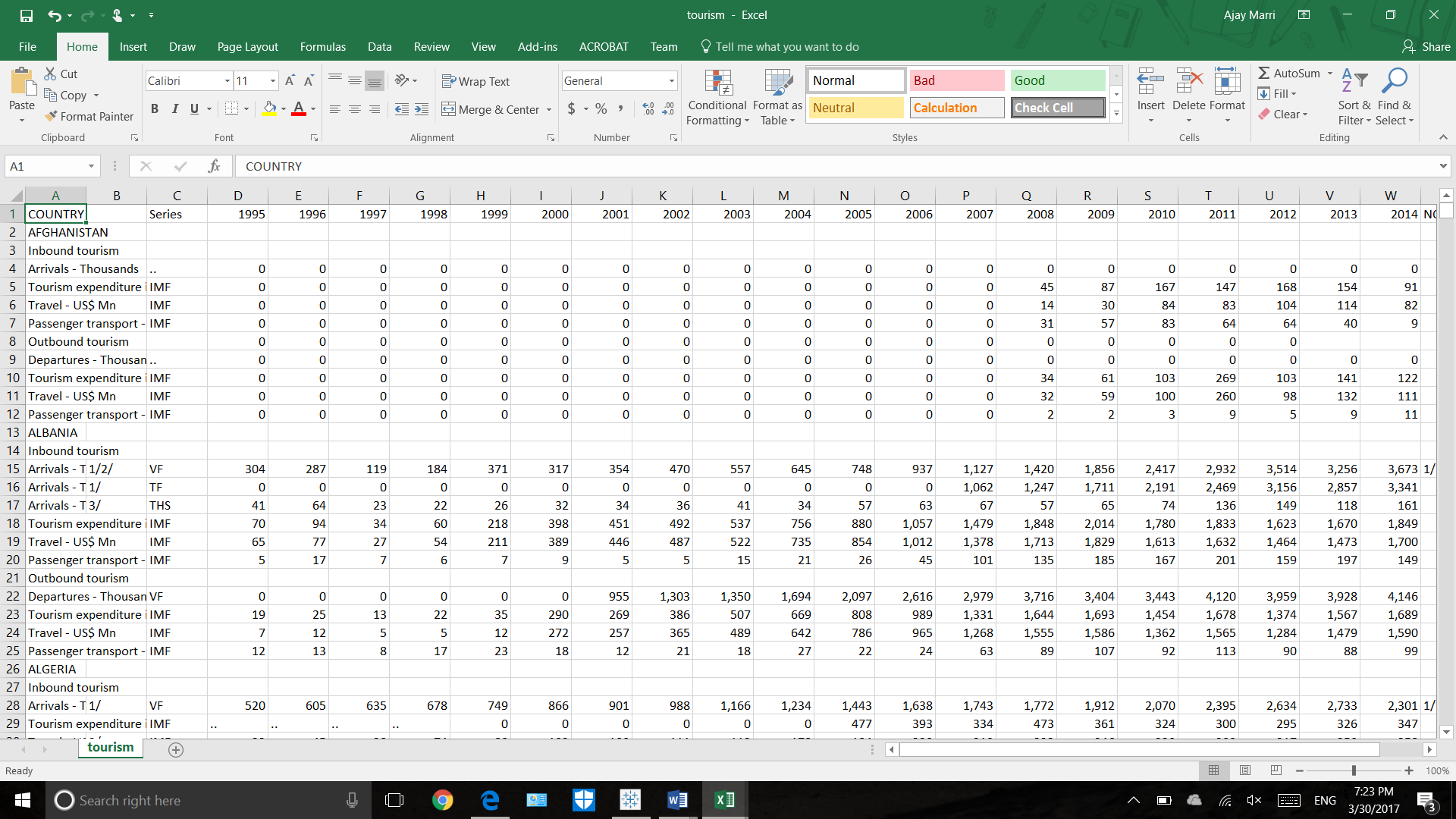
Before:



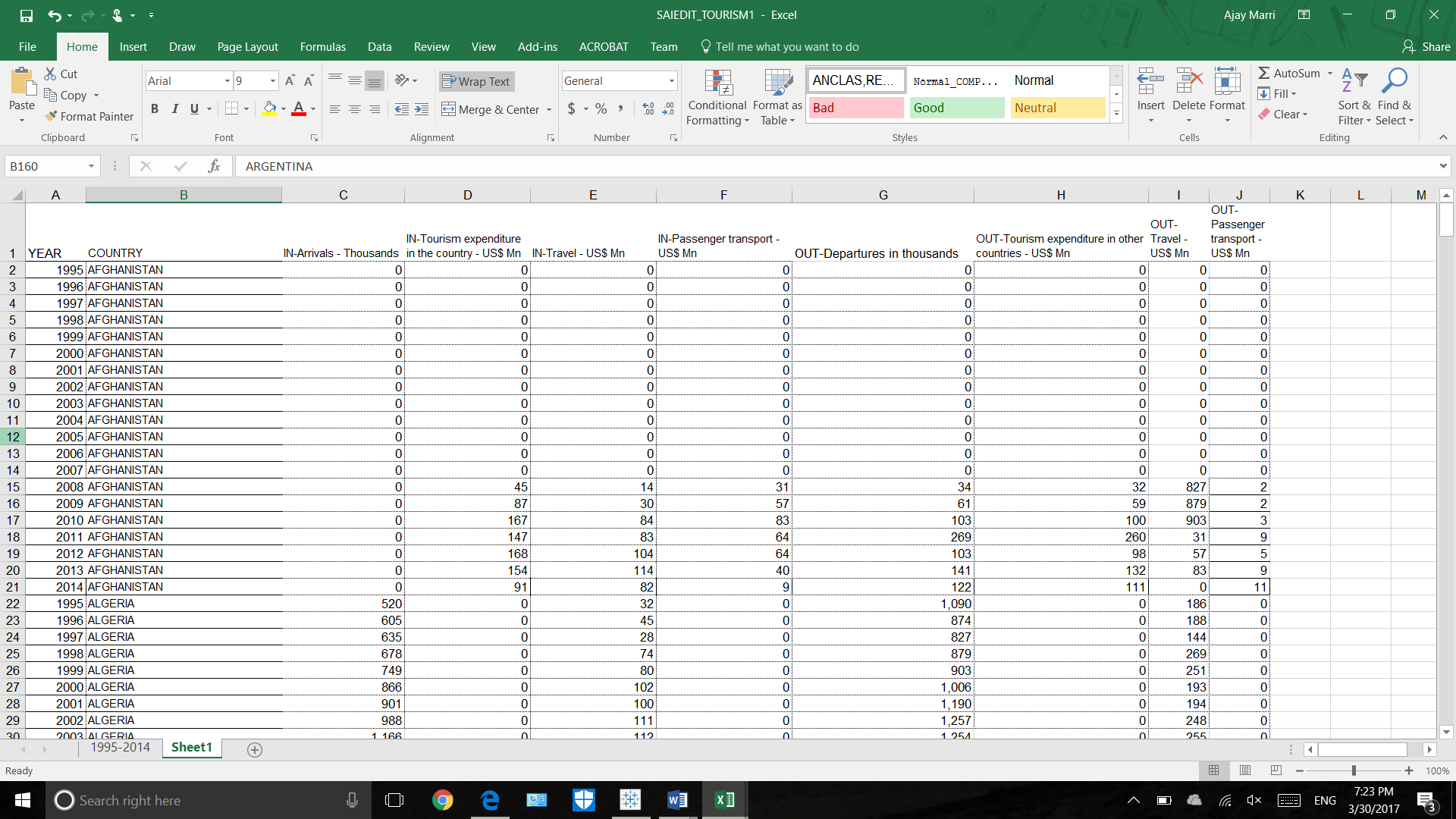
After:

1. Changing data accordingly to fit in tableau:

Explanation: Initially raw dataset incudes rows and columns in a zigzag fashion. Like all inbound tourism data and outbound tourism data as rows and only years a s colums. This doesn’t fit in tableau tp work with. So Transposing rows and columns accordingly will od that.

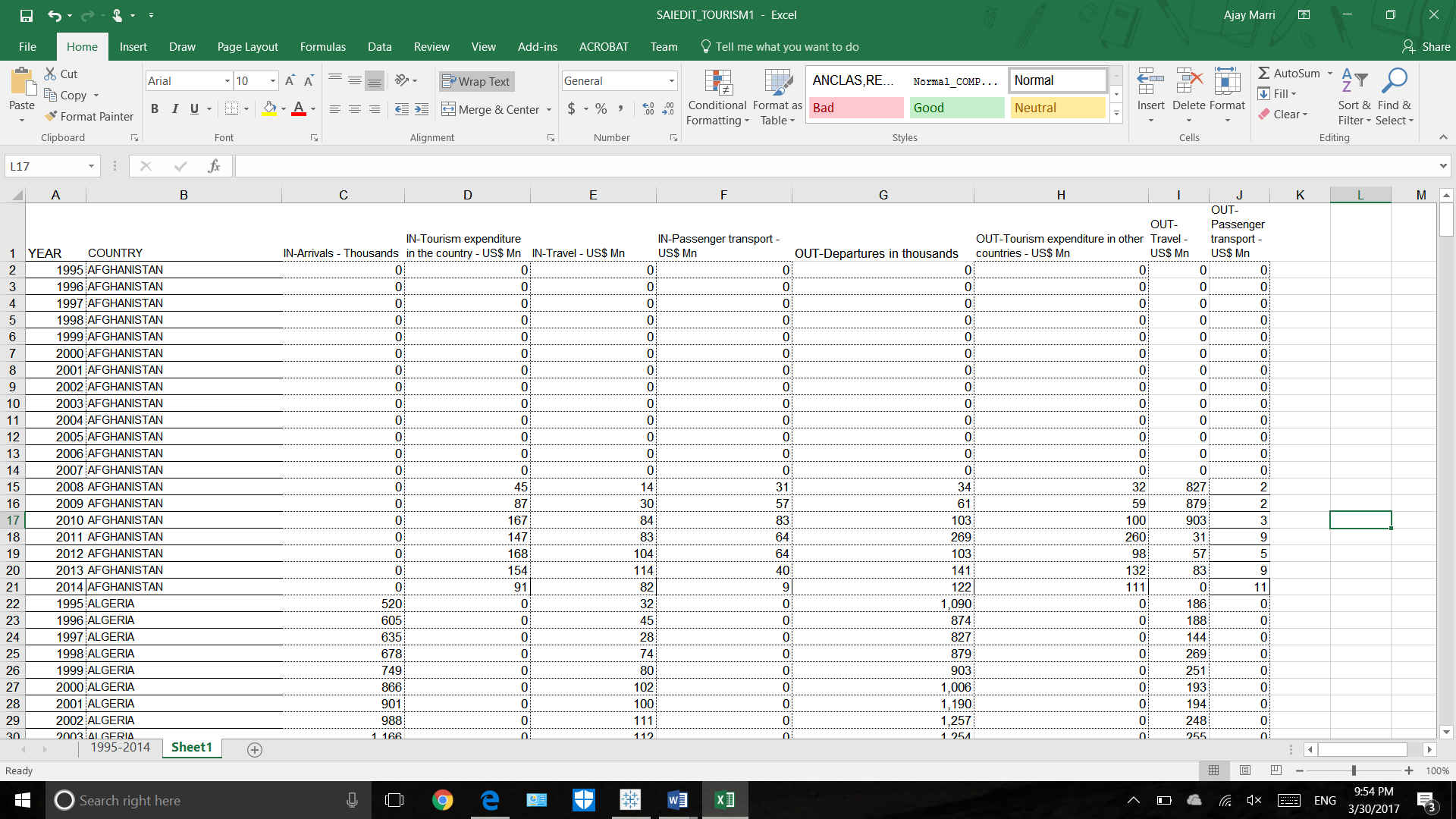
:Before

After:

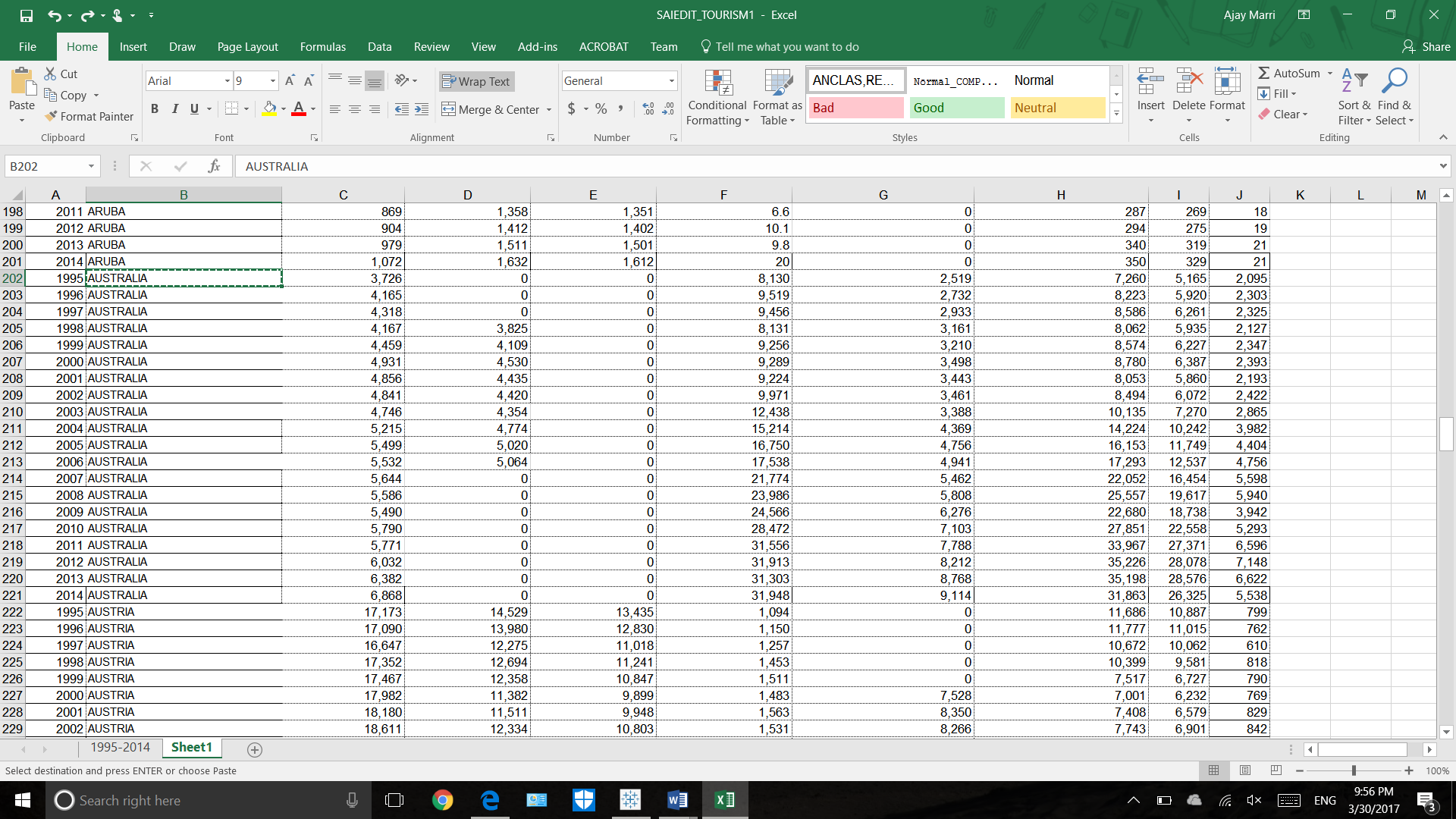


1. Tables are joined:

Explanation: Here I am using the only one dataset. So no need to bother about joining of tables and checking whether the columns are matched or not. Because given dataset includes all the related information.



1. Miss-spellings: In this data set, words are also correctly spelled. So no need to change spellings. Taken dataset satisfies this part of data cleaning. So, when we go across country names column headings we can see no data is miss spelled.



1. **Data Visualizations & explanations for the Research questions.**

This project visualizations includes:

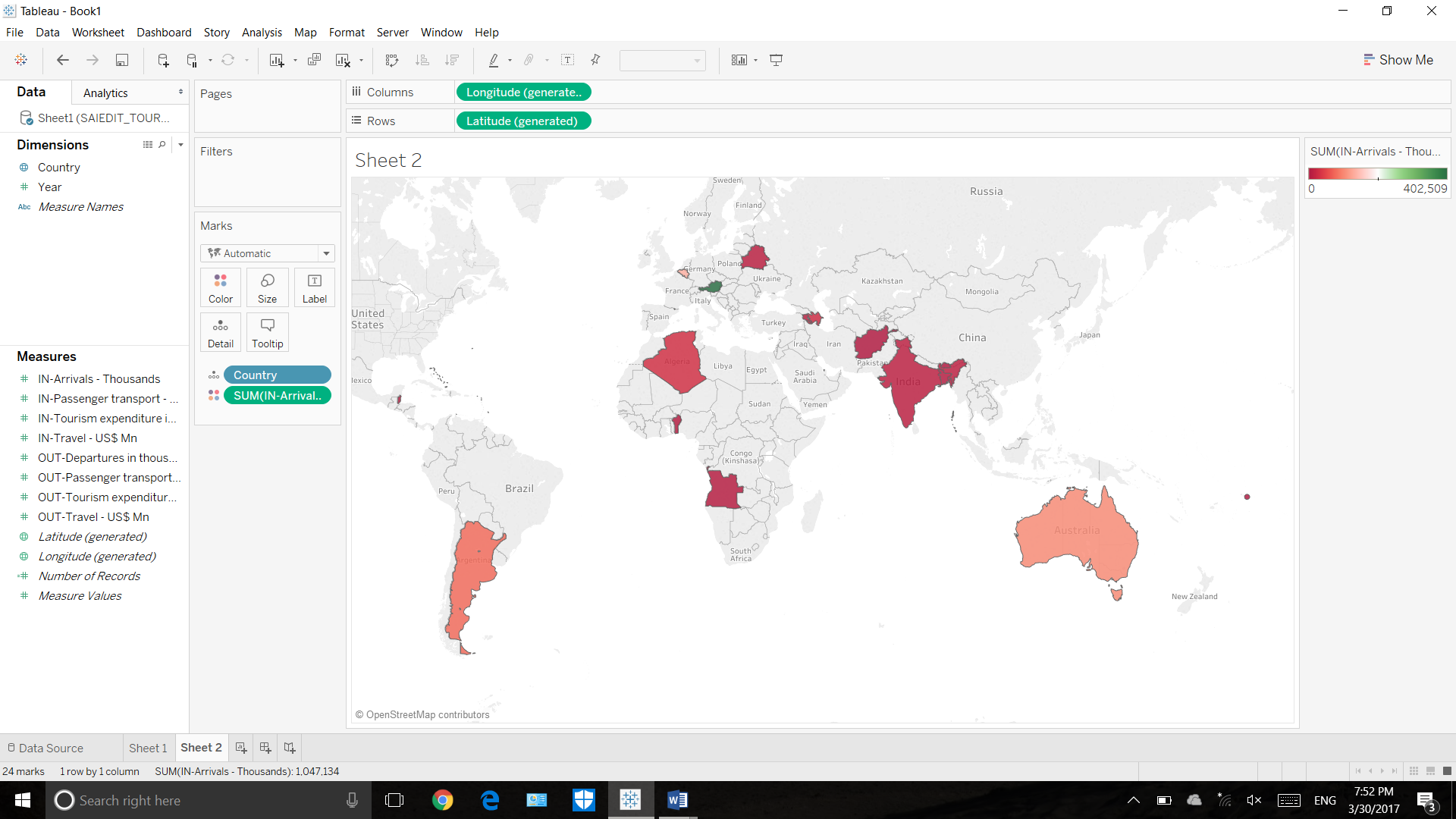
1. Geographic maps
2. . Rank
3. . Groups
4. . Scatter plot
5. . Calculated fields
6. Reference line.
7. Box & Whisker plot
8. Dual axis charts

1.Which country has highest inbound tourists.

Explanation:

This sheet shows the **geographical maps** of the countries and their summation of inbound arrivals. The range scale on right shows that countries from lowest inbound arrivals to counties with highest inbound arrivals. We can beautifully picturize that using colors card on the marks. From the graph it is clearly shown Austria has highest inbound tourists, because it is green in color.

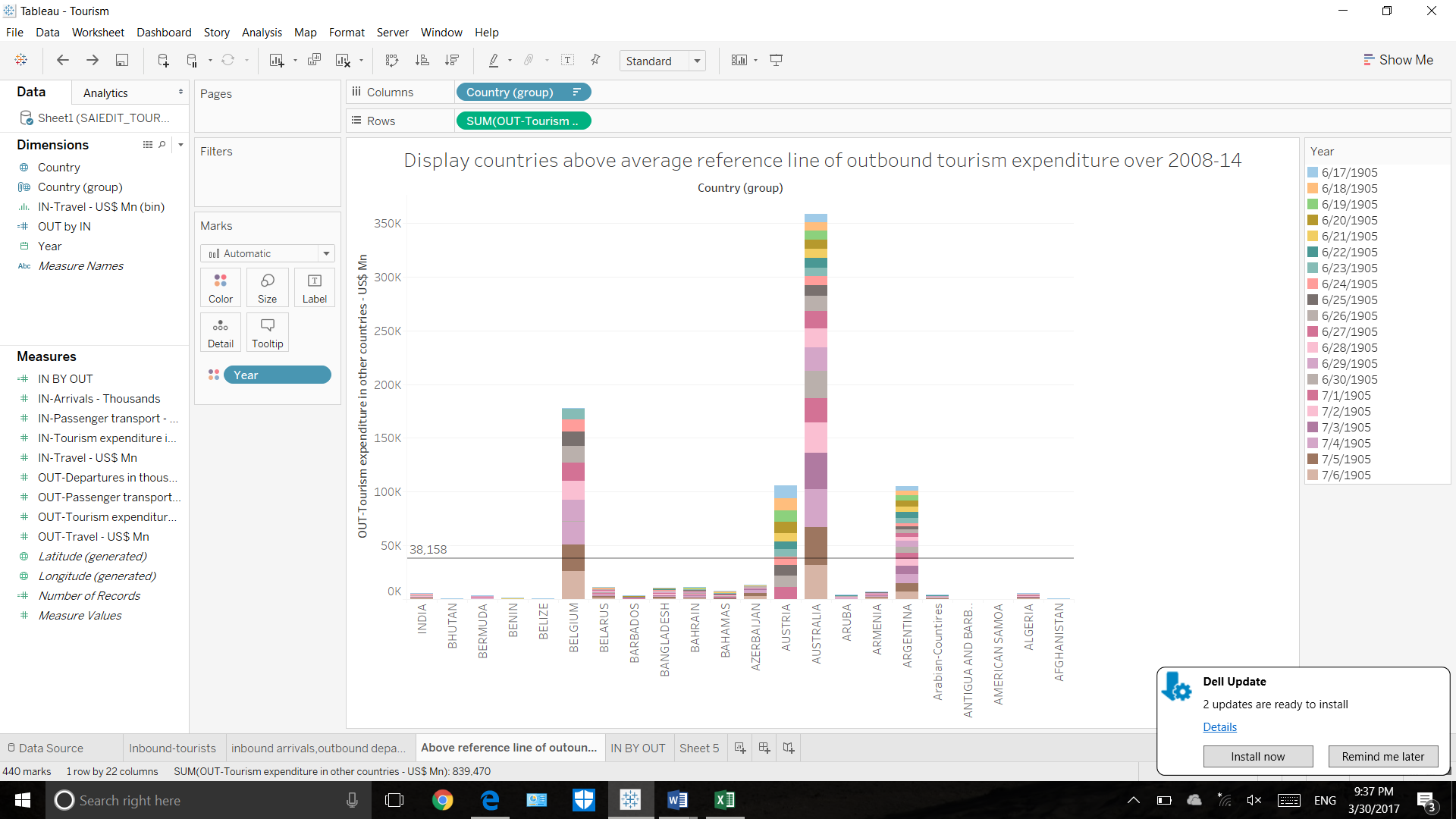
Screenshot:



1. Find the country which has maximum outbound tourism expenditure in years between 2008 to 2014 and find countries which are above average reference line.

Explanation:

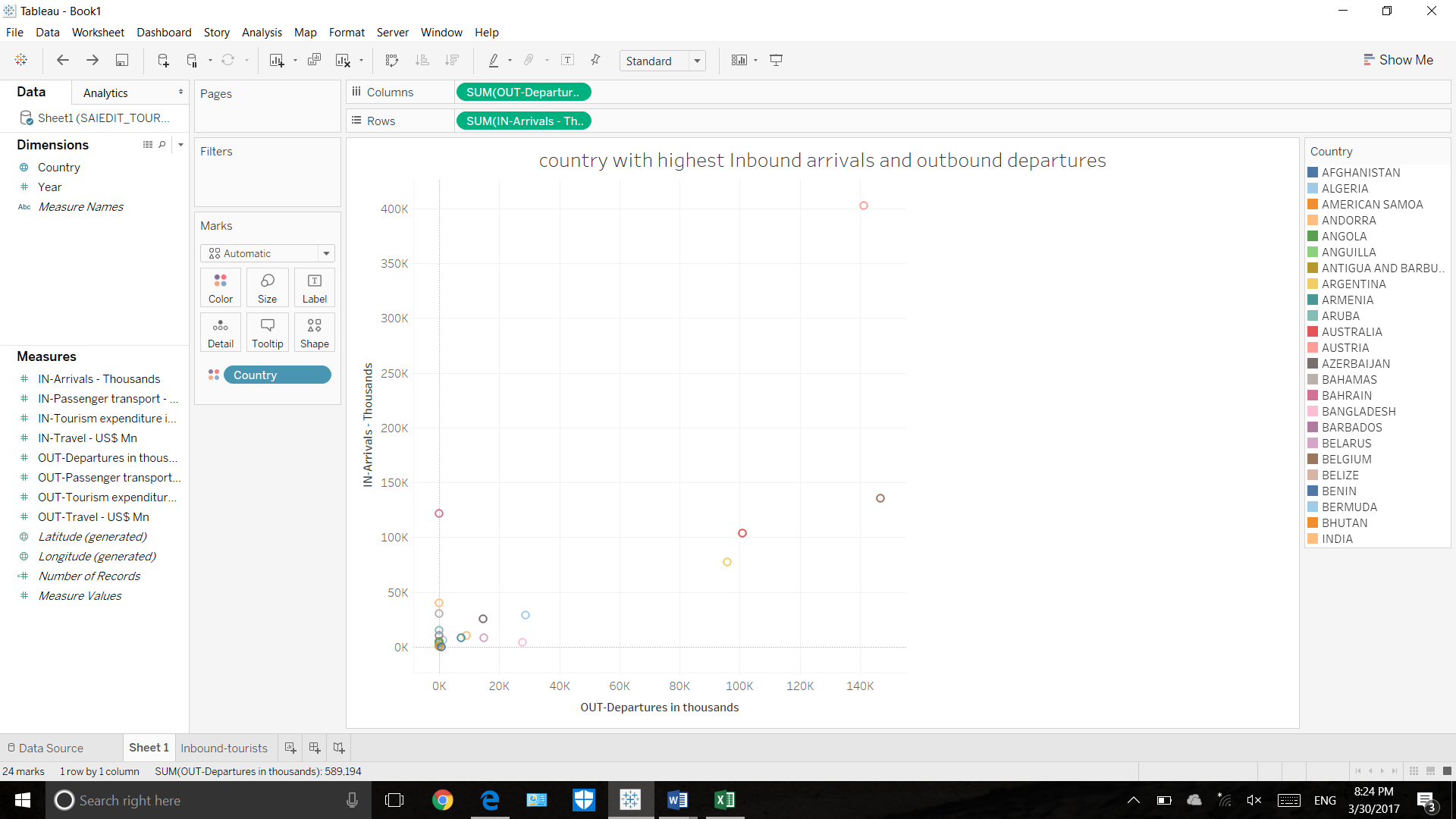
Here we are displaying countries with the sum of their outbound tourism expenditures for the year 1905. From the graph it is clear that Australia has highest outbound tourism expenditure. Here I have **grouped** ANGOLA, Angulia, Andorra to one and aliased that to Arabian countries. **Reference line** is also created to show which countries falls above average.



1. Which country has highest inbound arrivals and country with highest outbound departures.

Explanation:

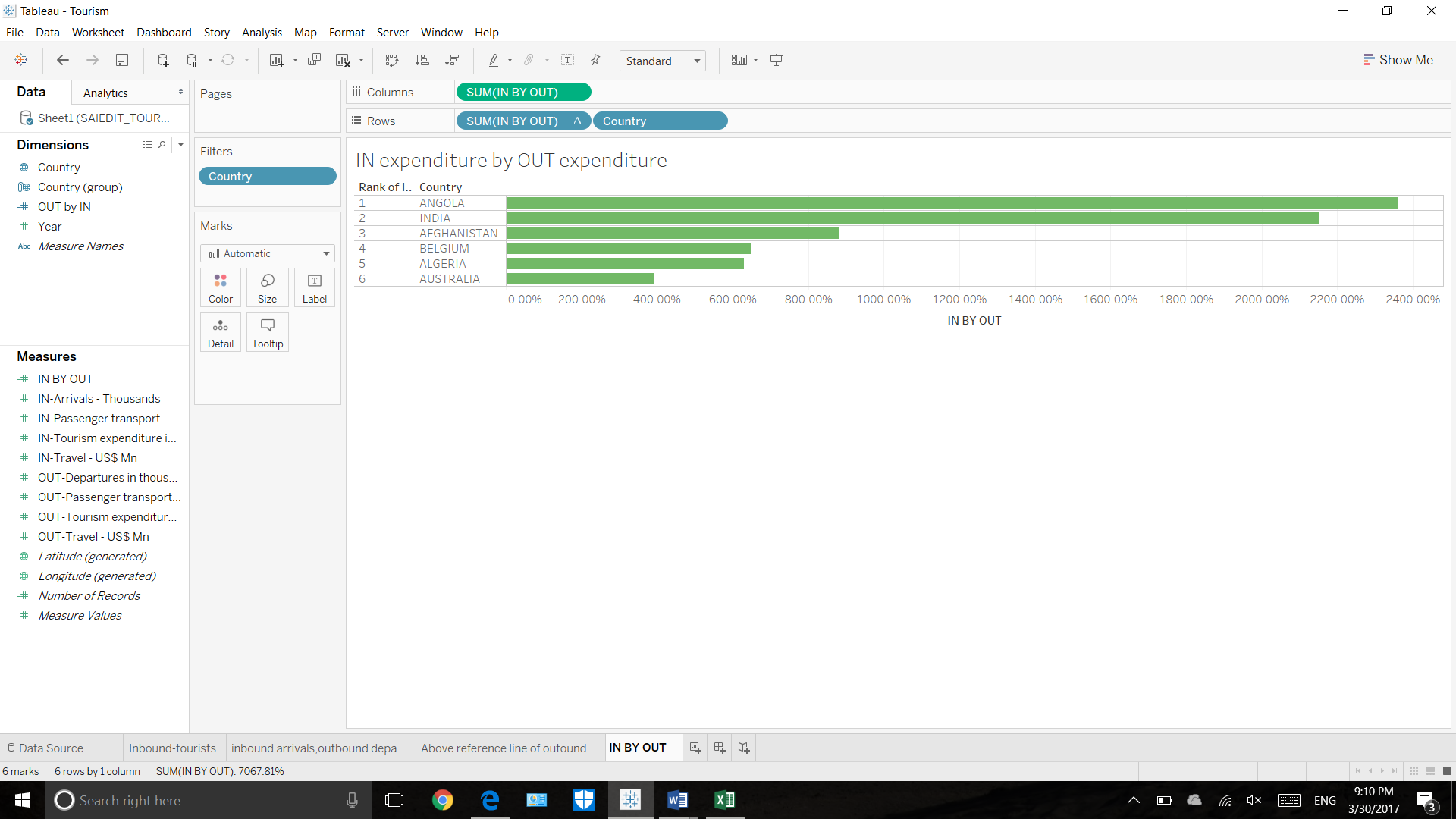
This shows countries with highest inbound arrivals and countries with highest outbound departures. I used **scatter plot** to show this visualization. From the graph it is clear that, Austria has highest inbound and Belgium has highest outbound departures. Countries are given different colors which are mentioned on the right hand side of visualization.



4. inbound expenditure is what percent of outbound expenditure for the following countries :AFGHANISTHAN,ALGERIA,ANGOLA,AUSTRALIA,BELGIUM,INDIA and rank them.

Explanation:

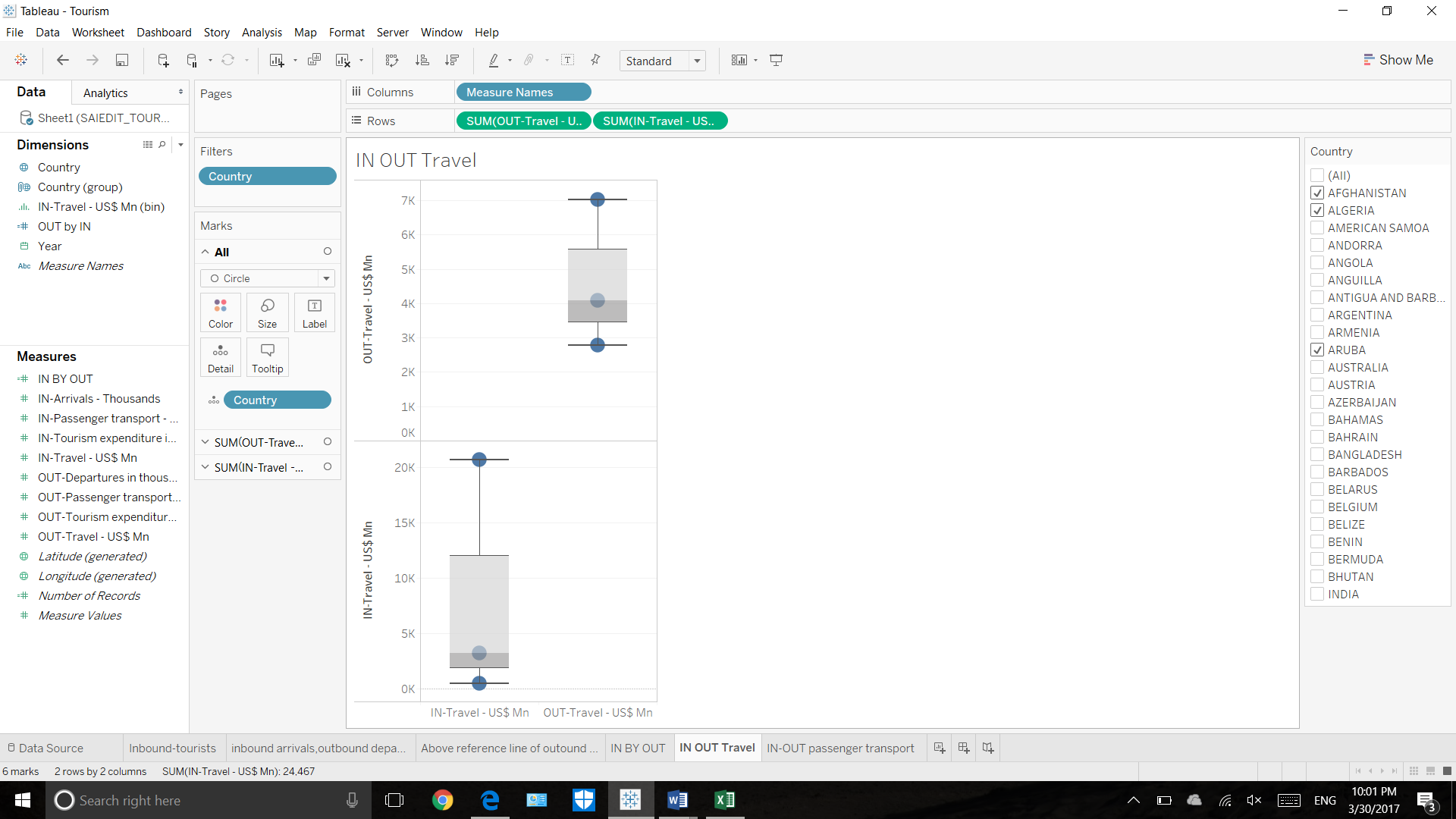
This visualization gives what percent of inbound expenditure is outbound expenditure for the countries Afghanistan, Algeria, Angola, Australia, Belgium and India. In order to do that, We have to create c**alculated field** by the formula . inbound expenditure / . outbound expenditure. Finally change this to percentage number format. Here given **ranks** also based on their percentages from highest to lowest.



5. display in and out travel US$ for countries: Aruba, Afghanistan, Algeria.

Explanation:

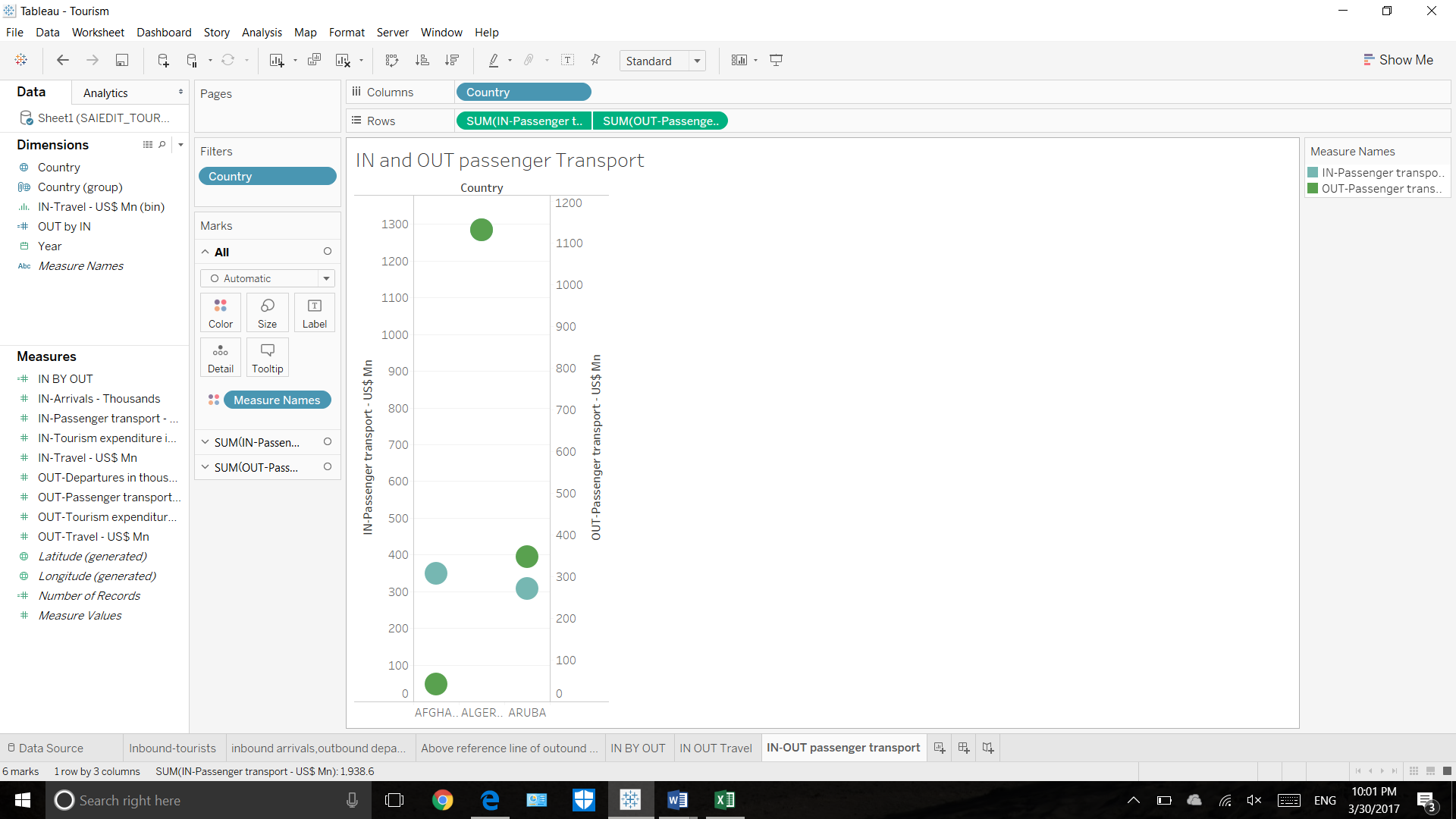
Here given the comparison display of in and out travel US$ for countries: Aruba, Afghanisthan, Algeria. Te comparsion is showed using the Box-Whisker plot. Measure names are given in order for better visualzation. On the right hand side we can see the filter for the countires.



6 .Display in and out passenger transport for Afghanisthan, Algeria,aruba.

Explanation:

Here we are going to display IN passenger transport to the countries and OUT passenger transport for the countries Afghanisthan, Algeria and Aruba. This gives a clear idea of number of people coming and going from the country and to the country.



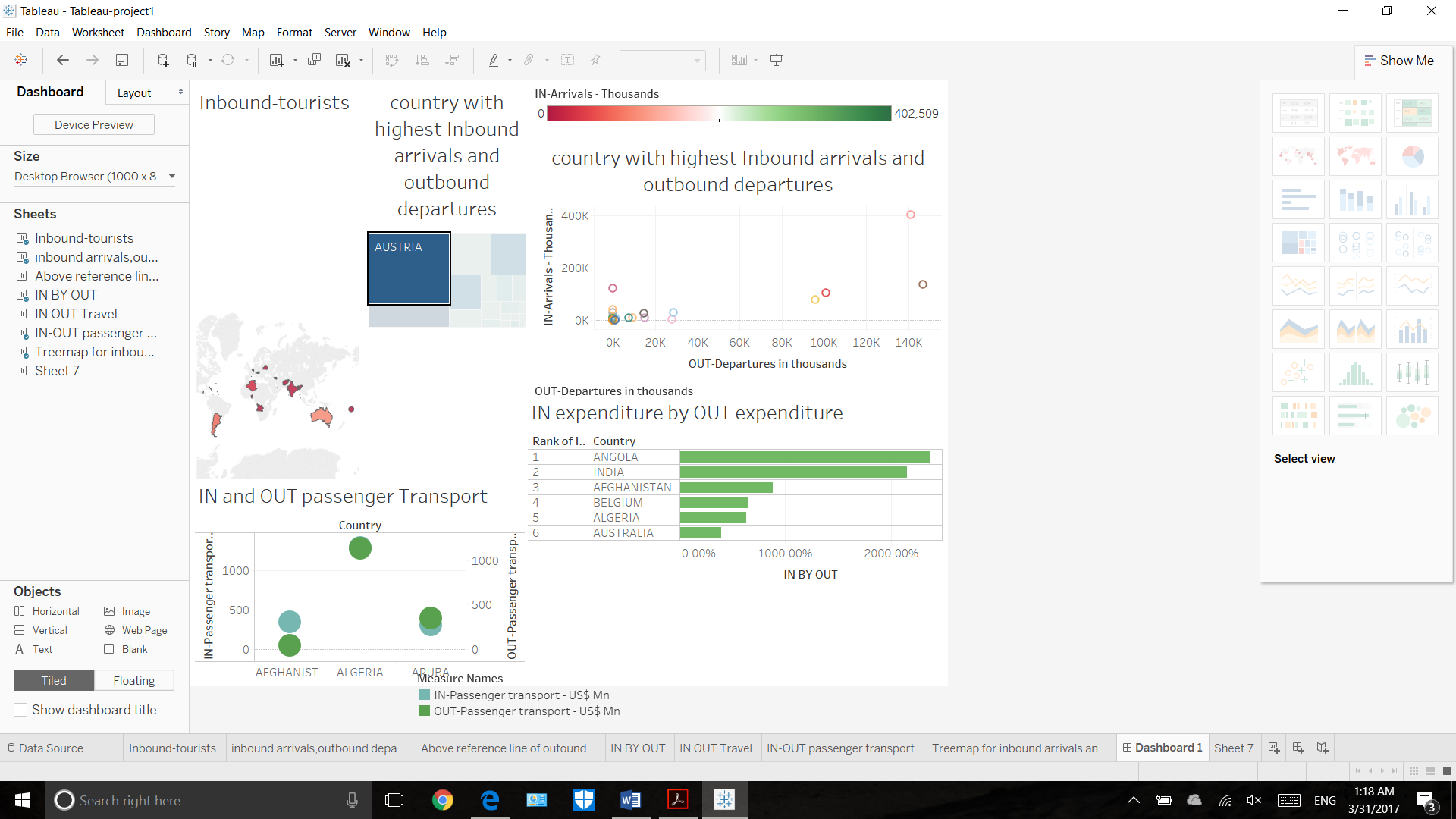
1. **Dashboard:**

Explanation:

This dashboard includes five graphs that includes

1. inbound tourists for the countries,
2. on and out passenger transport,(additional question)
3. IN expenditure by OUT expenditure percentage
4. country with highest inbound arrivals and outbound departures,
5. the country with highest inbound arrivals and outbound departures.(different visualization-additional question).

This dashboard is actually a accumulation of different visualizations. Just dragging the left hand side sheets will prepare you with dash board.



1. **STORY TELLING:**

This data set is about World Tourism Organization's Compendium of Tourism Statistics and Yearbook of Tourism Statistics. It gives statistics arrivals of nonresident tourists/visitors, departures and tourism expenditure in the country and in other countries. It also provides comprehensive information on international tourism worldwide and offer a selection of the latest available statistics on "arrivals of non-resident tourists/visitors" and "tourism expenditure". The best analytical project include data cleaning and visualization.

**Some information about the project:**

The dataset which deals with the tourism and gives a detailed information on the tourists that are travelled from the year 1903 to present except for some countries[1]. The details of the project, the incoming passengers, outgoing passengers, the geographical locations where they travelled , the brief description of the tourism data and also details on the expenders made are provided with this data set. The project “WORLD TOURISM DATASET” aims in analyzing the tourists activities that has happened in and around the world. This analysis would help in bringing into to picture the various unknown travel information around the world, which would help us analyze which countries are more attracted by tourists and which has yet to gain attention [1]. The article [1] gives us a brief idea on the best localities to visit in world.

Data Cleaning: Initially dataset is filled with empty spaces and some redundant rows. This makes difficult to analyze and view the data. I used data cleaning approaches like filling missed numeric values with 0 and deleting the redundant rows which makes data inconsistent. Data provided in data set is raw so we need to transpose some rows to columns to work with the dataset in tableau. Joining of tables and matching rows is not done as it is a single dataset. Finally, words are also correctly spelled. Visualization: Now that you have a narrative of your data, you can add interactivity to it. Interactivity is a critical element in tourism data. It lets people follow new records that are added to the dataset and it lets them test your assumptions – lending credibility to the story and to the data itself. Put together your narrative and the ability to explore, and you give your readers a data story[2].

**Why tourism and how it is surveyed:**

In 1976 Tourism Society of England defined it as "Tourism is the temporary, short-term movement of people to destination outside the places where they normally live and work and their activities during the stay at each destination[3]. It includes movements for all purposes." In 1981 International Association of Scientific Experts in Tourism defined Tourism in terms of particular activities selected by choice and undertaken outside the home environment. Domestic tourism, involving residents of the given country traveling only within this country;   
Inbound tourism, involving non-residents traveling in the given country; Outbound tourism, involving residents traveling in another country. UN also derived different categories of tourism by combining the 3 basic forms of tourism:

* Inbound tourism, National tourism, which comprises domestic tourism and
* Internal tourism, which comprises domestic tourism.
* Outbound tourism, International tourism, which consists of inbound tourism and outbound tourism.

**Analysis from the given data set:**

So we can analyze the tourism data with the help of world Tourism Organization[1]. This Statistics and Tourism compiles and disseminates tourism statistics of countries all over the world, while advancing and standardizing the methodological frameworks for measuring tourism and expanding its analytical potential.



From the given world tourism dataset, we can find number of tourists visiting each country over the years 1905 to 2014. Apart from this it also included the expenditure they are going to afford in each country and how much the country is earning. In the same way, They calculated for outbound tourists also[1]. Like, number of people visiting other countries and how much they are spending there, Finally each country is going to earn how many US$. By using the analytics tool, we can calculate number of tourists visiting our country and number of tourists going each year to other countries. We can also find out which is the best viewed country over the period. We can fnd some fun math calculations also like how much more is inbound tourists expenditure when compared to outbound tourists expenditure[3]. Which country has highest inbound tourists likewise analytical questions we can find from this dataset.

Understanding, for each country, where its inbound tourism is generated is essential for analysing international tourism flows and devising marketing strategies, such as those related to the positioning of national markets abroad. Deriving from the most comprehensive statistical database available on the tourism sector, the **world tourism data set** focuses on data related to inbound tourism

When a person reads the this dataset with appropriate cleaning, he will get a clear idea of which places are most likely to get visited and which are historic places (this can be done by analyzing the dta of 1905 travels). This way, information is provided to the viewers in clear cut manner with better visualization schemes using analytical tools. Travel data nerds can rejoice, because today they have an economy-sized bucket of stats to pore over regarding the cities getting the most tourists and tourist dollars.

URL’s :

[1] : https://www.wttc.org

[2] : <http://www2.unwto.org/content/data>

[3]: http://www2.unwto.org/content/who-we-are-0