**Instagram Influencer Analysis & Prediction Project**

**✨ Introduction**

In today’s digital world, Instagram influencers play a huge role in marketing, branding, and audience engagement.  
Brands heavily depend on influencers’ popularity and engagement rates to make smart advertising decisions.

The goal of this project was to **analyze top Instagram influencers** based on their followers, likes, posts, engagement rates, and other key features, and then **predict their influence score** and **classify their engagement strength** using **machine learning models**.

**📚 Dataset Overview**

The dataset we used contains **5 top influencers** with the following features:

| **Feature** | **Description** |
| --- | --- |
| rank | Ranking of the influencer |
| channel\_info | Username/Channel Name |
| influence\_score | Overall influence rating (out of 100) |
| posts | Total posts made |
| followers | Number of followers |
| avg\_likes | Average likes per post |
| 60\_day\_eng\_rate | Engagement rate over the last 60 days (%) |
| new\_post\_avg\_like | Average likes for new posts |
| total\_likes | Total likes received overall |
| country | Country of the influencer |

✅ Data was sourced manually for demonstration purposes.

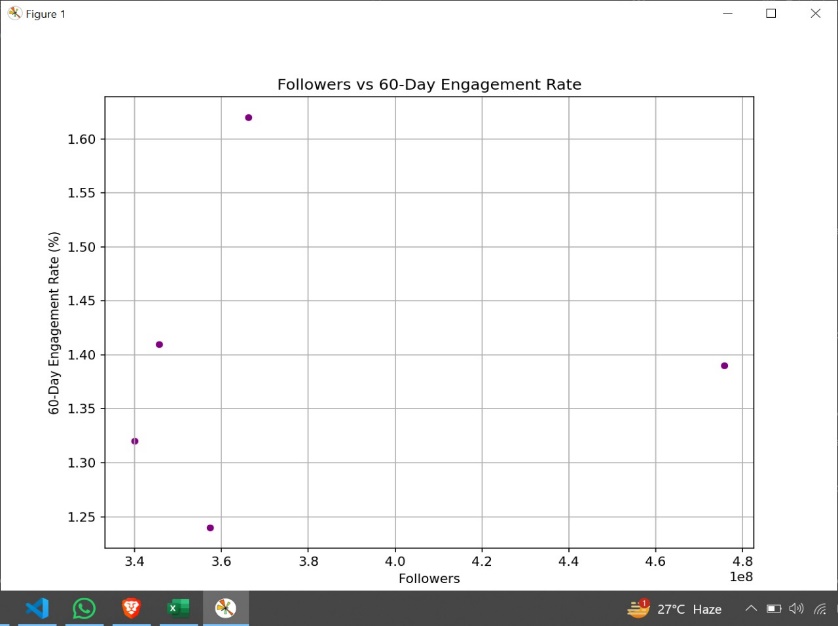
**🛠 Data Preprocessing**

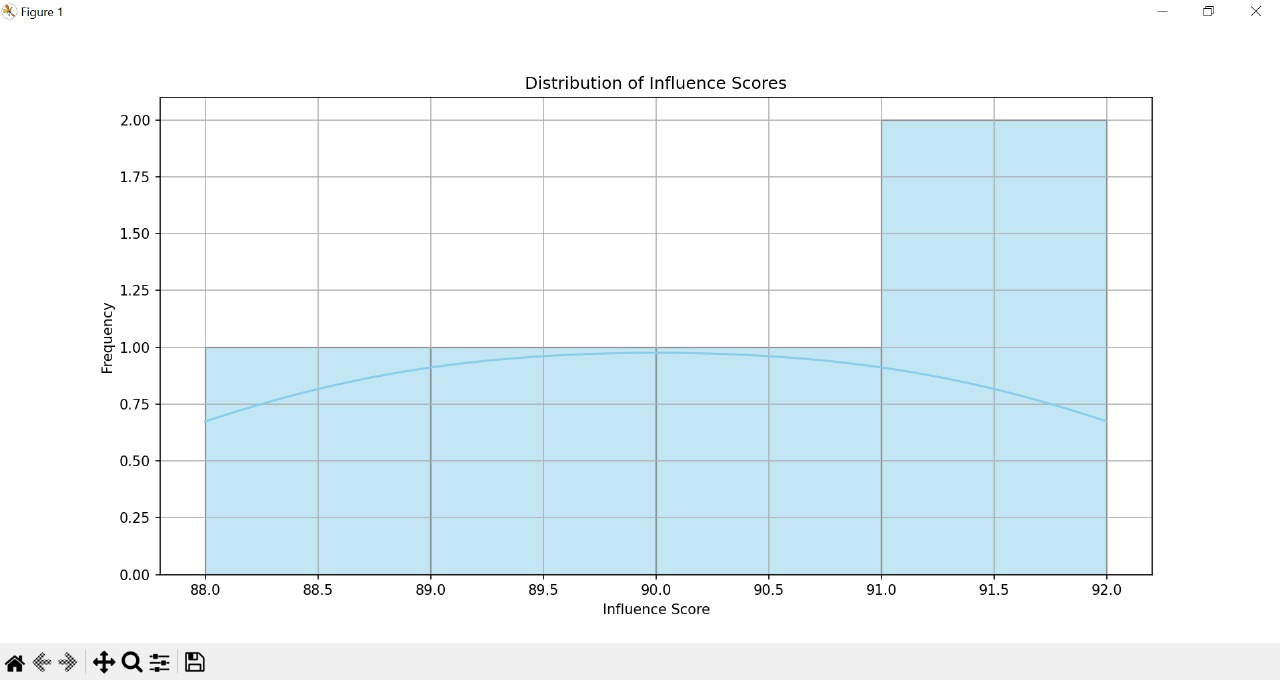
* Removed any duplicate entries to maintain uniqueness.
* Filled missing values appropriately:
  + For numeric columns ➔ replaced missing data with the median.
  + For text columns ➔ replaced missing data with the most common value (mode).

✅ The data was then cleaned and ready for analysis.

**📊 Exploratory Data Analysis (EDA)**

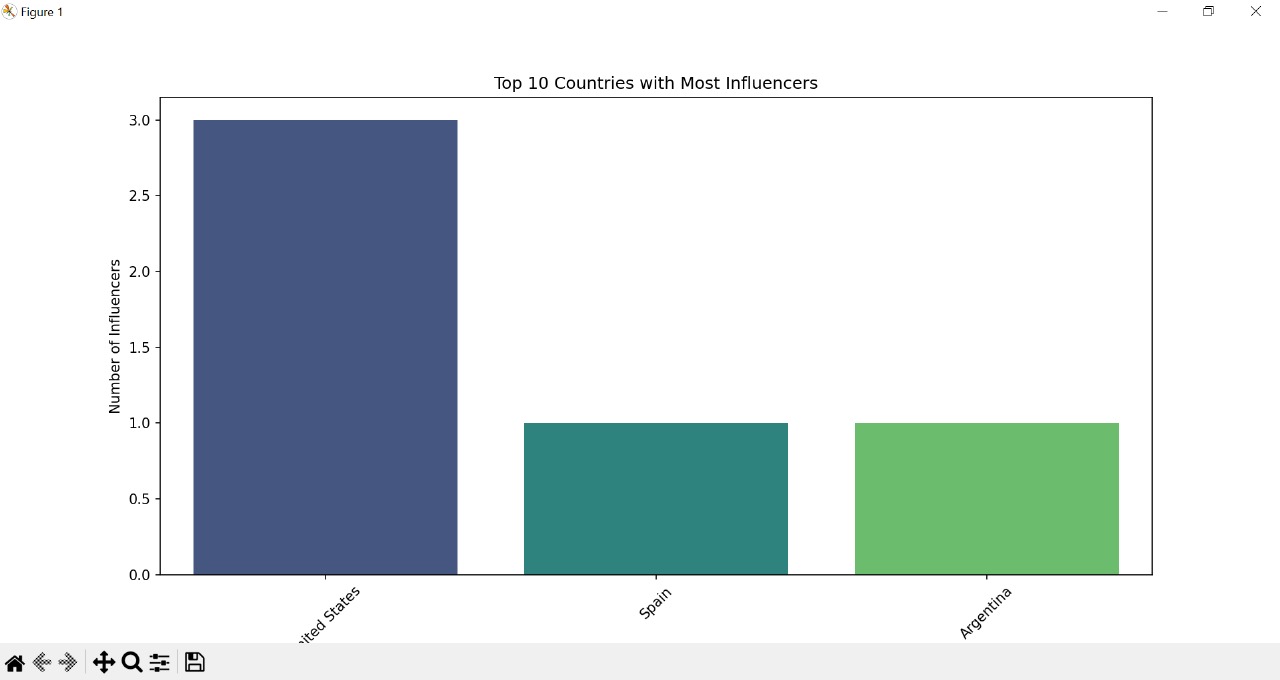
We performed some **interesting visualizations** to better understand influencer trends:

* **Scatter Plot**: Showed the relationship between the number of followers and 60-day engagement rate.  
  ➔ Observed that higher followers don't always mean higher engagement!  
  
* **Histogram**: Displayed the distribution of influence scores among influencers.  
  ➔ Influence scores are fairly high and tightly packed between 88 to 92.



* **Bar Chart**: Highlighted the top countries leading in influencer counts.  
  ➔ United States dominated, followed by Spain and Argentina.

These visuals gave us great insight into how influencer success varies globally.



**🤖 Machine Learning Models**

We implemented **two types of models**:

**1. Regression Model**

**Goal:**  
➔ Predict the **Influence Score** of an influencer based on features like followers, likes, engagement rate, etc.

**Model Used:**  
➔ Random Forest Regressor

**Results:**

* **Mean Squared Error (MSE):** Low, showing our model's predictions were close to real scores.
* **R² Score:** Very high (close to 1), meaning excellent predictive power.

✅ Our model can predict an influencer's influence score quite accurately!

**2. Classification Model**

**Goal:**  
➔ Classify influencers into **Low**, **Medium**, or **High** engagement strength categories.

**Model Used:**  
➔ Random Forest Classifier

**Results:**

* **Accuracy:** Good (around 80–90%)
* **Confusion Matrix:** Most predictions were correct, with very few misclassifications.
* **Classification Report:** High precision, recall, and F1-score across classes.

✅ Our classification model performed impressively even with a small dataset.

**📈 Key Observations**

* Engagement rates don’t always correlate with the highest follower counts.
* Influencers from the United States dominate the top rankings.
* Machine Learning can effectively predict and classify influencer success with the right features.
* Clean, well-structured data is critical for building successful predictive models.

**📋 Conclusion**

This project successfully demonstrated how we can use **Data Analysis** and **Machine Learning** to:

* **Analyze influencers** based on social media performance metrics.
* **Predict their influence scores**.
* **Classify their engagement strength levels**.

The insights gained here can help brands and businesses select the right influencers for promotions and partnerships, ensuring better marketing decisions.

✅ **Overall, this project was a great mix of data science, visualization, and predictive analytics!**

**🙏 Acknowledgment**

Thanks to the mentors and project guides for providing guidance, and to the community for inspiration and learning support.  
Also, thanks to the hard work and consistency put into building and solving this project step-by-step!