**Problem Definition / Use Case Form**

**Use this form to log the problem Details & Definition**

|  |  |
| --- | --- |
| **Problem Theme** | Data Analytics and modelling |
| **Technology** | Big Data |
| **Name of the Problem** | EDA of Netflix Contents |
| **Problem Statement** | This dataset consists of tv shows and movies available on Netflix as of 2019. The dataset is collected from Flixable which is a third-party Netflix search engine.  In 2018, they released an interesting report which shows that the number of TV shows on Netflix has nearly tripled since 2010. The streaming service’s number of movies has decreased by more than 2,000 titles since 2010, while its number of TV shows has nearly tripled. It will be interesting to explore what all other insights can be obtained from the same dataset. |
| **The Environment**  What is the ecosystem in which the problem exists?  Who are the Key Users? | Integrating this dataset with other external datasets such as IMDB ratings, rotten tomatoes can also provide many interesting findings.</p>  Inspiration   * Understanding what content is available in different countries * Identifying similar content by matching text-based features * Network analysis of Actors / Directors and find interesting insights * Is Netflix has increasingly focusing on TV rather than movies in recent years.   It is used by Netflix’ content creator |
| **The Current State**  What is the current Impact of the problem? |  |
| **Use Case/ Test Cases** | Test should be available using pytest/behave testing framework |
| **Output Expected** | * Complete Analytical Dashboard with different granularity attributes * No garbage Data * Data Quality and validation check * Complete Data warehouse * Orchestrated * Complete end to end pipeline |
| **Data Repository** |  |