**Case study 1**

**Data Identification for Analyzing the Relative Positions of Smaller Players in Sports Betting and Online Casino Industry (take FanDuel Group, BetMGM, and DraftKings as examples)**

[ABSTRACT 2](#_Toc126258821)

[DATA IDENTIFICATION 2](#_Toc126258822)

[Company Data 3](#_Toc126258823)

[Industry Data 4](#_Toc126258824)

[Customer Data 6](#_Toc126258825)

[Data Prioritization and Weights Design 7](#_Toc126258826)

[ANALYSIS 8](#_Toc126258827)

[INTEGRATION WITH CS 11](#_Toc126258828)

[System Design 11](#_Toc126258829)

[Fetch data 12](#_Toc126258830)

[Compare relative position 12](#_Toc126258831)

[Frontend 12](#_Toc126258832)

[Algorithm (in built) 14](#_Toc126258833)

[Code 14](#_Toc126258834)

[Data Scraping 14](#_Toc126258835)

[Pseudo Data Scraping 15](#_Toc126258836)

[ML 19](#_Toc126258837)

[REFERENCE 19](#_Toc126258838)

# 

Case study 1

Trends/data Identification before Analyzing the Relative Positions of Players in Sports Betting and Online Casino Industry (take FanDuel Group, BetMGM, DraftKings and small players as examples)

# ABSTRACT

Sports Betting and Online Casino takes around 33.3% of online gambling in 2022 and will be experiencing a compound annual growth rate (CAGR) of approximately 12% during the forecast period 2022 to 2030 [2] as the sports industry keeps growing [3] [4] and the online gambling industry in the U.S. becomes more mature and more profitable [5] [6] [7].

This case study aims to identify trends and data as prerequisites to conduct future analysis on the relative positions of smaller players in Sports Betting and Online Casino Industry. This case study mainly focuses on identifying and extracting necessary data while analyzing smaller players' relative positions. To provide a more structured and detailed analysis, data were categorized into three buckets: company, market/industry, and customer.

Data gathered from different sources were usually messy, so they were cleaned, processed, and transformed into a structured dataset. Then, a machine learning model (for the sake of time, I used DL) was designed and trained to predict the future relative positions of companies. A brief analysis was conducted based on the data gathered by extracting websites and open-source databases, and the Market Position Index and Relative Position Index showing the relationship among smaller players in Sports Betting and Online Casino were proposed.

# DATA IDENTIFICATION

While analyzing the relative positions of smaller players, data were firstly summarized into 3 buckets based on the definition of company. Companies make money by selling specific goods or services (in one or more industries/markets) to consumers, which means all three aspects should be equally considered.

Diagram, schematic

Description automatically generated

Current market position roughly depends on the size of the industry and the market share of the company. Data gathered from customers' behaviors can be treated as indicators and if combined with moneywise data, they can be used to predict future position of the company. Then the relative positions of companies can then be concluded.

## Company Data



Companywise data reflect the current situation and the profitability of the company itself. They can normally be found in the annual report or other public records published by the company. Also, many indexes can show the company’s current situation on a scale of percentage and work as indicators of the company’s overall financial health. So, I categorized companywise data into direct data (revenue, assets, liability, costs, etc.) and indirect data (net profit margin, D/E ratio, operating margin).

Direct data are often straightforward, and can are easy to compare. These include financial metrics such as revenue, assets, liabilities, and costs. These numbers give a clear picture of the company's financial position and can be used to compare the company's performance to that of its competitors. Indirect data, on the other hand, are more complex and require more interpretation. These include net profit margin, D/E ratio, and operating margin. These metrics give an idea of the company's profitability and efficiency, and can help to identify trends and patterns in the company's financial performance over time.

An important parameter is the Growth Rate, which can be calculated by comparing the company's current performance to its past performance. This can give an indication of whether the company is growing or declining. Other important financial metrics to consider include Return on Equity (ROE), Return on Assets (ROA), and Return on Investment (ROI). These ratios measure how efficiently the company is using its assets and equity to generate profits.

Another important aspect of companywise data is the analysis of the company’s cash flow. This can be done by examining the company’s cash flow statement, which shows how cash is being generated and used by the company. Positive cash flow is important for a company’s survival and growth.

Overall, companywise data is a valuable tool for investors, analysts, and management teams to evaluate the performance of a company and make informed decisions. The combination of both direct and indirect data can give a comprehensive view of the company’s current situation and its potential for future growth.

## Industry Data



Industry Data were categorized into 2 buckets as well, which are data related to the company itself and the data of the competitors. One key metric for analyzing a company's market position is market share. Market share is simply calculated by taking the company's sales over a given period and dividing it by the total sales of the industry over the same period. This metric gives a clear picture of how much of the total market the company is capturing and can be used to compare the company's performance to that of its competitors. TAM, SAM, SOM are also data from the company’s perspective. Distribution data, product and pricing data should be taken into consideration since they can also be used to calculate and compare the company’s current situation with its competitors and reflect the market position of it.

In additional, data from competitors should also be gathered while calculating the relative positions of companies. Competitive intelligence data provides information on a company's competitors, including their market share, product offerings, pricing strategies, distribution channels, and other key metrics. This data can be used to identify a company's strengths and weaknesses in comparison to its competitors, as well as to identify potential threats and opportunities in the market.

Market data, on the other hand, provides information on the overall state of the market, such as market size, growth rates, and trends. This data can be used to identify industry trends, determine the size of the market, understand the competitive landscape, and forecast future growth opportunities. It can also be used to benchmark a company's performance against industry standards and make strategic business decisions.

Overall, company data, competitors data, and market data are all important in analyzing a company's position in the market. By understanding the company's market share, sales, pricing, and distribution data, as well as industry trends, competitor performance and market data, a company can make informed decisions and develop a strong market position.

Understanding these metrics can help us know the company’s new opportunities for growth and expansion in order to analyze the position of the company more precisely.

## Customer Data



One important aspect of customer data is demographics. This includes information such as age, gender, income, and location of customers. This information can help companies to identify their target market and tailor their marketing and sales strategies accordingly. For example, if a company's customer base is primarily made up of young adults, they may want to focus their marketing efforts on social media platforms that are popular with this age group.

Another important aspect of customer data is purchasing habits. This includes information such as the products or services customers purchase, how often they purchase, and how much they spend. This information can help companies to understand what products or services are most popular and identify opportunities for new product development or expansion. For example, if a company finds that customers are frequently purchasing a particular product, they may want to consider adding more options within that product category.

Customer feedback is also an important aspect of customer data. This includes information such as customer complaints, reviews, and ratings. This information can help companies to identify areas where they need to improve and make changes to their products or services. For example, if a company receives a large number of complaints about the durability of a particular product, they may want to consider improving the quality of that product.

In conclusion, customer data is crucial for companies to understand their customers and make informed decisions. By analyzing demographics, purchasing habits, and feedback, companies can identify their target market, understand what products or services are popular, and make improvements to their offerings. This, in turn, can help companies to increase customer satisfaction and loyalty, and ultimately drive sales and growth.

## Data Prioritization and Weights Design

The Market Position Index and Relative Position Index are proposed in order to show the relative positions of companies in the market.

Where represents the identical weights of each parameter (revenue, profit, margin, etc.), represents the parameters. The weights are found and optimized by ML algorithm that is shown later in this article.

When is larger than 1, which means the Market Position Index of the denominator is larger than the Market Position Index of the numerator. This means the denominator has a relative stronger competitive market position than the numerator. The larger the , the stronger the competitiveness the denominator has compared to the numerator, vise versa.

In order to get a more accurate , each data should be assigned with a different weight to show its importance and the relationship between it and the position of the company. Normally, the set of weights assigned to the data can be trained and optimized by ML if the training model is appropriate. A numerical Market Position calculating method proposed by Chen and Uysal [1] depicted a way to calculate correlation indexes for different objects (states in the article). In order to get a brief understanding of how the data should be weighted, Forbes 500 list was firstly analyzed and applied. Since the companies are ranked based on the total revenue. The largest weight then can be assigned to revenue.

|  |
| --- |
| Rise in Adoption of Internet and Growth in Online Betting |
| Increase in Betting in Sports |
| Growth of E-sports Industry |
| Rise in Fixed-odds Betting |

# ANALYSIS

Sports betting is becoming more common around the world as the number of sports enthusiasts grows. The gambling industry is one of the fastest-growing sectors of the global economy, and many countries are contemplating legalizing sports betting which is expected to drive market growth in the forecast period [7]. Major online casino and sports betting operators are: FanDuel, BetMGM, DraftKings Sport Book, Caesars Sport Book, Points Bet, and BetRivers.

As discussed above, Fortune 500 only used revenue to rank all the companies [8]. In order to have a more straightforward and clear way to analyze the relative position of the companies, revenues are firstly compared. According to the report of Legal Sports Betting, the total revenue of sports betting industry is $ 14,288,454,382 by Jan 20, 2023 [9], whereas the total revenue of FanDuel is 1.9 billion [10] in 2021 (total was 4.33 billion USD in 2021 [5]), which takes more than 13% of the total revenue in the US.

I first went through all the annual reports of these major competitors in sports betting and online casino market and wrote a piece of code to scrape the keywords from the PDF files. The market share regarding to the revenue is shown below:

Chart, pie chart

Description automatically generated

A strong market position of FanDuel, owned by Flutter Entertainment can be observed by this pie chart. The brand is estimated to hold 46.6% of the Gross Gaming Revenue (GGR) in states where online sports betting is available. DraftKings comes in second with 20.2% of the GGR, while BetMGM and Caesars hold 14.1% and 7.5% respectively.

FanDuel is a strong market leader with a market capitalization of $11.5 billion as of February 1, 2023. The company has established itself as a top player through strategic partnerships and marketing campaigns, and has a strong reputation for offering a wide range of betting options and a user-friendly experience. DraftKings was the leader in the market in 2020 according to the annual reports and it’s revenue was $704 million in 2020. It has a market capitalization of $14.5 billion as of February 1, 2023. Its strong brand recognition and innovative technology have helped it establish a strong customer base and position itself as a key player in the market.

Additionally, BetRivers, Barstool Sportsbook, and PointsBet also contribute with 2.9%, 2.8% and 2.6% of the GGR, respectively. Out of the 60 operational online sports betting operators, only seven brands have a market share greater than 1%, with the remaining 53 firms competing for the remaining market share.

These seven brands account for a combined 96.7% of the GGR, leaving just 3% for the other 53 companies.

Chart, bar chart

Description automatically generated

In terms of growth, the sports betting industry has been growing rapidly in recent years. FanDuel's revenue has increased by 60% since 2018, while DraftKings' revenue has increased by 85% over the same period. betMGM has also reported significant growth since entering the sports betting market, has also seen rapid growth. In 2020, betMGM reported a 100% year-over-year increase in online sports betting revenue and a 400% increase in mobile sports betting handle. These numbers indicate the significant growth potential for companies in the sports betting industry.

Chart, bar chart, waterfall chart

Description automatically generated

Besides direct financial performance of companies, customers experience is also an important indicator to be considered. Website visits can indicate a level of interest in a company, it is one factor among many that can impact a company's market position. Additionally, monthly changes, is useful for tracking long-term performance, is a great help while analyzing the comprehensive picture of a company's overall market position apart from companies’ financial performance.

Website Visit of top 200 companies in online sports betting and online casino (Free data accessed from data40.com) was scraped as below:

Table

Description automatically generated

After data cleaning and deleting unnecessary data, the monthly change of website was then analyzed:

Chart, waterfall chart

Description automatically generated

Then, the correlation between the global rank and the website volume was calculated.

# INTEGRATION WITH CS

## System Design

Diagram

Description automatically generated

I also designed an APP specifically for relative positions analysis and the system design is shown above. The whole APP consists of 2 main parts as any other app, which are frontend and backend. For the sake of simplicity, I purposely put DB inside the scope of Backend.

This APP-beta can achieve 2 simple functionalities. The user story goes like below:

Diagram

Description automatically generated

### Fetch data

1. User opens the APP
2. User types out the name of the company
3. If this company’s data are in the DB, then Lambda function 1 will be invoked to fetch the data
4. If the data are not in the DB, Lambda function 2 will be firstly invoked to scrape data from the internet and then Lambda function 1 will be invoked to return the data

### Compare relative position

1. User opens the APP
2. User types out the names of 2 companies
3. If the companies’ data are in the DB, then Lambda function 1 will be invoked to fetch the data
4. If the data are not in the DB, Lambda function 2 will be firstly invoked to scrape data from the internet and then Lambda function 1 will be invoked to return the data

## Frontend

Graphical user interface, text, application, chat or text message

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, application, Teams

Description automatically generated

## Algorithm (in built)

Parameters:

* Company: data scraping from open-source website (public companies: yahoo finance, google finance)/annual reports
* Market:
* Customer: third-party (similarweb, CrunchBase)

Y:

* Initial index/ranking

## Code

### Data Scraping

import PyPDF2 as pdf2  
import re  
  
# Open the PDF file  
pdf\_file = open('WYNNRESORTS.pdf', 'rb')  
pdf\_reader = pdf2.PdfReader(pdf\_file)  
  
# Extract text from each page of the PDF  
text = ""  
for page\_num in range(pdf\_reader.numPages):  
 page = pdf\_reader.getPage(page\_num)  
 text += page.extractText()  
  
# Search for the revenue information in the text  
financial\_info = ''  
  
# revenue\_all = []  
lines = text.split('\n')  
for line in lines:  
 if r'revenue' in line.lower() and bool(re.search(r'\d', line)):  
 financial\_info += line  
 # if 'margin' in line.lower():  
   
 # margin += [int(s) for s in line.split() if s.isdigit()]  
  
# Write the revenue information to a file  
with open('financial\_info.txt', 'w') as f:  
 f.write(financial\_info)  
  
# Close the PDF file  
pdf\_file.close()

Instructions:

* Put the PDF file waiting to be scraped under the same folder of this file.
* Go to Line 5 and change the name in the open function to the PDF’s name
* Go to Line 20, change the ‘revenue’ to any keyword you want to scrape
* Run the code and you will see a fetched file containing all the info needed

### Pseudo Data Scraping

import pandas as pd  
from bs4 import BeautifulSoup  
import urllib.request as ur  
from datetime import date  
import os  
# import requests  
  
  
  
# Enter a stock symbol  
index= 'FLTR.L'  
# URL link   
url\_is = 'https://finance.yahoo.com/quote/' + index + '/financials?p=' + index  
url\_bs = 'https://finance.yahoo.com/quote/' + index +'/balance-sheet?p=' + index  
url\_cf = 'https://finance.yahoo.com/quote/' + index + '/cash-flow?p='+ index  
  
req\_is = ur.Request(url\_is, headers={'User-Agent': 'Mozilla/5.0'})  
read\_data\_is = ur.urlopen(req\_is).read()  
soup\_is= BeautifulSoup(read\_data\_is,'html.parser')  
  
req\_bs = ur.Request(url\_bs, headers={'User-Agent': 'Mozilla/5.0'})  
read\_data\_bs = ur.urlopen(req\_bs).read()  
soup\_bs= BeautifulSoup(read\_data\_bs,'html.parser')  
  
req\_cf = ur.Request(url\_cf, headers={'User-Agent': 'Mozilla/5.0'})  
read\_data\_cf = ur.urlopen(req\_cf).read()  
soup\_cf= BeautifulSoup(read\_data\_cf,'html.parser')  
  
# create directory  
try:  
 os.makedirs('pesudo\_data/' + index)  
except FileExistsError:  
 # directory already exists  
 pass  
  
  
# Income Statement scrape  
  
ls= [] # Create empty list  
# print(soup\_is)  
for l in soup\_is.findAll('div') or soup\_is.findAll("div", {"class" : "D(ib) Va(m) Ell Mt(-3px) W(215px)--mv2 W(200px) undefined"}):  
  
 if l in soup\_is.findAll("div", {"class" : "D(ib) Va(m) Ell Mt(-3px) W(215px)--mv2 W(200px) undefined"}):   
 ls.append(l.span.text)  
   
 else:  
 ls.append(l.string) # add each element one by one to the list  
new\_ls = list(filter(None,ls))  
new\_ls = new\_ls[new\_ls.index("Quarterly")-1:]  
del new\_ls[1:3]  
is\_data = list(zip(\*[iter(new\_ls)]\*new\_ls.index("Total Revenue")))  
  
Income\_st = pd.DataFrame(is\_data[0:])  
Income\_st.columns = Income\_st.iloc[0]   
Income\_st = Income\_st.iloc[1:,].T   
  
Income\_st.columns = Income\_st.iloc[0]   
Income\_st.drop(Income\_st.index[0],inplace=True)   
Income\_st.index.name = ""   
Income\_st.rename(index={"ttm": date.today().strftime("%m/%d/%y")},inplace=True)   
Income\_st = Income\_st[Income\_st.columns[:-5]]  
Income\_st.to\_csv('pesudo\_data/' + index +'/Income\_st.csv')   
  
  
# Balance Sheet scrape  
  
ls= [] # Create empty list  
# print(soup\_bs)  
for l in soup\_bs.findAll('div') or soup\_bs.findAll("div", {"class" : "D(ib) Va(m) Ell Mt(-3px) W(215px)--mv2 W(200px) undefined"}):  
  
 if l in soup\_bs.findAll("div", {"class" : "D(ib) Va(m) Ell Mt(-3px) W(215px)--mv2 W(200px) undefined"}):   
 ls.append(l.span.text)  
   
 else:  
 ls.append(l.string) # add each element one by one to the list  
new\_ls = list(filter(None,ls))  
  
new\_ls = new\_ls[new\_ls.index("Quarterly")-1:]  
  
del new\_ls[1:3]  
  
bs\_data = list(zip(\*[iter(new\_ls)]\*new\_ls.index("Total Assets")))  
  
balance\_St = pd.DataFrame(bs\_data[0:])  
  
balance\_St.columns = balance\_St.iloc[0]   
balance\_St = balance\_St.iloc[1:,].T   
  
balance\_St.columns = balance\_St.iloc[0]   
balance\_St.drop(balance\_St.index[0],inplace=True)   
balance\_St.index.name = ""   
  
balance\_St.to\_csv('pesudo\_data/' + index +'/balance\_St.csv')   
  
  
# Cash Flow scrape  
  
ls= [] # Create empty list  
# print(soup\_bs)  
for l in soup\_cf.findAll('div') or soup\_cf.findAll("div", {"class" : "D(ib) Va(m) Ell Mt(-3px) W(215px)--mv2 W(200px) undefined"}):  
  
 if l in soup\_cf.findAll("div", {"class" : "D(ib) Va(m) Ell Mt(-3px) W(215px)--mv2 W(200px) undefined"}):   
 ls.append(l.span.text)  
   
 else:  
 ls.append(l.string) # add each element one by one to the list  
new\_ls = list(filter(None,ls))  
new\_ls = new\_ls[new\_ls.index("Quarterly")-1:]  
  
del new\_ls[1:3]  
  
cf\_data = list(zip(\*[iter(new\_ls)]\*new\_ls.index("Operating Cash Flow")))  
  
cash\_Fl = pd.DataFrame(cf\_data[0:])  
  
cash\_Fl.columns = cash\_Fl.iloc[0]   
cash\_Fl = cash\_Fl.iloc[1:,].T   
  
cash\_Fl.columns = cash\_Fl.iloc[0]   
cash\_Fl.drop(cash\_Fl.index[0],inplace=True)   
cash\_Fl.index.name = ""   
cash\_Fl.rename(index={"ttm": date.today().strftime("%m/%d/%y")},inplace=True)   
cash\_Fl.to\_csv('pesudo\_data/' + index +'/cash\_Fl.csv')

### ML

SVM and KNN will be used.

# REFERENCE

|  |  |
| --- | --- |
| [1] | J. J., "Global Online Gambling Market & Betting Market Size Worth $ 145.6 Billion by 2030 at a 12% CAGR: Custom Market Insights.," www.globenewswire.com, 12 09 2022. [Online]. Available: https://www.globenewswire.com/en/news-release/2022/09/12/2513697/0/en/Global-Online-Gambling-Market-Betting-Market-Size-Worth-145-6-Billion-by-2030-at-a-12-CAGR-Custom-Market-Insights.html. [Accessed 26 Jan 2023]. |
| [2] | "Sports Outlook 2022 From media rights to NFTs, PwC unpacks the hottest industry trends.," [Online]. Available: https://www.pwc.com/us/en/industries/tmt/library/assets/pwc-tmt-sports-outlook.pdf. [Accessed 26 Jan 2023]. |
| [3] | P. Giorgio, "2023 sports industry outlook," Deloitte United States, [Online]. Available: https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/sports-business-trends-disruption.html. [Accessed 26 Jan 2023]. |
| [4] | R. Vanderford, "As online gambling grows, so does the financial crime risk," Wall Street Journal, 29 August 2022. [Online]. Available: https://www.wsj.com/articles/as-online-gambling-grows-so-does-the-financial-crime-risk-11661765402. |
| [5] | A. G. A. (AGA), "Interactive U.S. map: Sports betting," [Online]. Available: https://www.americangaming.org/research/state-gaming-map/. [Accessed 26 Jan 2023]. |
| [6] | "Internet Sports Betting Market Insights, 2021-2031," Transparency Market Research, 2021. |
| [7] | M. U. Joseph S. Chen, "MARKET POSITIONING ANALYSIS A Hybrid Approach," *Annals of Tourism Research,* vol. 29, no. 4, pp. 987-1003, 2002. |
| [8] | "Fortune 500 Companies 2022," Fortune, [Online]. Available: https://fortune.com/ranking/fortune500/2022/search/. |