Database Schema Design and Rationale

**Tools Used: Lucid Chart**

**1. Overview**

The database schema is designed to support a dashboard that provides insights into customer viewing habits based on age, location, and genre preferences. The purpose is to assist Streamflix in transitioning from a free to a paid subscription model by identifying trends that influence customer engagement.

**2. Tables and Schema Design**

The database consists of three key tables: Users, Movies, and Ratings. These tables are structured to efficiently store and retrieve data for analysis and visualization.

**Copy\_of\_Users Table**

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| UserID (PK) | STRING | Unique identifier for each user |
| Age | INT | Age of the user |
| Gender | STRING | Gender of the user |
| Country | STRING | User's country |
| SubscriptionStatus | STRING | Whether the user is on a Free or Paid plan |
| TotalWatchTime | INT | Total minutes watched by the user |
| Device | STRING | Device used for streaming (e.g., Mobile, Smart TV) |

**Movies Table**

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| MovieID (Primary Key) | INT | Unique identifier for each movie |
| UserID (FK) | STRING | User identifier for movies watched |
| Title | STRING | Name of the movie |
| Year of Production | INT | Year movie was Published |
| Genres | STRING | Movie genres (comma-separated) |
| Language | STRING | Language of the movie |
| Country | STRING | Country of origin |
| TotalViews | INT | Total number of times the movie has been viewed |

**Copy­\_of\_Ratings Table**

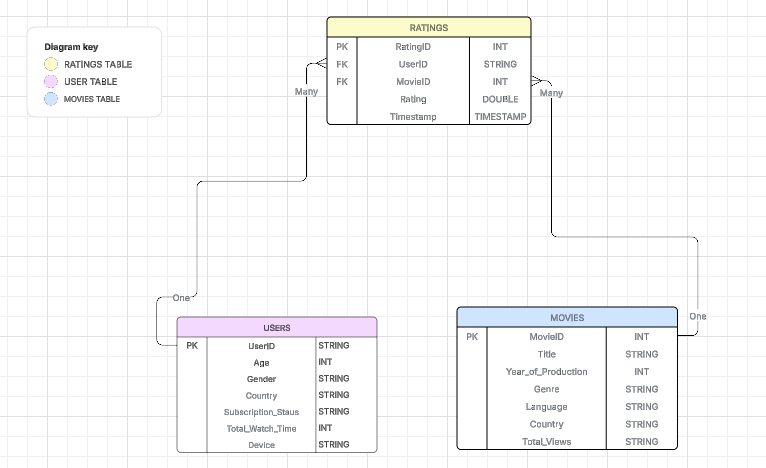
|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| RatingID (PK) | INT | Unique identifier for each rating entry |
| UserID (FK) | STRING | References Users.UserID |
| MovieID (FK) | INT | References Movies.MovieID |
| Rating | DOUBLE | Rating given by the user (scale of 1-5) |
| Timestamp | STRING | Date and time when the rating was given |

**3. Entity-Relationship Diagram (ERD) and Rationale**

An ERD visually represents the relationships between the tables in the database.

Diagram Overview:

* A User can rate multiple Movies, establishing a one-to-many relationship between Users and Ratings. (Users (1) ↔ (M) Ratings)
* A Movie can be rated by multiple Users, forming a one-to-many relationship between Movies and Ratings. (Movies (1) ↔ (M) Ratings)
* The Ratings table acts as a bridge between Users and Movies, enabling many-to-many interactions.



**4. Rationale for Schema Design** 1. Efficiency: The schema is optimized for querying large datasets, enabling quick retrieval of user preferences and trends.  
2. Normalization: Data redundancy is minimized by using separate tables for Users, Movies, and Ratings.  
3. Scalability: The design allows for easy integration of additional features like watch history, recommendations, and subscription behavior analysis.  
4. Power BI Compatibility: The structure aligns with Power BI’s data modeling, allowing for

**Next Steps**

1. Implement the schema in a database management system (e.g., MySQL).
2. Populate the tables with sample data for testing.
3. Write SQL queries to validate relationships and constraints.

**Final Notes**

* The schema is designed to efficiently store and retrieve user-generated movie ratings.
* Any modifications should maintain normalization and data integrity.
* Feedback and refinements are encouraged based on application needs.