Social Basketball Analytics: Predicting Excitement

We The Hack

Raymond Chou, Alison Chow and Victor Denisov

Background

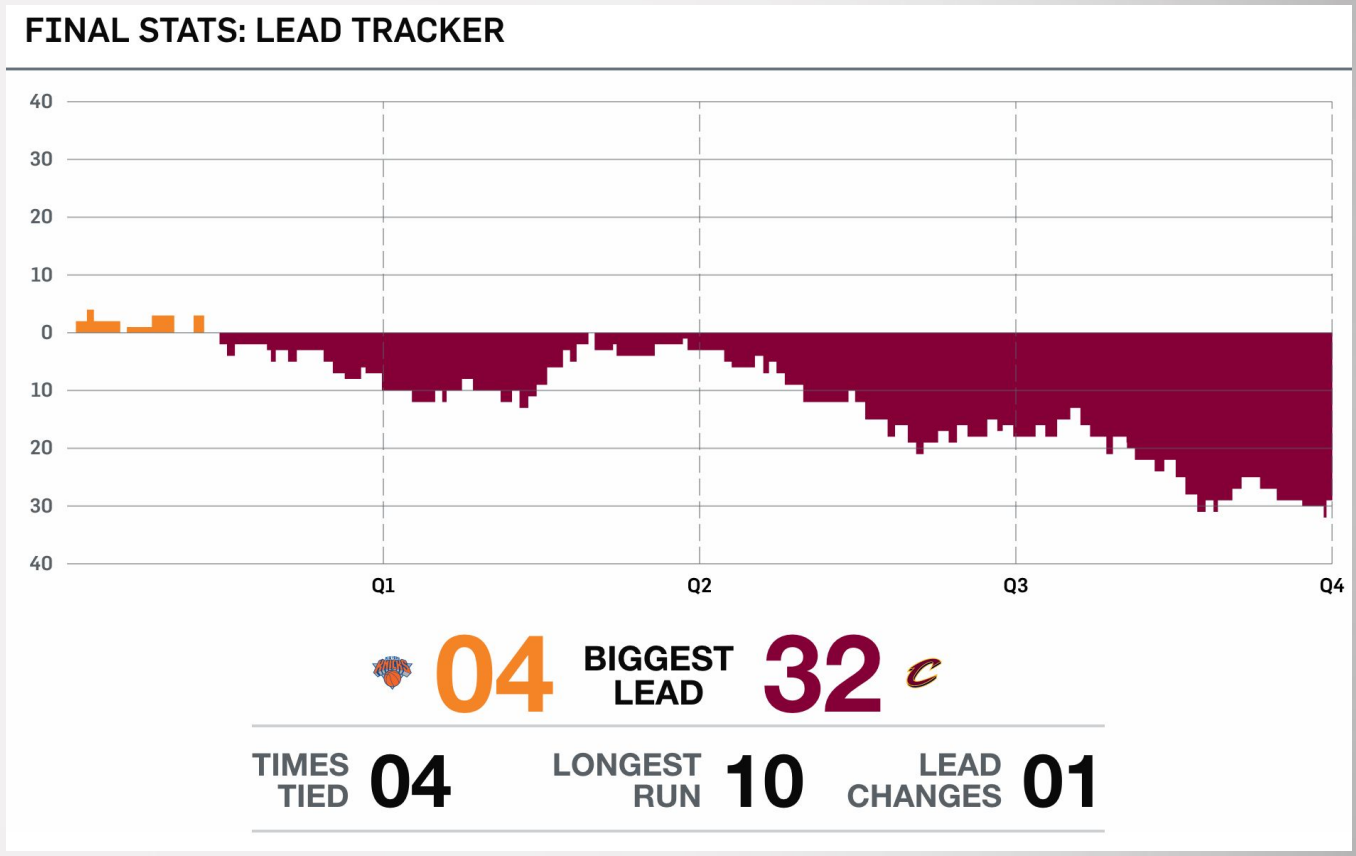
What is a 'run'?

- 8 consecutive unanswered points to surpass the opponent
- Based on a play-by-play estimation and validated with historical box score data

What is an 'exciting' run?

- Capture "exciting" runs through Twitter API data

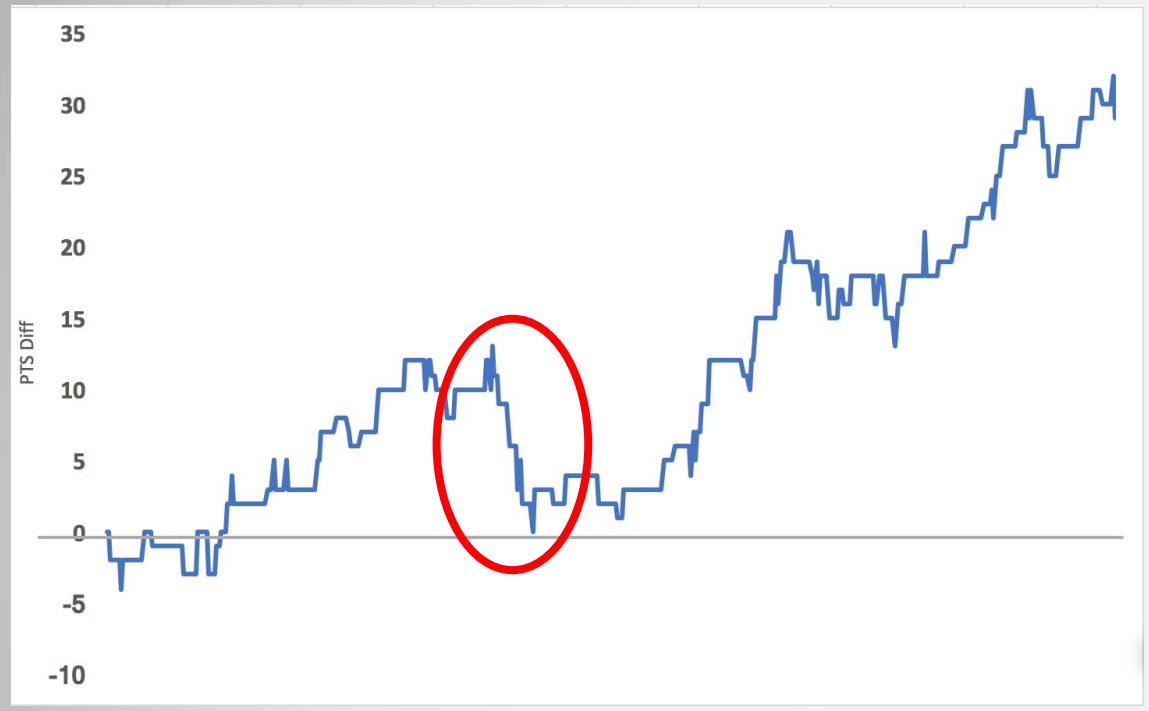
Cleveland Cavaliers(H) vs. New York Knicks(A) Point Differential



- Long runs lead to devastating point differentials

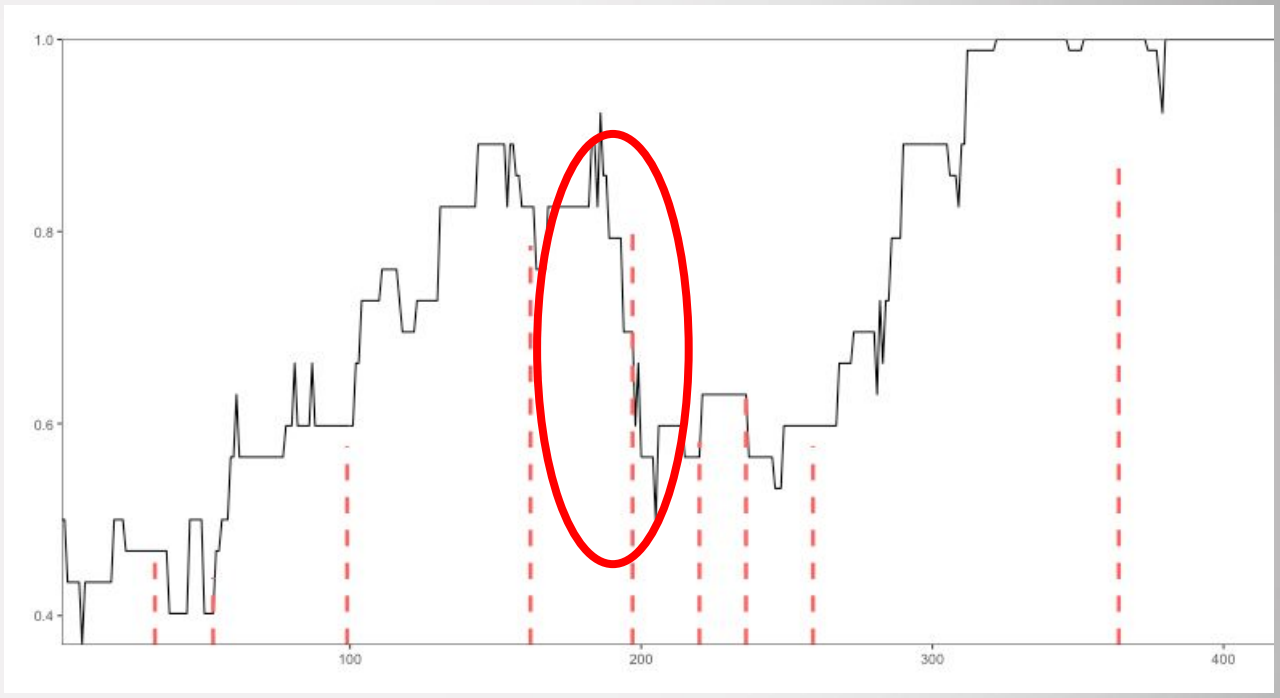
Methodology & Analysis

Point Differential Vs. Time



- Linear regression model predicted win percentage with 92.9% accuracy
- **Point Differential** as the most significant feature

Win Percentage Vs. Time



- Prediction of breakout points signified in red dotted line
- Each breakout point signifies an “exciting” run in a game

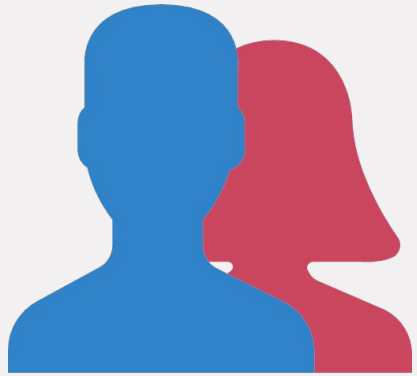
Social Analysis with Twitter API and Sentiment

TwitteR

RSentiment

Uses social media interactions as a measure of excitement

Performed sentiment analysis on sample tweets near breakout points in win percentages



Our model can retrieve number of tweets per minute by calling Twitter's REST API in R

Our model uses a natural language processing (NLP) based package in R

Output: Very Positive

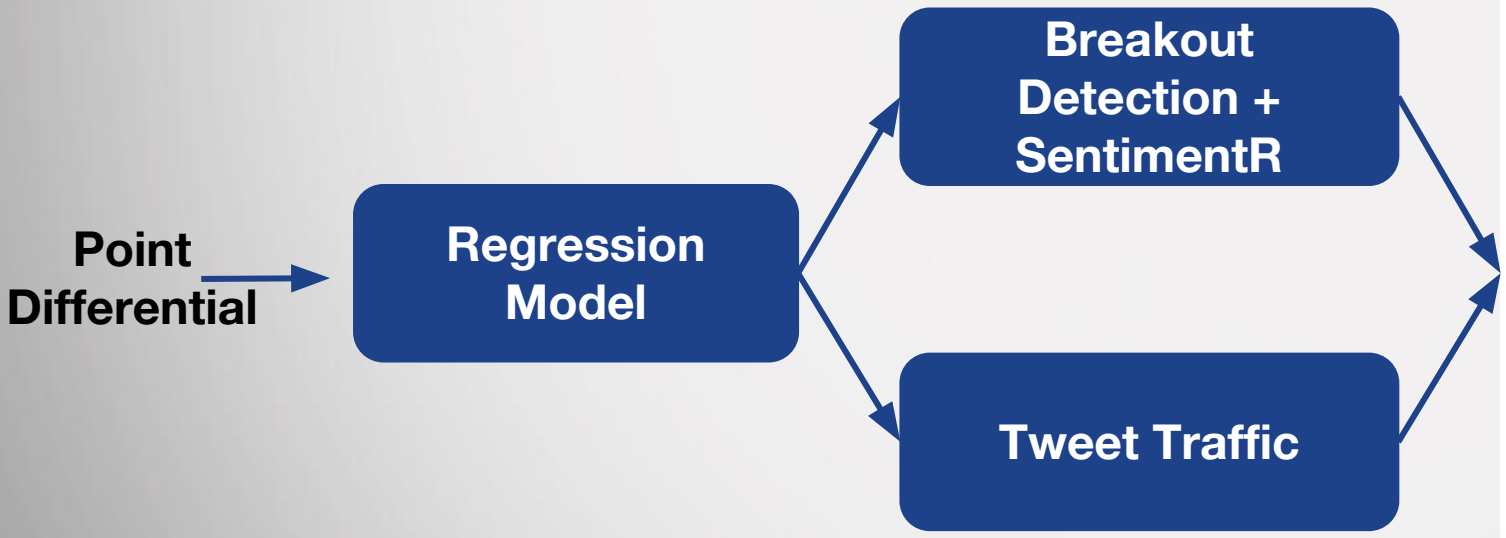
Overview

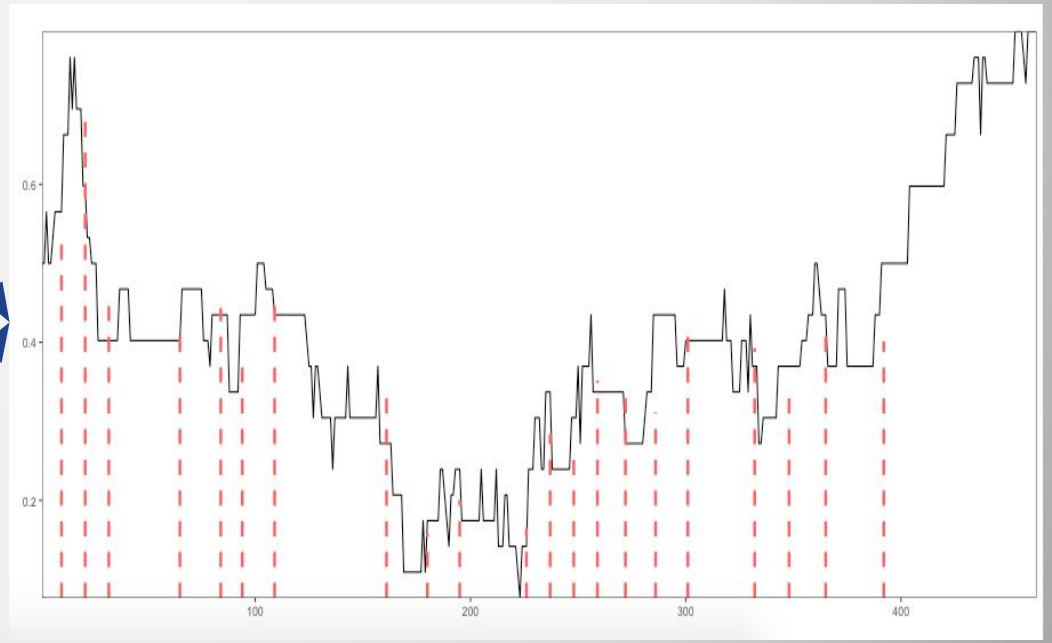
- When is Golden State going to go on a run?
- Which teams are the most “exciting” to watch?
- When is a game boring/statistically over?

Home Away

Vs.

Win Percentage Vs. Time






NBA