COTS DATA-DRIVEN DECISION REPORT

STACY CLAIRE ANZEMO

DATE: 11/03/2024

CHAPTER 1: INTRODUCTION AND PROJECT PLAN

1.1   Introduction

Data is becoming more and more important in helping to guide strategic decisions in today's rapidly changing business environment (Smith, 2019). Utilizing insights from in-depth data analysis to inform organizational goals and actions is known as data-driven decision-making, and it is frequently cited as a cornerstone of contemporary business practices (Jones et al., 2018).

In this report, an analysis of Café On The Sea's (COTS) current business performance will be presented, focusing on three key areas: sales value and volume analysis, product offering analysis, and the impact of home delivery service on sales performance. These analyses aim to provide valuable insights for the Corporate Strategy Manager's 3-year Strategic Plan project.

1.2   Purpose of the report

This study provides actionable insights that are essential for guiding strategic planning activities and acts as a thorough guide for decision-makers within Café On The Sea (COTS). This research intends to support informed decision-making processes by exploring several aspects of COTS's operations and helping the company map out a path for long-term success and growth.

The report is organized to provide a thorough analysis. It consists of several sections: an introduction outlining the project plan, a detailed analysis of sales value and volume, a product offering analysis, an evaluation of the effect of home delivery services, and a conclusion outlining important conclusions and suggestions. The report aims to provide COTS with valuable insights through a methodical methodology, facilitating well-informed decisions that promote organizational growth and success.

1.3   Project Plan

Data analytics framework

In this project, a six-stage framework will be used. The CRISP-DM (Cross-Industry Standard Process for Data Mining). The stages involve, Business data understanding, collection of the data, data preparation, data modelling, evaluation and implementation. By following this structure, we can systematically address the crucial business inquiries presented to us and obtain valuable outcomes.

1. **Understanding the business:** Establishing a thorough understanding of the project's demands, limitations, and goals is stressed in the first phase, known as "Business Understanding." This phase's goal is to outline the primary business concerns that the analysis will try to address in order to align the analytical activity with corporate goals.
2. **Data understanding:** Understanding data involves studying and becoming familiar with available sources of data, Data Understanding. This procedure involves gathering information, analyzing the data, and conducting initial data exploration to obtain more understanding of the data's characteristics and integrity.
3. **Preparation of data:** Data preparation occurs following this, shifting the focus to cleaning, transforming, and merging data for analysis during the Data Preparation phase. This step is essential for ensuring data is accurate, complete, and consistent, and also lays the foundation for insightful analysis.

4. **Modelling:** To answer the previously specified business questions, analytical models are developed and put into use during the modeling phase. To extract insights and patterns from the data, this step may involve a variety of statistical, machine learning, or predictive modeling techniques (Jawed and Sajid, 2022).
5. **Evaluation:** The Evaluation phase evaluates the models' efficacy and performance after they are developed. This entails assessing the predicted accuracy and dependability of the models by testing them against validation datasets.
6. **Deployment:** Ultimately, during the Deployment phase, the organization's decision-making processes include the operationalized insights and results from the analysis. This could entail producing dashboards, reports, or visualizations to effectively inform stakeholders of the findings.

1.4 Value Addition and Business Performance Improvement

COTS can boost corporate performance and its decision-making processes by utilizing data analytics. Through the examination of key performance indicators (KPIs) including revenue, sales volume, and customer satisfaction, COTS will be able to pinpoint opportunities for expansion and improvement (Strohbach *et. al.* 2015). Better analytics will enable data-driven decision-making, which will improve COTS's coffee shops' operational effectiveness, resource allocation, and profitability in the end (Elgendy *et. al.* 2022).

The goal of this study is to show how data analytics may offer useful information that will guide strategic choices and propel COTS's coffee shops' commercial success. chances for growth and optimization through thorough data analysis and interpretation will be found, helping COTS accomplish its strategic goals.

2   CHAPTER 2: DATA PREPARATION QUALITY ISSUES AND REMEDIES
2.1   Data Preparation Quality Issues
Data inaccuracy

Errors or discrepancies in the data that differ from the true or expected values are referred to as data inaccuracies. One way to detect inaccuracies in data is to contrast it with reliable and established sources. Should disparities be discovered between the dataset and the reliable source, it suggests possible errors. Also, Inaccuracies can be quickly found by implementing data validation tests during data processing. To assure data integrity and accuracy, these checks entail comparing the data to predetermined criteria or limitations.

Resolution:

 Manual repair- When errors are found, manual correction could be required. This entails reviewing the inaccurate data points and manually updating or fixing them in light of correct information.

Updating Outdated Information - Outdated information can sometimes lead to data inaccuracies. Resolving errors brought about by out-of-date information can be accomplished in part by updating the dataset with the most current and pertinent data.

Data Quality Control Measures - By putting in place these safeguards during the data entry and collection procedures, errors can be avoided before they start. This could entail creating uniform protocols for data entry and gathering, training staff members, and implementing data validation regulations.

Missing Values
Summary statistics, such as the count and percentage of missing values for each variable, can be used to identify missing data (Demchenko *et. al.* 2013).

Resolution:
Imputation techniques like mean, median, or mode imputation or the use of sophisticated techniques like predictive modeling to estimate missing values based on other variables are two options for resolving missing data. Furthermore, if possible, data can be gathered afresh for any cases that are missing values (Childs *et. al.* 2005).

Outliers
Outliers can be found visually with box plots, histograms, or scatter plots; alternatively, they can be found statistically with z-score or IQR (interquartile range) computations.

Resolution:
To deal with outliers, one can either eliminate them if they are the result of data input errors,

convert the data using methods such as log transformation, or use statistical procedures that are robust and less susceptible to outliers.

Duplicate Records

To identify exact or almost exact matches, combinations of data can be compared.

Resolution:
Removing or merging duplicates according to predetermined standards, including identifying the first or last instance or combining data from many records, is possible.

Inconsistent data

The process of identifying inconsistent data formatting involves scrutinizing the variables' values and structure and seeking deviations in formatting norms (Jin *et. al.* 2015).

Resolution:
Formatting standards can be applied or data normalization technologies can be used to achieve data formatting standardization. For instance, changing all date formats to the same format or changing all text to lowercase.

Problems with Data Integrity
By cross-referencing data with outside sources or domain expertise and searching for contradictions or discrepancies, data integrity problems can be found.

Resolution:
Manual validation, dataset reconciliation, or the application of data validation constraints and rules during data entry are some methods for resolving data integrity problems.

Data Bias
The distribution of variables among various demographic groups or categories and an analysis of the sample data's representativeness can be used to identify data bias.

Resolution:
Using statistical approaches to account for bias, gathering more representative or varied data, or utilizing bias-aware algorithms during analysis are some ways to address data bias.

2.2 Data Problems in The COTS Dataset

Missing values: There are missing values in the "Sales Volume" column, represented by dashes ("-"). For example, in the entry for February 2020 under "Sales Volume" for "Hot drinks" in Southampton, there is a dash indicating a missing value

Misspelling of column names: The naming convention is not followed by column names like "Colddrinks" and "Cold drinks", "Pazztry" and "Pastry". It is more difficult to work with the dataset programmatically because of this discrepancy. Sandwich 'instead' of 'sandwiches'

Inconsistency Currency format: The "Sales Value" field has an inconsistent currency format. Although the sterling pound (£) symbol is used in the majority of entries, there are few cases when negative numbers are used instead, which could be signs of errors or refunds.

Inconsistent city names: In several entries, the city name "Southam" occurs instead of "Southampton". Correcting this discrepancy is necessary for a reliable analysis. In the sales value column, there were also negative values

Inaccurate data: there is a year named 2032 in the dataset which is not supposed to be in that data set, from the description, COTS Dataset has 3years

To address these problems:

Correct misspelled entries: Manually Correct misspelled product categories like "Pazztry" to "Pastry" to ensure consistency and accuracy in analysis

Standardize the inconsistent city names: Standardize city names to ensure consistency throughout the dataset. This involves finding and replacing all inconsistent city names with the correct ones. that is replacing Southamp with Southampton.

Verify and complete missing data: Check to see if the data is indeed missing or if it only applies to specific situations, such as zero sales. If at all possible, fill in the gaps in the data or treat it appropriately given the circumstances.

Handle missing values: Depending on how missing data will affect the analysis, choose a method for handling missing values, such as imputation or removing rows with missing values.

Standardize currency: Make sure that every value in the "Sales Value" column has the same currency format by standardizing the format. Refunds represented by negative values ought to be treated differently or marked as needing extra care.

Remove the inaccurate data that is not supposed to be in the dataset. the 2032 year, or confirm with the management if it's supposed to be 2023

3 CHAPTER 3: DATA ANALYSIS AND COMMENTARY

3.1 Table A

Sales volume and value by month, by year and across the 3 years period

The below table shows the sum of sales volume and sales value by month, by year, and the total sum of sales volume and sales value which incorporates the three years

Month	2020 Sum of Sales Volume	2020 Sum of Sales Value (£)	2021 Sum of Sales Volume	2021 Sum of Sales Value (£)	2022 Sum of Sales Volume	2022 Sum of Sales Value (£)	Total Sum of Sales Volume	Total Sum of Sales Value (£)
1	1,230	4,022	1307	4,219	1,300	4,190	3,837	12,431
2	1290	3264	1396	4486	1456	4717	4141	12467
3	1530	4878	1541	5015	1630	5294	4701	15187
4	1736	5456	1914	6127	2014	6399	5663	17982
5	1513	4870	1459	4355	1675	5313	4647	14538
6	1596	5104	1724	5488	1805	5663	5120	16255
7	1755	5773	1766	5717	2120	6781	5641	18270
8	1810	5846	1935	6246	2409	7746	6163	19839
9	1508	4838	1588	5180	1882	5941	4978	15959
10	1522	4915	1660	5373	1883	5936	5064	16225
11	1480	4651	1672	5333	1937	6230	5088	16215
12	1615	5126	1718	4962	1501	4757	5194	14,845
TOTAL	**18589**	**58742**	**19678**	**62502**	**21609**	**68967**	**60235**	**191392**

Table A: Sales volume and Sales value Analysis

From the above analysis, the sales volume increased from 2020.in 2020 the total sales volume was 1858, it then increased to 19679 in 2021 then to 21609 in 2021. likewise to sales value which also increased as the years progress

3.2 Table B

comparisons of product groups performance covering sales volume and value by quarter, by year and across the 3 years period

This table shows product group performance quarterly, annually, and across the three years

Quarter	Product Category	2020 Sales Volume	2020 Sales Value (£)	2021 Sales Volume	2021 Sales Value (£)	2022 Sales Volume	2022 Sales Value (£)	Total Sales Volume	Total Sales Value (£)
Q1	Cakes	376	1880	401	2003	405	2025	1182	5908
	Coffee	12478	4074	1273	5090	1363	5450	3883	14614
	Cold drinks	813	2033	899	2246	844	2109	2555	6388
	Hot drinks	562	1124	678	1355	594	1188	1834	3667
	Pastry	813	1625	735	1470	914	1827	2461	4922
	Sandwiches	238	1428	260	1557	267	1602	765	4587
Total (Q1)		**4049**	**12164**	**4244**	**13721**	**4386**	**14201**	**12678**	**40085**
Q2	Cakes	446	2230	435	2175	501	2505	1382	6910
	Coffee	1526	6102	1606	6422	1556	6225	4687	18749
	Cold drinks	998	2495	915	1905	1139	2846	3052	7246
	Hot drinks	722	1444	765	1529	842	1684	2329	4657
	Pastry	848	1353	1080	2160	1154	2308	3082	5821
	Sandwiches	301	1806	297	1779	301	1807	899	5392
Total (Q2)		**4840**	**15430**	**5097**	**15970**	**5493**	**17375**	**15429**	**48775**
Q3	Cakes	475	2375	510	2550	542	2708	1527	7633
	Coffee	1568	6270	1591	6364	1913	7654	5072	20288
	Cold drinks	968	2419	1080	2699	1356	3390	3403	8507
	Hot drinks	767	1534	787	1574	888	1775	2442	4884
	Pastry	994	1987	993	1986	1332	2664	3318	6637
	Sandwiches	312	1872	329	1971	380	2278	1020	6121
Total (Q3)		**5084**	**16457**	**5289**	**17144**	**6410**	**20468**	**16781**	**54068**
Q4	Cakes	422	2108	430	2148	474	2372	1326	6628
	Coffee	1338	5352	1577	6306	1546	6182	4461	17840
	Cold drinks	903	2258	984	1926	1052	2629	2939	6813

	Hot drinks	687	1374	712	1424	821	1641	2220	4439
	Pastry	1001	2002	1055	2109	1118	2237	3174	6348
	Sandwiches	267	1599	293	1755	310	1862	870	5216
Total (Q4)		**4617**	**14692**	**5049**	**15668**	**5321**	**16923**	**14990**	**47284**
GRAND TOTAL		**18589**	**58742**	**19678**	**62502**	**21609**	**68967**	**59876**	**190212**

<div align="center">Table B: Product categories quarterly, annually, and across the three years</div>

3.3   Table C

comparisons of sales volume and value between coffee shops by quarter, by year and across the 3 years period

This table compares the performance of the three coffee shops using sales volume and sales values across the 3year period

Quarter	Coffee Shop	2020 Sales Volume	2020 Sales Value (£)	2021 Sales Volume	2021 Sales Value (£)	2022 Sales Volume	2022 Sales Value (£)	Total Sales Volume	Total Sales Value (£)
Q1	Blackpool	1844	5883	2014	6428	1890	6191	5748	18502
	Portsmouth	1371	3410	1232	4102	1454	4655	4056	12167
	Southampton	834	2871	998	3191	1042	3354.5	2874	9417
total (Q1)		**4049**	**12164**	**4244**	**13721**	**4386**	**14201**	**12678**	**40085**
Q2	Blackpool	2077	6587	2189	6781	2486	7913	6752	21281
	Portsmouth	1659	5291	1736	5439	1763	5582	5157	16311
	Southampton	1104	3553	1172	3750	1245	3881	3521	11184
Total (Q2)		**4840**	**15430**	**5097**	**15970**	**5493**	**17376**	**15430**	**48775**
Q3	Blackpool	2324	7486	2320	7462	3195	10165	7839	25113
	Portsmouth	1598	5204	1818	5866	1944	6205	5360	17275
	Southampton	1161	3767	1151	3816	1271	4098	3583	11681
total (Q3)		**5083**	**16457**	**5289**	**17144**	**6410**	**20468**	**16781**	**54068**

Q4	Blackpool	2070	6608	2230	7160	2873	9117	7173	22885
	Portsmouth	1449	4580	1706	4892	1610	5435	4765	10907
	Southampton	1098	3505	1113	3616	748	2372	2959	9493
total (Q4)		**4617**	**14692**	**5049**	**15668**	**5321**	**16923**	**14987**	**47289**
Grand Total		**18589**	**58742**	**19678**	**62502**	**21609**	**68967**	**59876**	**190217**

Table C: Coffee shops performance comparison

From this analysis, Blackpool has always had a higher sales volume and sales value. this is a result of the home delivery strategy that was implemented in Blackpool and not the other coffee shops

4   CHAPTER 4: DATA CHARTING AND COMMENTARY

4.1   Chart A

Comparison of sales value trends across coffee shops over time

This is a cluster column chart used to visualize the sales value trends across the coffee shops over the three years
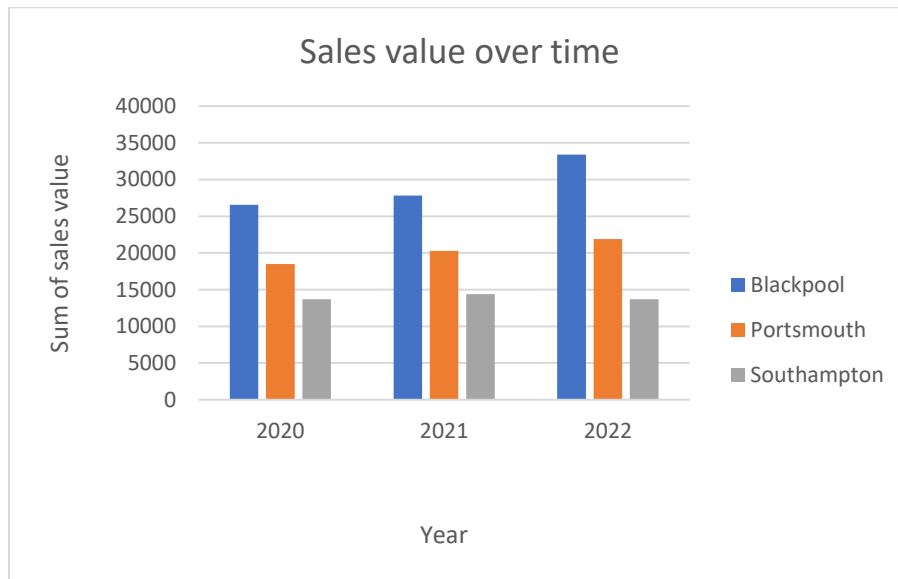

Chart A: Comparison of sales value over time

From this chart, the sales value has increased over the three years in all the coffee shops indicating that all the coffee shops are doing well in terms of performance

4.2 Chart B
Product category performance comparisons between coffee shops

The below chart shows each product performance in every coffee shop
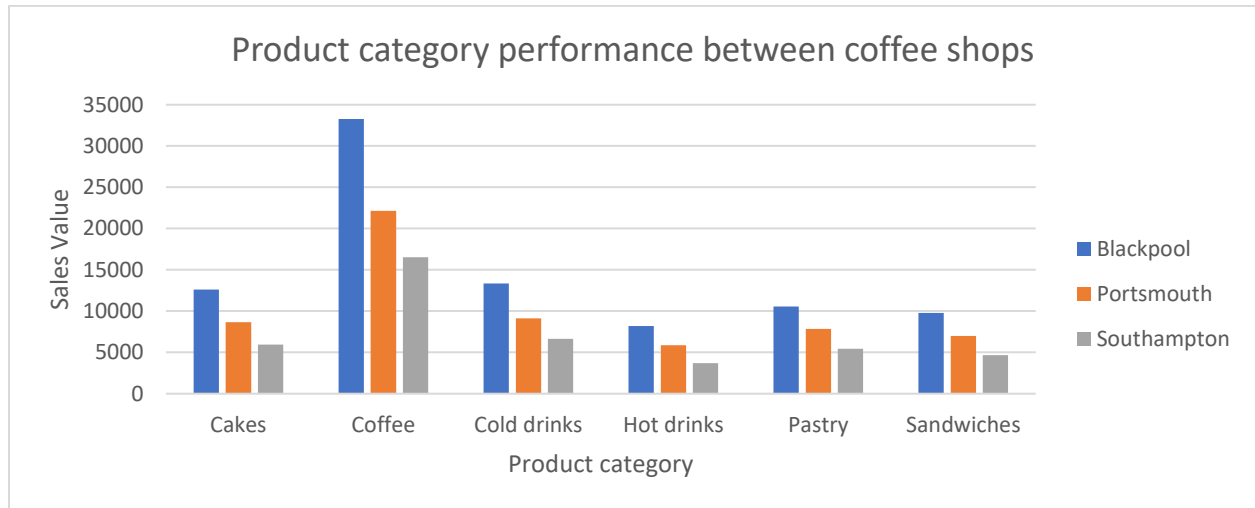
Chart B: Product category performance

Blackpool has the highest sales value of all the products. Coffee also has the highest sales value in all the shops implying that coffee is the best in terms of performance in all the coffee shops. Hot drinks are the least from the above chart.

4.3 Chart C

Impact of the home-delivery service offered in the Blackpool area, and in comparison, with the other two cities

This chart shows the general impact of the home delivery service implemented in Blackpool vs the other coffee shops that did not implement home delivery.
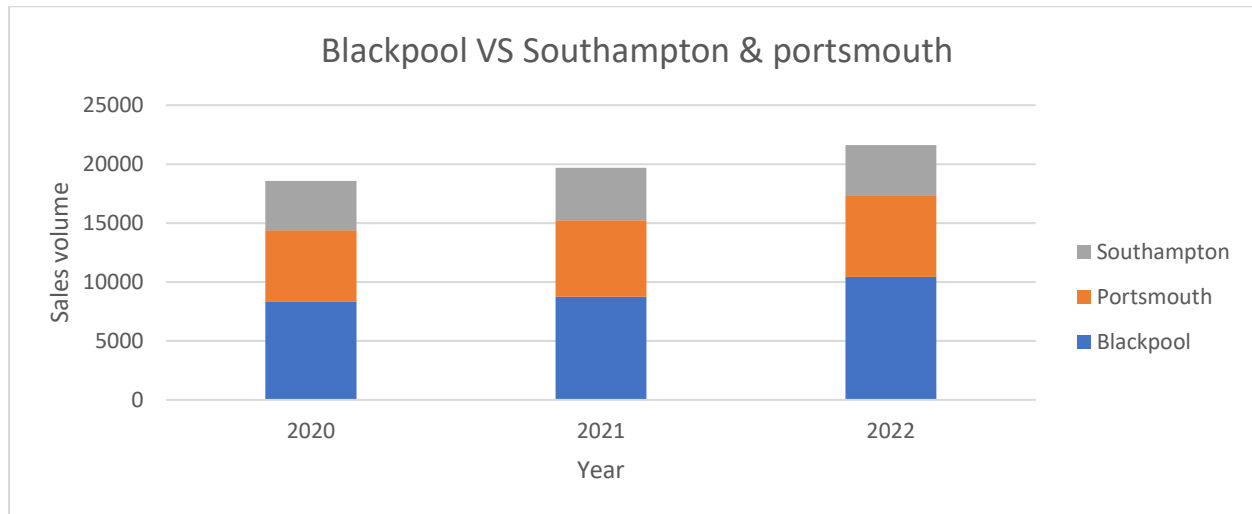

Chart C: Impact of the home-delivery service

The stacked column shows that home delivery had a positive impact on Blackpoll as it has a higher sales volume than the other shops. Also in 2022, the sales volume increased implying a positive impact from the home delivery.

5 CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

After a thorough examination of the operations and sales performance of Café On The Sea (COTS), several important findings have been made that address the three concerns brought out by the upper management.

Based on its good sales performance in comparison to Southampton and Portsmouth, the Blackpool coffee shop showed the most growth potential, according to the sales value and volume Analysis.

The review of the product offering revealed several underperforming products in several product categories, indicating areas for cost- and menu-cutting.

There may be room to expand this service to other locations given the favorable correlation that was shown between the home delivery service's impact and improved sales performance in Blackpool.

5.2 Recommendations

Suggestions for the CEO and Senior Management of COTS:

Simplify the menu by getting rid of any failing items that the product offering analysis revealed, as this will maximize profits and raise customer happiness.

Based on the success seen in Blackpool, think about extending the home delivery service to additional coffee shops, taking advantage of the growing trend of remote work and home-based consumption.

The Blackpool coffee shop should increase its floor area to meet the increasing demand from customers and take advantage of its good sales performance.

Use cutting-edge data analytics tools and methods to improve inventory control, target marketing campaigns, and increase the accuracy of sales forecasts.

Invest in staff training initiatives to raise data literacy and cultivate an organization-wide data-driven culture, enabling staff members to use data insights to make decisions.

To guarantee data security, quality, and integrity and to improve the dependability and credibility of analytical insights, establish data governance policies and procedures.

Invest in customer relationship management (CRM) systems to monitor consumer behavior and preferences, allowing for individualized marketing campaigns and offers.

Utilize digital technology to improve consumer engagement and loyalty, which will increase repeat business and revenue growth. Examples of these technologies are mobile ordering apps and reward programs.

Investigate forming strategic alliances with regional vendors to obtain premium ingredients at affordable costs, hence augmenting product offerings and profitability.
In conclusion, COTS can position itself for sustainable growth, competitive advantage, and ongoing success in the dynamic coffee shop market sector by putting the aforementioned advice into practice and successfully utilizing data analytics.

References

Childs, S., Blenkinsopp, E., Hall, A. and Walton, G., (2005). Effective e-learning for health professionals and students—barriers and their solutions. A systematic review of the literature—findings from the HeXL project. *Health Information & Libraries Journal*, *22*, pp.20-32.

Demchenko, Y., Grosso, P., De Laat, C. and Membrey, P., (2013), May. Addressing big data issues in scientific data infrastructure. In *2013 International conference on collaboration technologies and systems (CTS)* (pp. 48-55). IEEE.

Elgendy, N., Elragal, A. and Päivärinta, T., (2022). DECAS: A modern data-driven decision theory for big data and analytics. *Journal of Decision Systems*, *31*(4), pp.337-373.

Jawed, M.S. and Sajid, M., (2022), October. Cryptanalysis of Lightweight Block Ciphers using Metaheuristic Algorithms in Cloud of Things (CoT). In *2022 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 165-169). IEEE.

Jin, X., Wah, B.W., Cheng, X. and Wang, Y., (2015). Significance and challenges of big data research. *Big data research*, *2*(2), pp.59-64.

Jones, A., et al. (2018). Leveraging insights from data analysis for organizational goals. *Journal of Business Analytics*, 5(2), 123-136

Smith, J. (2019). *The role of data in guiding strategic decisions.* Harvard Business Review. Retrieved from

Strohbach, M., Ziekow, H., Gazis, V. and Akiva, N., (2015). Towards a big data analytics framework for IoT and smart city applications. *Modeling and processing for next-generation big-data technologies: with applications and case studies*, pp.257-282.