Data Dictionary

[Previous](https://olympus.mygreatlearning.com/courses/97125/modules/items/4652944?pb_id=16159)

[Next](https://olympus.mygreatlearning.com/courses/97125/modules/items/4652946?pb_id=16159)

**rating.csv:** It contains information on ratings given by the users to a particular movie.

user\_id: id assigned to every user

movie\_id: id assigned to every movie released

rating: the rating given by the user

timestamp: Time recorded when the user gave a rating

**movie.csv**: The file contains information related to the movies and their genre.

genre: genre of the movie - a movie can have multiple genres separated by pipe

movie\_id: id assigned to every movie released

movie\_name: Title of the movie

year: Year of release of the movie

**user.csv**: It contains information about the users who have rated the movies.

user\_id: id assigned to every user

age: Age of the user

gender: Gender of the user (M: Male, F: Female)

occupation: Occupation of the user

zipcode: Area code for the user's residence

Solution Approach

[Previous](https://olympus.mygreatlearning.com/courses/97125/modules/items/4652946?pb_id=16159)

[Next](https://olympus.mygreatlearning.com/courses/97125/modules/items/4652949?pb_id=16159)

**Solution Approach:**

Before answering the questions, the following actions should have been performed:

1. Load the 3 CSV files into the report
2. Ensure that the first row values are considered as Headers in all the 3 datasets
3. Model the data to ensure no circular relationships are present
4. Clean and transform the data to remove redundant columns
5. Each page in the report should have appropriate slicers and navigation actions

**Question 1:**Create a Card visual to display the Total number of movies released and Total number of distinct Genres separately

**Question 2:** You can use a card visual to display the number of users and a donut chart to visualize the gender distribution among users

Use COUNTROWS to calculate the number of users

**Question 3:** You can use an area chart to highlight the average ratings across movie genres, based on gender.  
  
Create a new Genre group that combines all similar genres into a single group, For example, the fields Action, Action|Adventure, Action|Thriller, should be grouped as “Action based”, the fields Adventure, Adventure|Sci-fi, Adventure|Fantasy, etc should be grouped as “Adventure based” and so on.  Use this group to create the visualization for Question 3.

**Question 4:** You can create a gauge chart that shows the average rating. For the second part of the question, you can use a stacked bar chart that shows the number of movies released across different genre groups   
  
Create a DAX measure to calculate the average rating. Use this measure to visualize the average rating.

Create a new Decades group that combines the values of the Year column into decades [1930 to 1939, 1940 to 1949, etc]. The values before 1920 can be combined with the 1920s decade.

Questions 1,2,3 and 4 should have a slicer that can help visualize the insights based on the decade chosen

**Question 5:** You can use a clustered column chart to visualize the distribution of users across age groups based on their gender. To visualize the location of these users, you can use the map chart, along with bubbles to indicate the number of users in each location.  
  
Create a new column age\_group that classifies users into 5 categories as below, based on their age. Use this column in further visuals.  
  
Age 18 and above,  and less than 25 = “18-24”  
Age 25 and above, and less than 35 = “25-34”  
Age 35 and above, and less than 45 = “35-44”  
Age 45 and above,and less than 55 = “45-54”  
Age 55 and above = “55+”  
Users with age less than 18, should be imputed with the median age corresponding to the occupation  
  
**Question 6:** You can use a matrix visual to visualize the average ratings across professions and age groups.

Questions 5 and 6 should have a slicer that can help visualize the values based on decade and occupation   
  
**Question 7:** You can create a treemap to visualize the Top N Movies and Genres with Occupation and User Age as slicers.  
  
Use the button feature to toggle between the two charts. [Hint - use bookmark along with button]

**Question 8:**You can create a scatterplot between user age group and average rating, with number of users as the size of the bubble. For the second part of the question, you can create a bar chart that visualizes the average rating of movies across years. Use analytics feature to show the average rating across each genre

Questions 7 and 8 should have a slicer that can help visualize values based on the decade chosen.  
  
Question 8 should have a slicer that can help visualize values based on the Genre chosen

**Make sure to include relevant slicers, filters, bookmarks, formatting and navigation features to enhance the report.**