

Tech Insight Case Study

Correlation between NBA teams popularity and profitability

Types of data we are looking for

- Popularity
 - Engagement on social media
 - Perception of the public
- Profitability
 - Revenue
 - Salary mass
- Both
 - Attendance

Data sources

- Forbes
 - Reliable magazine with a focus on business
- ESPN
 - Broadcast home of the NBA
- Twitter
 - One of the most popular social media

Data acquisition

- No dataset already available
- Scraping required to extract the data
 - Get API endpoint, query it and format the data
- Use existing public API (Twitter)

Data understanding

- Putting together individual datasets from different sources
- Add new calculated metrics (ex: average)
- Fill missing data

Other ideas

- Sentiment analysis on tweets
 - Lots of meaningless tweets (ads, tweets hard to read without context)
 - Possible, maybe as a separate study
- Find ticket prices
 - Different reseller for each team, so have to do the work 30 times
 - Websites protected against scraping
 - Possible with more time

Production environment

- Requirements
 - Mostly reliability
 - Scalability
- Cloud architecture
 - Google Cloud BigTable
 - NoSQL database