CAPSTONE - 2

Problem statement:

A streaming platform company allows different streamers to use its platform to conduct/deliver a live stream session. Now, the company wants to know which are its top streamers. But the company doesn't have a labeled dataset that indicates which are the top streamers and which are not.

This dataset is generated from the live-streaming platform. It consists of all the streams that were conducted on the platform and the information related to each stream - its duration, country it was streamed from, no. of comments received during the stream, no. of viewers who attended the live stream etc.

Tasks:

1. Analyze the dataset to come up with the top 20% streamers.

2. Label these top 20% streamers as "good" streamers, and the remaining as "bad" streamers. This will become your target variable.
   Now create a binary classification ML model that can classify whether any streamer is a good streamer or not.

Dataset:

1. **all_stream_info.csv:** All the streamer related data
2. **all_stream_info - data dictionary.csv:** To understand what each column represents (for reference purpose only)

Evaluation Metric:

Use F1-Score