DATAQUEST

Welcome to the Data Analytics Challenge!

Your data team at **G-Flix Analytics** has assigned you a crucial task—analyze and uncover valuable insights from a massive dataset riddled with inconsistencies. Only the sharpest data analysts can clean, structure, and interpret this information to drive impactful decisions!

About G-Flix

G-Flix is a premium **pay-to-watch** movie streaming platform that provides thousands of films and TV shows to its subscribers. Users pay for subscriptions, stream content across various devices, and generate extensive user data. However, this data is far from perfect—duplicate entries, missing values, and incorrect records create obstacles. Your mission? Organize, analyze, and extract meaningful trends to help G-Flix improve its services and business strategy.

The G-Flix Database

To understand user behavior and improve services, G-Flix maintains a rich database filled with key information, currently comprising data of 2000 users.

- **Users** – Subscriber details, including age, country, and registration date.
- **Movies** – A catalog of films with titles, genres, release years, and ratings.
- **Subscriptions** – Records of user plans, including start dates, renewals, and pricing.
- **Watch History** – Logs of what users watch, when, and on which devices.
- **Ratings** – User-submitted movie ratings, helping track audience preferences.

This data holds valuable insights, but errors and inconsistencies have made it unreliable. That's where you come in!

Mission 1: The Data Nightmare at G-Flix

G-Flix is facing a data crisis! Due to system glitches, user errors, and unexpected anomalies, our data has become unreliable. As a G-Flix Data Analyst, your job is to fix the chaos and restore order before key business decisions are made.

- **User Records Mayhem:** Some users have been duplicated, and some users have age values that are unrealistic, potentially setting world records. Invalid emails are also running wild, and people seem to come from unbelievable places! (remember that duplicate records are only present in the .csv files, not the .db file)

- **Subscription Confusion**: Some users have overlapping subscriptions, while others seems to be getting paid to watch movies-seriously?
- **Watch History Glitches**: Our system has picked up some strange patterns in the viewing records. Some users appear to be using unexpected devices. There are even cases where certain ids don't match our database. A closer investigation is required to make sense of these inconsistencies.
- **Ratings Gone Wrong**: Some users rated movies with impossible scores, and others have multiple conflicting ratings for the same film. Can we even trust this feedback?

With these errors—and potentially others that **have yet to be discovered**—business decisions are at risk. It's up to you to clean the data, uncover key insights, and help G-Flix regain control!

Mission 2: Creating the Ultimate Data Dashboard:

Once you've cleaned the data, your next task is to uncover meaningful trends and present them in a compelling dashboard. This dashboard will be presented to the G-Flix executive team, who will use it to make high-stakes business decisions. Look for key trends and unusual patterns—sometimes, the most unexpected insights drive the biggest decisions!

BONUS SQL CHALLENGES: The Ultimate Data Mystery

Mission: J-Flix Data Support

J-Flix, an established streaming platform, has approached G-Flix for help with analyzing their user and content data. Their database is clean, structured and have the same schema as G-flix, but they need precise SQL queries to extract key insights. As a data analyst at G-Flix, it's your job to deliver six essential queries that will help J-Flix optimize their platform

Stay tuned—your SQL mystery cases will be dropped soon in the group. Be ready to think on your feet, adapt quickly, and crack the data puzzles as they come.

Reminder:

Your analysis should be purely data-driven. External assumptions about movie popularity may not apply in this competition.

Guidelines for participants

1. Documentation Requirement:

- Participants must provide clear and relevant documentation for their code. This can be included as comments within the code or as a separate README file.

- The use of AI-generated documentation is strictly prohibited and will result in immediate disqualification.

2.Dataset Submission Format:

- The shared drive folder containing the uploaded datasets or database must be named in the following format "**<team_name>_submission"**

If the dataset is in **.csv format**, the files must be named as follows:

- users.csv
- movies.csv
- subscriptions.csv
- watch_history.csv
- ratings.csv

If the data is in **.db** format, it should be named in the format **"<Team_name>_database".**

Conclusion

The fate of G-Flix's data now rests in your hands! As a data analyst, you have the power to transform messy, unreliable information into actionable insights that drive real business decisions. Whether it's uncovering hidden patterns, fixing critical errors, or solving complex SQL mysteries, your skills will determine how G-Flix moves forward.

Are you ready to take on the ultimate data challenge? Step up, clean the chaos, and showcase your analytical expertise—because only the best will emerge as true DataQuest champions!

Here, you can find all the required files - **ALL THE BEST :)**