**\*DISCLAIMER**

    -This project is a sole Intellectual property of mine i.e. **GANDHARV GUPTA.** Kindly acknowledge before copy or use of this project. The dataset is    open and is a property of Kaggle which holds all proprietary rights over it

**NETFLIX-**Region Wise Revenue-Generation analysis with ML

**Data Set:** Netflix Revenue generation data between 2019-2024

**Model Description and Data Dictionary:**

* -The CSV File contains Netflix revenue data from 2019 to 2024 for each of its financial year (from Jan 1st to 31st Dec.)
* -Abbreviations:
  + -Global Revenue: Revenue collected Worldwide
  + -UCAN: United States and Canada
  + -EMEA: Europe, Middle East and Africa
  + -APAC: Asia-Pacific
  + -LATM: Latin America
  + -ARPU: Average revenue per member (for each quarter)
* -All revenue is in US Dollars.

**Problem Statement:**

* -Analysing -Which Region contributes MAXIMUM to GLOBAL revenue generation (that maximizes the profit)
* -Recommending-Which region has the highest potential for revenue generation

**Predictor/Independent Variable::**

* -UCAN: United States and Canada
* -EMEA: Europe, Middle East and Africa
* -APAC: Asia-Pacific
* -LATM: Latin America
* -ARPU: Average revenue per member (for each quarter)

**Dependent/Target Variable:**

* -Global Revenue: Revenue collected Worldwide

**Model Selection:**

* -The problem already has labelled data and requires prediction of a **CONTINOUS NUMERICAL Variable**. Hence, we will make a **SUPERVISED LEARNING** based **REGRESSION MODEL**
* -Also, Descriptive analysis will be done to know the revenue maximizing region

**Data Source:**

* -A Kaggle based UNCLEEANED Netflix\_revenue\_data
* -File Type: A FLAT FILE(.CSV)

**Libraries:**

* Pandas
* Numpy
* Mysql.connector
* Matplotlib
* Seaborn
* Scipy.stats
* Sklearn.preprocessing.StandardScaler
* Sklearn.linear\_model.LinearRegression
* Sklearn.metrics.mean\_squared\_error,r2\_score
* Sklearn.model\_selection.train\_test\_split
* Sklearn.model\_selection.cross\_val\_score,cross,val,predict
* Sklearn.linear\_model.Ridge
* Warnings

**Procedure:**

* Data Collection
  + Data is read from a FLAT FILE(.CSV) using pd.DataFrame() API
* Data Wrangling, Data cleaning, Data Preprocessing, Data Transformation:
  + -Treating Missing values
  + -Data Standardization
  + -Feature Extraction (Extract Year from Date columns)
  + -Converting Dollars-to-Cr
  + -Rounding off values
  + -Data Normalization by Feature Scaling(Z-Score)
  + -Dropping columns that are not required(ARPU Columns )
* Exploratory Data Analysis (EDA) and Visualization:
  + Continuous numerical variables:
    - Finding correlation coefficients
    - Finding p-values
    - Creating correlation matrix
    - Plotting Heatmap of correlation
    - Plotting Pair Plots
    - Plotting Regplot
    - Plotting Residual Plots
    - Plotting Box plots
    - Plotting Bar plots
* Model Building and Predicting:
  + SLR(Simple Linear Regression)
  + MLR(Multiple Linear Regression)
  + Polynomial Regression
* In-Sample Evaluation:
  + MSE(Mean squared error)
  + R2(R-squared)
  + Regression plot for SLR
  + Residual Plots
  + Distribution plot for MLR
* Model Testing(using test-data)
  + Train\_test\_split
* Out-Sample Evaluation(Using test-data)
  + MSE
  + R2
  + Regression plot
  + Residual Plot
  + Distribution plots
* Model Refinement:
  + Cross validation with
    - Cross\_val\_score
    - Cross\_val\_predict
* Making Final Predictions by taking input from User and translating results
* Recommending and translating results to a **Business Idea** (NETFLIX) to maximize its profit

**DEDUCTIONS:(Analysing Trends)**

**### \*\*Year-By-Year BOX plot Trends\*\*:**

i)Year V/S Global Revenue trends:

* -In 2009, Global revenue varied largely b/w 40k CR to 50k CR[Large variation] with few revenues outlying around 60kCR
* -Between 2000 and 2021 Global revenues ranged largely between 50kCR to 60KCr and 60KCR to 70Kcr resp.
* -2022 and 2024 global revenue generation did not saw any largevariations and lied below 70kCr and above 80k Cr resp.
* -significant revenue generation variation seen in the year 2023

ii)Year V/S USA and CANADA streaming revenue generation:

* -The revenue generation between 2019 to 2024 ranged from 20k Cr to 37k Cr-HIGHEST among any of the member countries
* -Despite high revenue generation, The total contribution by UCAN members were still less than the EMEA member countries

iii)Year V/S Europe Middle-East and Africa streaming revenue generation:

* -The revenue generation jumped ACUTELY between 2019 and 2020 by a margin or around 4k Cr.
* -The revenue generation fell in the year 2022 and caught up in 2023
* -Although the maximum revenue generation was around 26k Cr, the overall contribution to Global revenue is MAXIMUM from this group

iv)Year V/S Latin America streaming revenue generation:

* -The revenue generated lied between 6k Cr and 10k Cr between the years 2019 to 2024, which increased steadily
* -Despite lower revenue generation, it contributed highly (though less then EMEA) to global revenue

v)Year V/S ASIA Pacific streaming revenue generation:

* -The revenue generated lied between 3k Cr and 9k Cr between the years 2019 to 2024, which increased steadily
* -The revenue generation from ASIA contributed almost the least, as compared to other member counties, to global revenue

vi)Maximum Contributors to Global Revenue:

* EUROPE, MIDDLE-EAST,AFRICA--->LATIN AMERICA--->USA,CANADA--->ASIA,PACIFIC

**Model's USE CASE**

1. -The model understands the strength of correlation among: revenue generated by specific regions of the world and its influence on Global revenue generation of NETFLIX
2. -The model can recommend(along with descriptive analysis),the most profit maximizing regions and poor performing regions
3. -The model can project NETFLIX global revenues based on inputs from revenue generated from various regions of the world
4. -The model invariably can tell, which regions need to improve more, or performed poorly, or any other kind of irregularity in revenue increase
5. -The above relevant info. can greatly help NETFLIX(or any org.) to maximize its profit or revenues

**Final Remarks and Conclusions:**

1. -USA and CANADA are the highest revenue generating regions
2. -Despite USA and CANADA's high revenue generation, EUROPE, MIDDLE-EAST, AFRICA and LATIN AMERICA influence NETFLIX GLOBAL Revenues the most
3. -It becomes a matter of company's discretion, whether to exploit the highest revenue generating regions or to invest more in those regions that have greater share of contributions overall in their global revenues
4. -ASIA-PACIFIC region has the contributes least among all the regions to NETFLIX global revenues
5. -Given the high population in these areas, greater investment needs to be done in ASIA PACIFIC region to improve their contribution share
6. -Finally, all regions saw a gradual and steady(except few) increase in revenue over time

**\*DISCLAIMER**

-This project is a sole Intellectual property of mine i.e. **GANDHARV GUPTA. Kindly** acknowledge before copy or use of this project. The dataset is    open and is a property of Kaggle which holds all proprietary rights over it.

END.