How accurately can machine learning predict future sales for a small coffee shop in Edinburgh?

Gleb Sokolovskyi
Candidate Number: 8279
P304

3,193 words

Abstract:

This project looks at the possibilities of machines to aid businesses in predicting sales. I will be looking into whether a computer program can identify the causes of changes in sales in the past, and learn to predict sales for the future. The project takes the form of a computer program that is able to read and analyse sales data from a coffee shop, which belonged to my parents, and analyse the underlying factors that could explain sales. For example, whereas a human would expect that sales of water would be higher on a warmer day, machines must be taught this. My conclusion is that, as outlined in the testing stage, my program's predictions were reasonably accurate for the data that I was using. However, further data, which is unfortunately unavailable, would be required to fully test the robustness of this conclusion.

Content

(1) Introduction

My interest in this project was drawn from the increasing popularity of technology and its application to solving everyday problems. Specifically, machine learning is an increasingly popular topic, with many businesses and even governments exploring its possibilities. In this particular field, I was initially inspired by a book called "The Essence of Artificial Intelligence" by Alison Cawsey. The book discusses how human intelligence can be reduced to the complex manipulation of symbols, and that it does not matter which medium is used to manipulate these symbols; it must not necessarily be a biological brain. Also discussed is how humans unknowingly perform many complex and sophisticated knowledge and reasoning processes, which are very hard to replicate on a machine.

When my parents were running a coffee shop, "Le Petit Repas", they were manually calculating roughly how much of each product to order for delivery the following week. If done so randomly or simply based on the previous week's sales, they might have either overordered, resulting in costly excess waste, or underordered, leaving customers dissatisfied. Given that 60% of new UK businesses fail within 3 years[1], having a grasp of the finances is crucial for small business owners, something which can be supported by sound forecasting.

For large corporations, however, this doesn't represent as much of a problem, as they put a lot of resources into management, of which stock management is a big area. An example is "Coca Cola", which uses artificial intelligence in social media to predict where and when sales of their drinks will be sold most.[2] Although such giants in the market are special cases, making up only 0.1% of private sector businesses, I believe their use of AI and machine learning for commercial gain represents a powerful opportunity for the remaining 99.9%.[3]

To study the usefulness for small businesses of harnessing the idea of sales prediction, I wrote a program that takes in and processes real-life data from the coffee shop. It does so by: finding the relationship between different factors and sales over a given period of time; learning from this information to form conclusions; and outputting an equation for each product, which can be used to predict sales.

A problem that might have occurred in the process of this project was the programming involved, with which I had no prior experience. This aspect task requires a lot of data analysis programming, which in turn requires know-how. However, I overcame this challenge by exploring the resources available online, a course called "Kaggle" that teaches AI programming in Python, and the above mentioned book "The Essence of AI" by Alison Cawsey.

My original hypothesis was that sales data could be explained well using my model, breaking sales down into the fundamental factors driving it. Despite some inaccuracy in my model, as will be outlined, my hypothesis was largely confirmed by my results. In order to test the robustness of this in a further study, I would ideally use new data from the same setting and see whether predicted sales matched up with these.

My study is split into six stages. Section 2 outlines the 'Processing data' stage, involving data collection, inputting and manipulation in Excel and Python. Section 3 postulates the various factors that could affect sales. Section 4 explores the code that I wrote for my program. Section 5 introduces and implements the concept of Multiple Linear Regression in my program. Finally, Section 6 discusses the testing of my results, for example in terms of confidence levels and anomalies, and Section 7 explores the limitations of the project.

---

[1] "Start-ups across the UK are going bust - they need more careful management for our economy to boom", *The Telegraph,* 24th January 2019, https://www.telegraph.co.uk/politics/2019/01/24/start-ups-across-uk-going-bust-need-careful-management-economy/

[2] "How Coca-Cola is using AI to stay at the top of the soft drinks market", *Artificial Intelligence News*, 7th May 2019, https://artificialintelligence-news.com/2019/05/07/how-coca-cola-is-using-ai-to-stay-at-the-top-of-the-soft-drinks-market/

[3] "UK Small Business Statistics - Business Population Estimates for the UK and Regions in 2018", *National Federation of Self Employed & Small Businesses (FSB)*, https://www.fsb.org.uk/media-centre/small-business-statistics

(2) Processing data

At the end of each week, data on all deliveries and sales at "Le Petit Repas" were inputted into a spreadsheet. Table 1 shows an example extract from such a spreadsheet, with products sorted by type and details on units .

STOCK COUNT		**Week ending 22.03.2015**					
PRODUCT	**What to count**	**Last Week**	**Delivery**	**This Week**	**PRICE**	**##**	Consumption per week
JB FOODS - (0131 448 2888) – LET 014							
PRODUCT							
MUFFINS							
Triple Chocolate Flowerpot Muffin	EACH	27	24	23	£0.63	##	
Blueberry Flowerpot Muffin	EACH	16		5	£0.51	##	
Raspberry & White Choc Flowerpot Muffin	EACH	18		17	£0.48	##	
Strawberry & Cream Flowerpot Muffi	EACH	8		13	£0.52	##	
Lemon & Poppyseed Flowerpot Muffin	EACH	38		33	£0.46	##	
PASTRIES							
Delifrance Croissant	EACH	71	50	57	£0.38	##	
Delifrance Chocolate Twist	EACH	50		32	£0.35	##	
BREADS							
Panefresco Panini	EACH	84	90	86	£0.25	##	
Multi-Seeded Baguette	EACH	30	60	40	£0.29	##	
Jacksons Brown Sliced Bread 14+2	Per loa	15	10	12	£1.23	##	
Jacksons White Bread 14+2	Per loa	15	10	16	£1.19	##	
FIFE CREAMERY – (01592 655757) – PET 005							
PRODUCT							
MENU ITEMS							
Crispy Streaky Bacon	1kg	0.5	2	0.3	£11.46	##	
SANDWICH FILLINGS							
Sliced Ham 80%	500g	1.5	4	3	£2.27	##	
Tuna in Brine Pouch	1.02kg	3	3	3	£5.25	##	
CRISPS							
Real Crisps - Salt & Vinegar	EACH	28	48	63	£0.27	##	
Real Crisps - Cheese & Onion	EACH	24		16	£0.27	##	
Real Crisps - Ham & Mustard	EACH	28		14	£0.27	##	
Real Crisps - Sweet Chilli	EACH	46		34	£0.27	##	
PORRIDGE							

Table 1 - data extract from 22nd of March 2015

The column "Last Week" indicates the amount of the respective product remaining at the end of a previous week, and thus at the beginning of the present week. "Delivery" is how much of each item was delivered at the start of the week. Finally, "This Week" indicates how much of the given product was remaining at the end of the present week. Using all this data, consumption per week can be calculated for each product and stored. This key value will be used later to study the behaviour of sales with relation to different underlying factors.

(3) Factors affecting sales

There are a vast number of factors that can affect the sales of a particular product on a given day. These factors are also likely to have varying importance across different products. In order to study the importance of such determinants of sales, I first had to postulate the factors themselves.

From economic theory, the foremost determinant of the sales of a product is its price. As demonstrated by a typical demand curve, plotted as sales against price, there is an inverse relationship between the two. However, in my data set, the product prices remained constant over the period studied. As such, any potential 'price effects' on sales are isolated in this context.

Another key determinant of sales that I postulated is the weather, which I decided to narrow down to temperature. It is clear that the effect of high or low temperatures will vary across product types, for example with sales of cold drinks and ice cream expected to be higher on hot summer days, and the reverse expected to be true for hot drinks. Accessing such information for Edinburgh weather was relatively easy, with readings stored for many previous years online.[4] However, given the weekly (rather than daily) nature of my consumption data, I have used average temperature per week.

A further factor affecting sales that I considered is the time of year. In the context of Edinburgh, for example, the annual summer Fringe Festival draws hundreds of thousands of people from around the world to participate in and attend various arts performances. This naturally causes sales to be much higher for almost all products in the month of August, independent of factors such as weather, and was estimated to amount to a boost of over £140 million to the local Edinburgh economy in 2015.[5]

Furthermore, public holidays, including bank and religious holidays, might also be expected to see changes in sales. Again, independent from the given weather in such periods, the fact that many people do not work on such holidays would be expected to lead to a boost for sales. Similarly, I expected that school holidays will also see a boost to sales. As can be seen in 'Factors_updated.txt', there were holidays during 25 weeks from the whole year and the café remained open for all of these holidays, meaning that their effect on sales is theoretically identifiable.

Finally, given the café's close proximity to the centre, I postulated that sports games, including rugby and football, could also have a sizeable impact on sales. Such matches, and particularly those featuring international teams, tend to attract a large amount of supporters and thus potential customers.[6] Indeed, I learned anecdotally that this factor was crucial to account for in the café in practice, in order to prevent serious shortages on the supply-side.

I stored all the factors in a text document named "Factors_updated.txt", through which my program accessed the respective data.

[4] https://www.wunderground.com/history/weekly/EGPH

[5] "How Fringe Festival benefits businesses in Edinburgh", *TNT Direct,* 3rd September 2015, https://direct.tnt.co.uk/blog/how-fringe-festival-benefits-businesses-in-edinburgh

[6] "Economic impact of sporting events", *The Independent*, 20th May 2015, https://www.independent.co.uk/student/shu/economic-impact-of-sporting-events-10260570.html

(4) Writing the code

I originally planned to approach this project from a "Neural Networks" perspective. This would consist of having multiple layers of "Neurons", each of which having an input and an output of either 0 or 1. Input, increasing its importance, has a specific "weight". A neuron would have an output of one if sum of all products of inputs and weights is bigger than a certain number. I thought this approach was appropriate because different factors have different "importance" for sales, so the idea of weights would be useful. However, after some research and initial programming, I decided to instead pursue on a "Multiple Linear Regression" (MLR) approach, which will be outlined below.

This decision was made based on a number of reasons. Firstly, Neural Networks are better suited for nonlinear relationships between variables, while the relationships involved in this context are mainly direct proportionalities. In addition to this, as will be further discussed in the "Limitations" section, I only have one year of data to study. This represents a relatively statistically small dataset from which to make accurate conclusions, and the MLR approach is much more suited to using small amounts of data to establish relationships between variables due to a simpler parameter estimation process.[7] Furthermore, MLR is suited to working with multiple independent variables (those outlined in Section 3) affecting the dependent variable (sales) simultaneously. For example, the Fringe Festival takes place during the summer holidays, at which time the temperature is typically above average. This approach allows one to statistically separate the effect of such factors acting simultaneously, leading to isolated direct estimates of the effect of Fringe, for example, on sales.

My code calculates the effect of different factors on sales of individual products through the following method:

1. Calculate the effect of holidays on sales, but by only using data points for which there were no other factors such as the Fringe and matches.
2. Calculate the effect on sales during the Fringe Festival. As this is during the holidays, however, the code isolates the effect of the Fringe by stripping out the effect of holidays from this combined estimate.
3. The effect of matches is calculated in the same way.
4. Temperature, unlike the other binary variables, is a continuous variable. Its effect on sales is estimated by calculating linear equations for each combination of values for the binary variables.
5. Finally, the accuracy of the estimated equation is given by the 'difference', calculated by subtracting the predicted value for sales from the actual value.

As can be seen, the challenging part of these calculations is to not include more than one factor at one time. For example, when calculating how much holidays affect sales, it's important to include weeks where neither the Fringe nor matches are occurring, meaning that the only binary variable affecting sales is holidays. In addition to this, it is even more challenging to calculate by how much Fringe affects sales. This is because Fringe always takes place during the summer holidays, so 'partialling out' the effect of the former requires the calculation outlined in point 2.

I used the programming language Python to write my code, which I chose for its simplicity and popularity amongst machine learning programmers. Python takes a short development time in comparison to other languages and has a wide variety of libraries[8], of which I used *matplot*, *datetime, statistics, numpy* and *scipy* in my program.

[7] "What is Multiple Linear Regression?", *Statistics Solutions,* https://www.statisticssolutions.com/what-is-multiple-linear-regression/

[8] "Why is Python the Best-Suited Programming Language for Machine Learning?", *GeeksforGeeks,* https://www.geeksforgeeks.org/why-is-python-the-best-suited-programming-language-for-machine-learning/

(5) Implementing Multiple Linear Regression (MLR)

Multiple Linear Regression is a statistical estimation approach in which the output depends on multiple input variables. To show mathematically:

$$Y = a + m1*x1 + m2*x2 + \ldots + mn*xn$$

Where "a" is the y-intercept, m is the marginal effect on the output variable of the respective input variables x.

This model is very much suited to describing and predicting the behaviour of sales. To use an example, Figure 1 charts the sales of cold drinks over time. Sales start at a relatively low level of consumption per week in spring, increasing over summer and subsequently falling into winter.

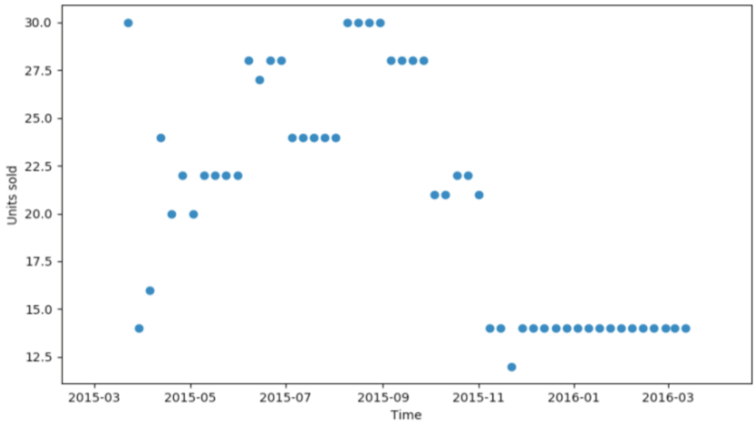

Figure 1 – Sales of cold drinks over time

However, this relationship is unlikely to depend on just one variable that changes with the seasons, such as temperature. To further demonstrate this, Table 2 lists the values of all factors during this period.

Date Week Ends On	Avg Temp F	Avg Temp C	Max Avg Temp F	Max Avg Temp C	Match (1/0)	Fringe (1/0)	Holidays (1/0)
3.05.15	40	4.44	43	6.11	0	0	0
10.05.15	46	7.78	50	10.00	0	0	0
17.05.15	49	9.44	55	12.78	0	0	0
24.05.15	50	10.00	54	12.22	0	0	0
31.05.15	49	9.44	53	11.67	0	0	1
7.06.15	50	10.00	55	12.78	0	0	1
14.06.15	53	11.67	58	14.44	0	0	0
21.06.15	54	12.22	58	14.44	0	0	0
28.06.15	54	12.22	60	15.56	0	0	0
05.07.15	61	16.11	68	20.00	0	0	1
12.07.15	57	13.89	62	16.67	0	0	1
19.07.15	56	13.33	55	12.78	0	0	1
26.07.15	54	12.22	58	14.44	0	0	1
2.08.15	53	11.67	56	13.33	0	0	1
9.08.15	57	13.89	62	16.67	0	1	1
16.08.15	57	13.89	62	16.67	0	1	1
23.08.15	58	14.44	64	17.78	0	1	1
30.08.15	59	15.00	66	18.89	1	1	1
6.09.15	53	11.67	56	13.33	0	1	0
13.09.15	55	12.78	57	13.89	0	0	0
20.09.15	52	11.11	55	12.78	0	0	0
27.09.15	53	11.67	56	13.33	0	0	0
4.10.15	49	9.44	52	11.11	0	0	0
11.10.15	50	10.00	58	14.44	0	0	0
18.10.15	45	7.22	50	10.00	0	0	0
25.10.15	50	10.00	55	12.78	0	0	1
1.11.15	49	9.44	53	11.67	0	0	1
8.11.15	47	8.33	52	11.11	0	0	1

Table 2 – Small part of "Factors_updated.txt", which shows all factors taking place before, during and after the Fringe Festival

As can be seen, the temperature in August isn't that much higher than in September, however sales in August exceed the ones in September by a massive margin. The other factor to be accounted for, which is likely to explain the difference in this case, is the Fringe Festival running through almost the whole of August. The holidays for students at this time may also have had a decisive influence.

Studying and visualising the data in such a way is an interesting exercise, but lacks rigour. My statistical method essentially follows a similar approach, but does so comprehensively and consistently, studying each data point to identify the effect of the multiple variables, which often act simultaneously.

To implement MLR, I wrote a program "Finding_Fringe_Matches_Holidays_Percentages.py". This program produces an output table listing all the different factors together with their relationship with units sold, as well as the final MLR equation. The program asks which product's equation and table you want to see. Choosing for example Still Water, the output is as follows.

```
San Peligrino - Aranciata Rosso 35
Irn Bru Plastic (12x500ml) 36
Irn Bru Sugar Free Plastic 37
Strathmore  Still Water (24x500 ml) 38
Strathmore Sparkling Water 39


Choose which product info you would like to see: 38
                Coeficient                   Intercept
Fringe          4.75221151856163
Holidays        -1.1561562082156398
Matches         6.041274782958041
Temp            1.208412694028377            6.988665698726562


The Multiple Linear Regression equation for the chosen product is:
Y =  6.99  +  1.21 *temp  +  4.75 *fringe  +  -1.16 *holidays  +  6.04 *match

Note: fringe, holidays and match can only be either 1 or 0
>>> |
```

Figure 2 – Program output for Still Water

The MLR equation indicates that sales for Still Water are predicted to:

- Increase by 1.21 units for every single degree increase in temperature,
- Increase by 4.75 units during the Fringe Festival,
- Decrease by 1.16 units when there is a holiday,
- Increase by 6.04 units when there is a sports match, and
- Take a value of 6.99 units when the temperature is zero and neither Fringe, holidays nor matches are occurring.

Such a regression output table can be produced for each individual product. This both allows the shop to account for historical variations in sales, as well as importantly predict sales for future periods.

(6) Testing

The extent to which insights can be drawn from the MLR equation results depends crucially on their accuracy. In order to test how accurate that equation is, "Testing.py" plugs all the non-sales factors (explanatory variables) from the actual data into the equation, and then outputs how accurately each product prediction is. Unfortunately, due to lack of amount of data available for this project, I had to use data that I used to teach the program the pattern to test it. This is not ideal, but I only had one year of data and I needed all of it to see the complete variation in factors.

The accuracy equation used in this program is as follows:

$$\text{Accuracy} = \left(100 - \frac{\textit{Predicted Sales} - \textit{Actual Sales}}{\textit{Actual Sales}} * 100\right)\%$$

Where a value close to 100% indicates a highly accurate MLR equation. The values for each product are listed in Table 3.

```
Triple Chocolate Flowerpot Muffin                    69.31 %
Blueberry Flowerpot Muffin                           64.45 %
Raspberry & White Choc Flowerpot Muffin              26.77 %
Lemon & Poppyseed Flowerpot Muffin                   72.57 %
Delifrance Croissant (1x50))                         96.08 %
Delifrance Chocolate Twist (1x60)                    18.22 %
Panefresco Panini (1x30)                             95.9 %
Multi-Seeded Baguette (1x30)                         84.0 %
Jacksons Brown Sliced Bread 10+2                     96.16 %
Jacksons White Bread 10+2                            94.43 %
Crispy Streaky Bacon                                 96.59 %
Sliced Ham 80%                                       88.47 %
Tuna in Brine Pouch                                  99.97 %
Real Crisps - Salt & Vinegar (1x48)                  67.81 %
Real Crisps - Cheese & Onion                         82.18 %
Real Crisps - Ham & Mustard                          -11.3 %
Real Crisps - Sweet Chilli                           73.69 %
Stoats Classic Scottish Original                     83.81 %
Stoats Apple & Cinnamon                              77.96 %
Stoats Cranberry & Blueberry                         84.95 %
Plain Scone (1x6)                                    90.83 %
Sultana Scone                                        82.95 %
Tea Cakes (1x45)                                     84.31 %
Empire Biscuits (1x8)                                83.0 %
Sliced White Cheddar                                 97.42 %
Pulled Pork                                          99.48 %
Grated Mozzarella (100%)                             92.67 %
Baked Potatoes (Frozen                               47.69 %
Coke Plastic (24x500 ml)                             94.28 %
Diet Coke Plastic                                    94.67 %
Capri Sun Tropical Juice (40x200 ml)                 94.36 %
Appletiser (Apple                                    94.15 %
Bundaberg Ginger Beer (12x375 ml)                    94.68 %
San Peligrino - Aranciata (24x330ml)                 90.1 %
San Peligrino - Limonata                             93.17 %
San Peligrino - Aranciata Rosso                      94.56 %
Irn Bru Plastic (12x500ml)                           92.27 %
Irn Bru Sugar Free Plastic                           93.89 %
Strathmore  Still Water (24x500 ml)                  94.11 %
Strathmore Sparkling Water                           94.21 %
>>>
```

Table 3 – Results returned from "Testing.py"

From Table 3, it can be seen that the computed accuracy level is generally quite high, with an average of 81.62% and many results over 90%.

Notable exceptions, however, include the products: Raspberry & White Choc Flowerpot Muffin with 26.77 % accuracy; Delifrance Chocolate Twist with 18.22 %; and Real Crisps - Ham & Mustard with -11.3 %, indicating an extreme overprediction. The reasons for the low accuracy of these products are not particularly clear and, given that they are relatively few in number, I have assumed they are anomalies. The fact that the vast majority of products have been predicted accurately, however, seems to indicate an overall strength of my program.

(7) Review

Despite the accuracy indicated in Section 6, there are a few limitations to note with regards to my program and the general approach of my project.

Firstly, I was unfortunately only able to access data for one year (2015-16). This sets limitations in particular on the reliability of my results, regardless of how accurate they may be for the year studied. That is, extrapolation to make predictions for sales in further years may lead to misleading conclusions. The fact that the data is now almost five years old also might imply outdated estimated relationships, if one assumes that these change over time. This does not however represent a flaw in my program, which could simply be fed in further data. Indeed, even with one further year of data, my 'Fringe' estimates for example could be much more insightful, as I could incorporate the size of the festival in each year when estimating the effect on sales.

Secondly, as referred to in Section 3, there is no change in price for all products over the period studied. From economic theory, this is the most crucial determinant of sales, and as such my conclusions can only be compared with similar periods with no price variation. Furthermore, estimates of the effect of price would in turn allow the calculation of price elasticity of demand, both of which numbers would be very useful for a business in choosing its optimal pricing strategy. Thus, the remit of my study is limited in this sense. It should however be stressed that this lack of price variation implies a certain statistical 'cleanliness' to my estimates for the other factors, which are unusually undistorted by potential own- or cross-price effects.

In a further project, I would also explore the significance of having constant prices in the broader context of a positive general inflation rate. In theory, having constant prices at a time of general rising prices should make the constant-price business comparatively competitive, thereby increasing sales. Although data are not available for Scotland or Edinburgh alone, the UK-wide inflation rate published monthly by the Office for National Statistics could be a useful indicator. Interestingly, looking at the data for the period covered in my study (March 2015- March 2016), monthly inflation remained remarkably close to zero percent, with the price level only rising a meagre 0.5% over the period, a quarter the Bank of England's 2% target.[9] Thus, omitting a consideration for inflation in my study does not present an overly large challenge to my estimates, as inflation itself was close to zero, but would be important to consider in further studies.

Another economically-important factor that I have not considered here is household disposable income. Rising incomes over time typically lead to increased consumption and thus sales. Looking at per-person gross disposable household income (GDHI) figures for Edinburgh, relatively high growth of 4.4% was seen between 2014 and 2015, and 2.1% for 2015 to 2016.[10] Thus my study is limited in that it has not considered this important factor and, unlike inflation, income growth was non-negligible in the period studied. This is problematic in MLR models to the extent that income cross-correlates with the other variables studied, leading to an "omitted variable bias" in the estimates for the other variables.[11] This would be interesting to explore in further studies, improving both the accuracy and reliability of my estimates.

[9] "Consumer price inflation, UK: March 2016", *Office for National Statistics*, 17th May 2016, https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/consumerpriceinflation/mar2016

[10] "Regional gross disposable household income, UK: 1997 to 2016", *Office for National Statistics,* 24th May 2018, https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/bulletins/regionalgrossdisposablehouseholdincomegdhi/1997to2016#analysis-of-nuts1-regions

[11] "Omitted Variable Bias: A Comprehensive Econometrics Review", *Albert*, 20th September 2016, https://www.albert.io/blog/omitted-variable-bias-econometrics-review/

(8) Bibliography

Albert Schools; "Omitted Variable Bias: A Comprehensive Econometrics Review", 20th September 2016, https://www.albert.io/blog/omitted-variable-bias-econometrics-review/

Barker, Shane; "4 Of The Biggest Factors That Affect Customer Purchase", 17th May, 2018, https://www.searchenginepeople.com/blog/4-biggest-factors-affect-customer-purchase-decisions.html

Bhasin, Hitesh; "Factors affecting sales of products", 21st February, 2018, https://www.marketing91.com/factors-affecting-sales-product/

harkiran78; "Why is Python the Best-Suited Programming Language for Machine Learning?", *GeeksforGeeks,* https://www.geeksforgeeks.org/why-is-python-the-best-suited-programming-language-for-machine-learning/

IBM; https://www.wunderground.com/history/weekly/EGPH

Marr, Bernard; "How Coca-Cola is using AI to stay at the top of the soft drinks market", *Artificial Intelligence News*, 7th May 2019, https://artificialintelligence-news.com/2019/05/07/how-coca-cola-is-using-ai-to-stay-at-the-top-of-the-soft-drinks-market/

May, Rob; "Start-ups across the UK are going bust - they need more careful management for our economy to boom", *The Telegraph,* 24th January 2019*,* https://www.telegraph.co.uk/politics/2019/01/24/start-ups-across-uk-going-bust-need-careful-management-economy/

National Federation of Self Employed & Small Businesses (FSB), "UK Small Business Statistics - Business Population Estimates for the UK and Regions in 2018", https://www.fsb.org.uk/media-centre/small-business-statistics

Philip Gooding; "Consumer price inflation, UK: March 2016", *Office for National Statistics*, 17th May 2016, https://www.ons.gov.uk/economy/inflationandpriceindices/bulletins/consumerpriceinflation/mar2016

Philip Gooding; "Regional gross disposable household income, UK: 1997 to 2016", *Office for National Statistics,* 24th May 2018, https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/bulletins/regionalgrossdisposablehouseholdincomegdhi/1997to2016#analysis-of-nuts1-regions

Statistics Solutions; "What is Multiple Linear Regression?", https://www.statisticssolutions.com/what-is-multiple-linear-regression/

The Independent, "Economic impact of sporting events", 20th May 2015, https://www.independent.co.uk/student/shu/economic-impact-of-sporting-events-10260570.html

TNT; "How Fringe Festival benefits businesses in Edinburgh", *TNT Direct,* 3rd September 2015, https://direct.tnt.co.uk/blog/how-fringe-festival-benefits-businesses-in-edinburgh