Topps Now NBA Print Run Prediction Project

Machine Learning & Basketball Card Analytics Case Study

Prepared by: Jeff Adams

Executive Summary

This project applies machine learning techniques to predict print runs for the Topps Now NBA card series. The work includes data preparation, feature engineering, model development, performance evaluation, and insight generation to understand the factors that drive collector demand.

1. Data Overview

The dataset consists of Topps Now NBA card releases, including player attributes, team information, card-level features, and print runs. Additional engineered features were created to better capture collector behavior and print run patterns.

Key Columns

- Date
- Player
- Team
- Status (Rookie/Vet)
- Print Run
- Short Print Indicators
- Auto/Relic Flags
- Superstar Tier
- Team Tier

2. Feature Engineering

Several domain-informed features were engineered to strengthen model performance:

- **IsRookie**: Binary flag indicating rookie cards.
- **IsSuperstar**: Categorizes players based on marketability.
- **TeamTier**: Groups teams into 7 tiers based on market size and engagement.
- **HasAutoRelic**: Flags autograph or relic cards.
- **SPInserts**: Captures scarcity indicators.
- **LogPrintRun**: Log-transformed target variable to stabilize variance.

3. Modeling Process

A Random Forest Regressor was selected due to its robustness in handling non-linear relationships and mixed variable types. A log-transformed target was used to address skewness in print run distribution.

Model Performance

- R^2 Score: 0.585: explains 59% of the variance
- Log MAE: 0.741: predictions are within .74 units of true value
- Log RMSE: 0.983: large misses do occur but low enough to support pattern discovery.

These results indicate the model is capturing a meaningful portion of the variance in print runs and generate reasonable predictions when working with limited historical data. As more cards are released and added to the model, the stability and accuracy are expected to continue improving.

Feature Importance

The strongest predictors of print run volume were:

1. Team Tier
2. Autograph/Relic Indicator
3. Superstar
4. Rookie
5. Short Print Flag

4. Visualizations

Chart 1: Actual vs Predicted Log Print Runs

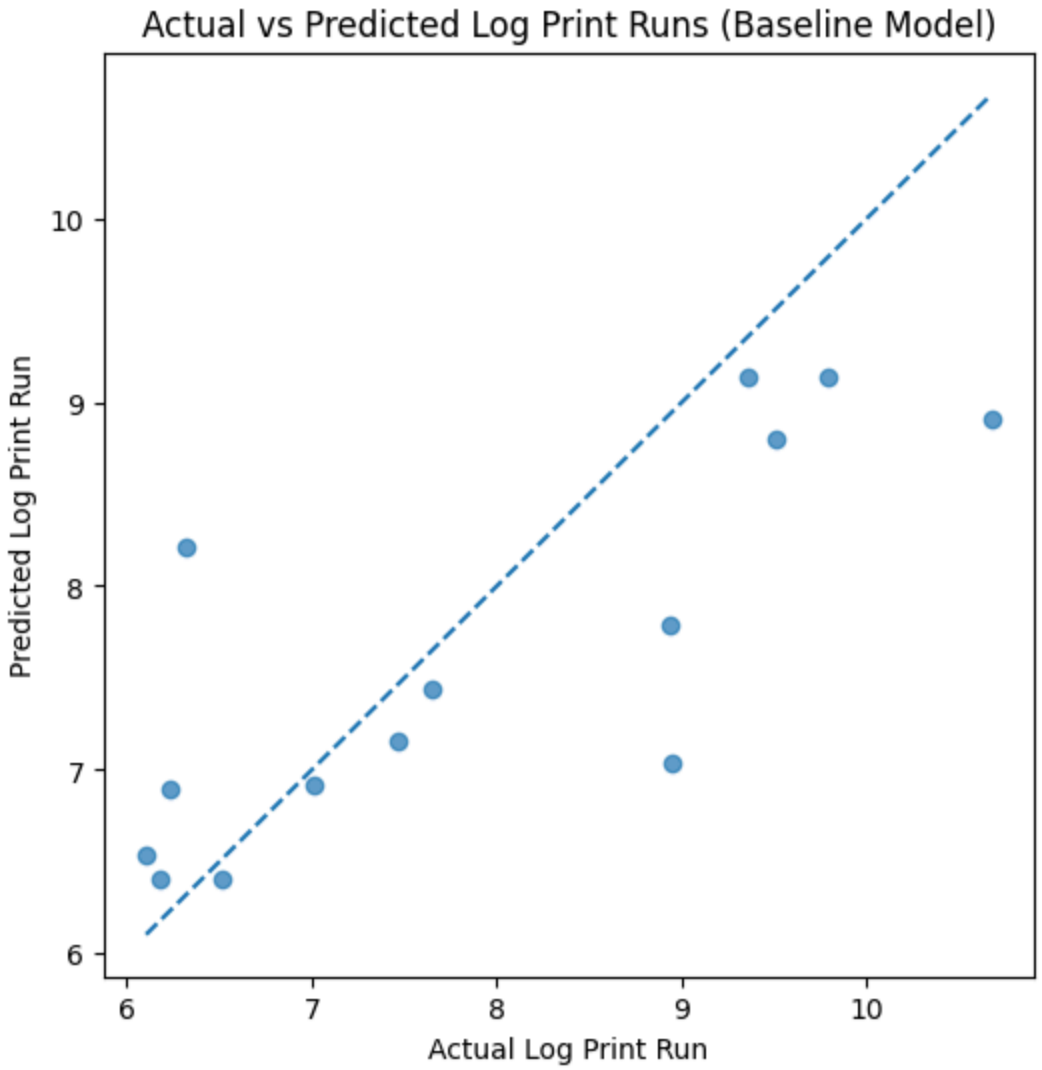

Chart 2: Feature Importance Bar Chart

	feature	importance
4	TeamTier	0.442979
2	HasAutoRelic	0.240069
3	IsSuperstar	0.152065
0	IsRookie	0.146177
1	SPInserts	0.018710

5. Insights & Interpretation

Key behaviors observed in the data:

- Rookie cards consistently produce higher print runs.
- High-tier teams drive greater collector interest.
- Auto/Relic cards cause predictable demand spikes.
- Log transformation significantly improved model behavior.
- Marketability factors (Superstar Tier, Team Tier) outperform numeric stats.

6. Next Steps

Opportunities to improve accuracy as more cards are released and data becomes available:

- Incorporate game performance metrics (milestone moments, outlier results , debut performance).
- Add Google Trends search volume as a player popularity signal.
- Experiment with gradient boosting models (XGBoost, LightGBM).
- Build a Power BI dashboard to monitor predictions.
- Introduce player streak indicators (recent highlight plays, award momentum).

Conclusion

This project demonstrates end-to-end analytics capabilities—data engineering, modeling, visualization, and actionable insights. As additional Topps Now NBA cards are released, the model will continue to evolve and improve in accuracy.

https://www.notion.so/Topps-Now-NBA-Print-Run-Prediction-Project-2c417528456280c0a300fffeab6e3f31?source=copy_link#2c417528456280138b4ff86795d8c136