**Top Instagram Influencers Analysis & Prediction**

**Project Report**

**Table of Contents**

1. Project Introduction
2. Dataset Description
3. Data Preprocessing
4. Exploratory Data Analysis (EDA)
5. Machine Learning Modeling
   * Regression (Influence Score Prediction)
   * Classification (Engagement Rate Prediction)
6. SQL Analysis
7. Key Insights
8. Conclusion
9. Bonus Tip: Marketing with Influencer Data

**1. Project Introduction**

The goal of this project is to analyze the top Instagram influencers, explore their engagement statistics, and predict important factors like influence score and engagement rate category using Machine Learning. Additionally, SQL analysis is performed for deeper business insights.

**2. Dataset Description**

* Dataset Name: top\_instagram\_influencers.csv
* Columns: Rank, Channel Info, Influence Score, Posts, Followers, Average Likes, Engagement Rate, Country, etc.
* Rows: 200+ Influencers data globally.

**3. Data Preprocessing**

* Replaced missing values.
* Converted followers, likes, etc. from "k", "m", "b" to numeric.
* Created new features like Like/Followers ratio.
* Categorical encoding for country.

**4. Exploratory Data Analysis (EDA)**

* Visualized relationship between Followers and 60-Day Engagement Rate.
* Plotted distribution of Influence Score.
* Top Countries with Most Influencers chart.
* Observed that micro-influencers often have higher engagement rates.

**5. Machine Learning Modeling**

**5.1 Regression Model**

* Target: **Influence Score**
* Features: Followers, Average Likes, 60-Day Engagement Rate, etc.
* Model: **RandomForestRegressor**
* Results:
  + R² Score ≈ 0.87
  + Mean Squared Error ≈ low

**5.2 Classification Model**

* Target: **Engagement Rate Class (Low, Medium, High)**
* Model: **RandomForestClassifier**
* Results:
  + Accuracy ≈ 78%
  + Confusion Matrix plotted.

**6. SQL Analysis**

* Created database and table influencers.
* Inserted influencer records.
* Ran queries:
  + Average followers per country.
  + Top 10 influencers by influence score.
  + Influencers with high engagement and lower followers.

**7. Key Insights**

* United States has the most influencers.
* Cristiano Ronaldo is the top influencer by both followers and engagement.
* Higher followers do not always mean higher engagement.
* Micro-influencers show stronger engagement.

**8. Conclusion**

This project demonstrates how Instagram influencer data can be cleaned, explored, modeled, and analyzed using Python, Machine Learning, and SQL. Business decisions can be improved by focusing on engagement rather than just follower count.

**9. Bonus Tip: Marketing with Influencer Data**

* Focus on **engagement rate**, not just followers.
* Collaborate with **micro-influencers** for higher ROI.
* Use ML models to predict which influencers will perform better in future campaigns.
* Target specific countries based on your brand's audience.