

**School of Computer Science and Engineering**

# J Component Review-1

**Programme : B. Tech (CSE: CORE)**

**Course Title : Foundation of Data Analytics**

**Course Code : CSE3505 Slot : F2**

**Title: Streaming Content Dashboard**

**Team Members: KASHISH BAJAJ (20BCE1790)**

**AKSHIT JAIN (20BCE1818)**

**AKASH RAJ BEHERA (20BCE1829)**

**ABSTRACT**

**As we all know in today’s world data analysis and visualization is becoming important thing because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments. Now a days people do not want to waste any time on viewing bad shows and they first look at the ratings and later they decide what to see. According to this situation we designed our project to make streaming content dashboard which will enable us to visualize all the famous shows in every aspect we can understand in a clear way. We also clustered the combined data from Netflix, Hulu, Disney Plus and Amazon Prime using K-Means and created a recommendation system to find similar movies to what the viewer has watched.**

**OBJECTIVE**

**To create visualize various patterns and trend in OTT content from Netflix, Hulu, Amazon prime and Disney Plus.**

**Cleaning the dataset containing the data and pre-processing it.**

**Applying EDA to find Patterns in the data and visualize them.**

**To create interactive dashboard using tableu about OTT content from Netflix, Hulu, Amazon prime and Disney Plus**

**To create a Recommendation System**

**Recommender Systems (RSs) are characterized by the capability of filtering large information spaces and selecting the items that are likely to be more interesting and attractive to a user.**

**OTT Platforms are the biggest users of recommendation systems. So, in this Project we aim to visualize the content library of top OTT Platforms like Netflix, Disney Plus, Hulu and Amazon Prime. While doing this we will also discover correlations and recurring patterns in the dataset with interesting inferences.**

**Finally, we will see how the recommendation engine works to deliver similar content as quickly as possible.**

**REQUIREMENTS**

* **PYTHON 3.8**
* **MATPLOTLIB**
* **PLOTLY**
* **PANDAS**
* **NUMPY**
* **WORDCLOUD**
* **SKLEARN**
* **SEABOARN**
* **TABLEAU 2021.4 or later**
* **SCIPY**

**MATERIALS**

**For this visualization and analysis, we use feature attributes from the dataset, namely,**

**Type**

**Title**

**Director**

**Cast**

**Country**

**Date Added**

**Release Year**

**Rating**

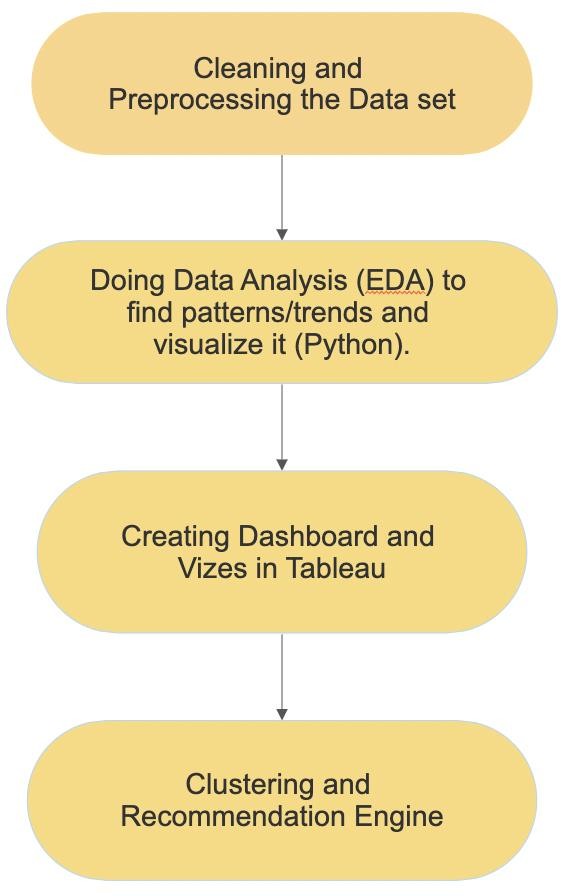
**Duration**

**Listed In**

**Description**

**Each individual dataset contains all the following attributes. During the project we will combine all 4 datasets into one and then we will append a column denoting the OTT platform.**

**METHODS**



**FINDINGS**

**Our Recommendation System takes a movie or show name as input and then narrows its search space to the cluster that they belong to. Then it runs a similarity check on the description of the entered title with every entry on the cluster.**

**It then returns a list of similar movies and which OTT platform you can watch that content.**

**CONCLUSION**

**From the Visualization we gained a lot of Inferences. Like how each platform values movies more than tv shows. We also found that Amazon and Netflix has the biggest content library with Disney & Hulu slowly building their catalogues. We also saw how US is the biggest producer of OTT Content with India coming at a close Second. We also inferred how the growth of OTT Content libraries has been meteoric in recent years, almost growing exponentially. We also saw the rating distribution between the OTTs and how they favor older teens/Adult markets as their main customer segment.**

**Finally, we created and tested the recommendation engine. We can see how such engines use clustering to reduce runtime dramatically while producing high quality results. This also highlights the importance of clustering data in large corporate environments like multinational OTT providers.**

**This proves that clustering isn’t just a mere visualization tool but also a very important machine learning implementation that reduces runtimes in such demanding workloads drastically.**