**Project: Identifying and Leveraging Social Influencers**

As part of my MSBA coursework, I worked on a group project aimed at detecting social media influencers, quantifying their financial impact, and leveraging their influence for business applications. The project was divided into two key parts:

**Part I: Predicting Social Media Influencers Using Network Analytics**

* Analyzed a Kaggle dataset containing social influence data, where each observation compared two individuals (A and B) based on Twitter activity (followers, retweets, network characteristics, etc.).
* Built a predictive model using **logistic regression** to classify which individual in each pair was more influential.
* Explored feature transformations (e.g., A/B or A-B) to improve model interpretability and performance.
* Normalized all data to ensure values were between 0 and 1 for consistency.
* Evaluated model performance using a **confusion matrix** and identified key predictors of influence.
* Assessed how businesses can utilize the model to strategically engage with influencers.

To quantify the **financial value of influence**, I calculated the **lift in expected net profit** for a retailer using analytics to identify influencers versus a non-analytical approach.

* Considered a scenario where an influencer's tweet increases the probability of purchases.
* Compared the expected profit under two strategies:
  + A flat $5 per tweet for all users (no analytics).
  + A targeted $10 per tweet for identified influencers (using the model).
* Conducted a **profitability analysis** to determine the increase in net profit from using analytics, and compared this with an ideal "perfect" model scenario.

**Part II: Identifying Influencers from Twitter Data**

* Collected a set of tweets and structured them into a network by mapping interactions such as **retweets, mentions, and replies**.
* Created a **network dataset** in CSV format, where each row represented a connection between two users, categorizing them by interaction type.
* Processed the dataset to calculate **degree, betweenness, and closeness centrality** for each node using network analysis techniques.
* Used **Python (NetworkX)** to analyze the influencer network.
* Applied the coefficients from Part I’s predictive model to compute an **influence score** for each user, incorporating key network characteristics.
* Generated a **ranked list of the top 20 influencers** based on influence scores.

**Key Takeaways and Business Implications:**

* Developed a data-driven approach to identifying social media influencers.
* Demonstrated how businesses can optimize influencer marketing strategies using predictive analytics.
* Showed the financial benefits of using an analytical approach to influencer selection.
* Applied network analysis techniques to real-world Twitter data, leveraging centrality metrics to refine influencer rankings.

This project provided hands-on experience in **predictive modeling, network analytics, and financial analysis**, offering valuable insights into leveraging social media influence for business growth.