Model Card: Regression Model – XGBoost for Deposit Forecasting

Goal:
Predict each player's expected deposit over the next 30 days using behavioral and transactional features.

Data Window:

- **Historical period:** 3 months of player data
- **Target period:** Last 30 days before each player's final activity (defined per player)

Key Features Used:

- Average session length, number of active days, total deposits
- Historical deposit patterns, engagement frequency

Performance Metrics:

- **R^2 score:** -0.147
- **RMSE:** 41$
- **Error Pattern:** Low prediction variance, behaved like baseline model.
- **Top drivers (by gain):**
    - Deposit_count, avg_deposit, avg_session_length

SHAP Insights:

- High avg_deposit and active_days → increase predicted value
- High total_deposit and num_sessions → decrease prediction (saturation/churn risk)

Caveats:

- Limited variance capture for extreme deposit behaviors
- Player behaviors may be nonlinear or segmented in ways XGBoost doesn't fully capture
- Feature encoding and scaling may affect interpretability

Next Steps:

- Explore more interpretable models or model ensembles

Model Card: Clustering Model – KMeans Behavioral Segmentation

Goal:
Segment players into behavior-based clusters to enable personalized product strategies and comparative analysis.

Data Window:

- **Historical period:** 3 months of player data
- **Target period:** Last 30 days before each player's final activity (defined per player)

Key Features Used:

- avg_session_length, active_days, total_deposit, deposit_count, etc.
- Scaled numerical features only

Performance Metrics:

- **R^2 score:** 0.002
- **RMSE:** 39$
- **Error Pattern:** Low prediction variance, behaved like baseline model.

Methodology:

- KMeans clustering with optimal K=11 (selected via Elbow Method)
- Added cluster labels to player data

Insights:

- Calculated average next-30-days deposit per cluster
- Clusters had very similar average 30 days deposit. This indicating low model performance chances.

Caveats:

- Clustering assumes fixed group structure (K fixed)
- Cluster labels are not inherently interpretable

Next Steps:

- Profile clusters with additional metadata (e.g., geography, acquisition channel)
- Use clusters as features in future models

Time Series Model (ARIMA) – Not Applicable

Reason:
Although ARIMA (or ARIMAX) would typically be useful for time series forecasting per user, in this case, each player has **insufficient data points** to train a meaningful personal time series model. Most players don't have long enough deposit histories to fit or validate ARIMA-type models.