BIG DATA ANALYTICS FOR BUSINESS - Part one (EBUS633)

Student Name: Soujanya Krishnamurthy
Student ID: 201585697
Student email ID: s.krishnamurthy@liverpool.ac.uk

Motivation

Big data's availability disrupts conventional decision-making practices, allowing businesses to make strategic decisions leveraging data analytics techniques (Monino, 2016). While the word ‘business acumen’ is still considered a great skill, it has to be supported by data analysis to make valuable decisions. Depending upon the type of data available and the organisation’s requirement, different analytical tools and techniques can be applied to make strategic decisions (Chong and Shi, 2015). The organisation in the case study is a sports retailer well established in the Indian market and wants to improve its brand image to increase customer loyalty. The company invests heavily in print and social media advertising, it feels the need to step up the game to gain an advantage over its competitors. Hence it is suggested that the company invest in celebrity endorsement as there have many instances where celebrity advertising has proven successful among many sports retail giants such as Nike, Adidas, Uniqlo etc. However, the investment in celebrity branding is very high and is only worth it if it provides the investment return. The best way to solve this dilemma is through data analytics by looking at case studies where brand endorsement has helped organisations make profits. One such tool would be regression analysis between celebrity endorsement deals and their impact on sales and profits. While the regression analysis is used as a reference at this point, a similar analysis can be made using the company’s sales data before and after celebrity endorsement to determine the effectiveness. The company also wants to look at the best celebrity fit, which can be analysed through data visualisation of the top trending celebrities in the sports retailing sector.

Literature review

Regression Analysis

The Internet serves as a massive data storage facility, with massive volumes of data being generated every second. 2.7 Zettabytes of data exist in today's digital environment, according to the IBM Big Data Flood Infographic. Furthermore, a Facebook study predicted 100 Terabytes of data updated every day, with an estimated 35 Zettabytes of data generated annually by 2020, resulting in a lot of activity on social networks (Chen et al., 2014). This data is useful only when leveraged by using predictive analytics tools such

as regression analysis to answer the question ‘What should we do in future?’ (Thi Thi Zin, Jerry Chun-Wei Lin, Online Service, 2019)

Regression analysis is a technique that takes raw data from different sources, analyse the main features and approximate the relationship that exists among these features through mathematical function (Draper and Smith, 2011). The simplest form of the regression equation is where one dependent variable and one independent variable is considered through the below mathematical form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

While the above equation forms the basis of regression, complex relationships can also be examined through a regression where there is more than one independent variable through multiple regression analysis. In today’s world where the data obtained is very large in size and subject to a lot of changes, multiple regression helps in analysing the correlation between these variables and the significance of relationships so that the organisation can focus on the variables which have a significant effect. This helps organisations make faster decisions and save costs. While Linear regression examines discrete variables, logistic regression is used where the data is continuous in nature (Allen, 2004). Regression analysis is extensively used across many industries such as manufacturing, retail, service, transport etc. Some of the common examples of regression analysis are (Allen, 2004):

Delivery time for a specific number of cases of soft drinks to be delivered to plan the routing and scheduling.

A football team can use regression analysis to examine the performance of the team through the number of yards gained by rushing the opponents and its effect on the number of games won by the team.

Data Visualisation

In simple terms, data visualisation is a picture or graphic display of data. However, data visualisation also helps understand data relations and helps organisations in narrowing the focus of their analysis. It helps identify hidden patterns that affect consumer behaviour and sales and make future decisions (Zhwan M. K. & Subhi R.M Zeebaree, 2021). The below

diagram shows the various steps involved in the data visualisation process (Zhwan M. K. & Subhi R.M Zeebaree, 2021).

Figure 1

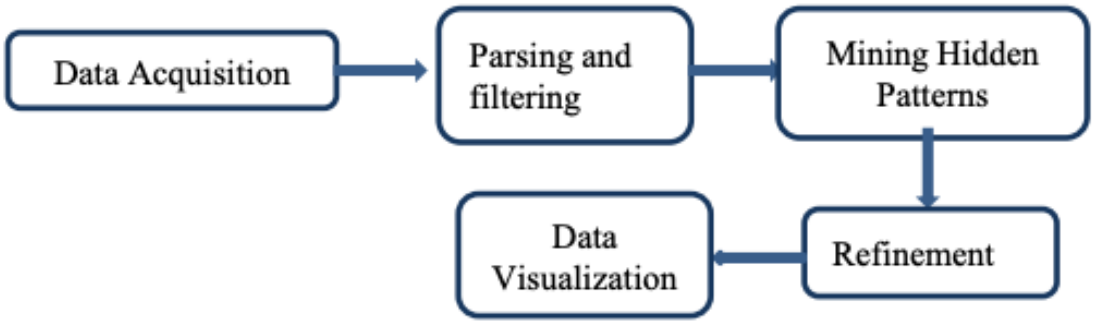


Figure1: Big Data Visualization Process

The data, which is collected from various sources is first collated and converted in a structured format. The next step is to remove unimportant and redundant data. Once the data is ready, it can be represented into useful diagrams and charts based on the desirable outcomes. These charts and diagrams help the organisation identify useful and hidden patterns leading to effective decision making. Based on the patterns identified, the organisation can further analyse data using predictive analytical tools.

Big data is characterised by 4 V's Variety, Volume, Velocity and Veracity, Data visualisation tools such as Tableau, Power BI etc can visualise these large datasets occurring constantly through dynamic data visualisation. Another issue addressed by visualisation is scalability. A large number of data reduction techniques such as sampling, aggregating, data ranges are applied to make the data more readable and accurate (Sherif Sakr, Bikakis, 2019).

Case examples

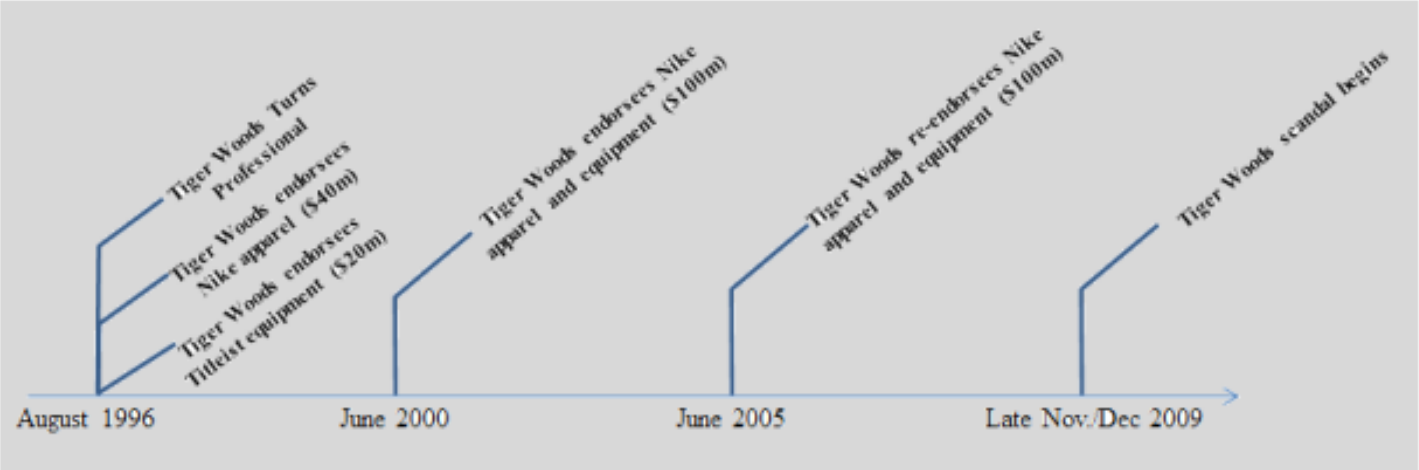
Regression analysis

The golf industry in the USA is larger than the film and video industries, with a combined value of $76 billion. In 2005, the sector was worth $ 7.8 billion, with 7 primary sections ranging from facility operations to real estate, golfer equipment/supplies, and golf

endorsements combined. The 3 main types of equipment of the game are golf balls, clubs and bags. There are around 1000 gold balls types listed in the US golf association based on various characteristics such as the number of layers, skin and core part. Like many other industries, the golf industry has celebrities endorsing the golf balls.

The selection of a celebrity is based on their popularity based on the ranking and their trustworthiness (Knoll and Matthes, 2017). The study focuses on the endorsement of Nike golf balls by world golf champion Tiger Woods. Figure 2 (Chung, Derdenger and Srinivasan, 2013) shows the timeline of Tiger Woods's contract with Nike.

Figure 2: Tiger Woods endorsement timeline



In 1996, Woods started his association with Nike by endorsing their apparel in the US while he was endorsing Titleist golf equipment. In 2000, he extended his contract to endorse Nike's golf equipment which lasted for many years until his negative publicity in 2009.

Competitors

Before the data is analysed, it is important to understand the market position of Nike in the golf industry. The study takes into account 550 on-course shops (Lawn) and 250 off course shops(retail) shops in the US. These shops together sell 669 unique products represented by 26 brands. Some of the prominent brand selling golf balls are Titleist, Top-Flite, Pinnacle, Bridgestone, Nike etc. Below figure 3 (Chung, Derdenger and Srinivasan, 2013) shows the market share of the top brands in 2000 and 2010 for off-course golf balls. We can see that the market share for Nike increased from 1.5% to 10% in 2010. This

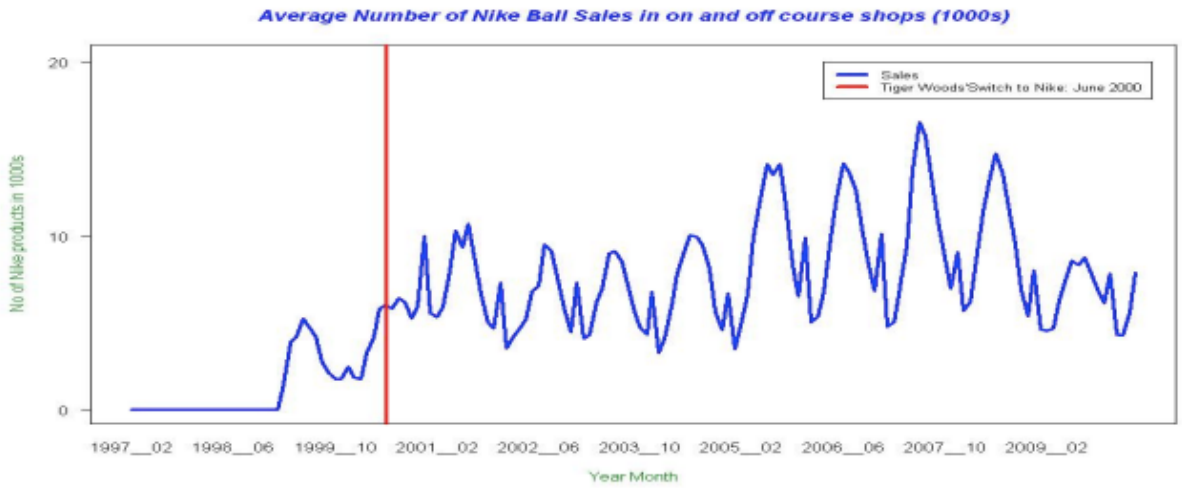
corroborates the claim that there was a positive impact on Nike during the endorsement period with Tiger woods.

Figure 3: Market share of brands for Off course balls

2000				2010			
Brand	Share	Avg Price [Max Min]	#Products	Brand	Share	Avg Price [Max Min]	#Products
Titleist	23.51	$28.17 [$34.46 $22.32]	9	Titleist	23.08	$23.64 [$36.34 $15.34]	7
Top-Flite	22.74	$13.50 [$18.01 $10.23]	9	Callaway	11.75	$20.87 [$34.27 $8.37]	12
Pinnacle	11.63	$12.08 [$14.91 $9.75]	7	Bridgestone	10.57	$26.90 [$34.49 $17.67]	8
Precept	7.52	$22.82 [$31.44 $18.08]	9	Nike	10.00	$15.76 [$25.48 $8.60]	10
Wilson	7.52	$19.67[$26.94 $9.22]	11	Taylor Made	7.50	$21.32 [$36.29 $12.51]	6
Maxfli	6.23	$25.72 [$34.85 $14.28]	7	Top-Flite	7.14	$9.70 [$14.24 $6.20]	6
Strata	4.91	$24.79 [$25.64 $24.26]	3	Srixon	6.72	$21.02 [$32.35 $10.61]	7
Callaway	3.67	$39.68 [$40.30 $39.07]	2	Pinnacle	5.99	$9.60 [$12.22 $7.94]	6
Taylor Made	2.36	$31.36 [$33.97 $26.98]	3	Precept	5.79	$13.99 [$18.21 $8.02]	4
Spalding	2.35	$8.83[$26.98 $8.83]	1	Other	4.01	$8.61 [$8.61 $8.61]	1
Other	2.08	$13.24[$13.24 $13.24]	1	Wilson	3.00	$8.47 [$8.47 $8.47]	1
Dunlop	1.83	$11.59[$14.38 $10.04]	4	Noodle	2.26	$12.08 [$12.23 $11.93]	2
Nike	1.59	$26.48 [$36.39 $18.14]	4	Volvik	1.77	$15.79 [$15.79 $15.79]	1
Cobra	0.95	$21.22 [$22.85 $18.93]	4				
Srixon	0.51	$24.17 [$24.17 $24.17]	1			Note: Price is adjusted to 1997	
HHI	1417.03			1137.89			
Four Firm CR	65.40			55.40			

Furthermore, Nike also saw an increase in the sales for its on course and off course balls from the time it started its endorsement contract with Tiger Woods as shown in figure 4 (Chung, Derdenger and Srinivasan, 2013). Meanwhile, it should be noted that the sales are spiked in nature due to seasonality trends as Golf is a seasonal sport.

Figure 4: Nike Ball sales data from 1997 to 2010



The study shows the relationship between Nike's golf ball sales and Woods's contract with Nike. For the purposes of analysis Nike's sales data from 1999 to 2001 has been examined when Tiger Woods endorsed Nike's golf equipment. This was measured against the advertisements shown by Nike with Tiger Woods endorsing the golf balls. This is done by taking the regression log of monthly sales with endorsement variable ranking on a reduced form analysis using the OLS method on R software. Below figure 5 (Chung, Derdenger and Srinivasan, 2013) shows the results of the analysis.

Figure 5: Regression analysis of Nike golf ball sales

Regression of Sales on Parametric Endorsement Variable $\left(\frac{1}{rank}\right)^{\alpha}$						
	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0.4$	$\alpha = 0.3$	$\alpha = 0.2$	$\alpha = 0.1$
(Intercept)	23.406 (21.317)	-3.550 (31.895)	-17.209 (37.558)	-40.059 (47.233)	-85.887 (67.004)	-223.627 (127.294)
1/rank of Tiger Woods	41.033 (17.682)**	67.997 (29.590)**	81.659 (35.622)**	104.511 (45.710)**	150.340 (65.941)**	288.083 (126.740)**
Unplanned Exposure	0.893 (3.642)	.8270 (3.644)	0.831 (3.644)	0.835 (3.645)	0.839 (3.645)	0.843 (3.646)
AdjustedR^2	0.9030	0.9029	0.9028	0.9028	0.9028	0.9028

Note: Signif. codes: 0 < ** < 0.05

The figure suggests that the ranking of Tiger Woods from No.2 to No.1 leads to an increase of around 20,000 additional golf ball sales.

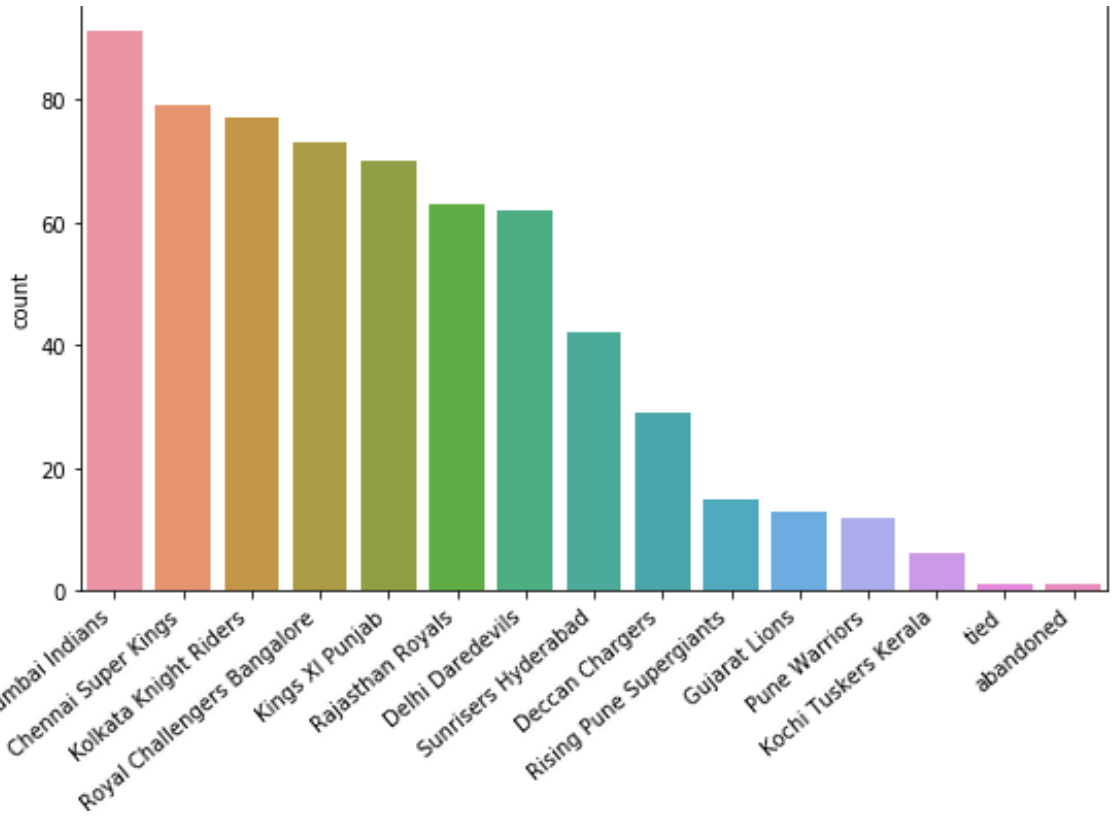
Data Visualisation

Based on the case study of regression analysis, it has been decided by the company that it should consider investing in celebrity endorsement. As discussed earlier, data visualisation of stats of sports personalities can be applied to pick the best celebrity fit.

Cricket has always been the most popular game in India watched by millions of Indians. Hence the company has decided that it should start its celebrity scout with cricket. Recently, a game that has gained huge popularity is the Indian Premier League and hence the company has decided to look at the stats of the IPL players and teams to decide some of the top eligible celebrities for endorsement. For the study, different datasets have been used based on statistics on teams, matches, players etc from 2010 to 2020 by conducting visualisation on Tableau.

While the loyalty towards the game is mainly regional. Some teams with maximum wins are watched and liked by all Indians. Figure 6 shows the number of wins secured by the teams in the last 10 years. It can be inferred that Mumbai Indians have the maximum number of wins followed by Chennai Super Kings and Kolkata Knight readers.

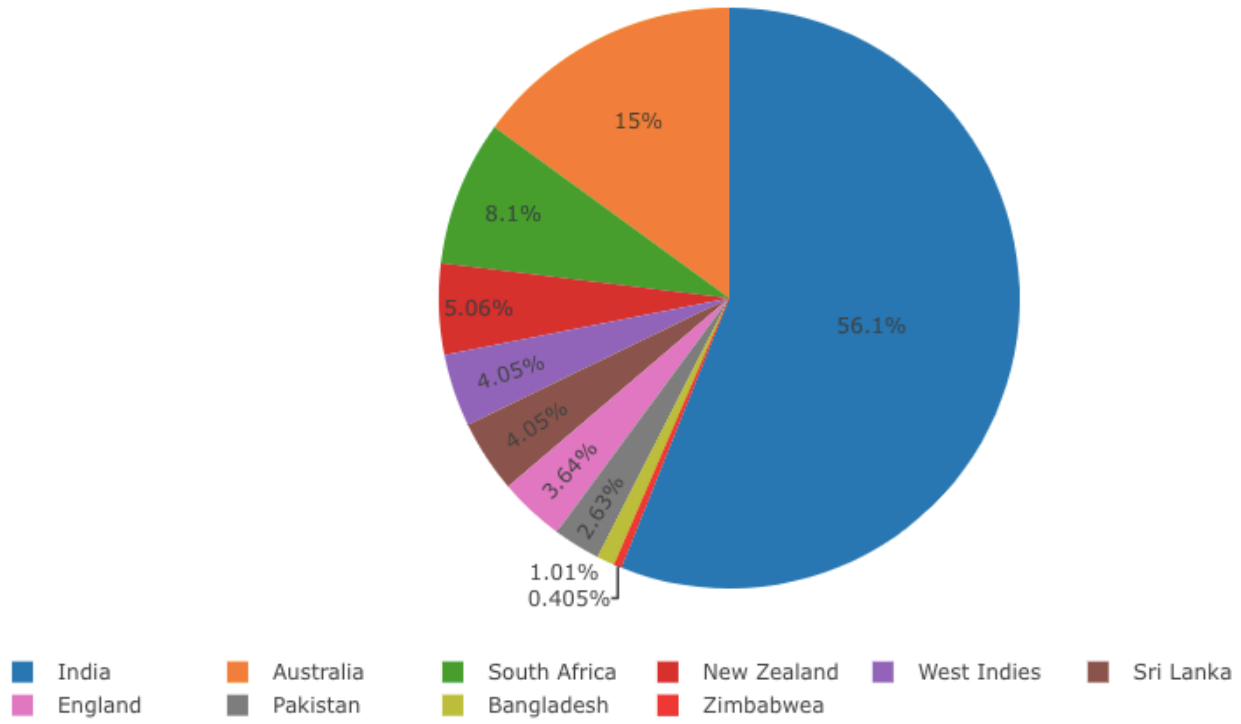
Figure 6: Number of wins in IPL



This gives the company an idea of what teams should the company consider while selecting celebrities.

While the majority of players playing in IPL are from India, there are a number of international players who have gained popularity in India and have signed many endorsement contracts due to their popularity. Figure 7 shows the countries from where players are hired.

Figure 7: Countries from where players are hired



It can be observed that India and Australia followed by South Africa have the maximum number of players playing for IPL. Hence the company should focus on players from these countries.

Now that the company has looked at statistics on teams and the countries, it is also important to consider players' statistics such as runs, teams played etc.

Figure 8: IPL Players dashboard

The players' dashboard reveals that some of the players who have scored maximum runs have also scored maximum centuries in IPL. Players from Delhi Daredevils and Chennai super kings have the maximum century scoring players compared to other teams with Virat Kohli as an exception. The charts also reveal that there is an equal distribution of both Indian and International players who have obtained maximum runs and scored centuries.

Barriers

Regression Analysis

While the above study shows an effective analysis between celebrity endorsement and impact on sales. There could be many deviations when applied to the company and case study in question. Firstly, Since Tiger Woods is a professional golfer and the study is on

golf balls, there is a direct relationship between these two variables. The sports retailer which is planning to sell considers celebrity endorsement sell many items and cannot endorse particular apparel or equipment which could lead to an extrapolated analysis. Secondly, the study only considers the sales numbers and not other aspects of consumer behaviour that may impact the sales such as price, general brand loyalty etc.

Finally, merely a regression analysis or data visualisation may not be sufficient to assess the effects of celebrity endorsement. An important factor in decision making is the budgeting of the celebrity contract and its effect on pricing and market share (Farrell et al., 2000).

Data Visualisation

While the data on IPL provides some insights on players and team performances. Today's consumer behaviour is also heavily influenced by the social media image the celebrity possesses. Also, the data does not take into account the age and gender factor as the popularity of IPL is more among the youth than the older people. The fanbase of men can be different and based on statistics versus those of women who might favour a player for his personality or social media presence (Knoll and Matthes, 2017). All these factors need to be considered while selecting celebrities.

Recommendations & Roadmap

The company should conduct a sensitivity analysis on the effectiveness of the regression model by using alternative model designs and under other regression models such as KNN.

Based on the above case study, It is suggested that the company make the decision of using a celebrity to endorse the company but sufficient analysis needs to be done on the type of celebrity, his/her attractiveness/popularity and sales forecast post signing the contract. The two possible approaches to celebrity endorsement could be either to sign a particular player or to sign a top-performing team with top players to endorse the brand.

Channel of advertising

Traditionally, the main channel of celebrity marketing has been the TV and print media. However, recently a lot of companies such as Levis, Adidas have started posting celebrity

ads on Instagram. Hence the company can perform separate regression analysis and data visualisation on the best channel to advertise the product as this will also impact the advertising costs since TV is considered more expensive than other channels (Knoll and Matthes, 2017). Additionally, many celebrities have their own social media accounts and endorse products through those accounts. Hence the company can initially start with the celebrity influencing the product on his social media page and eventually invest in advertisements and photoshoots.

Seasonality

Elaborating on the previous point, IPL is a seasonal game where the preparation for the match starts months before through player auctions, contract deals etc. Hence Another effective strategy for the firm can also be to sign TV advertisement deals with the celebrity during the IPL season and have print/influencer deals for the rest of the year.

Figure 9: shows the roadmap the company should follow to plan an effective strategy.

Figure 9: Roadmap

Determine the overall celebrity budget of the company
Select top 3 players/teams based on data visualisation dashboard
Select the best fit based on budget, length of the contract, and perceived ROI
Determine an effective channel of strategy and the type of advertisement to be deployed using the testing the waters strategy
Monitor changes in sales based on celebrity endorsement and conduct predictive analysis based on one-year sales data.

Conclusion

The primary motivation of the case study was to use a more data-driven and mathematical approach to formulate an effective branding strategy rather than using intuition. In this aspect, the example of Nike's branding strategy for their golf balls has shed light on how effective celebrity endorsements can be. However, the company should also be cautious and take a holistic approach while formulating its own celebrity endorsement strategy. The selection of celebrities through data visualisation can also be coupled with other traditional tools such as customer polls, social media analysis etc. Overall, Data analysis tools help leverage the massive data collected by organisations and make more meaningful decisions.

References

Allen, M.P. (2004). Understanding regression analysis. New York: Springer Science+Business Media.

Chen, M., Mao, S., Zhang, Y. and Leung, V.C.M. (2014). Big Data. Cham Springer International Publishing.

Chong, D. and Shi, H. (2015). Big data analytics: a literature review. Journal of Management Analytics, 2(3), pp.175–201.

Chung, K.Y., Derdenger, T.P. and Srinivasan, K. (2013). Economic Value of Celebrity Endorsements: Tiger Woods' Impact on Sales of Nike Golf Balls. Marketing Science, 32(2), pp.271–293.

Draper, N.R. and Smith, H. (2011). Applied regression analysis. New Delhi: Wiley-India.

Farrell, K.A., Karels, G.V., Montfort, K.W. and McClatchey, C.A. (2000). Celebrity performance and endorsement value: the case of Tiger Woods. Managerial Finance, 26(7), pp.1–15.

Iain Pardoe (2012). Applied regression modelling. Hoboken, N.J.: Wiley.
Knoll, J. and Matthes, J. (2017). The Effectiveness of Celebrity endorsements: a meta-analysis. Journal of the Academy of Marketing Science, 45(1), pp.55–75.

Knoll, J. and Matthes, J. (2017). The Effectiveness of Celebrity endorsements: a meta-analysis. Journal of the Academy of Marketing Science, 45(1), pp.55–75.

Monino, J.-L. (2016). Data Value, Big Data Analytics, and Decision-Making. Journal of the Knowledge Economy.

Sherif Sakr, Bikakis, N. and Springerlink (Online Service (2019). Encyclopedia of Big Data Technologies. Cham: Springer International Publishing.

Thi Thi Zin, Jerry Chun-Wei Lin and Springerlink (Online Service (2019). Big Data Analysis and Deep Learning Applications : Proceedings of the First International Conference on Big Data Analysis and Deep Learning. Singapore: Springer Singapore.

Zhwan M. Khalid & Subhi R.M Zeebaree, 2021. "Big Data Analysis for Data Visualization: A Review," International Journal of Science and Business, IJSAB International, vol. 5(2), pages 64-75.