Saira Barlas
C12809080

Case Study: Netflix using big data analytics to recommend movies/shows

Netflix is a data driven company which targets Users in regards to their watching history. It uses the data collected of each User to determine what recommendations can be made to the User and what shows are popular on Netflix by the Users. I chose this case study because it applies certain strategies by Netflix in order to keep the Users entertained and from leaving Netflix, hence it is a very strategic business tactic and is quite unique in the application of business intelligence. It also utilises Big Data to successfully keep Netflix on top of its competitors.

Grain

The main focus of this model is to show the recommendations based on what the User from a Profile has seen and recommend a title based on how similar it is to their history of shows/movies seen.

1. User_Dim Table

This table represents the details needed to register with Netflix at the start. **Email** and **Password** are needed for registration. **CardDetails** requires information for the payment of the Netflix account. **FirstName** and **LastName** are needed for the User's identification. **Devices** is a Netflix feature which allows the User to select up to various devices which the User will use to view Netflix e.g. Playstation 4, SmartTV, Smartphone etc. **UserKey** is the primary key of this table which identifies the unique User.

2. Profile_Dim Table

This table represents the various profiles that a User can make with their account. A **ProfileKey** identifies the different profiles owned by one User, as a User can have more than one Profile. **Language** allows the Profile to be viewed in that specified language. **Watched** is a category for a Profile, which shows all the shows/movies that have been seen by the Profile. **CurrentlyWatching** is a category for all the shows/movies which are currently being viewed. **MatureSettings** allows a User to configure a Profile in depending on who will be using that Profile.

3. Time_Dim Table

This is a table that keeps track of the days and hours the User has logged in. **Second, Minute, Hour, Day, Month,** and **Year**. It also keeps track of User watch history i.e. what they watched on what day. **TimeKey** is the primary key of this table which identifies each day and hour separately.

4. Recommendations_Dim Table

This is a table which shows how Netflix uniquely uses data to recommend shows/movies to the User. **RecommendationGenre** shows recommendations based on a similar genre to what the User has already watched. **RecommendationType** gives recommendations either for all Shows or Movies or Documentaries etc. **RecommendationTitle** is the title of the show/movie that is being recommended. **RecommendationKey** is a unique primary key for all the recommendations.

Saira Barlas
C12809080

5. Membership_Dim Table

This table allows a User to pick the membership type of their account. Netflix currently offers 4 types of memberships, ranging from free trial to ultra HD with 5 profiles being used at once! Users have the option to pick from four of these choices and subscribe. **MembershipType** represents the type of membership. **MembershipKey** is the unique primary key for each of these types.

6. Show_and_Movie_Dim Table

This table is a representation of what Netflix has to offer i.e. shows and movies. **Title** is the name of the show/movie and **Type** defines whether it is a show/movie. **Summary** shows what the show/movie is about and it also displays the **Genre** of the show/movie. **Subtitles** are offered for them and **Quality** is also viewable i.e. 1080p. **ShowandMovieKey** is the unique key which identifies each movie and show separately.

7. Content_Fact Table

This is a fact table which displays the foreign keys of all the other tables i.e. **UserKey, ProfileKey, TimeKey, ShowandMovieKey, MembershipKey, RecommendationKey.** It also has **EpisodeAvailable** which shows the number of episodes the movie/show currently has. **Ratings** shows the rating of each show/movie. **Duration** is there to display how long the show/movie is. **WatchingHistorySimilarity** is a unique identifier which is used to calculate how similar a show is from the one the User has seen and the one that the User should be recommended.

Star Schema

I chose a star schema because it's simple in structure and is easily understandable. The queries are also very effective as there were a few tables that are needed. It is also widely supported by a lot of business intelligence tools. Star schema is also the most commonly used data warehousing method.

Query

```
SELECT Title

FROM Show_and_Movie_Dim

WHERE ContentHistorySimilarity FROM Content_Fact >= ContentHistorySimilarity;
```

The above query selects a title from the Show_and_Movie_Dim table and makes a recommendation based on the similarity of the history of the User.