Project Scenario

You are working as an **entry-level data analyst** at a digital strategy agency that manages content across multiple social media platforms. The agency has collected thousands of posts from various clients and now wants to understand what factors drive higher engagement. Your team relies on you to analyze a dataset that includes each post's **platform, post type, sentiment score, posting time, and engagement metrics**, such as likes, shares, and comments.

Business Context

Clients depend on the agency to optimize their social media presence, but their content teams often operate on intuition rather than data. They create posts across different platforms—Instagram, Twitter, Facebook, LinkedIn—and vary the content format (text, image, video, announcement, promotional post, etc.).

Your agency wants to replace guesswork with **evidence-based insights**. The goal is to understand:

- Which platforms generate the most engagement
- What types of posts attract the strongest reactions
- Whether sentiment (positive, negative, neutral) impacts engagement
- How posting time and day affect user interaction

To answer these questions, you will use Python to analyze the dataset, identify trends, create visualizations, and build a predictive model that estimates total engagement based on content and posting characteristics.

Project Objectives

- Discover which platforms and post types consistently drive higher engagement
- Analyze how sentiment, posting time, and post length influence user behavior
- Build a predictive model that estimates engagement using post features
- Provide clear, data-driven recommendations to improve client performance
- Help the agency refine content strategy to increase reach and audience interaction

---

Your Challenge

As the analyst, you are responsible for:

- Cleaning and preparing the dataset for analysis
- Engineering new features (e.g., total engagement, post length, posting patterns)
- Visualizing trends across platforms, post types, and sentiment levels
- Training and evaluating machine learning models that predict engagement
- Communicating insights and recommending strategies to stakeholders

By the end of the project, you will deliver a comprehensive report that helps marketing teams understand what makes a post successful — and how clients can use data to consistently boost their social media performance.

Project Introduction

This project explores the factors that drive social media engagement using a real dataset containing posts from multiple platforms. By analyzing attributes such as platform, post type, sentiment score, posting time, and engagement metrics, the goal is to uncover trends that help marketing teams understand what makes content successful.
Using Python, I cleaned, transformed, and analyzed the dataset, engineered new features, visualized engagement patterns, and built predictive models to estimate the engagement a post

is likely to receive.
The outcome is a data-driven framework that social media strategists can use to optimize content planning, posting schedules, and audience targeting.

Methods & Tools Used

- **Python** for data cleaning, feature engineering, and modeling
- **Pandas & NumPy** for data preparation and transformation
- **Matplotlib & Seaborn** for visualizations and trend analysis
- **Scikit-learn** for building regression models and evaluating performance
- **One-Hot Encoding & Scaling** for preprocessing categorical and numerical features
- **Machine Learning Models**: Linear Regression, Random Forest, Gradient Boosting
- **Feature Importance Analysis** to identify key engagement drivers

---

Insights & Recommendations

Based on the analysis and model results, several actionable insights emerged:

Key Insights

- **Platform matters**: Some platforms consistently show higher engagement, suggesting clients should prioritize where their audience is most active.
- **Post type influences engagement**: Visual or interactive post types generate more reactions than plain text or announcements.
- **Sentiment has predictive power**: Posts with more positive sentiment tend to attract more engagement; highly negative tones underperform.

- **Timing impacts visibility**: Certain posting times and days correlate strongly with higher interaction rates.

- **Longer posts don’t always perform better**: Engagement depends more on content type and platform than length.

Recommendations

- Focus posting efforts on **high-engagement platforms** identified by the analysis.

- Use **image- or video-based post types** when possible to boost reactions.

- Maintain a **positive and engaging tone** in content to improve interactions.

- Schedule posts during **peak engagement windows** identified from historical patterns.

- Use predictive modeling outputs to **prioritize content themes** with the highest likelihood of success.