COMMUNICATING THE RESULTS OF DATA ANALYTICS

Sumedh Ambokar [50207865] | Kaushik Ramasubramanian [50207352]

History of Modern Olympic Games

Introduction:

Modern Olympics Games started in the 1896 with the first being held in 1896 in Athens, Greece. Over the years there have been thousands of athletes participating in the games hailing from hundreds of countries. We made an attempt to check the history of Modern Olympic games and perform data analytics with help of official data set of all the Olympic medalists spanning from the 1896 Olympic Games held in Athens to the London Olympics Games in 2012.

Data Set:

The project is based on a data set comprising of two excel files which provide comprehensive details of the medal winning athletes and the participated countries. The data used is from open source data set downloaded from Kaggle[1].

File 1: summer.csv

- This File comprises of all the Olympic Medalists from 1896-2012 Olympic Games.
- Each Medalist is supported by the attributes:
  - **Year:** Year in which the Olympic Games were held.
  - **City:** City in which the Olympic Games were held.
  - **Sport:** This Identifies the Sport Domain in which the medalist won e.g. Aquatics, Athletics etc.
  - **Discipline:** This identifies the Discipline under the sport in which the medalist won e.g. Swimming, Running etc.
  - **Athlete:** Provides name of the medalist.
  - **Country:** Represents country code of the medalist e.g.: HUN-Hungary
  - **Gender:** Gender of the medalist.
  - **Event:** This identifies the actual event in which the medalist won.
  - **Medal:** Type of the medal won by the medalist.
- This data file can be used to extract information related to the medal winning athletes, medal winning countries, events of every Olympic Games, details of a respective Olympic Games and much more.
- A detailed description of how the dataset has been used to perform analytics is given later.

File 2: dictionary.csv

- This File comprises of all the countries which have participated in the Modern Olympic Games.
- Each country record is supported by following attributes:
  - **Country:** Name of the country.
  - **Code:** Country Code of respective country.

  - **Population:** Official total population in 2012.
  - **GDP per Capita:** Gross domestic product used to estimate country's economic status.
- This data sheet provides additional information about the participating countries in Olympic Games. This data has been used to support the main data set of Olympic medalists.

Data Analysis:

- The data provided is clean and does not require further cleaning or modifications. Thus, we have used the excel files directly as our data sets.
- Both the files were added to the Tableau application by using the feature exporting from an excel file.
- We wanted to analyze only the data for athletes and countries which have won at least one medal in the Olympic Games.
- Thus, we created an inner join between the data sets provided by the two files with the 'Country' field from summer.csv linked with 'Code' field from dictionary.csv.
- The combined data from both the files was used for creating interactive dashboards based on informative charts, also to analyze and predict the data.

Dashboards:

For this project, we have created two dashboards and many worksheets which are displayed as the part of dashboards. Below is the detailed description of the dashboards and the worksheets.

Dashboard 1: History of Modern Olympic Games

Summary:

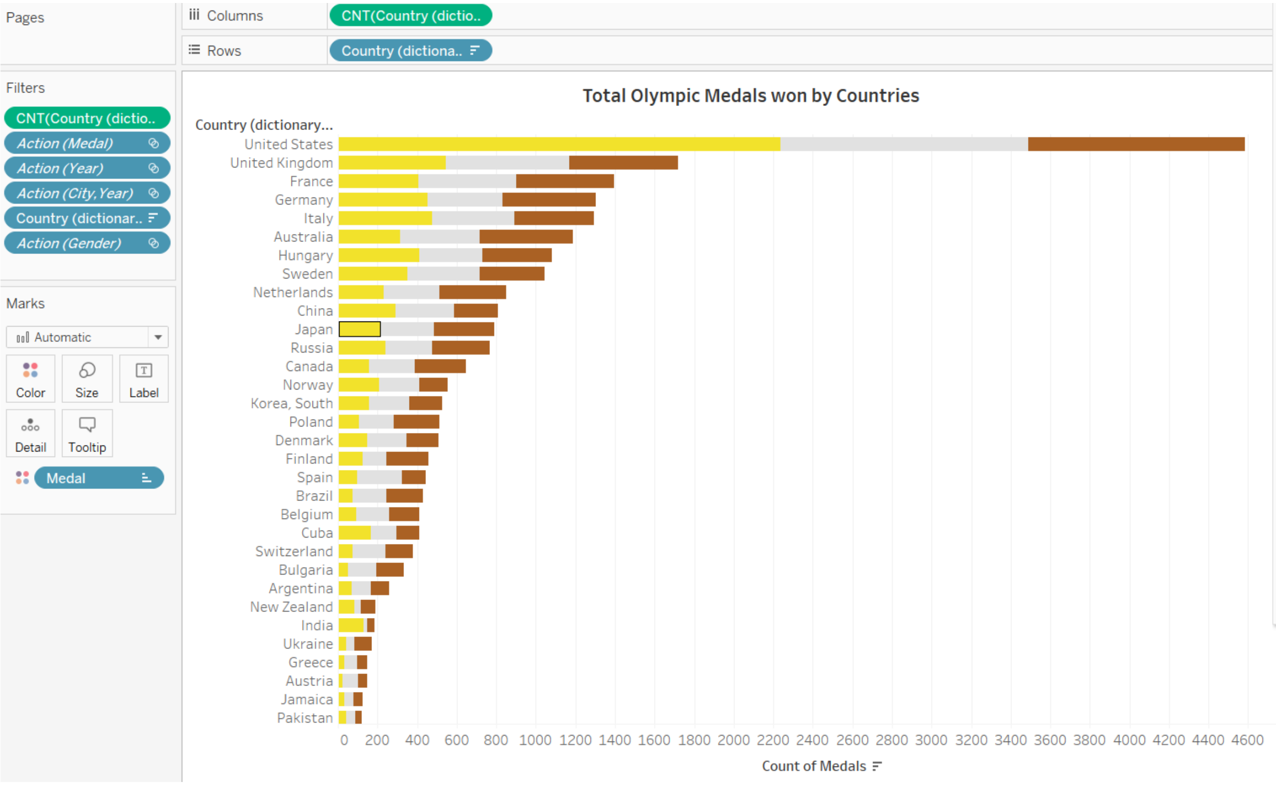
- This Dashboard has been created for Activity two. It represents the details of all the Olympic Games from 1896-2012 based on the dataset.
- We have focused on the entities Country, Athlete and Medals. Each of the worksheet represents distinct information related to the specified entities and more.
- The dashboard has been made interactive by adding various filters and selection options to refine the results and help users to access specific set of data from the complete dataset.

Worksheets:

There are total seven worksheets displayed on the dashboard. With four being used to display data and three for filtering the data on the dashboard. Below are the details of each worksheet displayed:

Display Worksheets:

1. Total Olympic Medals won by Countries:



Pages

Columns: CNT(Country (dictio..

Rows: Country (dictiona..

Filters:
- CNT(Country (dictio..
- *Action (Medal)*
- *Action (Year)*
- *Action (City, Year)*
- Country (dictionar..
- *Action (Gender)*

Marks: Automatic

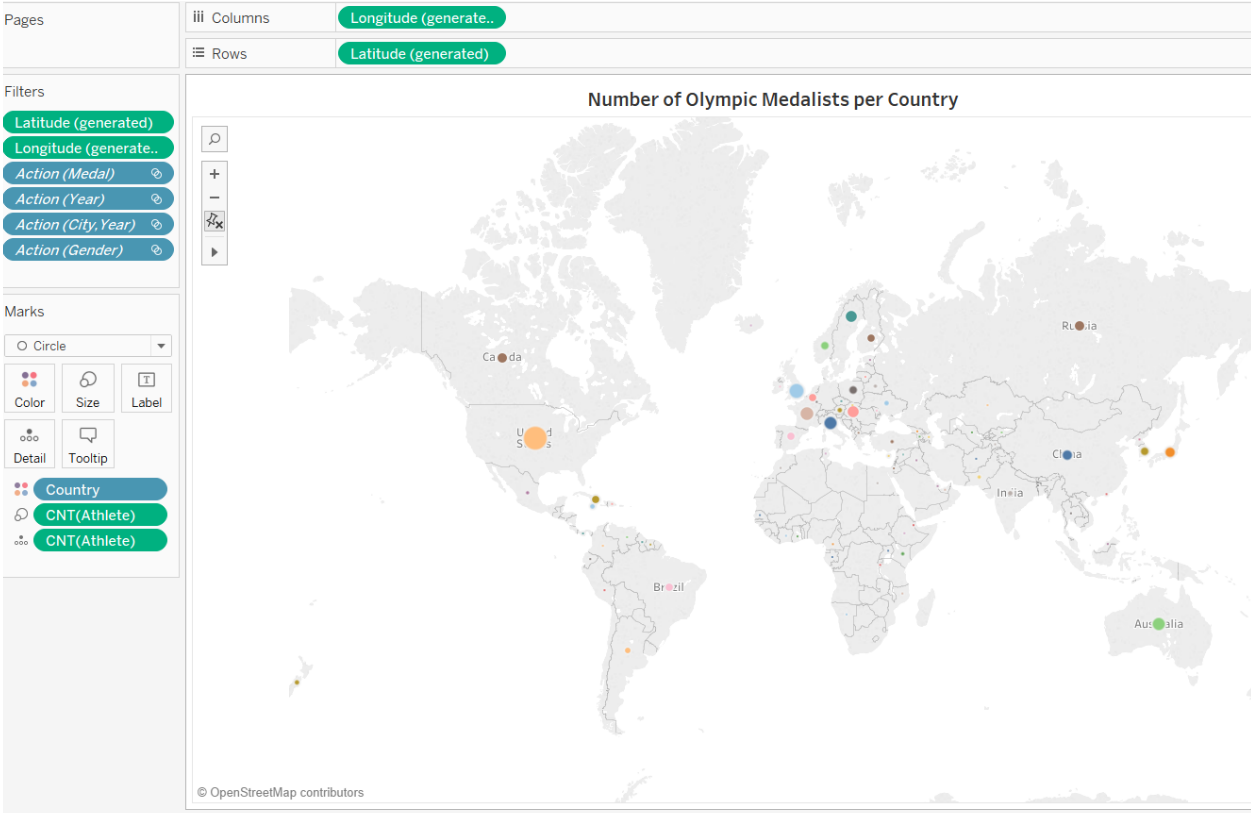
Color · Size · Label · Detail · Tooltip

Medal

Details:

- This worksheet represents the number of Olympic Medals won by each country. The Medal tally is specified for each type of medal i.e. Gold, Silver and Bronze.
- **Rows:** Rows are represented by the country name. We have used the Country field from the dictionary.csv file as it contains the complete name of the country and not just the code, to represent the countries on the Y-axis.
- **Columns:** Columns have been set to the count of the country field as we need total number of medalist records grouped by country attribute.
- **Filters:**
  - We wanted data for all the countries to be represented in this worksheet, so we have not added any filters to this worksheet.
- **Marks:**
  - **Color:** We wanted to represent the tally of every type of medal separately or specified by a distinct color. Thus, we included 'Medal' field in the color Marks and assigned appropriate colors to the type of the medals.

2. Number of Olympic Medalists by Country:



Pages

Columns: Longitude (generate..

Rows: Latitude (generated)

Filters
- Latitude (generated)
- Longitude (generate..
- *Action (Medal)*
- *Action (Year)*
- *Action (City,Year)*
- *Action (Gender)*

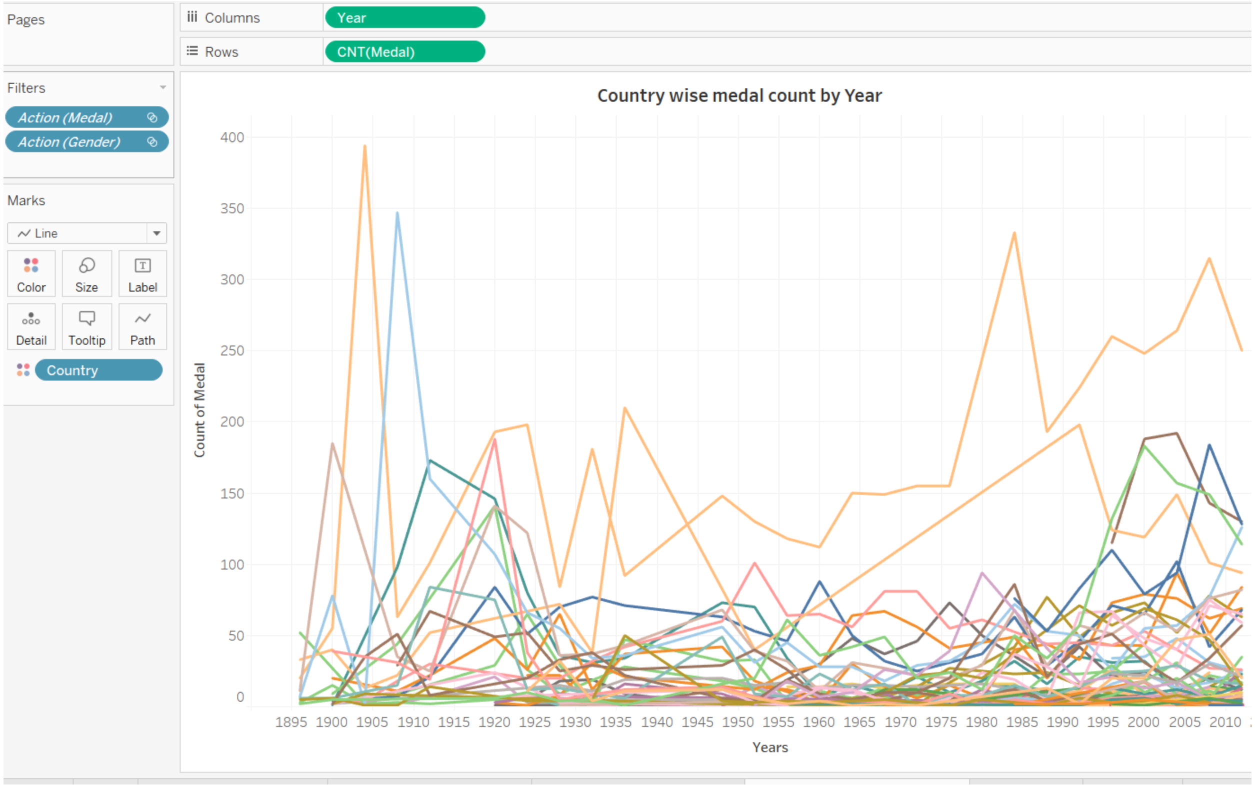
Marks
- ○ Circle
- Country (Color)
- CNT(Athlete) (Size)
- CNT(Athlete) (Detail)

Number of Olympic Medalists per Country

Details:

- This worksheet, on a World Map, shows the number of Olympic medalists belonging to every country that participated in the Olympic Games. This is based on cumulative data from all the Modern Olympic Games organized.
- **Rows:** Latitude of the location of individual countries auto generated in Tableau.
- **Columns:** Longitude of the location of individual countries auto generated in Tableau.
- **Filters:**
  - We have put filters on the Longitude and Latitude to avoid any null values.
- **Marks:**
  - **Color:** We have added Country field to the color Marks to represent data for every country in a distinct color.
  - **Size:** Count of Athletes is used to modulate the size of the bubble displayed against every country in the map. More the number of Athletes from the country who have participated in the Olympic Games bigger is the size of the bubble.

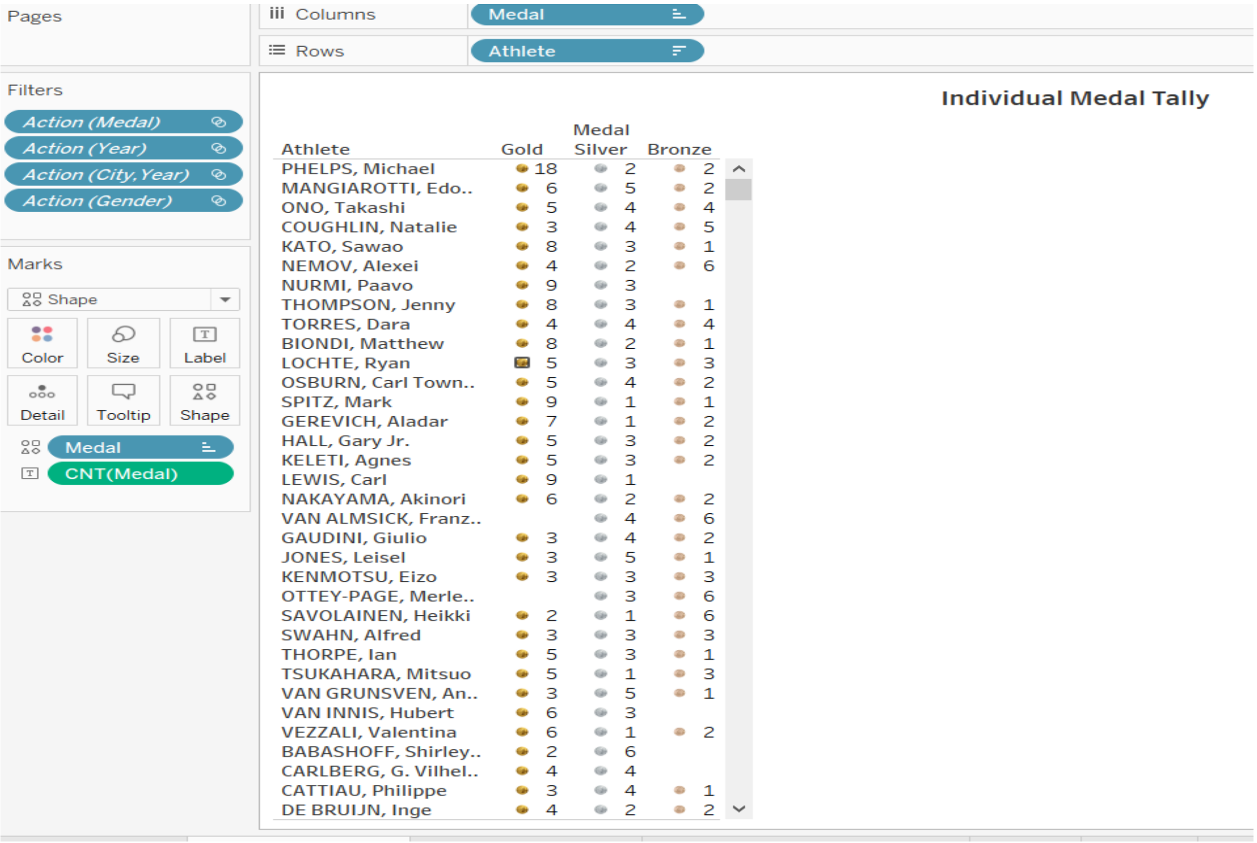
3. Country wise medal count by year:



Details:

- This worksheet shows a line graph representing the number of medals achieved by a country in a particular edition of Olympics. Each line represents the plot for a different country.
- **Rows:** For rows, we have used count of the Medal as we need to total number of medals for each country grouped by year.
- **Columns:** Columns have been set to the field 'Year', so that the data is grouped by year specified.
- **Filters:** No additional filters have been added specifically to this worksheet as the data filtering is taken care at the dashboard level.
- **Marks:**
  - **Color:** We have added field Country in the color mark so that every country is represented by a distinctly colored line.

4. Individual Medal Tally:



Pages

Columns: Medal

Rows: Athlete

Filters

- Action (Medal)
- Action (Year)
- Action (City, Year)
- Action (Gender)

Marks

Shape

- Shape: Medal
- Label: CNT(Medal)

Individual Medal Tally

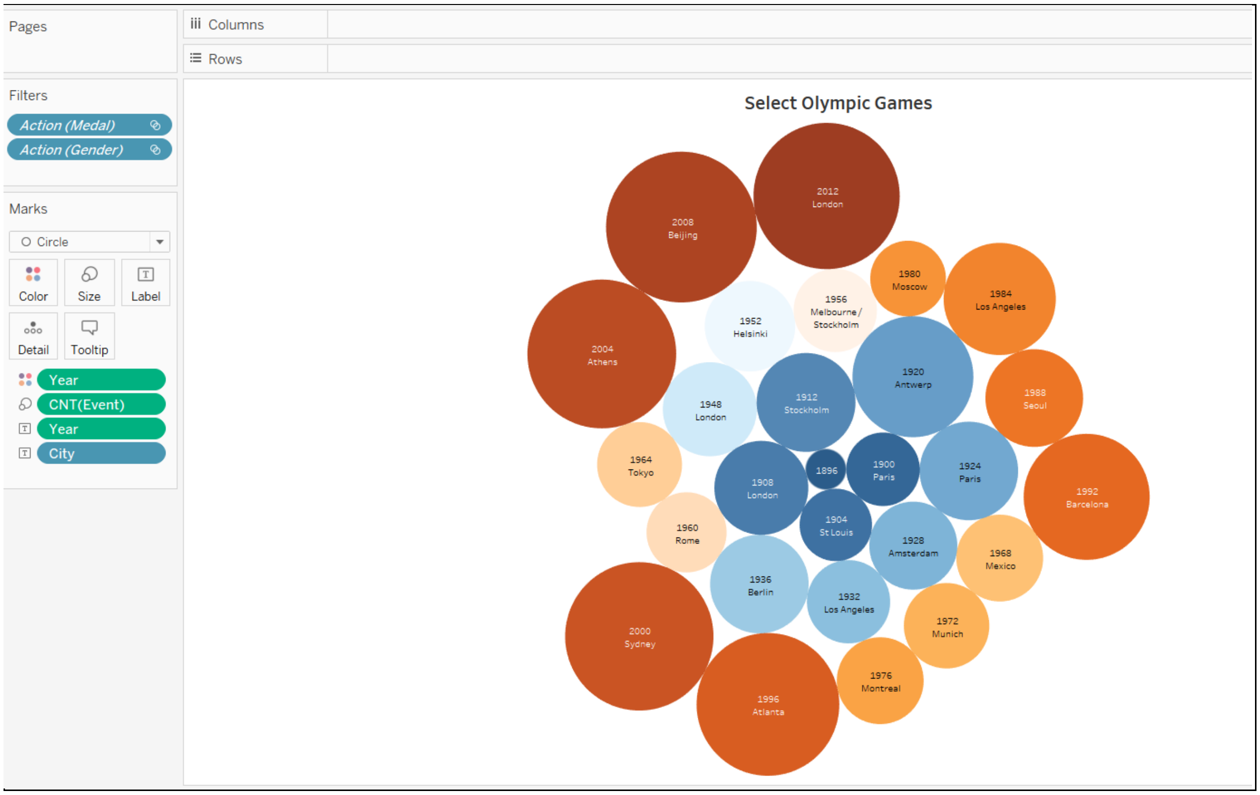
	Medal		
Athlete	Gold	Silver	Bronze
PHELPS, Michael	18	2	2
MANGIAROTTI, Edo..	6	5	2
ONO, Takashi	5	4	4
COUGHLIN, Natalie	3	4	5
KATO, Sawao	8	3	1
NEMOV, Alexei	4	2	6
NURMI, Paavo	9	3	
THOMPSON, Jenny	8	3	1
TORRES, Dara	4	4	4
BIONDI, Matthew	8	2	1
LOCHTE, Ryan	5	3	3
OSBURN, Carl Town..	5	4	2
SPITZ, Mark	9	1	1
GEREVICH, Aladar	7	1	2
HALL, Gary Jr.	5	3	2
KELETI, Agnes	5	3	2
LEWIS, Carl	9	1	
NAKAYAMA, Akinori	6	2	2
VAN ALMSICK, Franz..		4	6
GAUDINI, Giulio	3	4	2
JONES, Leisel	3	5	1
KENMOTSU, Eizo	3	3	3
OTTEY-PAGE, Merle..		3	6
SAVOLAINEN, Heikki	2	1	6
SWAHN, Alfred	3	3	3
THORPE, Ian	5	3	1
TSUKAHARA, Mitsuo	5	1	3
VAN GRUNSVEN, An..	3	5	1
VAN INNIS, Hubert	6	3	
VEZZALI, Valentina	6	1	2
BABASHOFF, Shirley..	2	6	
CARLBERG, G. Vilhel..	4	4	
CATTIAU, Philippe	3	4	1
DE BRUIJN, Inge	4	2	2

Details:

- This worksheet specifies the number of Olympics medals won by individual athletes.
- **Rows:** Rows is set to field Athlete, so that all the athletes are represented on the Y-axis.
- **Columns:** Columns is set to field Medal, so that the data is split into three columns representing each type of medal.
- **Filters:** No additional filters have been added specifically to this worksheet as the data filtering is taken care at the dashboard level.
- **Marks:**
  - **Shapes:** We wanted to display an image of respective type of medal. Thus, we added Medal to the shapes and selected custom images from the shapes option for individual medal type. Custom images were provided to the worksheet by adding the images in the ‘Local Tableau Repository’ under the ‘Shapes’ directory.
  - **Labels:** As we needed count of the medals against every Athlete, thus we added Count of Medals to the Labels.

Filtering Worksheets:

1. Select Olympic Games



Details:

- This worksheet represents the Olympic Games and its details like the number of events played, year in which it was held and location. We have used Bubbles to represent the Olympic Games held.
- **Marks:**
  - **Color:** We wanted to represent every Olympic Games distinctly using colors, thus added Year of the Olympic Games to the Color Mark.
  - **Size:** We wanted to represent Olympic Games with more events with a bigger bubble, thus Count of Event was added to the Size Mark.
  - **Labels:** Two labels Year and City have been added to the Labels Mark to display the details of the Olympic Games.

2. Select Medal

Details:

- This worksheet was created to refine data on the dashboard using the type of the medal.
- **Marks:**
  - **Shapes:** We wanted the list of types of medals to be displayed along with the image of medals. Thus, Medal field was added to Shapes Mark and custom images for the medals were selected for individual medal types.

3. **Select Gender**

Details:

- This worksheet was created to refine data on the dashboard using gender of the Athlete.
- **Marks:**
    - **Shapes:** We wanted the list of Genders to be displayed along with respective image. Thus, Gender field was added to Shapes Mark and custom images for individual gender were selected.

Additional Filters:

- **Filter by Country:** A country filter has been added to the dashboard
- **Filter by Athlete:** A filter based on Athlete name is also added.

Dashboard 2. Classification Dashboard

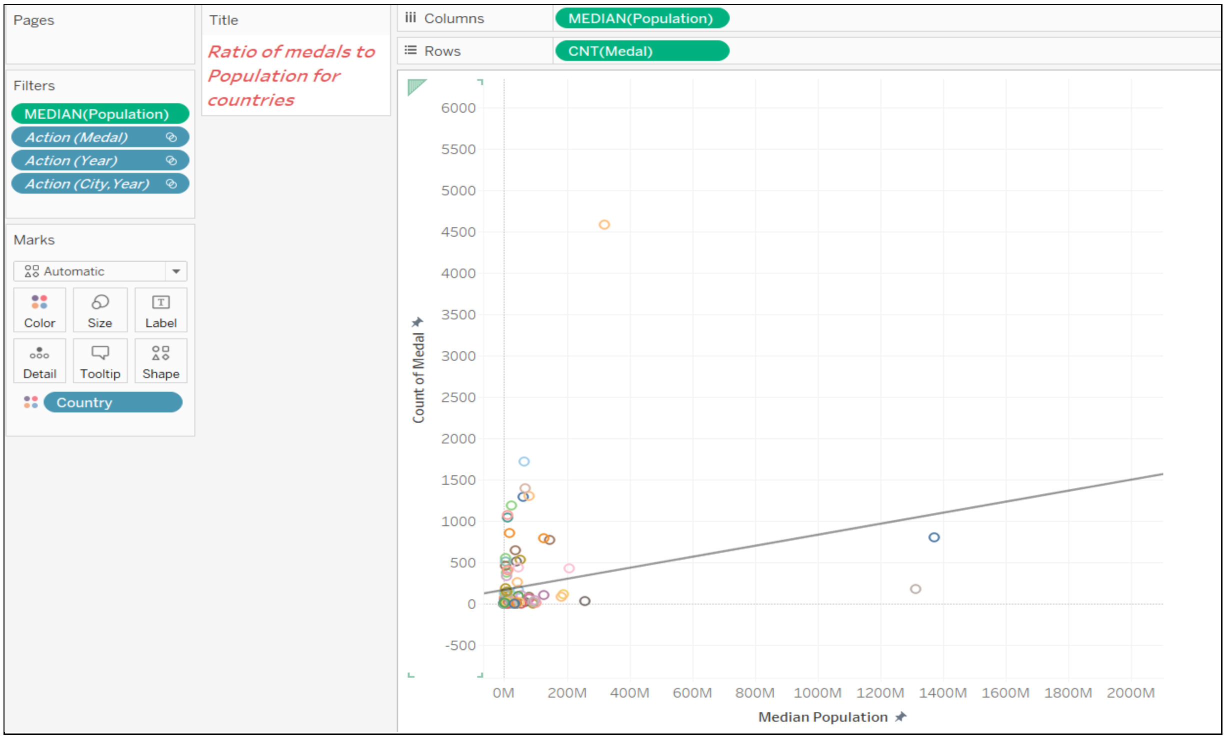
Summary:

- This dashboard shows the visualization for the classification performed on the dataset. This covers the Activity three for the Project.

Worksheets:

1. **Linear Regression Model:**
    - Linear regression is an approach for modelling the relationship between a scalar dependent variable Y and one or more explanatory variables denoted by X.
    - In this case, we performed a linear regression model in order to represent the proportion of medals secured by a particular country participating in Olympics.
    - As a result of this, it would be clear as to which countries have achieved more medals for the proportion of their population.
    - A scatter plot was formed based on the Ratio of medals to Population for various countries.
    - We could observe that after plotting, most of data was nearby origin and there were few outliers.
    - The outliers tend to change the orientation of the regression model.
    - For example, countries like India, China have high Population. But India hasn't secured many Olympic medals, but China has. This makes the orientation of the linear model to go downwards.
    - On the other hand, countries like USA, GBR and France have high medal count but they don't have as much of a population count.
    - Thus, it can be interpreted from this linear model that there are certain countries who have large population but aren't able to win as many medals. Thus, it could be inferred that these countries need to take special care in the kind of infrastructure they provide for the sports to flourish.

- They could probably take guidance from the countries with small population but having high medal tally, and look at the coaching facilities provided in these countries so that there could be significant improvement in their countries as well.
- Thus, using the analysis done in this data countries could very well improve the standard of sports in their countries.
- Further there was a LM model drawn for this dataset. It was observed that this model has **R-squared value of 0.0491899 and P-value of 0.0125631.**
- Using this R-squared value stated above it can be concluded that the model represents about 4% variation in the data. The greater R-squared value the better model.
- Since a lesser p-value shows that the model fits the data better. Here p-value of 0.0125631 indicates that the model is appropriate to the corresponding data.
- **Rows :** The plot was formed using the count of medals on the Y-axis.
- **Columns:** Median of the population was used to plot in the X-axis.
- **Marks:** The color marks on the scatter plot indicate the countries.



Pages

Filters
MEDIAN(Population)
Action (Medal)
Action (Year)
Action (City,Year)

Marks
Automatic
Country

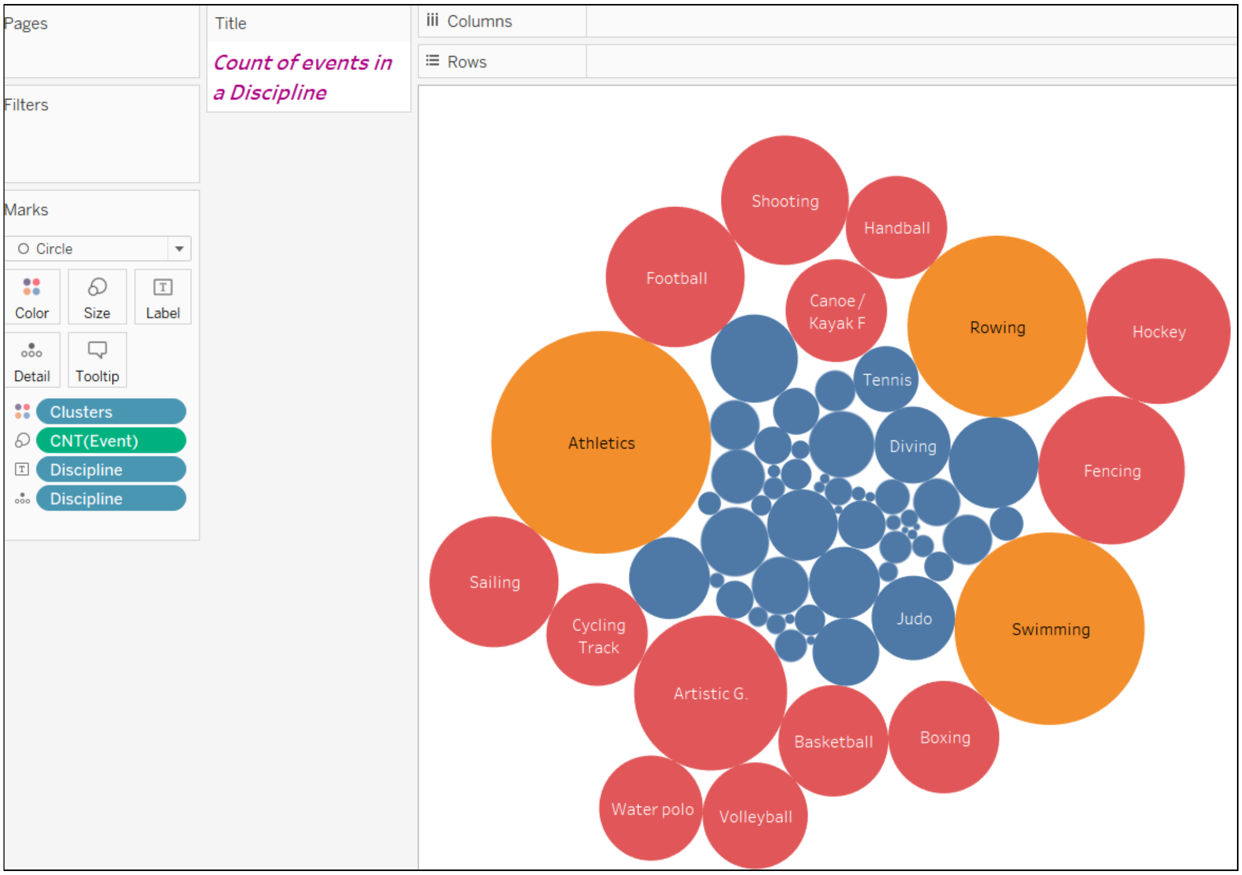
Title
Ratio of medals to Population for countries

Columns: MEDIAN(Population)
Rows: CNT(Medal)

2. **Clustering:**

- Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).
- In this case, we have used k-means clustering where 'k' indicates the number of clusters.
- Here, we tried to interpret the number of events corresponding to a particular discipline.
- As a result of this, we can analyze the different types of events present under a particular discipline in the Olympics.
- Using this, we can infer the popularity of a particular discipline, as there can be more number of disciplines only there are more people interested in the sport.

- For example, there are more events under swimming as that sport is more versatile.
- In this case, we used packed bubbles to plot the count of events corresponding to the discipline.
- It is evident from the plot, that Athletics, Rowing and Swimming have large number of events under them as these are more versatile events.
- On the other hand, events like tennis, Lacrosse, Golf don't have many events under them.
- Therefore, an unsupervised classification on the data using K-means clustering was done. The data containing the count of events in a discipline was divided into 3 clusters. However, the number of clusters can be modified by the user based on his/her needs as it is interactive.
- **Marks:**
    - **Size:** Size of the bubbles is set using the count of the events from a particular discipline.
    - **Label/Detail:** Discipline field has been added to the Label and Detail Marks.
    - **Color:** We have used colors to distinguish the cluster so the clusters have been added to the Colors Mark.



Title: *Count of events in a Discipline*

Marks: ○ Circle

Clusters

CNT(Event)

Discipline

Discipline

Dashboard Links:

- **History of Olympics games -** https://public.tableau.com/profile/kaushik5909#!/vizhome/HistoryofModernOympicGames/OlympicGamesDashboard
- **Classification -** https://public.tableau.com/profile/kaushik5909#!/vizhome/ClassificationDashboard/Classificationdashboard

STORY

We have consolidated the important worksheets to from a series displaying the medal count, medalist count and much more along with the dashboards related to entities like Country, Athletes and Olympic Events.

Link: https://public.tableau.com/profile/kaushik5909#!/vizhome/Story_123/Story

References/Links:

1. Download Dataset from: https://www.kaggle.com/the-guardian/olympic-games
2. **Dashboard 1: History of Olympics games -** https://public.tableau.com/profile/kaushik5909#!/vizhome/HistoryofModernOympicGames/OlympicGamesDashboard
3. **Dashboard 2: Classification -** https://public.tableau.com/profile/kaushik5909#!/vizhome/ClassificationDashboard/Classificationdashboard
4. **Story:** https://public.tableau.com/profile/kaushik5909#!/vizhome/Story_123/Story