

**Netflix Data Analysis With Python**

**Presenters:**

**Miss. Keerthana Pulugam\_2010030445**

**Miss. Pravallika\_201003004**

**Miss. Tahseen Begum\_2010030168**

**Miss. Sowgna\_2010030344**

**Guide: Chanda Raj Kumar Rao**

**ABSTRACT:**

The term “Exploratory and Sentiment Analysis” is a conjunction of two separately unique approaches present in the vast field of Data Science. The key to this project is to enhance the value of the Data being utilized, in our case it is Netflix Data – which is an Open-Source Data Set obtained from Kaggle – that was wrangled and exercised to derive maximum insights using EDA – Exploratory Data Analysis and Sentiment Analysis after the amalgamation of two additional sets – Geographical Latitudes & Longitudes and Netflix Title Critics/Reviews Data Set. The project is made using different utility analytical tools present in the Python Library of versatile packages. This paper introduces systematic and insightful usage of methods for Exploratory Data Analysis & Sentiment Analysis by utilizing various packages concerned.

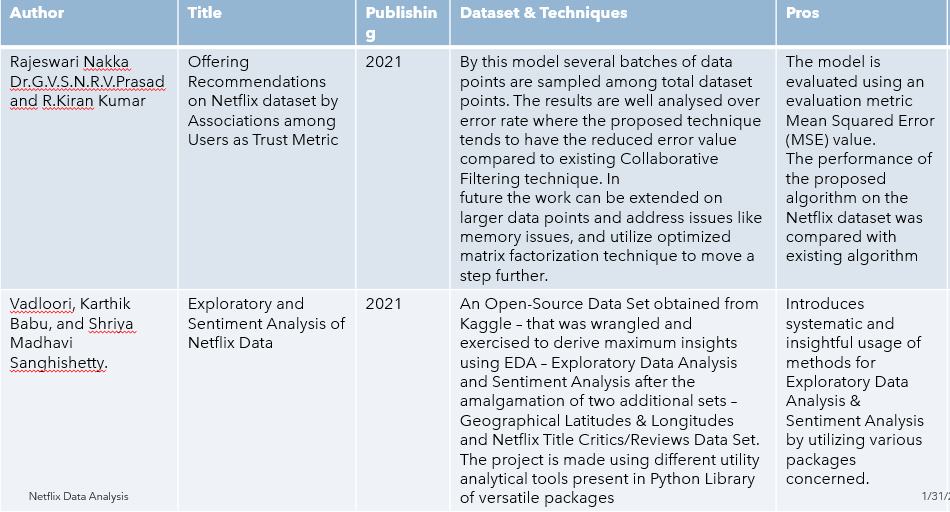
**INTRODUCTION:**

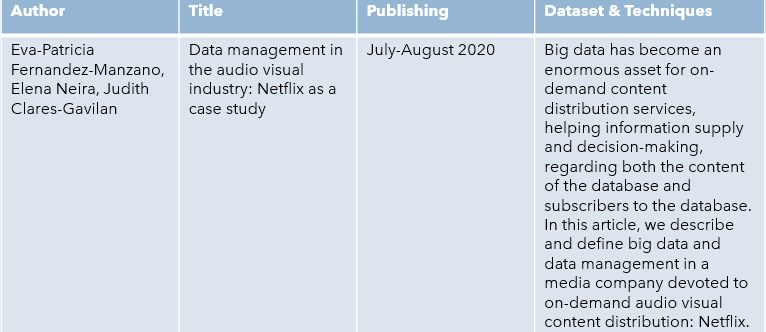
* The term “Data Analysis” is known to be rooted in the statistics space, which itself is known to have a long history. With the help of the statistical development techniques, we can derive interesting outcomes. The advancement of rapid technological implications in the world led to a consequent advent of Big Data; we are constantly being faced with enormous amounts of raw data which is subject to future enhancements based on the required parameters and criteria by an entity. Starting with the collection of data, the most common and subsequent step is to perform the analysis of it. Data analysis is hence known to be a scientific process solely focused on the data as its subject. It begins with retrieving data from various external-cum-internal sources and then performing intrinsic analysis with the data in order to discover and obtain beneficial information catering the needs of an entity.

**EXAMPLE:**

* For example, the analysis of population growth by district can help governments determine the number of hospitals that would be needed in a given area. When collecting the optimal data for analysis it must hold the minimum viability in terms of features and attributes suitable for our analysis. This can be represented in terms of bodily and health-oriented features like Health Status, Age, Male: Female Ratio, BMI etc., will provide much more issue specific insights over the population. It can enable a person to visually represent these features as per the requirements.

**LITERATURE REVIEW:**

****

****

**DATASET:**

****

**METHODS:**

It can enable a person to visually represent these features as per the requirements. Fundamentally, there are two primary methods for data analysis – based on the nature and characteristic of data - qualitative data analysis and quantitative data analysis techniques. These data analysis techniques have the scope to be utilized independently or in combination with other methods in order to gain access to some of the best business and intelligence-oriented insights for making better decisions over the already present data

**RESULTS:**

* We developed a correlation amongst the utility features and established guidance for our analysis.
* Built a plotting for the size of Series & Movies in the dataset and also plotted IMDb\_Scores based on their relevance count.
* Insights based on Super Hit, Hit, Average, and Flop box-office status of a Title using IMDb and Hidden Gem Score as interlinked criteria, decided based on their correlation.

Plotted the country-wise count of Netflix Titles using Funnel Plot and developed a Geospatial Plot using Folium based on the latter feature.   
• Built a Sentiment Analysis Model by fitting the Series/Movie Reviews dataset, which obtains the result by making use of the summary column in the

**CONCLUSION:**

Data Analysis is a fundamental step to address the various needs of a client in any professional spectrum. The varied range of insights that can be derived from a data is itself primarily valuable in nature as there are multiple businesses that are actively looking for futuristic, predictive and descriptive insights from the already present raw data generated by them. It helps the organizations to gain access to numerous concealed patterns, information and bits of knowledge after the analysis had been performed. The analysis that we have just performed using the Netflix data not only provides us with incentives to take smart and intelligent business decisions, but also contribute to the overall growth of the firm.