**PROJECT REPORT**

**BUSINESS INTELLIGENCE AND DATA VISUALIZATION**

**BOLLYWOOD MOVIES**

**MADE BY:-NIKHIL CHAWLA(18CSU140)**

**PALAK DHINGRA(18CSU151)**

**PRANAV GOEL(18CSU158)**

***ABOUT THE TOPIC***

Our project is based on the bolllywood movies released since 1992 to 2018. It tells us about the movie title, actors staring in the movie, its release date, earnings, director, producer and many more things.

***DATASET DESCREPTION***

Name of columns : Actors 1, Actors 2, Actors 3, Actors 4, Actors 5, Actors 6, Actors 7, Budget, Dialogue, Director, Directors, First Day, First Week, First Weekend, Genre, India Gross, Lyrics, Music, Overseas Gross, Producer, Production Banner, Release, Date, Runtime, Screenplay, Screens, Story, Title, Worldwide Gross.

We furthur created calculated fields.

***REASON FOR CHOOSING***

We found this dataset very interesting. It tells us about a lot of bollywood movies that are released since 1992. the dataset tells how these movies did on the box office. We also get to know the first day earnings, first week earning and also overall how much the movie earned. It tells about not only how well it did in india but also worldwide.

***VISUALIZATION ANALYSIS***

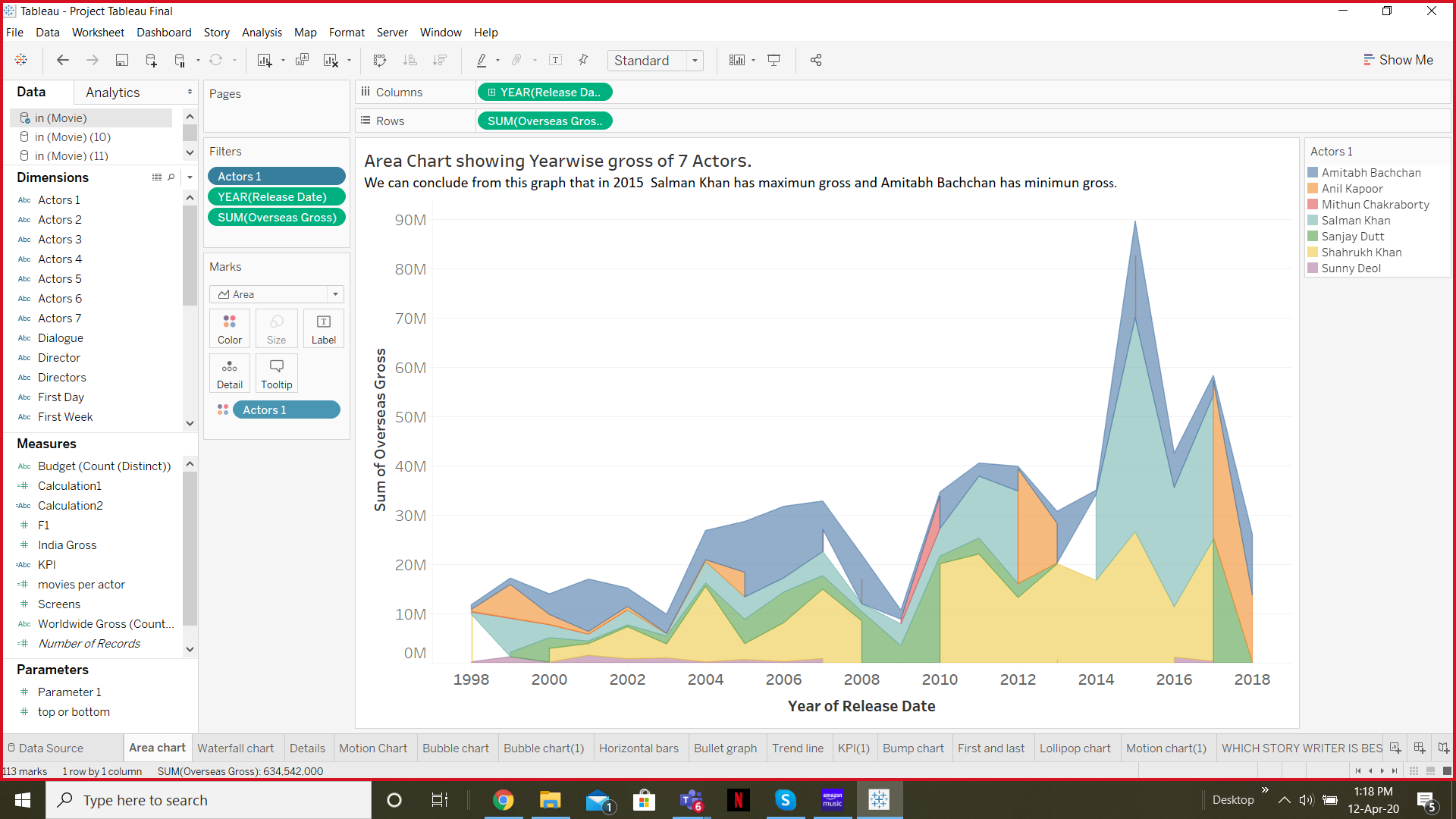
**Area Chart:-**

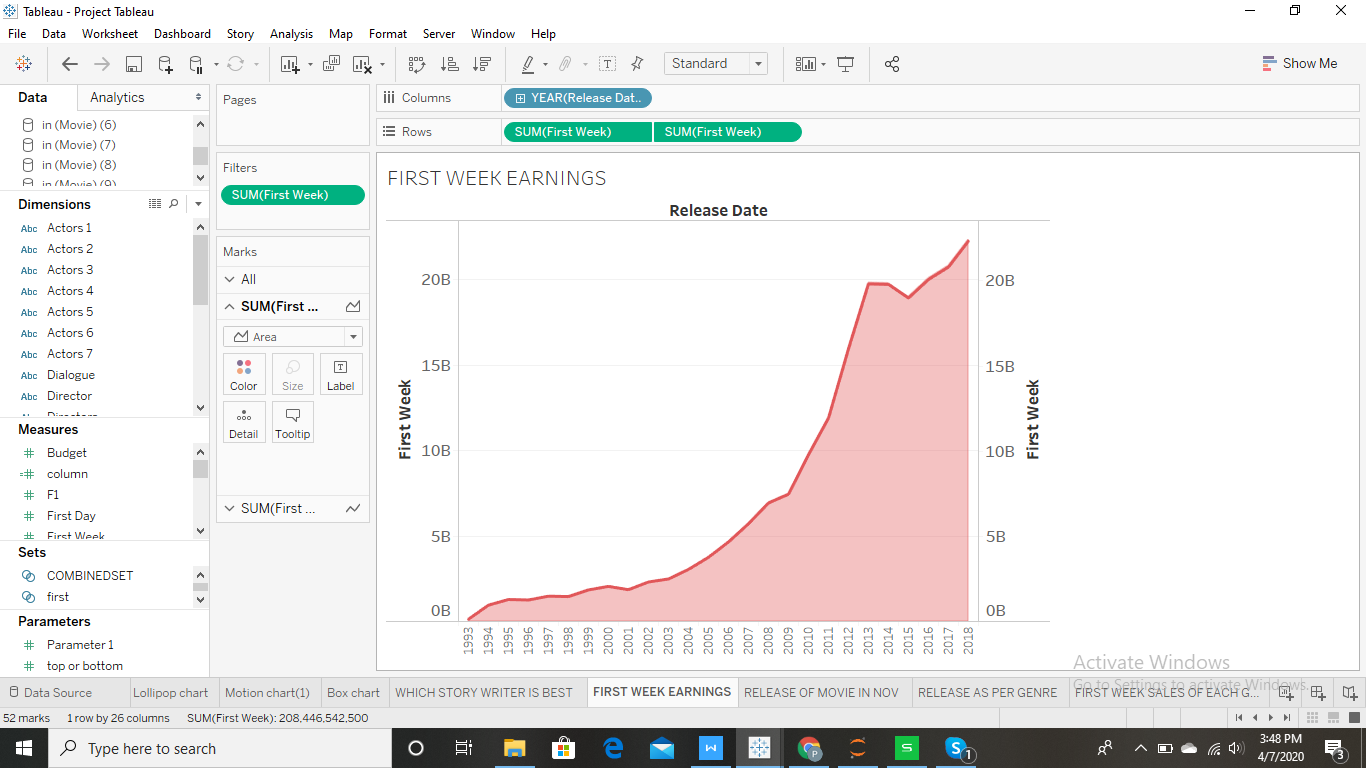
The **area chart** is a combination between a line **graph** and a stacked bar **chart**. It shows relative proportions of totals or percentage relationships.

Area chart showing year wise gross of actors

We can conclude from this graph that in 1998 mithun chakraborty has maximun gross and aamir khan has minimun gross.

Area chart showing first week collection of movies year wise. From this we can conclude that earlier people were not that much into movies but now a days the craze of movies has increased so much that as soon as a movie is released people go to there near by theaters to watch it.

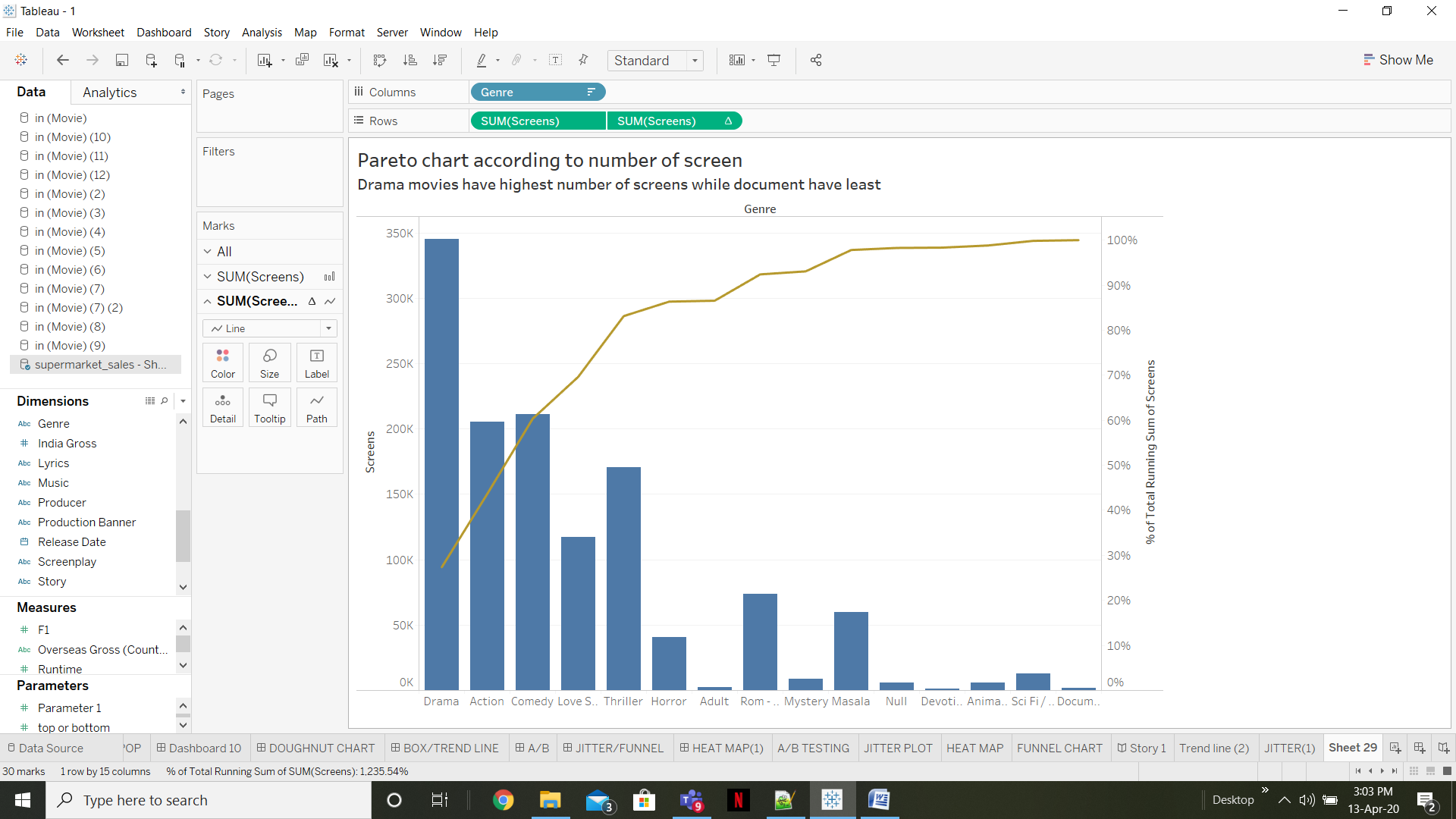




**Pareto Chart:-**

A **Pareto chart** is a type of **chart** that contains both bars and a line **graph**, where individual values are represented in descending order by bars, and the cumulative total is represented by the line.

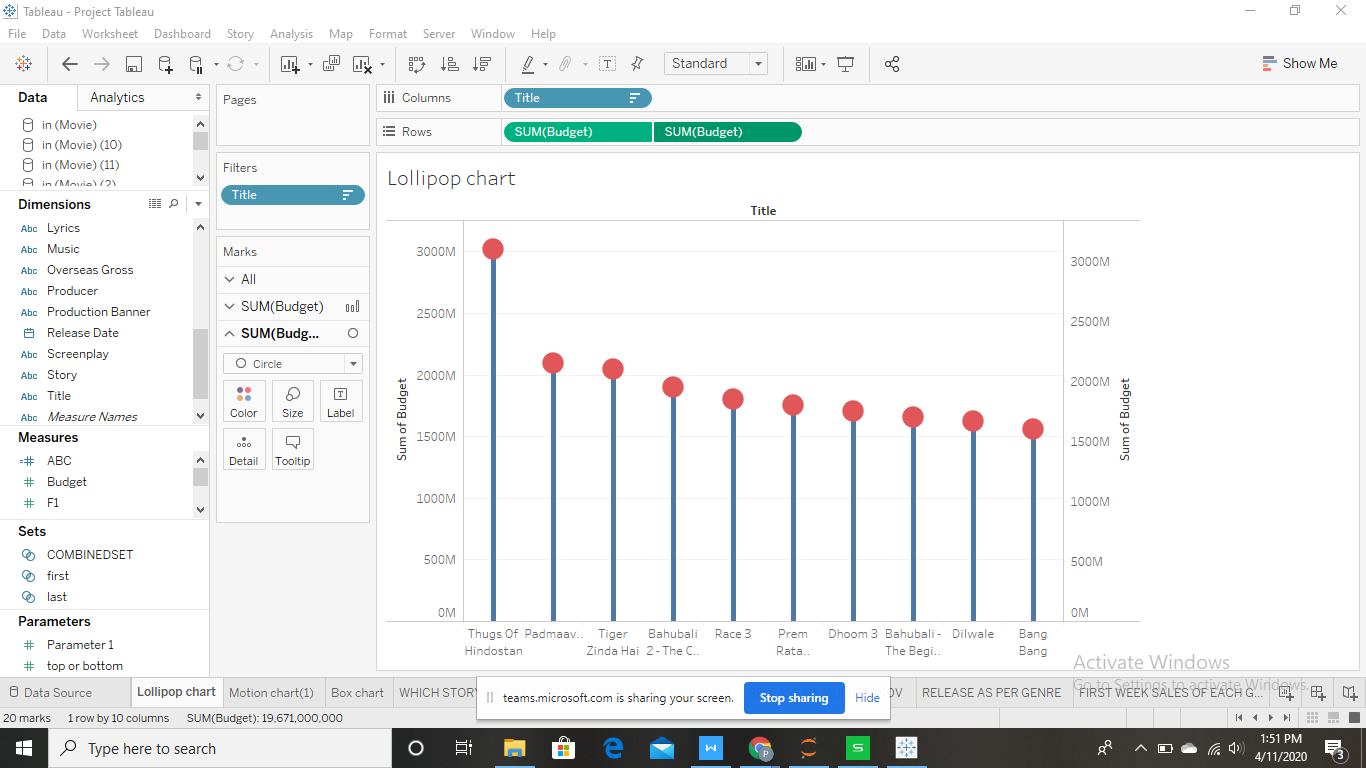
Pareto chart for genre according to budget. Drama movies have highest budgets while documentaries have lowest budget. It follows 80-20 principal.



**Lollipop Chart:-**

The **lollipop chart** is a composite **chart** with bars and circles. It is a variant of the bar **chart** with a circle at the end, to highlight the data value.

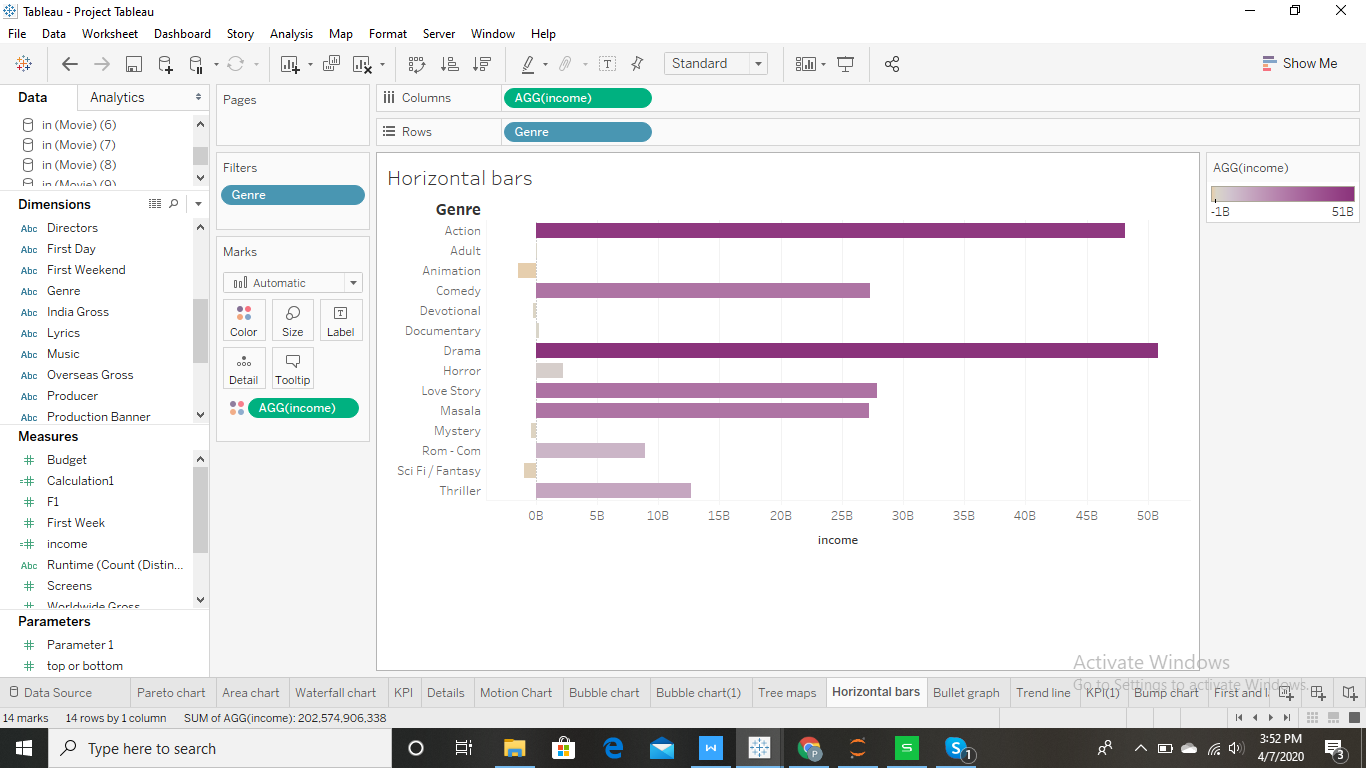
Top 10 movies with highest budgets. From this we can conclude that thugs of hindustan has the highest budget.



**Horizontal Bars:-**

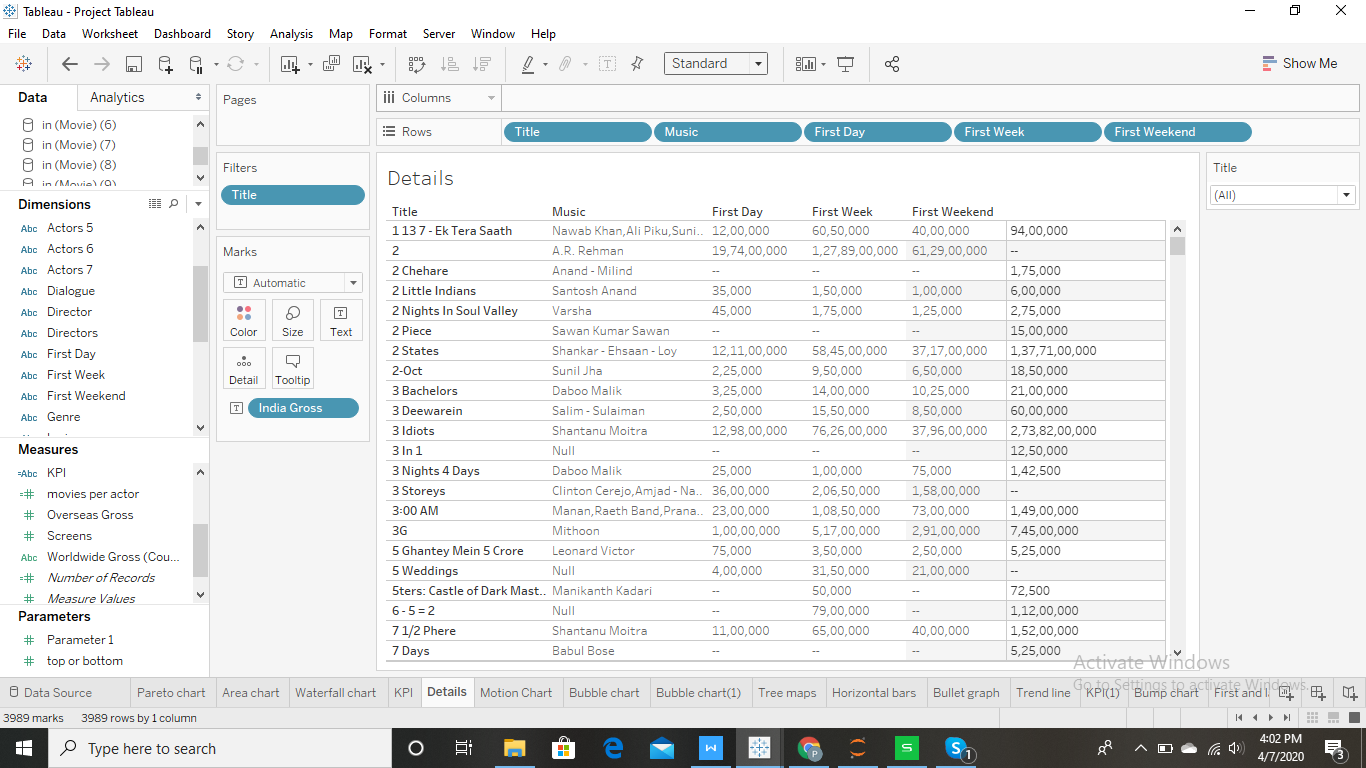
A bar chart or bar graph is a chart with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. The bars are used to show comparisons among categories.

It tells us about the income of each genre. We can see that drama movies have highest income followed by action and then love stories. While animated movies don’t happen to earn well.



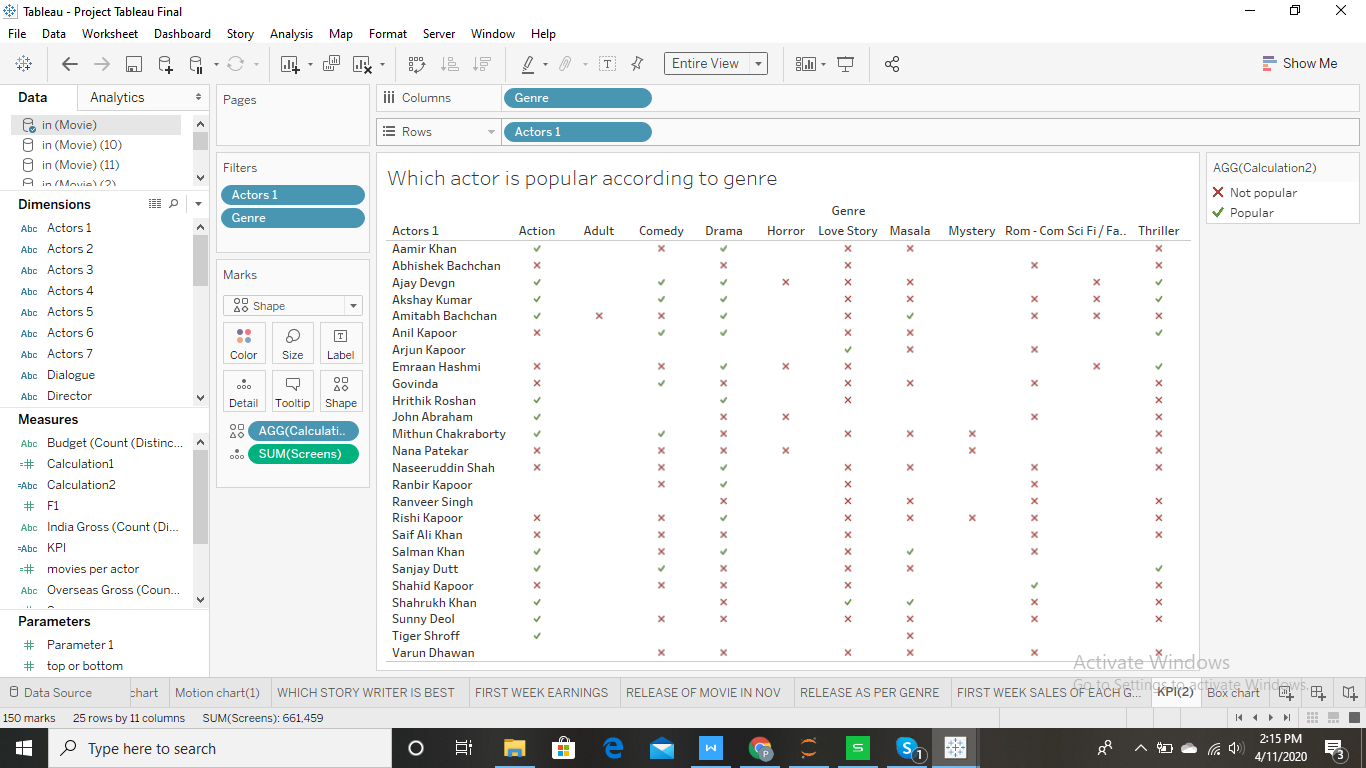
**Details:-**

This tells us about the dataset. This shows first day, first week, first weekend and india gross of each movie.



**KPI:-** Key Performance Indicator (**KPI**) is a measurable value that demonstrates how effectively a company is achieving key business objectives. Organizations use **KPIs** to evaluate their success at reaching targets. ... Each department will use different **KPI** types to measure success based on specific business goals and targets.

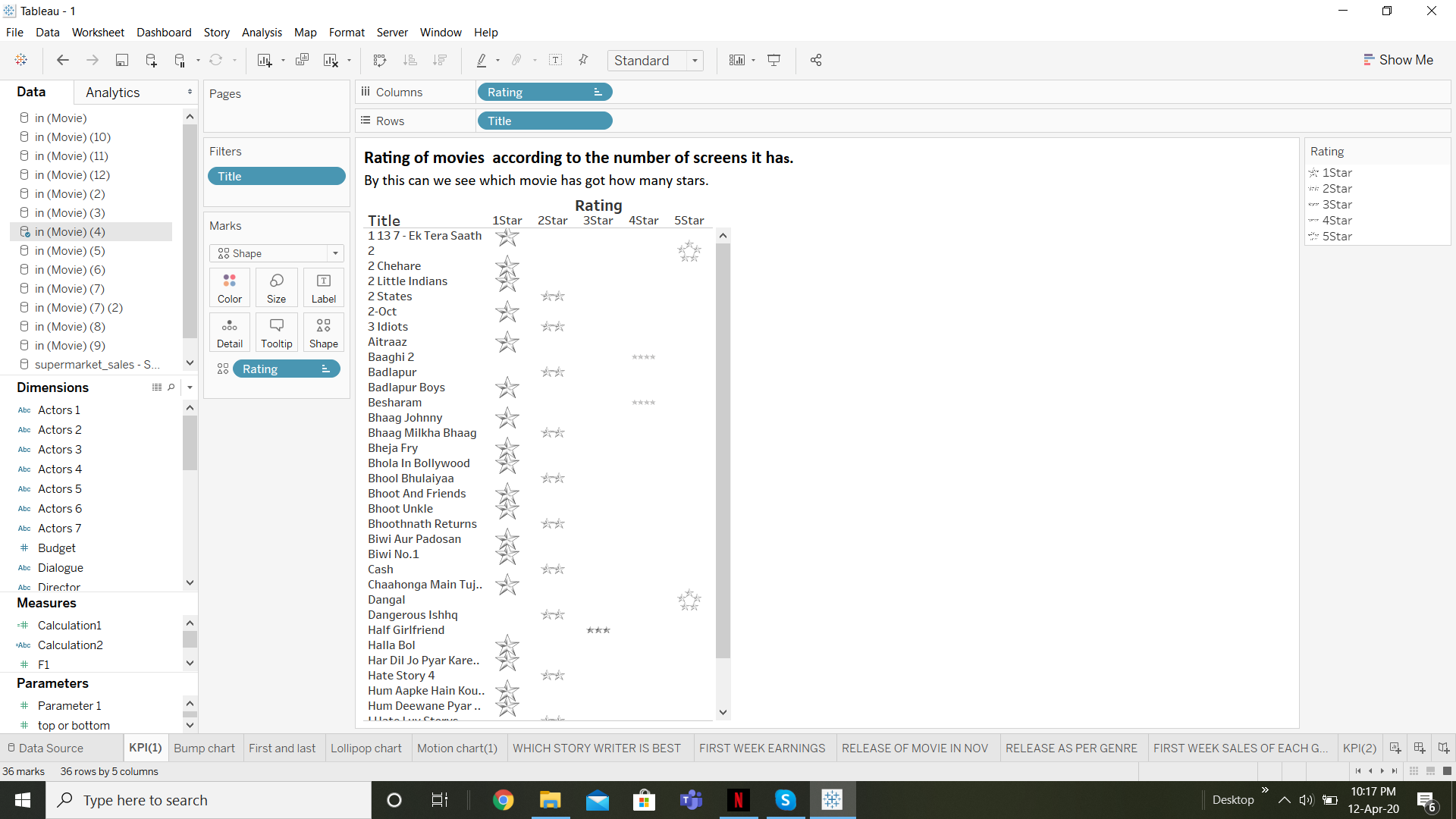
This kpi tells us which actor is popular in which genre according to the number of screens their movies have. If the movies have more than 5000 screens it is considered to be popular else unpopular. For example if we see aamir khan is popular in action and drama movies while unpopular in comedy, love story, masala and thriller.



**Group By:-**

Simplifies large numbers of dimension members by combining them into higher-level categories

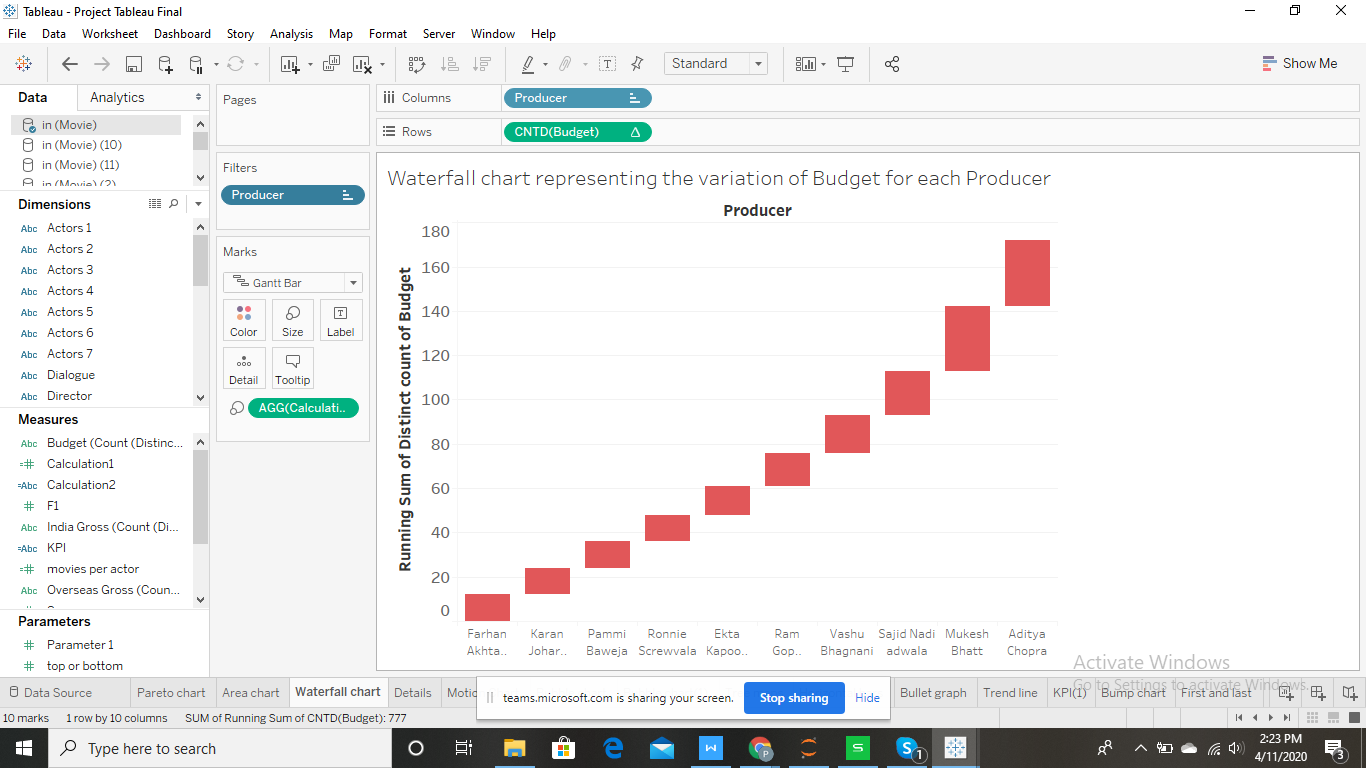
Rating movies according to the number of screens it has. If a movie has 1000 screens it will get 1 star if it has screens between 1000-2000 it will get 2 stars and so on. So from this we can say that 2- states has 2 stars so it has screens somewhere between 1000-2000. while “2” has 5 stars so it might have screens between 4000-5000 or above.



**Waterfall Chart:-**

**Waterfall charts** effectively display the cumulative effect of sequential positive and negative values. It shows where a value starts, ends and how it gets there incrementally.

Waterfall chart representing variation of budget for top 10 producer. According to this chart aditya chopra has makes movies with maximum budget.

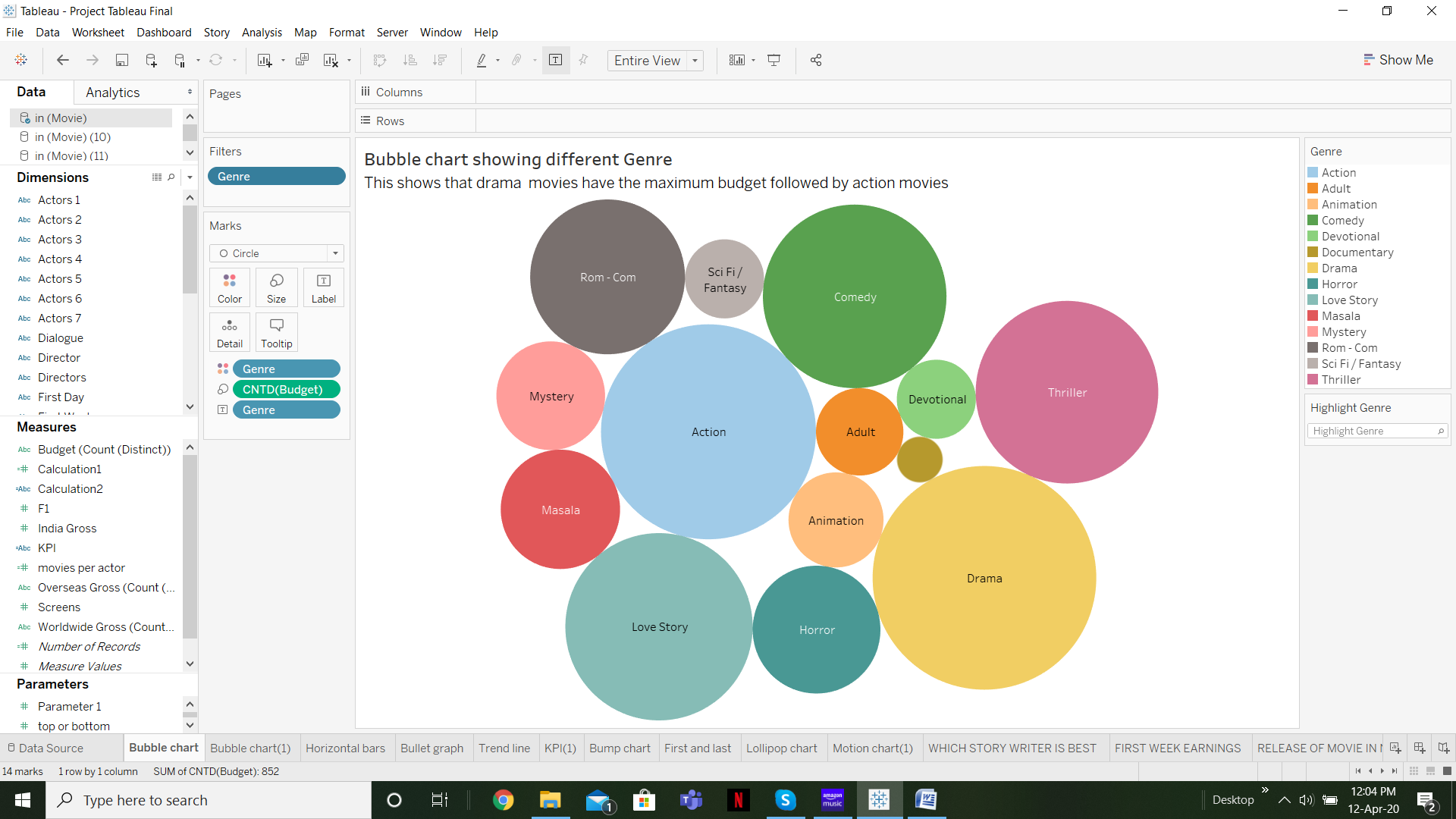
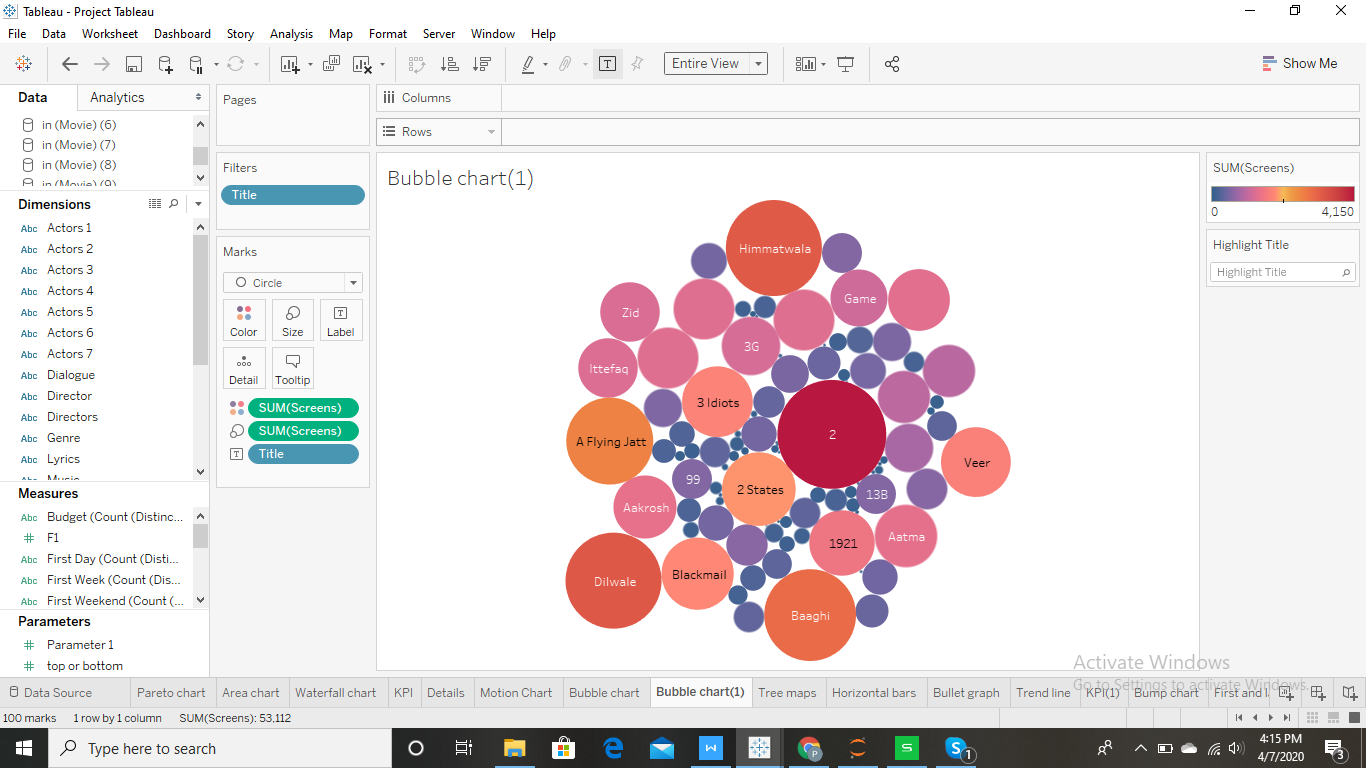


**Bubble Chart:-**

**Bubble charts** display data as a cluster of circles. Each of the values in the dimension field represents a circle whereas the values of measure represent the size of those circles.

Bubble chart showing different genres according to budget. This shows that drama movies have the maximum budget followed by action movies

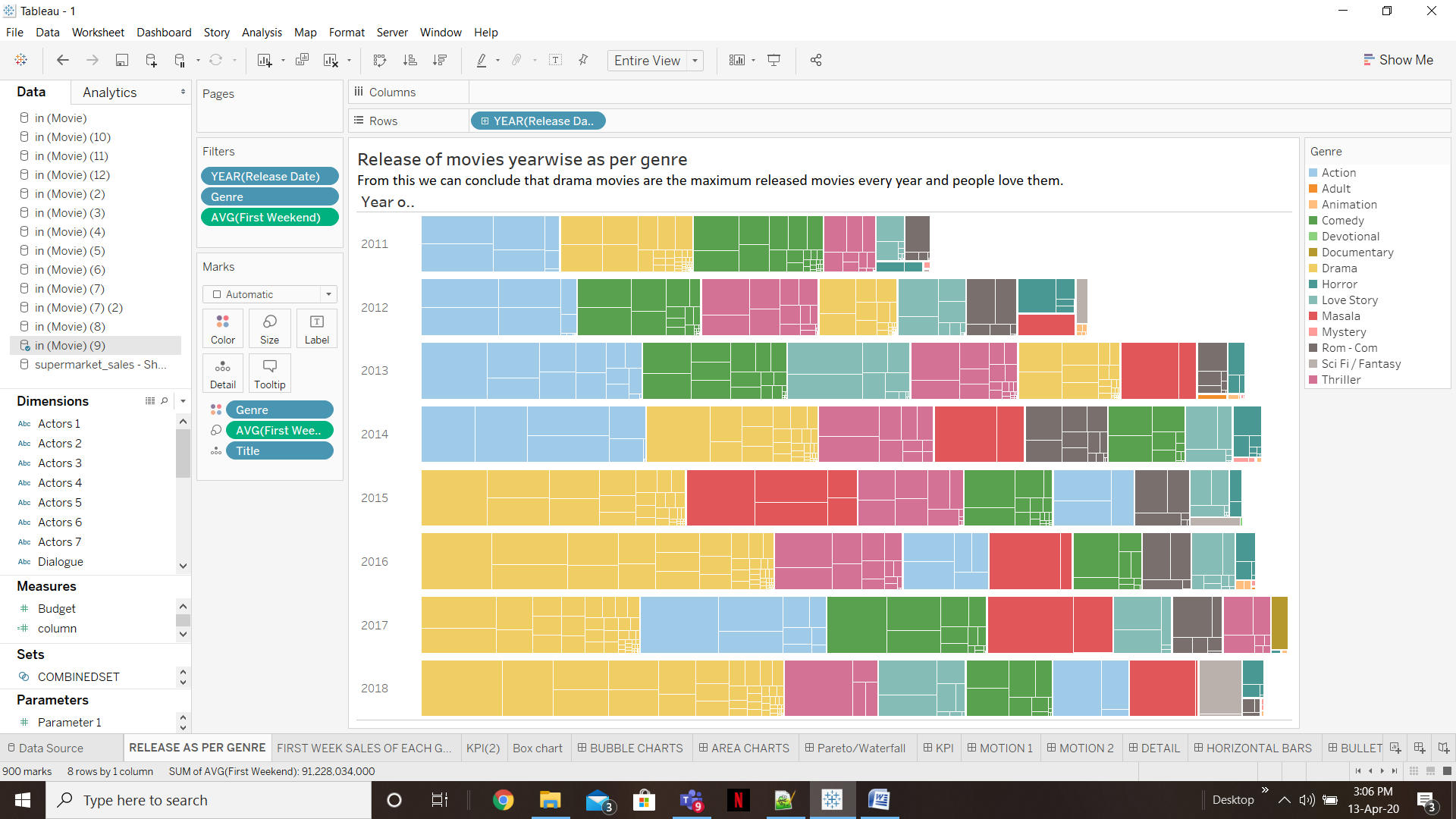
This chart represents which movie has maximum screens. We can see that 2 has the maximum screens.

**Tree Map:-**

The **tree map** displays data in nested rectangles. The dimensions **define** the structure of the **tree map** and measures **define** the size or color of the individual rectangle.

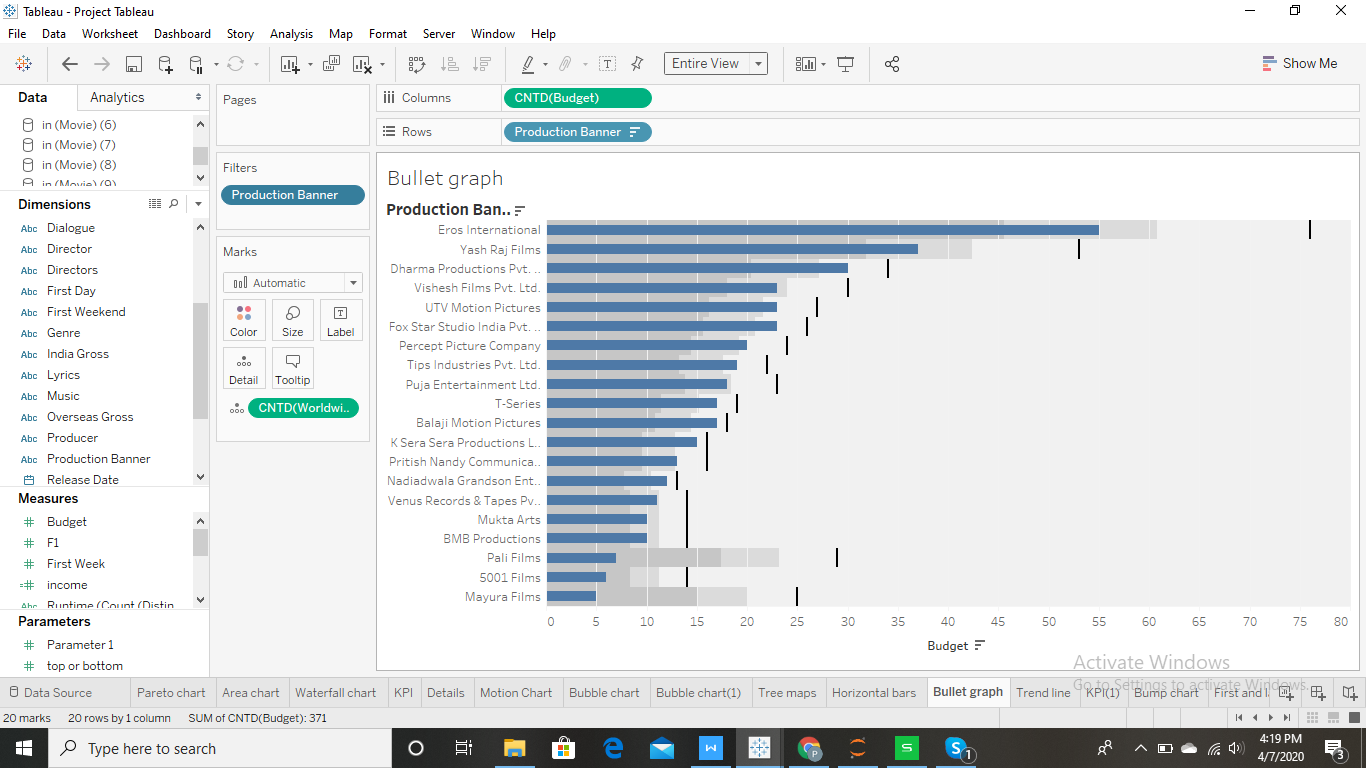
Release of movies year wise as per genre.this shows that drama movies are the maximum released movies every year.



**Bullet Chart:-**

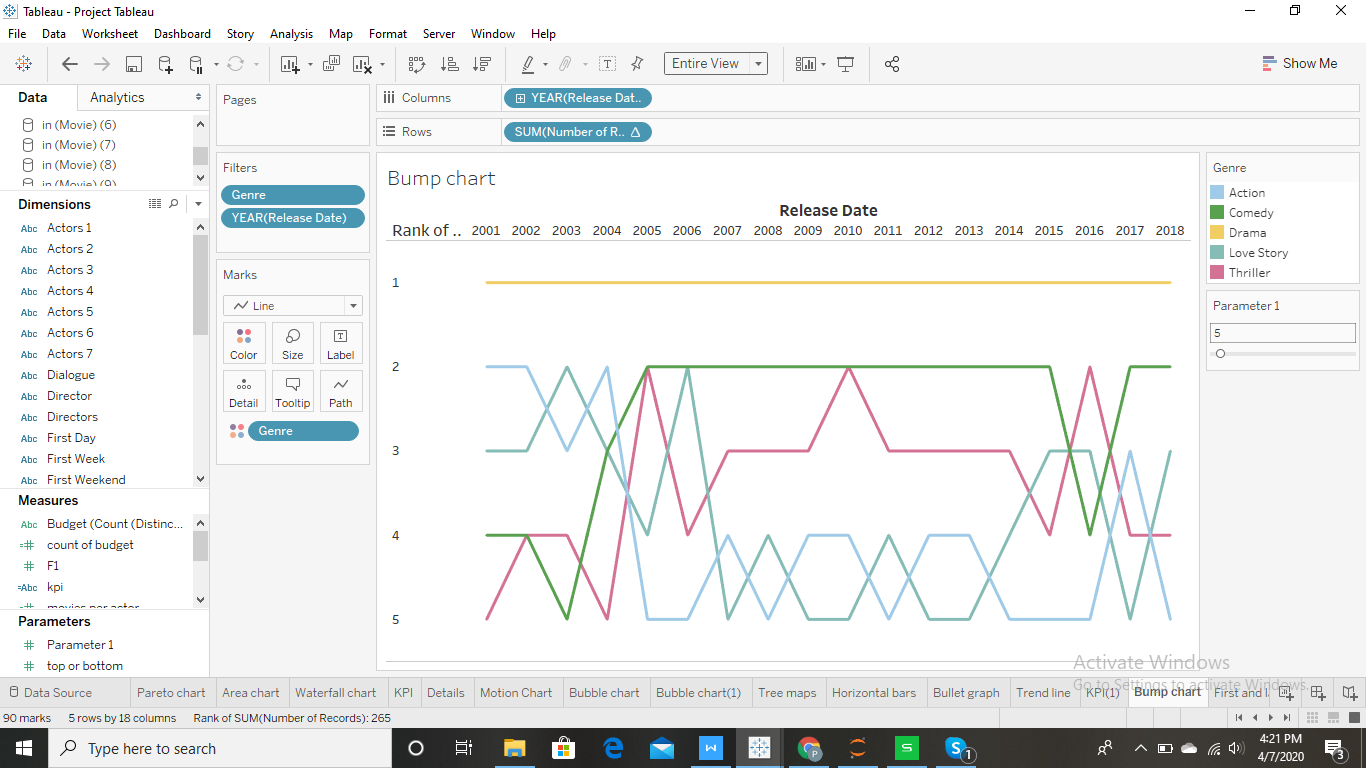
A **bullet chart** is an advanced sort of bar **chart** where we can compare two measures on a single bar. In a **bullet chart**, the primary measure is shown by the main dark bar on the front and the secondary measure is shown beneath the main bar as a reference line.

In this graph we have compared two measures that are budget and worldwide gross of top 20 production banners. We can conclude that movies released by eros internaional have maximum budget and they earn well through out the world.



**Bump Chart:-** Bump Chart is used to compare two dimensions against each other using one of the Measure value.  They are very useful for exploring the changes in Rank of a value over a time dimension or place dimension or some other dimension relevant to the analysis. The **Bump Chart** takes two dimensions with zero or moremeasures.

Ranking of each genre year wise. This shows that drama movies are always on top. This shows that for year 2018 drama is rank 1, rank 2 is comedy and rank 3 is love story and so on.



**Line And Bar Chart:-**

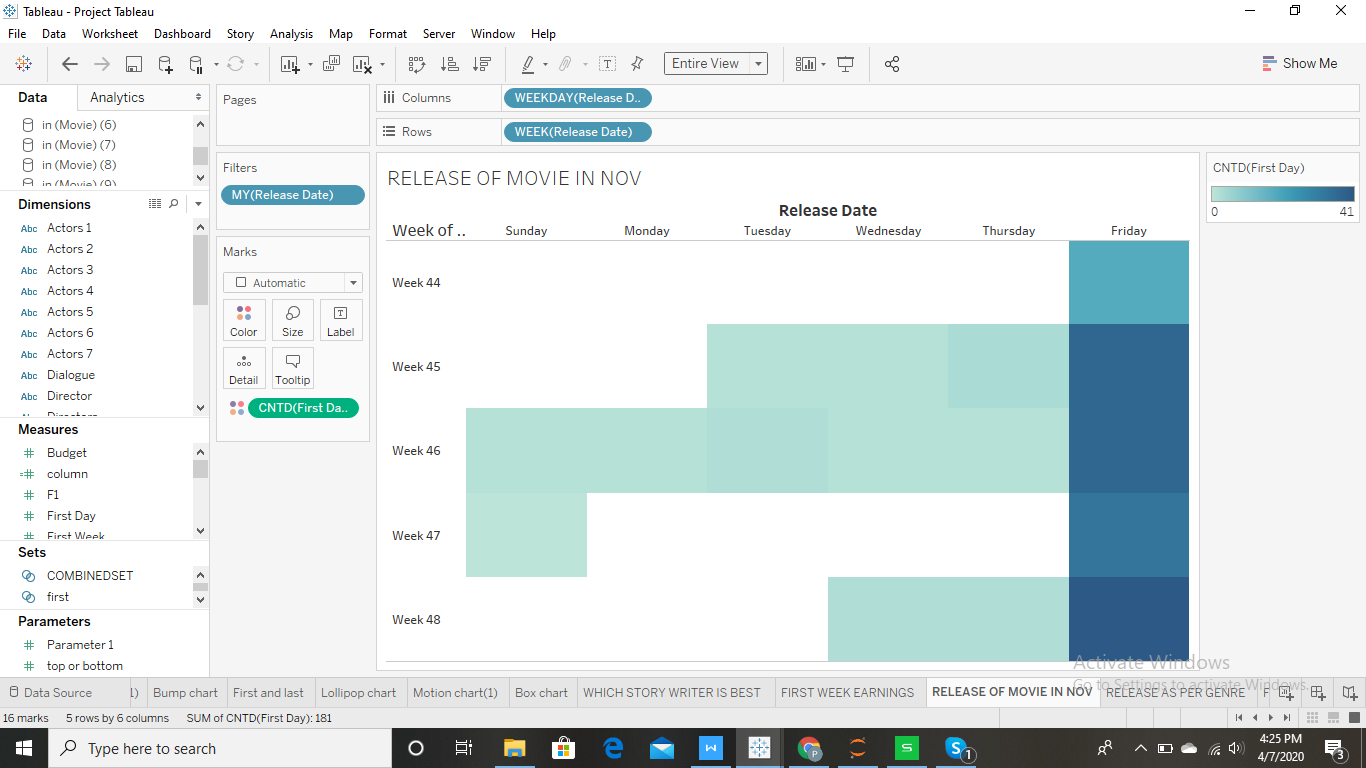
This chart tells us how story written by each writer earns in india and also worldwide. The bars show the world wide gross where as the line shows the india gross. Stories written by Bhashir babbar does great not only in india but also world wide.



**Highlighted Table:-**

Use **highlight tables** to compare categorical data using color.

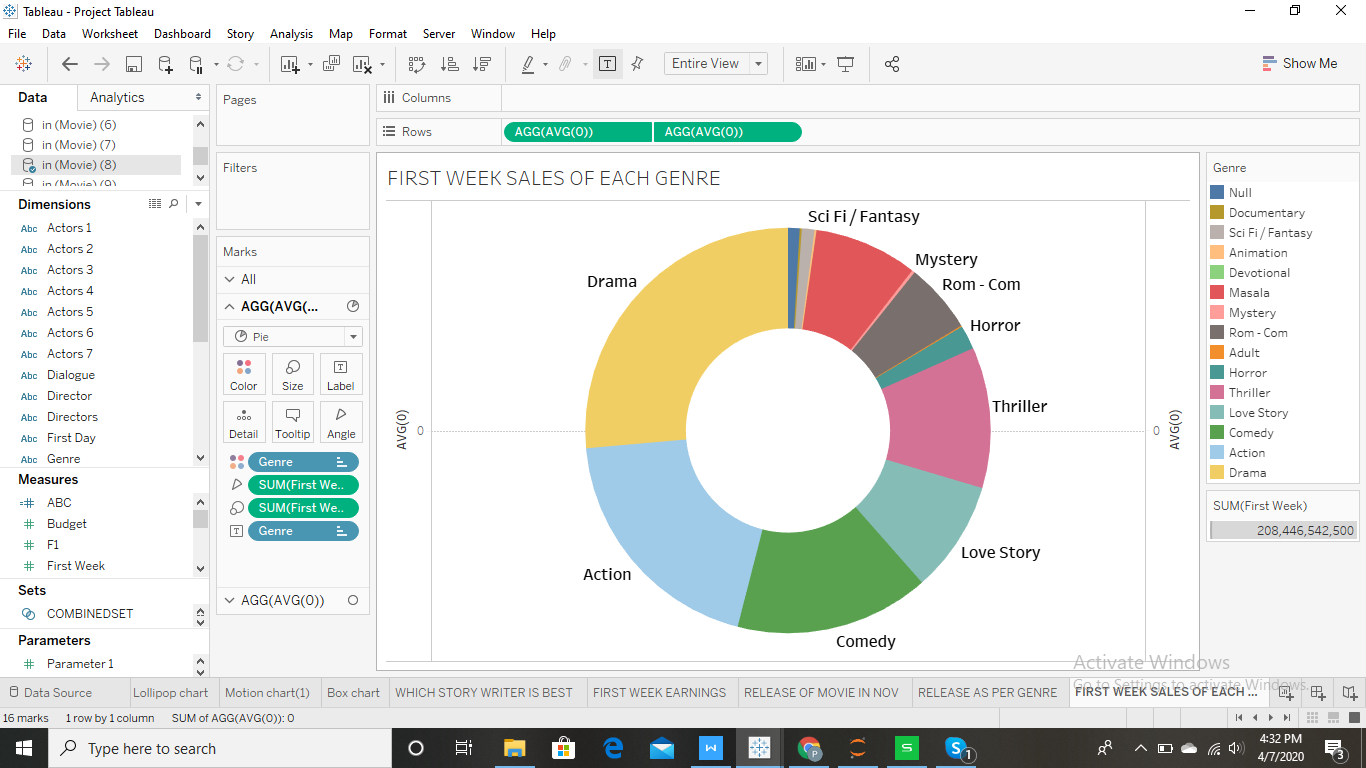
This shows the release of movies in the month of november for every year and their first day earnings. We can say that people are more likely to watch a movie on Friday as compared to any other day.



**Doughnut Chart:-**

However, it is similar to **pie chart** which divides the a circle in different sectors based on the value of their proportions. But, the difference between **Pie** and **Doughnut chart** is that in a **doughnut chart** there can be more than one data series added as rings to the **chart**.

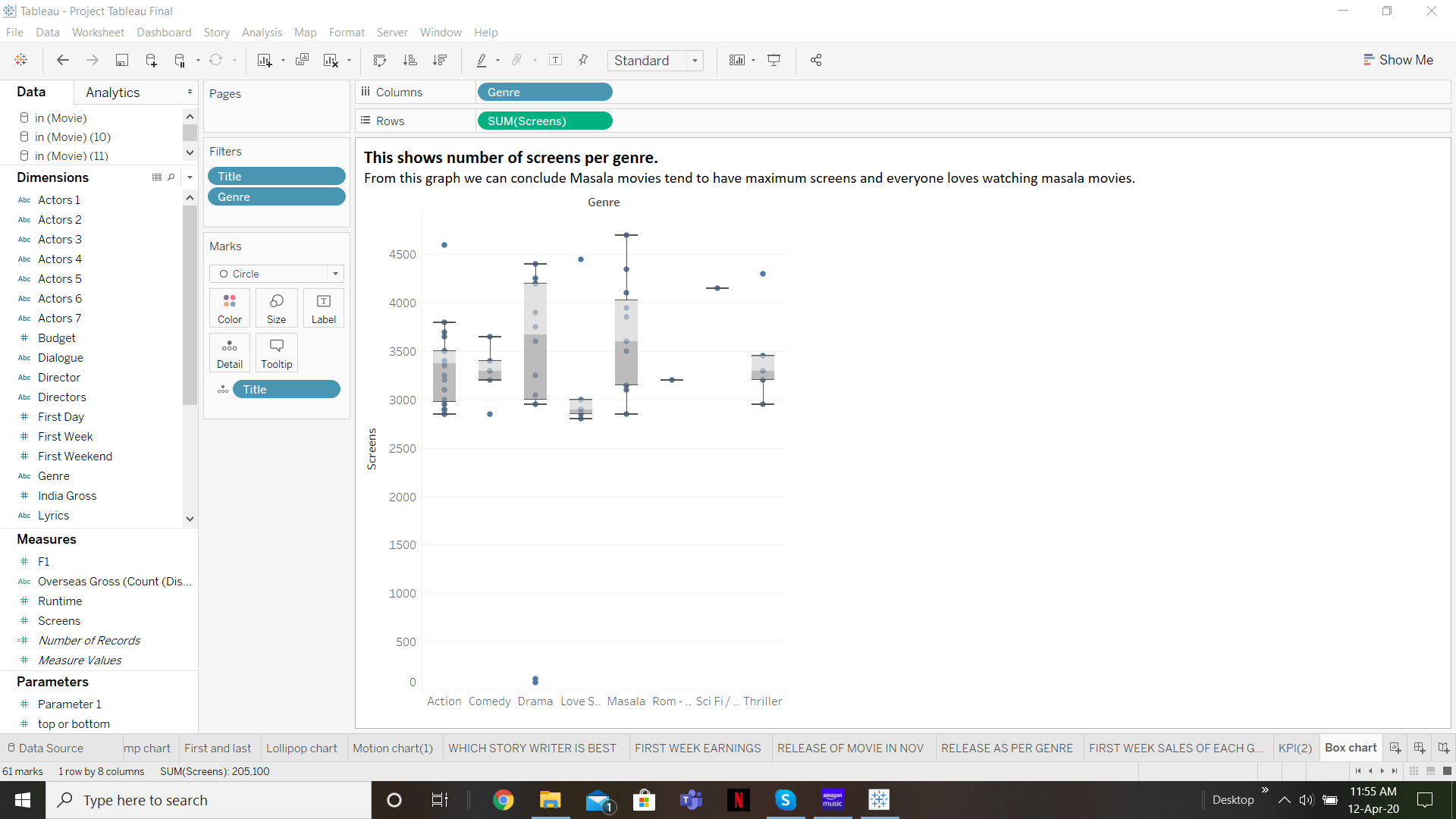
This chart shows the first week earnings of each genre. This tells that drama movies earn the most in first week followed by action and then comedy and so on.



**Box Plot:**-

The **box plots** are also known as a **box**-and-whisker **plots**. They show the distribution of values along an axis. **Boxes** indicate the middle 50 percent of the data which is, the middle two quartiles of the data's distribution.

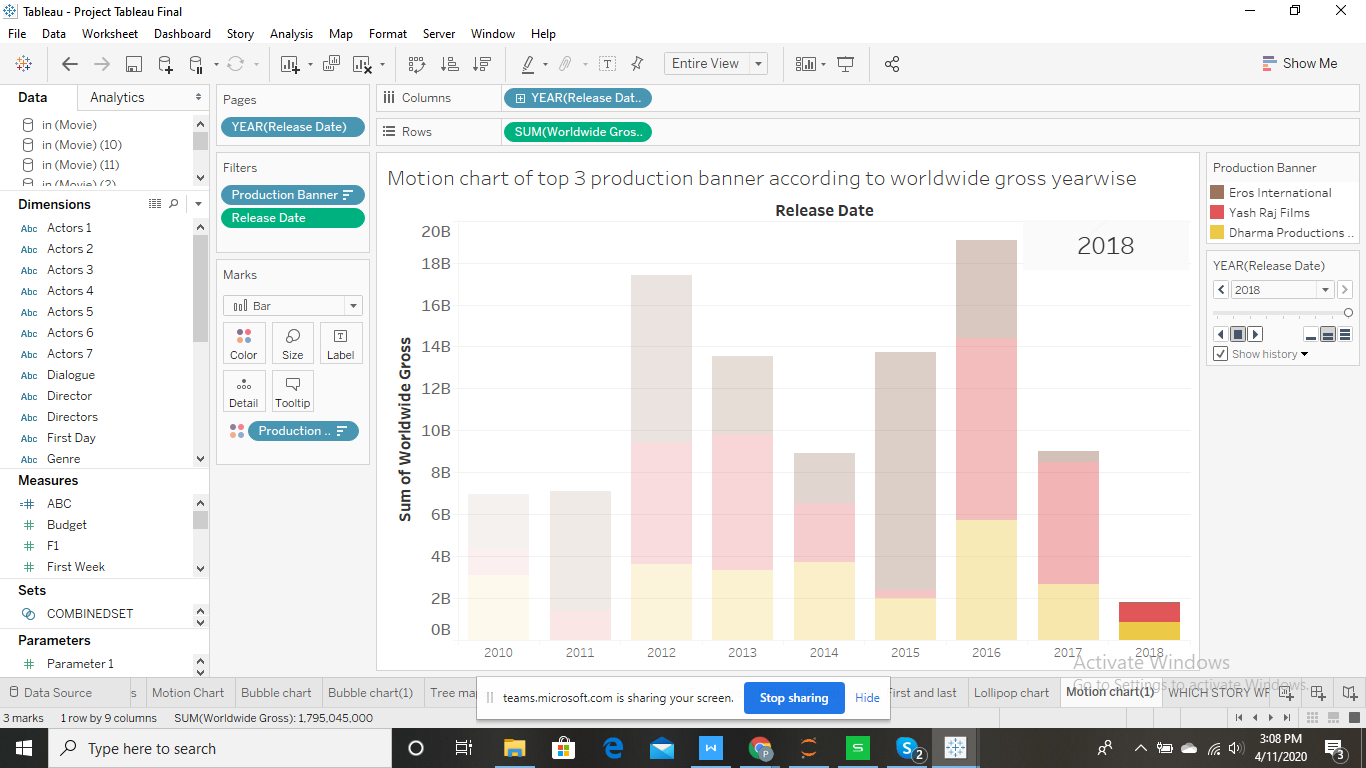
This shows number of screens per genre. Masala movies tend to have maximum screens.

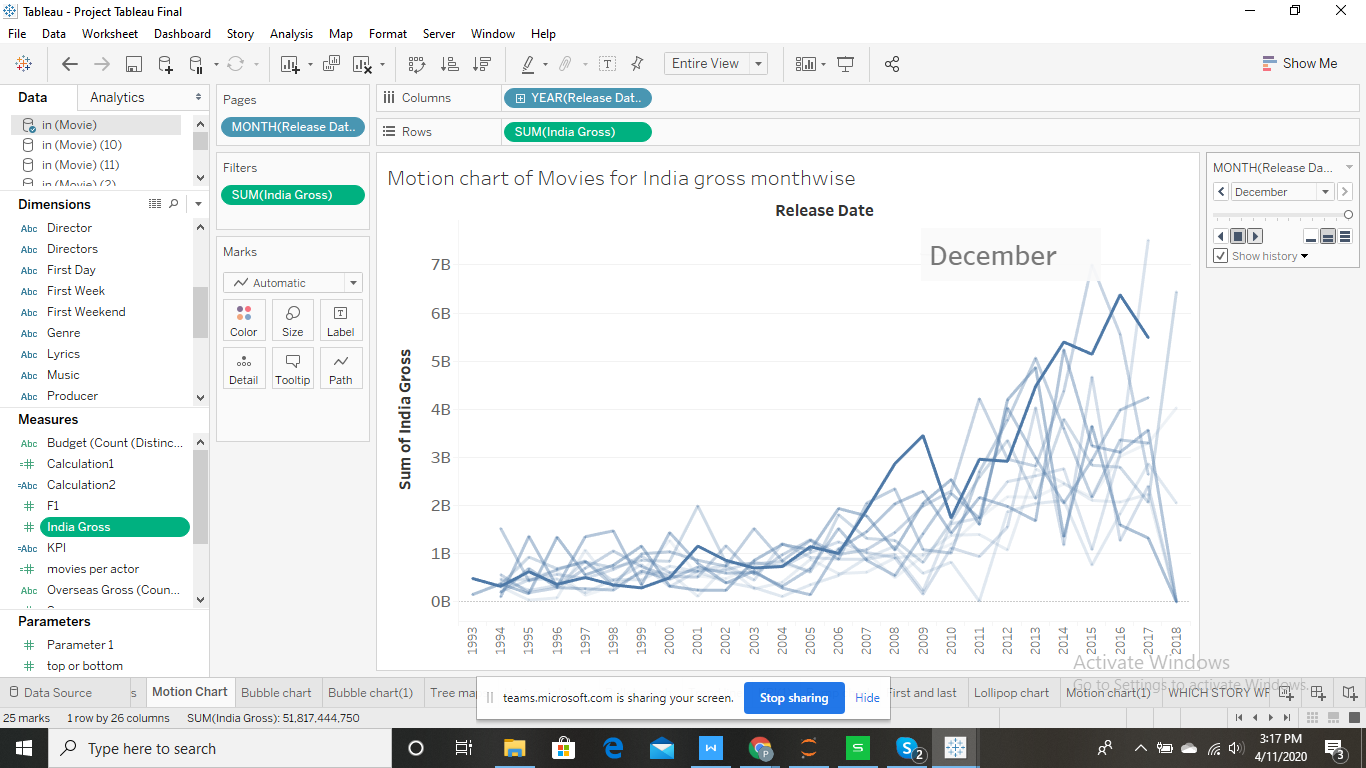


**Motion Charts:-** Motion charts provide mechanisms for mapping ordinal, nominal and quantitative variables onto time, 2D  coordinate axes, size, colors, glyphs and appearance characteristics, which facilitate the interactive display ofmultidimensional and temporal data.

Motion chart of top 3 production banner according to worldwide gross year wise. This shows that in the year 2018 maximum earnings was of yash raj production followed by dharma production and then eros international.

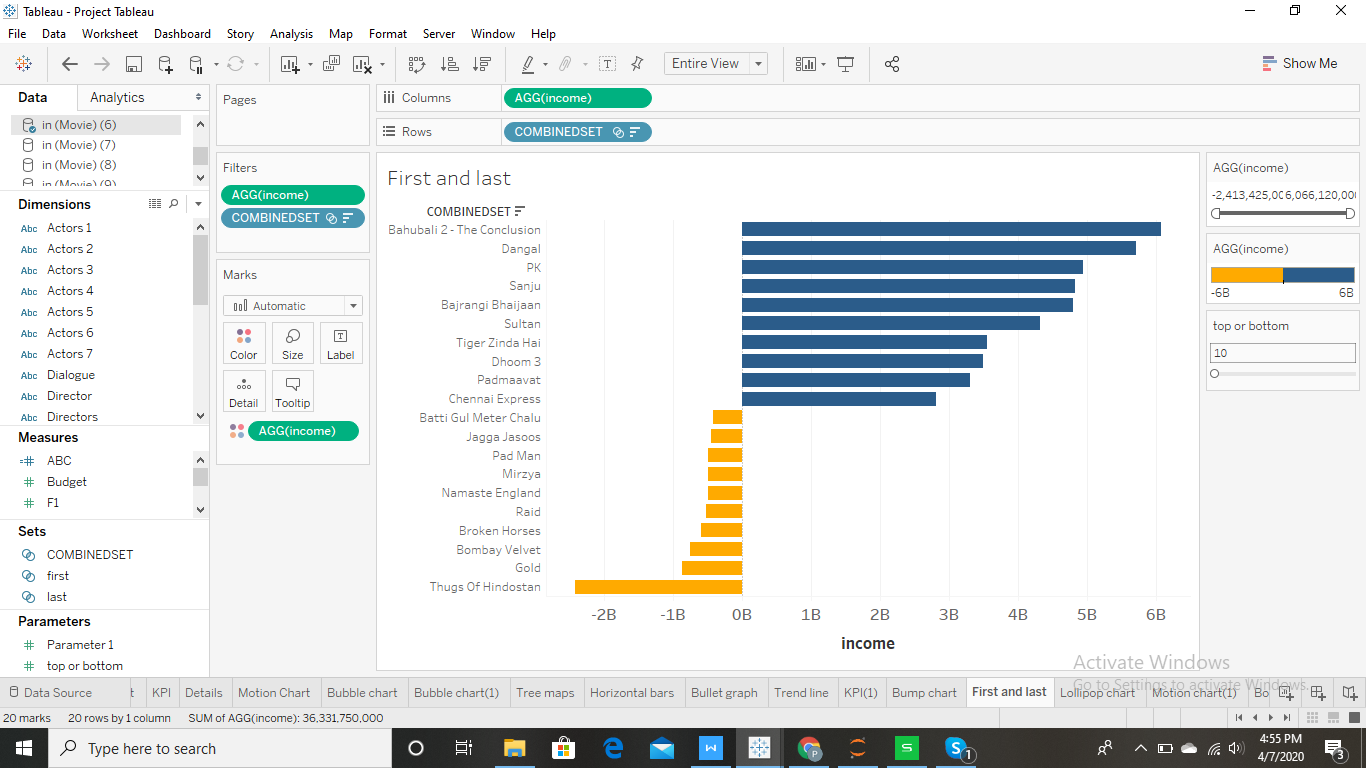
Motion chart of movies for india gross month wise for every year. Now as we can clearly see people now are more into movies as they were before and the india gross is more in the month of december.





**Bar Chart:-**

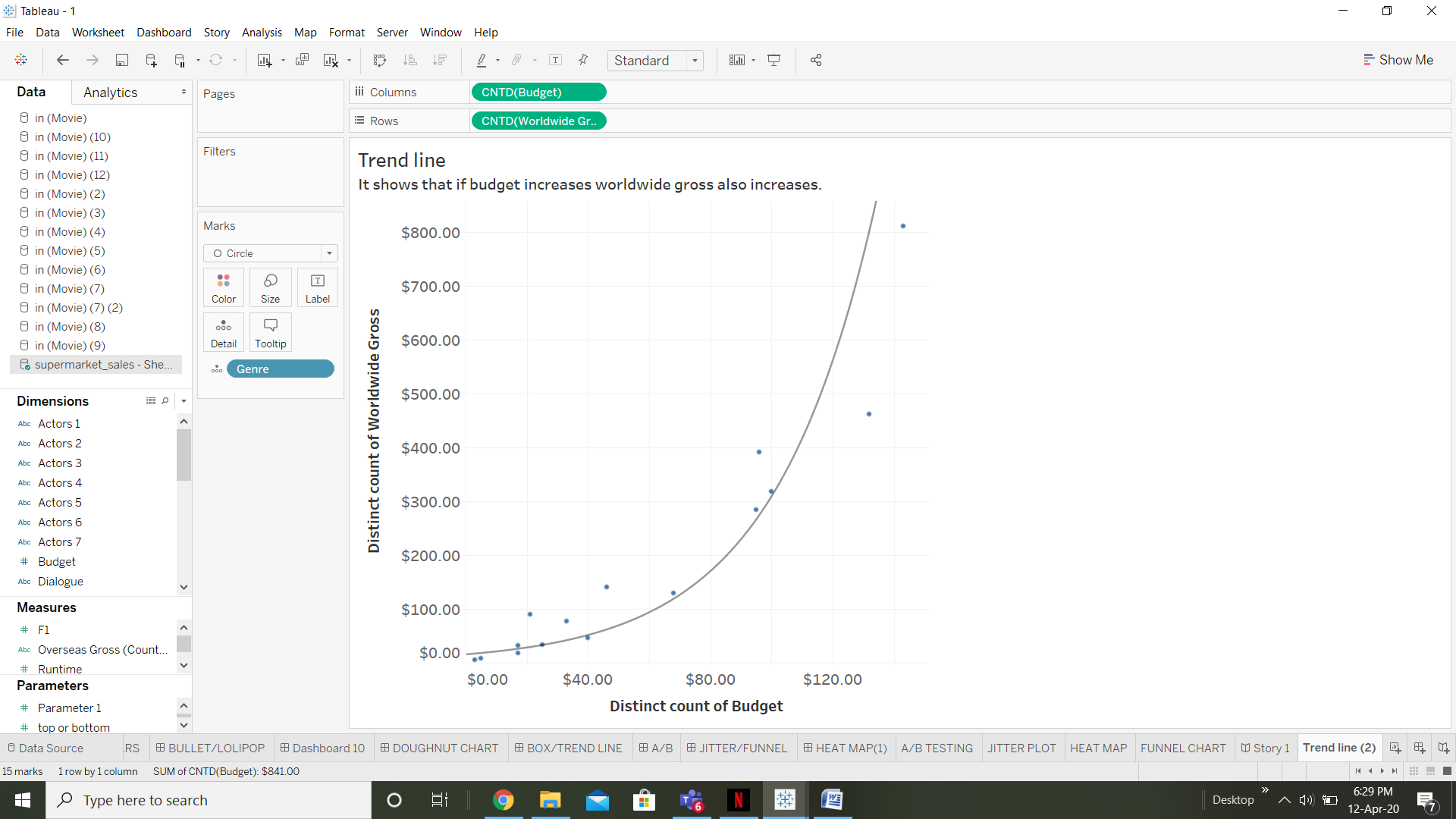
Top 5 and bottom 5 movies according to their earnings. Bahubali 2 earned the most while thugs of hindustan earned the least.



**Trend line:-**

**Trend lines** are used to predict the continuation of a certain **trend** of a variable. It also helps to identify the correlation between two variables by observing the **trend** in both of them simultaneously.

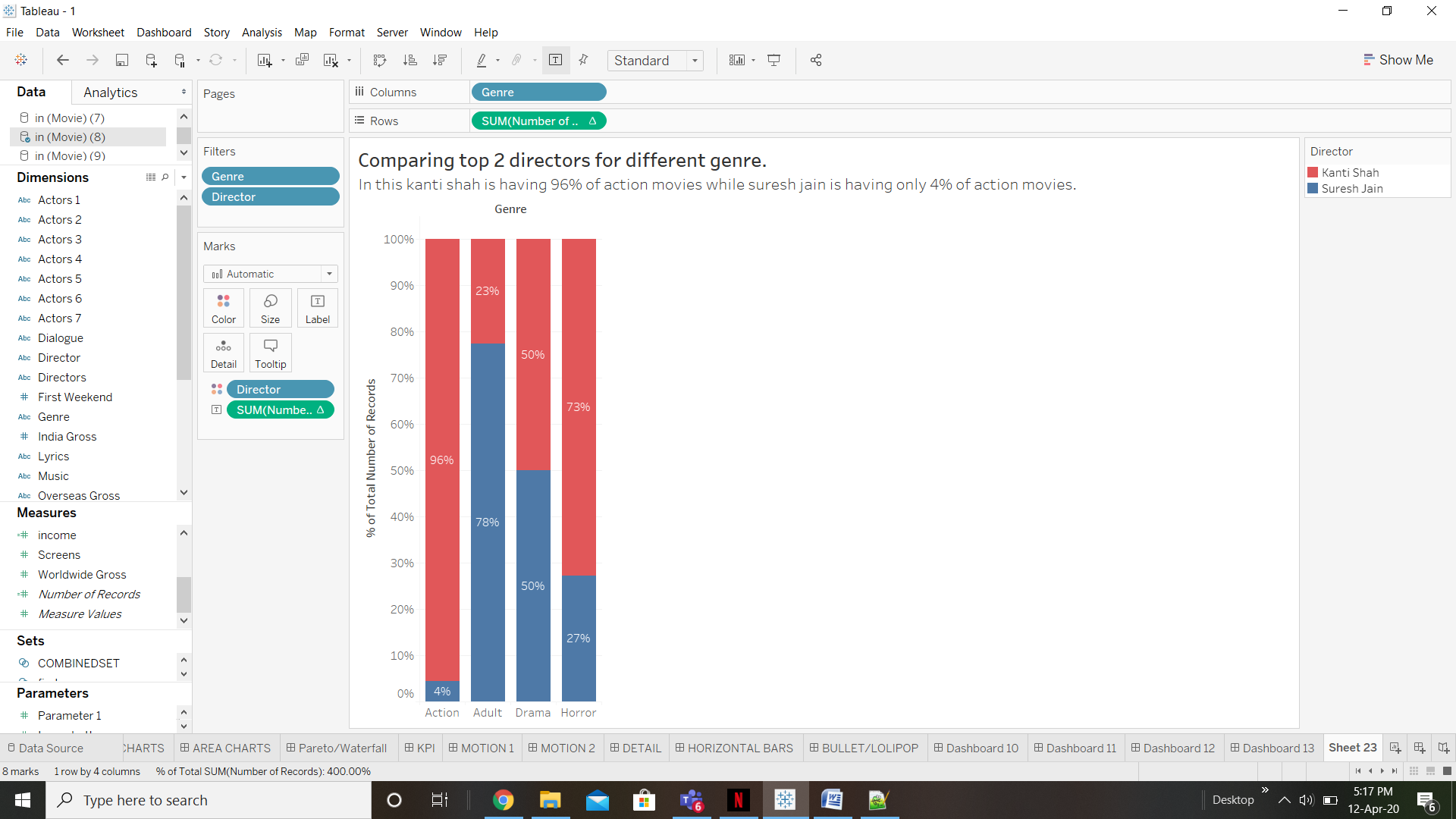
It shows that if budget increases worldwide gross also increases.



**A/b Testing:-**

A/B testing is the process of comparing two variations of a page element, usually by testing users' response to variant A vs variant B, and concluding which of the two variants is more effective.

Comparing top 2 directors for different genre.In this kanti shah is having 96% of action movies while suresh jain is having only 4% of action movies.

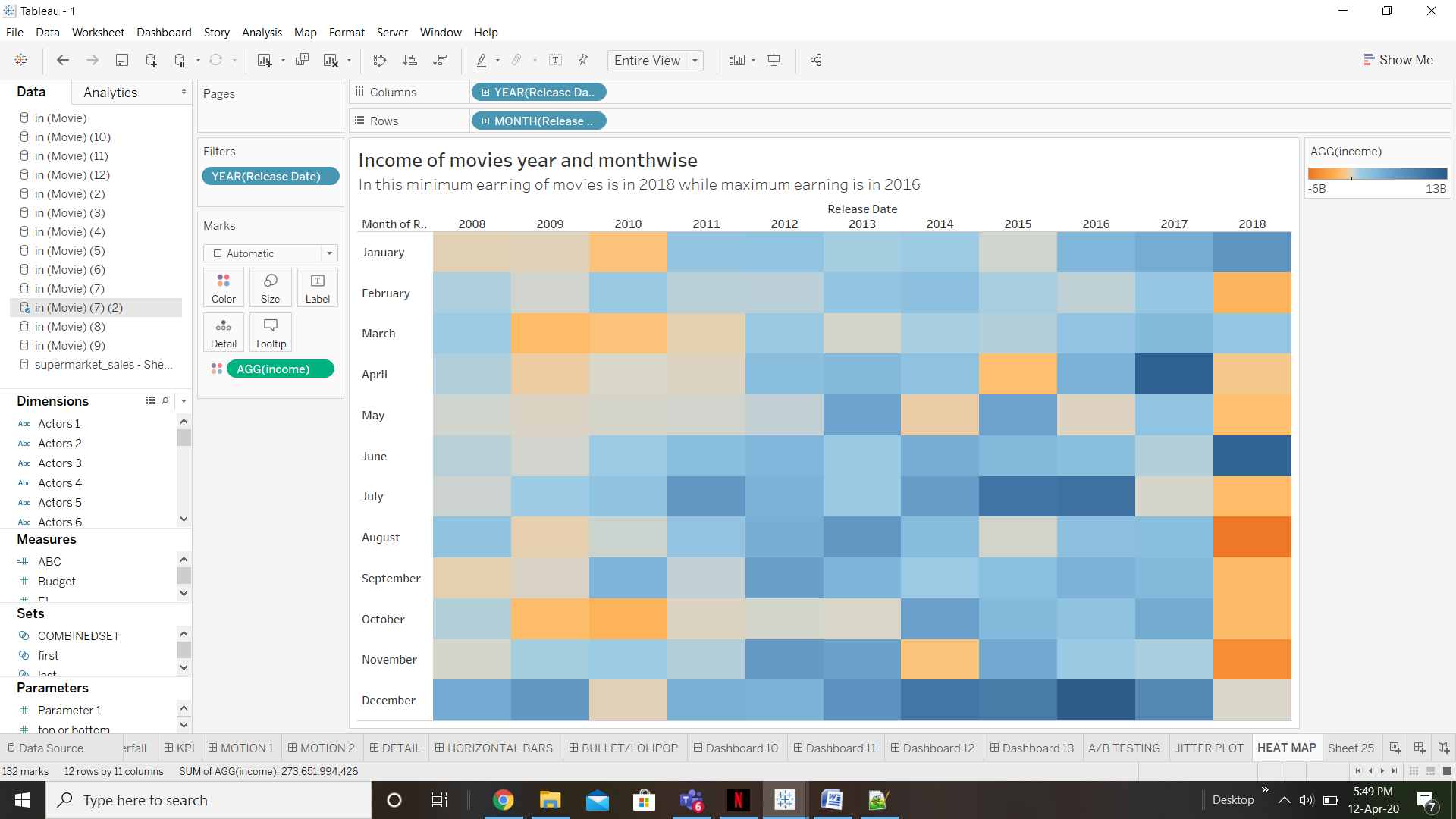
****

**Heat map:-**

**Heat maps** are a visualization where marks on a chart are represented as colors.

Income of movies year and monthwise

In this minimum earning of movies is in 2018 while maximum earning is in 2016

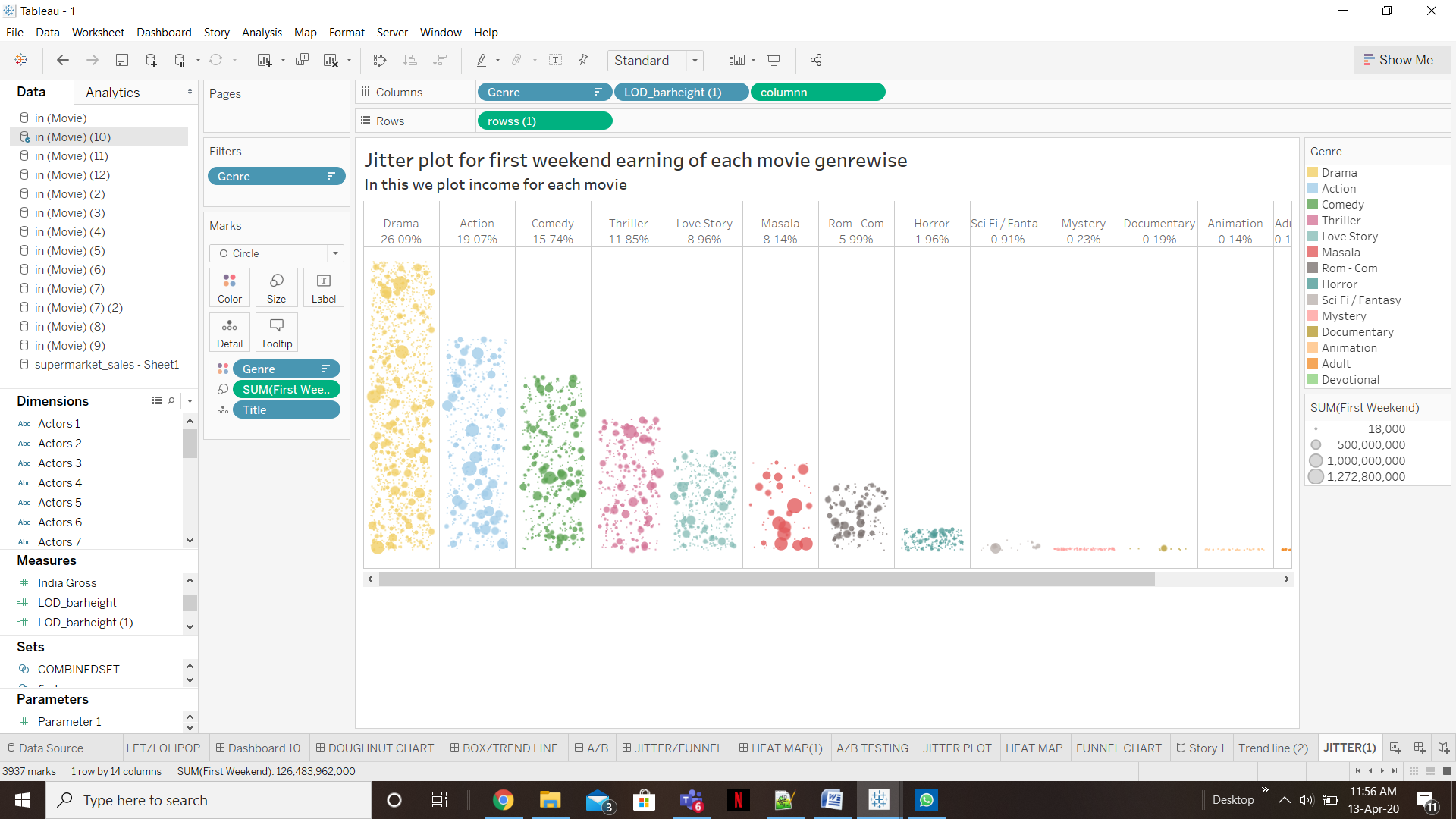


**Jitter plot:-**

“**Jittering**” is a technique for separating overlapping marks on a view. By giving marks some extra room by separating them into different columns, hidden data is often revealed and it is easier to visualize how the data points are distributed.

Jitter plot for first weekend earning of each movie genrewise

In this we plot income for each movie

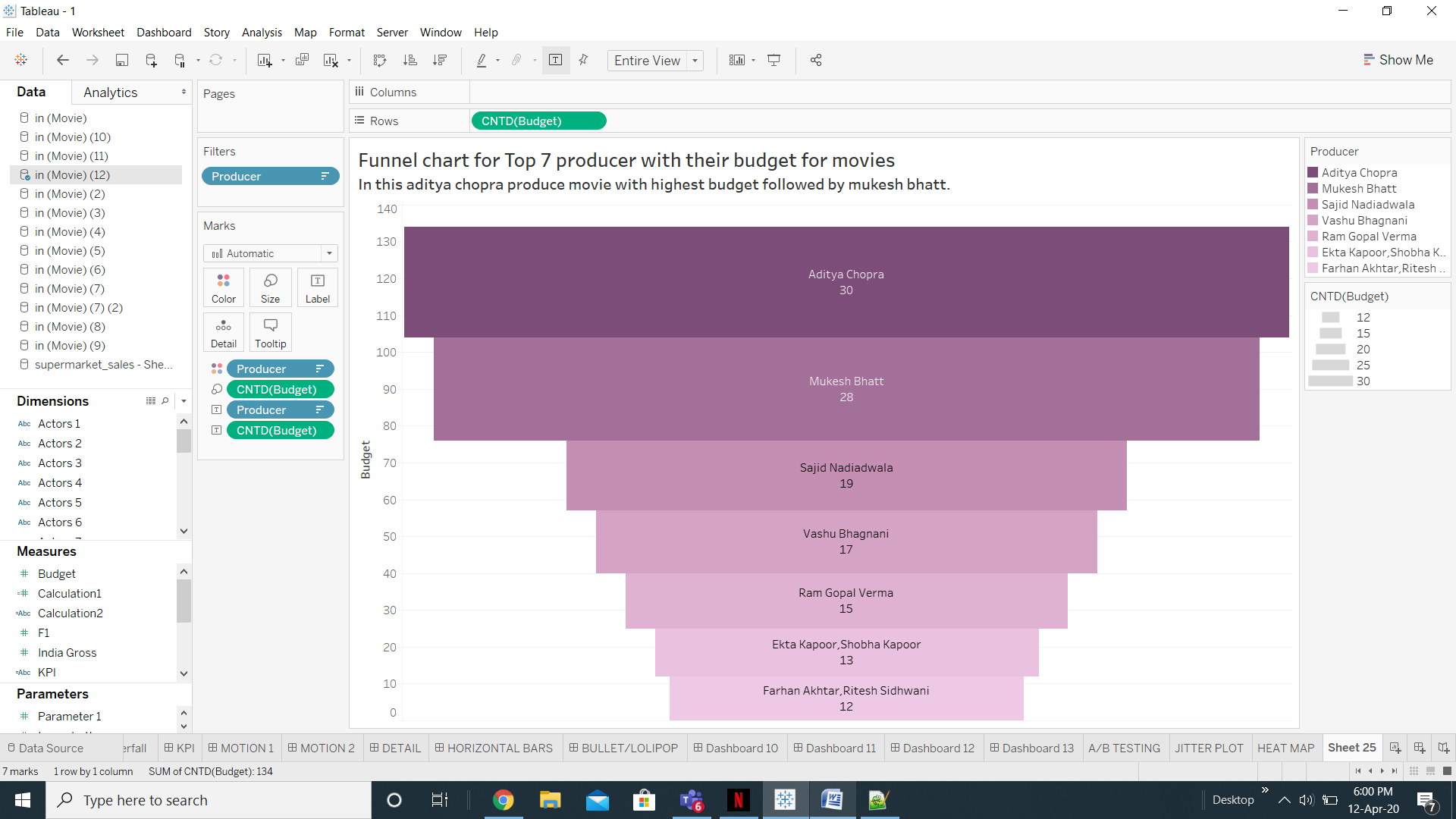
****

**Funnel chart:-**

**Funnel charts** are a great way to track progress through a workflow with multiple steps. They can highlight how well a process is being followed and easily identify any bottlenecks.

Funnel chart for Top 7 producer with their budget for movies

In this aditya chopra produce movie with highest budget followed by mukesh bhatt.

****

**Future Scope:-**

Bollywood is not just one of india’s most eye catching growth phenomena it is also the world’s biggest and fastest growing commercial film cluster. On a scale comparable to hollywood , bollywood finances , producers , market and distributes well over 200 films annually , including big budget blockbusters. Our dataset tells us about some of bollywood movies that have been released since 1992. On the basis of information like Indian gross , overseas gross and worldwide gross we are able to calculate how much a movie earns. From the data we are also able to predict what kind of movies are more loved by people. It also helps us to predict which music director gives the best music. With the help of this data bollywood can produce more movies which earn well on box office and hence help bollywood grow further.