

**BPP Business School****Coursework Cover Sheet**

Module name	Data Driven Decisions for Business
Programme	MSc Management
Student Reference Number (SRN)	
Assignment Title	DB Bank Dataset [CW1]

Declaration of Original Work:

I hereby declare that I have read and understood BPP's regulations on plagiarism and that this is my original work, researched, undertaken, completed and submitted in accordance with the requirements of BPP School of Business and Technology.

The word count, excluding contents table, bibliography and appendices, is 2503 words.

Student Reference Number:

Date:

By Brainwritings EXPERTS

## Contents

<b>Introduction &amp; Project Plan</b> .....	3
<b>Data Quality Issues and Remedies</b> .....	4
<b>Data Analysis</b> .....	9
<b>Data Charting</b> .....	15
<b>Conclusions &amp; Recommendations</b> .....	19
<b>References</b> .....	20

By Brainwritex EXPERTS

## Introduction & Project Plan

The reliance of businesses on data analytics to make key decisions has been a noticeable trend in the past few years. A study has shown that the market size of business intelligence and data analytics software applications is projected to grow from \$15.3 billion in 2021 to over \$18 billion by 2026 (Valishery, 2022). Owing to this, the assessment and decision-making of opening new branches, prioritizing certain money saving services, and their impact on the performance of the bank must be conducted using data inferences. Therefore, this report will aim to employ data analytics and charting techniques to dissect the feasibility of opening a new branch in one of Manchester, London, and Liverpool. Moreover, it will analyze the performance of various money saving services offered in these branches along with their impact on the deposit value of the branches to decide on the prioritization of a high-performing money saving service which will benefit the bank monetarily.

### Project Delivery Plan

The project delivery plan will be based on the SEMMA framework. This framework stands for sample, explore, modify, model, and assess (Wiemer et al., 2019). In the context of what is required for this project, this framework will be highly useful because of how it enables the employer to understand various analytical stages. The project will be specifically delivered based on this framework because it would ensure effective data analysis and assessment which would then lead toward action based on the inferences made through the data.

### Key Performance Indicators (KPIs) and the Role of an Analytical Framework

The first KPI for this project will be customer volume. This is primarily because customer volume remains vital for adjudging the feasibility of a city among Manchester, London, and Liverpool for

opening new branches. The second KPI would be the performance of different money saving services across these branches which will help in understanding the most feasible money saving service for the bank. Lastly, the impact of these services on the deposit value of the branch will help in assessing the overall implications of them. The implementation of the SEMMA framework would help in isolating all these KPIs and assessing them individually contributing to a space where better strategy can be crafted to enhance them.

### The Role of Data Analytics in Improving Business Performance for Darcy & Bennet (D&B)

A study by Aydiner et al. (2019) showed that the adoption of business analytics had a positive impact on business process performance ultimately resulting in better firm performance. Similarly, Darcy & Bennet can benefit from this since data analytics would help the company understand possible avenues for expansion. This will be both in terms of initiating new branches and prioritizing certain money saving services. Therefore, data analytics can improve business performance by inculcating a better decision-making process which would ultimately result in monetary success and the maintenance of a sound brand image.

### **Data Quality Issues and Remedies**

#### Generic Data Quality Problems, Their Identification, and Remedies

##### 1. Missing Data

Missing data is the data that could not be found in a study or was assumed to be collected by the end of the study but did not happen (Pham et al., 2024). This can be a major issue since it can lead to biased datasets leading to false inferences. It can be identified by performing checks for default

values or NaN. Nevertheless, profiling tools are used nowadays for this purpose. Data analysts or scientists can use imputation methods to predict missing data in the case of low percentage of missing data or remove the entire section to reduce bias (How to deal with missing data).

## 2. Hidden Data

Hidden data is the data that is not visible in the standard settings and can often be misleading by painting a different picture (Dubois, 2009). It can be identified by activating special settings. In many cases, it may not be visible but there will be traces in the form of slack space or the labeling of data as 'bad'. Using different extensions and software to decode hidden data can help in eliminating the false inferences that may arise as a result.

## 3. Duplicate Data

Duplicate data is the data that is repeated which leads to a detrimental effect on classification accuracy (Kolcz et al., 2003). The easiest way to identify the presence of duplicate data is by employing data de-duplication tools which automate the process since it can take a long time if done manually. Various de-duplication methods can be used to remove this data. These include Ovid multifile search, Refworks, Mendeley, and Endnote, all of which operate in different ways (Kwon et al., 2015).

## 4. Inconsistent and Unstructured Data

Inconsistent data refers to the contradictory nature of a dataset where the data provided throughout the dataset contradicts itself. In different systems, this can be identified differently by employing certain functions. For example, in R programming language, it can be done by using the recoding

function (Handling inconsistent data, 2023). This problem can be solved through standardizing the dataset and using normalization techniques.

Unstructured data refers to the data that is not compiled according to a predefined format. This form of data leads to difficulties in data analysis (Castellanos et al., 2017). The use of Natural Language Processing (NLP) and Machine Learning tools help in the identification of unstructured data (Castellanos et al., 2017). Techniques such as tokenization, stemming, n-gramming, and stop word removal can be employed to overcome the issue for different formats (Kotu & Deshpande, 2019).

#### 5. Outliers

They are the data that differs significantly from the other values present in a dataset (What are outliers in the data?, 2019). Outliers in the data can be identified using techniques such as scatter and box plots (What are outliers in the data?, 2019). It can also be identified when analyzing the data through personal observations and can be picked up by data analysts. Outliers can be removed or also separately analyzed in specific cases to account for them in final inferences.

#### 6. Erroneous Data

Erroneous data is the data that must be rejected by the system and is deemed invalid (Testing-Advanced programming techniques). The identification of this data can be conducted through the performance of various tests to ensure that such data doesn't exist in the dataset (Testing-Advanced Programming techniques). If such data is identified, it can be removed or corrected using appropriate input software.

#### 7. Obsolete Data

When a new stream of data in a particular field is established, the previous data becomes obsolete which is propelled by IoT systems (Alieksieiev et al., 2018). Identifying obsolete data is easy with the information available on the web and constant reviews can help establish this. Obsolete data can be removed and new data can be introduced into datasets to ensure relevant data analysis.

### Problems in the Dataset of D&B, Their Identification, and Resolution

#### 1. Negative Values in the Dataset

Problem: A glaring problem that was identified in the dataset of D&B bank were the negative values. The presence of negative values can be for many reasons but it has been found to have consequences for statistical analysis (Sadler, 2021).

Identification: This problem was identified during the data validation phase where it was found that these negative values existed in the “Customer Volume” and “Deposit Value” sections. An example of this is the presence of the “-” sign in the “Customer Volume” section and “£0” in the “Deposit Value” section in row 9 and 308 of the dataset among other examples.

Resolution: The rows bearing this problem were nullified by perceiving “-” as 0 and eliminating their role from the overall statistical analysis so that they don't have negative consequences on the overall process.

#### 2. Incorrect City and Year Entries

Problem: There were numerous instances where incorrect entries correlating to the year and the city name were found.

Identification: This problem was detected while reviewing the data where one problem was picked up and by reviewing the whole data, it was established that there many such problems that existed within the dataset. Examples of incorrect city names can be found by the repetitive listing of London as “Lon”. The rows 81,134, and 135 are a few examples of this trend. When it comes to the year entries, an example is the listing of 2022 as 2042 in five consecutive rows from 213 to 217.

Resolution: The problem was resolved by using the Find and Replace function in Excel and correcting the “Lon” entries to London while correcting the listing of 2042 to 2022.

### 3. Use of abbreviations for savings services categories

Problem: The problem of inconsistencies in naming was found in the dataset when review the savings services categories.

Identification: The problem was identified while reviewing the data to form an analysis where inconsistencies in naming were detected. For example, the category “Cash Management Accounts” was listed as Cash MA, the category “Money Market Accounts” was listed as Money MA. The abbreviation ISA was also used for “ISA Accounts”. The use of these abbreviations was observed in rows 102, 174, and 184 among many other.

Resolution: Using the Find and Replace option, all of these abbreviations were replaced with their full category names.



## Data Analysis

Table A: Annual Customer Volume and Deposit Value Trends by City from 2020-2022

Customer volume and deposit value are two key metrics to measure the success of a business. Therefore, the following table will reflect on these metrics by providing monthly and yearly data across a 3-year period ranging from 2020 to 2022.

		Customer Volume			Deposit Value (£)		
Year	Month	London	Liverpool	Manchester	London	Liverpool	Manchester
2020	1	322	390	593	£1,075	£1,280	£1,743
	2	330	423	614	£586	£1,380	£1,629
	3	371	475	777	£1,005	£1,723	£2,398
	4	394	553	738	£1,585	£1,934	£2,869
	5	377	441	664	£1,157	£1,776	£2,005
	6	358	494	721	£1,200	£1,810	£2,163
	7	393	592	803	£1,421	£1,547	£2,485
	8	451	557	838	£1,555	£1,760	£2,662
	9	366	438	679	£853	£1,119	£2,202
	10	377	539	603	£1,112	£1,679	£1,986
	11	309	528	615	£966	£1,770	£2,218
	12	348	554	809	£1,455	£1,368	£2,076
		<b>4396</b>	<b>5982</b>	<b>8454</b>	<b>£13,967</b>	<b>£19,146</b>	<b>£26,436</b>
2021	1	313	397	555	£1,020	£1,255	£1,877

	2	264	486	637	£1,020	£1,479	£1,872
	3	318	451	696	£1,263	£1,174	£2,538
	4	457	682	822	£1,206	£2,314	£2,601
	5	392	484	807	£1,032	£1,573	£2,526
	6	314	539	766	£1,037	£2,039	£2,667
	7	402	607	774	£1,596	£2,246	£2,382
	8	403	708	899	£1,186	£1,804	£2,821
	9	269	523	690	£1,274	£1,680	£2,341
	10	361	551	737	£1,157	£1,702	£2,454
	11	394	594	788	£1,006	£1,776	£2,486
	12	312	646	745	£1,131	£2,042	£2,765
		<b>4199</b>	<b>6666</b>	<b>8916</b>	<b>£13,925</b>	<b>£21,081</b>	<b>£29,330</b>
<b>2022</b>	1	396	440	598	£1,098	£1,210	£1,810
	2	355	431	607	£1,040	£1,470	£1,991
	3	446	497	717	£1,184	£1,720	£2,194
	4	517	677	875	£1,417	£2,091	£2,835
	5	388	646	718	£1,166	£1,570	£2,475
	6	447	546	806	£1,270	£1,960	£2,630
	7	432	617	1078	£1,439	£2,173	£3,481
	8	516	727	1207	£1,559	£2,439	£3,776
	9	349	582	892	£1,188	£1,770	£2,611
	10	383	540	914	£1,070	£1,603	£3,099
	11	467	570	980	£1,193	£1,953	£3,232
	12	343	584	891	£1,193	£1,542	£3,245
		<b>5039</b>	<b>6856</b>	<b>10284</b>	<b>£14,814</b>	<b>£21,501</b>	<b>£33,378</b>
<b>Across 3 Years</b>		<b>13634</b>	<b>19504</b>	<b>27654</b>	<b>£42,705</b>	<b>£61,728</b>	<b>£89,144</b>

Table A: Data and trends in customer volume and deposit value by month, by year and across 3 years period

- Annual Growth

Manchester showcased annual growth both in terms of customer volume and deposit value which indicates that branches in the city have attracted customers. Liverpool remains second in terms of this while London ranks last as it also showed decreasing trends in customer volume and deposit value from 2020 to 2021.

- Customer attraction and retention

The deposit value of Manchester increased from £26,436 in 2020 to £33,378 in 2022. This is an indication of how the branch has managed to attract customers and also deploy effective strategies to retain them. On the contrary, London has digressed in this regard since its deposit value decreased from 2020 to 2021 and only noticed a slight increase in 2022. Liverpool plays a mediating role in this regard as it has not progressed immensely as compared to Manchester but shown some growth.

Table B: Comparison of Saving Services Performance using Quarterly Results of Three Years

This table will provide quarterly data from 2020-2022 of the performance of money saving services offered by the bank based on customer volume and deposit value.

**Customer Volume**

	<b>Cash management accounts</b>	<b>Certificate of deposit</b>	<b>High-yield savings accounts</b>	<b>ISA accounts</b>	<b>Money market accounts</b>	<b>Traditional savings accounts</b>

2020	Q1	286	450	596	905	785	1274
	Q2	338	416	650	923	854	1559
	Q3	343	438	827	1058	876	1576
	Q4	276	433	665	1098	865	1346
		<b>1242</b>	<b>1737</b>	<b>2738</b>	<b>3983</b>	<b>3380</b>	<b>5754</b>
2021	Q1	320	307	684	718	877	1213
	Q2	385	492	746	953	1072	1616
	Q3	386	404	884	1004	1057	1541
	Q4	380	465	737	1010	936	1601
		<b>1470</b>	<b>1667</b>	<b>3050</b>	<b>3685</b>	<b>3941</b>	<b>5970</b>
2022	Q1	276	353	702	931	824	1402
	Q2	269	540	796	1213	1233	1569
	Q3	384	629	755	1349	1292	1991
	Q4	443	571	823	1111	1072	1651
		<b>1372</b>	<b>2093</b>	<b>3077</b>	<b>4604</b>	<b>4420</b>	<b>6613</b>
Across 3 Years		<b>4084</b>	<b>5496</b>	<b>8865</b>	<b>12271</b>	<b>11740</b>	<b>18336</b>

**Deposit Value (£)**

		Cash management accounts	Certificate of deposits	High-yield savings accounts	ISA accounts	Money market accounts	Traditional savings accounts
2020	Q1	£1,392	£1,890	£911	£1,607	£1,915	£5,104
	Q2	£1,757	£2,280	£1,559	£1,947	£2,722	£6,233
	Q3	£1,571	£2,124	£1,361	£1,816	£2,399	£6,331
	Q4	£1,742	£2,095	£1,375	£1,847	£2,372	£5,200
		<b>£6,462</b>	<b>£8,389</b>	<b>£5,206</b>	<b>£7,217</b>	<b>£9,407</b>	<b>£22,868</b>
2021	Q1	£1,628	£2,110	£1,292	£1,499	£1,935	£5,033
	Q2	£1,539	£2,229	£1,367	£2,669	£2,681	£6,508
	Q3	£2,099	£2,558	£1,597	£1,942	£2,787	£6,345
	Q4	£1,963	£1,874	£1,741	£2,118	£2,531	£6,292

		<b>£7,229</b>	<b>£8,770</b>	<b>£5,997</b>	<b>£8,228</b>	<b>£9,934</b>	<b>£24,178</b>
<b>2022</b>	<b>Q1</b>	£1,664	£2,079	£1,235	£1,728	£1,836	£5,175
	<b>Q2</b>	£2,025	£2,457	£1,797	£2,257	£2,941	£5,935
	<b>Q3</b>	£2,067	£2,734	£1,946	£2,507	£3,449	£7,733
	<b>Q4</b>	£1,868	£2,714	£1,812	£2,546	£2,562	£6,627
		<b>£7,624</b>	<b>£9,984</b>	<b>£6,790</b>	<b>£9,038</b>	<b>£10,787</b>	<b>£25,470</b>
<b>Across 3 Years</b>	<b>£21,315</b>	<b>£27,142</b>	<b>£17,993</b>	<b>£24,483</b>	<b>£30,128</b>	<b>£72,516</b>	

*Table B: Benchmark comparisons of saving services categories performance covering customer volume and deposit value by quarter, by year and across the 3 years period*

- Performance of money saving services

The most high-performing service is the traditional savings account as it has shown progress throughout 2020 to 2022 both in terms of customer volume and deposit value. While traditional savings account is far ahead than the rest in terms of customer preference, some other well-performing services include ISA accounts, money market accounts, and certificates of deposit as they have shown significant deposit value and customer volume. The sustainable trends across these services highlight their reliability and growing customer preferences for them.

- Quarterly analysis to analyze seasonal trends

The general trends have shown that the customer volume and deposit value tend to increase from Q1 in Q2 and Q3 and then undergo a decrease in Q4. The increase in revenue in Q2 indicates the strength of the foundation laid by the bank in Q1. The end of Q2 sees an evaluation of goal progression which if done effectively leads to good Q3 performances. Moreover, the performance

in this quarter also increases due to tax implications. The reason for the slump in Q4 is most likely the holiday season and other annual closures.

Table C: Quarterly Comparison of Customer Volume and Deposit Value by Bank Branches in Three Cities

The table will present quarterly data of customer volume and deposit value correlating to branches in three cities.

	City	Customer Volume			Deposit Value (£)		
		London	Liverpool	Manchester	London	Liverpool	Manchester
2020	Q1	1023	1288	1984	£2,665	£4,384	£5,770
	Q2	1129	1487	2123	£3,942	£5,520	£7,037
	Q3	1210	1587	2320	£3,828	£4,425	£7,349
	Q4	1034	1621	2027	£3,533	£4,818	£6,280
			<b>4396</b>	<b>5982</b>	<b>8454</b>	<b>£13,967</b>	<b>£19,146</b>
2021	Q1	895	1334	1888	£3,302	£3,908	£6,287
	Q2	1163	1705	2395	£3,274	£5,925	£7,794
	Q3	1074	1838	2363	£4,055	£5,729	£7,544
	Q4	1067	1790	2270	£3,294	£5,520	£7,705
			<b>4199</b>	<b>6666</b>	<b>8916</b>	<b>£13,925</b>	<b>£21,081</b>
2022	Q1	1197	1368	1922	£3,322	£4,400	£5,995
	Q2	1352	1869	2399	£3,852	£5,621	£7,940
	Q3	1297	1926	3177	£4,185	£6,382	£9,868
	Q4	1193	1694	2785	£3,455	£5,098	£9,576

		<b>5039</b>	<b>6856</b>	<b>10284</b>	<b>£14,814</b>	<b>£21,501</b>	<b>£33,378</b>
<b>Across 3 Years</b>		<b>13634</b>	<b>19504</b>	<b>27654</b>	<b>£42,705</b>	<b>£61,728</b>	<b>£89,144</b>

*Table C: Benchmark comparisons of customer volume and deposit value between bank branches by quarter, by year and across the 3 years period*

- Performance Insights

Manchester shows great results and growth corresponding to customer volume and deposit value which is also seen in Table A. Liverpool does not show huge numbers when it comes to customer volume compared to Manchester, however, it is still able to show promising deposit value numbers pointing to a valuable customer base. London shows no growth trends which is also visible in Table A which is indicative of the lackluster performance in the city.

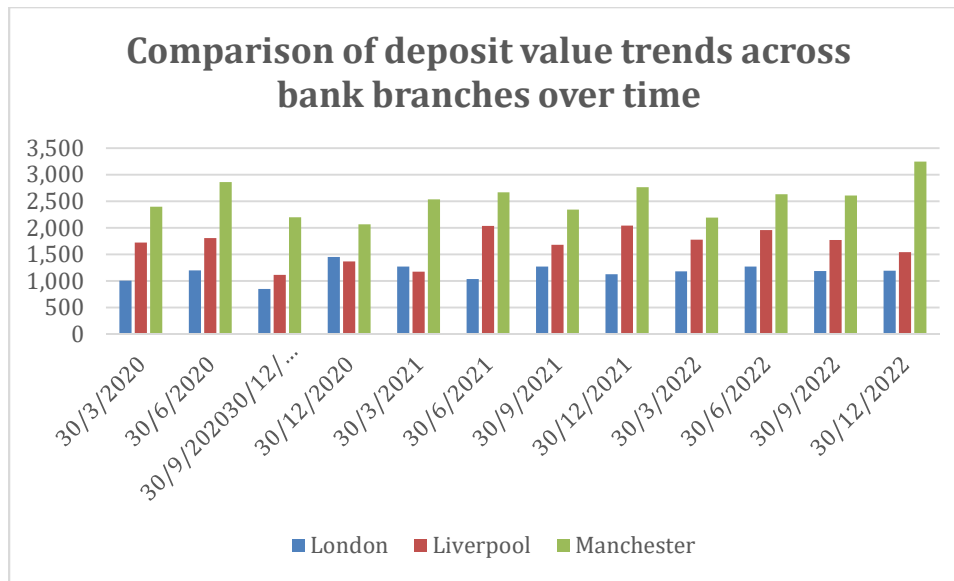
- Quarterly trends

Increase in performance across all cities can be observed in Q2 and Q3 while a decline is observed in Q4 which is down to the reasons discussed in inferences made for Table B while discussing quarterly trends.

### **Data Charting**

Chart A: Comparison of deposit value trends across bank branches over time

The following table will cover the deposit value of bank branches in Manchester, Liverpool, and London.



**Chart A: Comparison of deposit value trends across bank branches over time**

The chart shows that the deposit value of Manchester has grown between the period of 2020 to 2022 and also showed increasing trends within. Compared to this, Liverpool and London fall way behind. Although the deposit value of Liverpool branches has not taken off, it still remains a better-performing city than London.

**Chart B: Saving service category performance comparisons between bank branches**

Chart B will showcase the performance of different service categories across branches in the three cities under discussion.



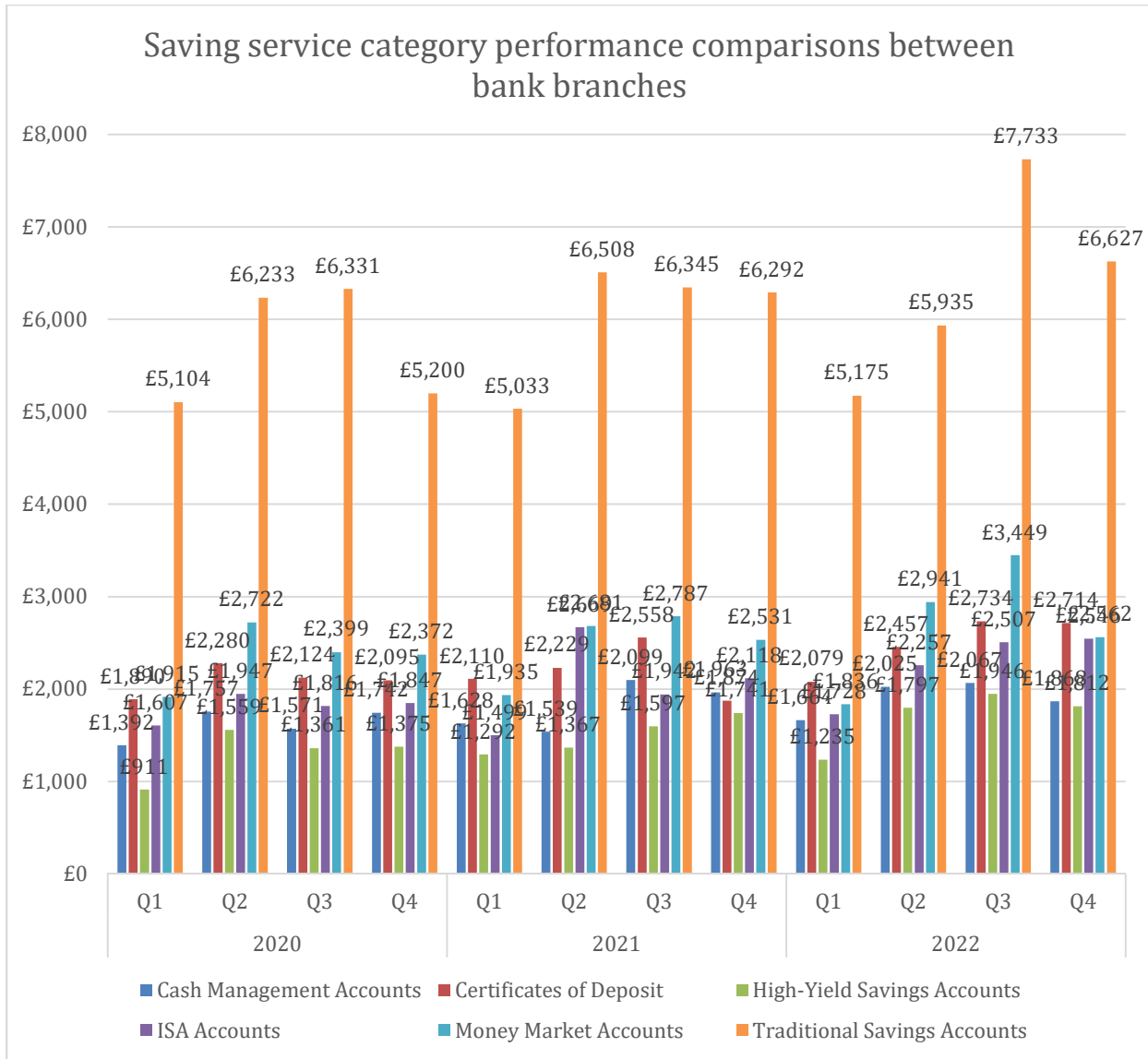


Chart B: Saving service category performance comparisons between bank branches

Crafted on the dataset of Table B, this chart highlights and further emphasizes the strong performance of traditional saving accounts from 2020 to 2022 which makes it the best performance money saving service for the bank. The ISA accounts, money market accounts, and certificates of deposit category showed strong customer uptake demonstrating customer trust in these services.

Chart C: Impact of the Expansion and Renovation Done in the Manchester Branch of the Bank and Comparison with Other Two Bank Branches

This chart will demonstrate the impact that renovation has had on the Manchester branch of the bank while comparing it with the other two branches.

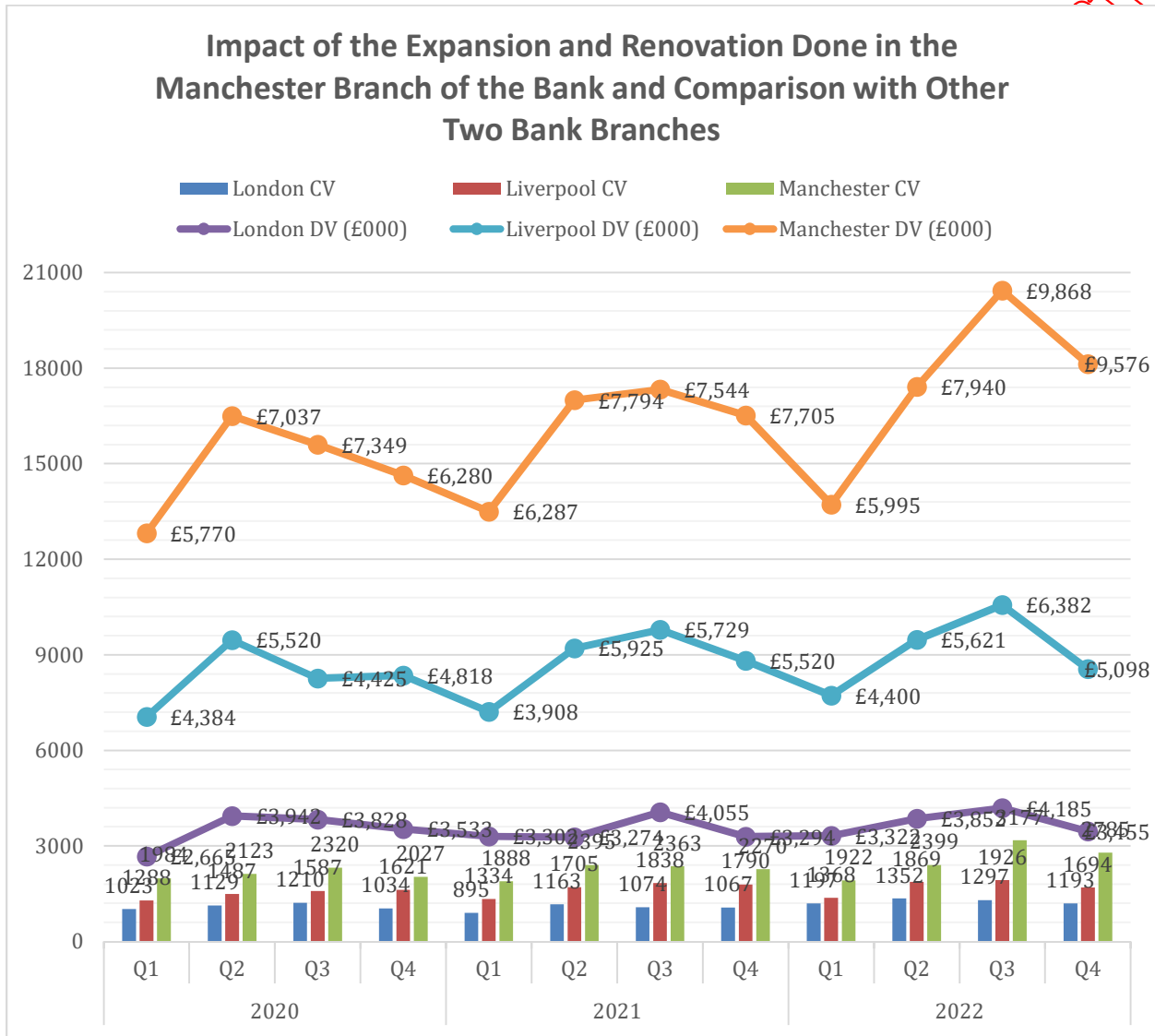


Chart C: Impact of the Expansion and Renovation Done in the Manchester Branch of the Bank and Comparison with Other Two Bank Branches

The chart shows that Manchester observed a steep increase in its customer volume and deposit value following the renovation. Liverpool and London also showed some increase during 2022 but it isn't hugely noticeable. Owing to this, the better performance of Manchester branches can be attributed to the renovation of these branches.

## Conclusions & Recommendations

To conclude, the most feasible city to open new bank branches is Manchester owing to the high customer volume and deposit value observed in branches across the city. The best money saving service inferred through tables and charts is traditional savings account due to its high deposit value. The renovation of branches in Manchester had a positive impact on their performance which can be incurred through the relevant tables and charts. It is recommended for D&B to enhance their data analytical frameworks to ensure better decision-making. It is also recommended to decrease their reliance on physical branches since it is vital to emerge according to modern trends. The bank must also focus on the renovation of its branches since it is proven to have a positive impact on performance.

By Brain writes

EXPERTS

## References

- Aydiner, A. S et al. (2019). Business analytics and firm performance: The mediating role of business process performance. *Journal of Business Research* 96: 228-37  
[https://scholar.google.com.pk/scholar\\_url?url=https://www.academia.edu/download/91060013/197623623.pdf&hl=en&sa=X&ei=nqysZuigM72Uy9YP74XdwAw&scisig=AFWwaebwHV8AMtq8NDjVgFLdOFx-&oi=scholar](https://scholar.google.com.pk/scholar_url?url=https://www.academia.edu/download/91060013/197623623.pdf&hl=en&sa=X&ei=nqysZuigM72Uy9YP74XdwAw&scisig=AFWwaebwHV8AMtq8NDjVgFLdOFx-&oi=scholar)
- Alieksiev, V et al. (2018). BIG DATA AGGREGATION ALGORITHM FOR STORING OBSOLETE DATA. *Industry 4.0*, [online] 3(1), pp.20–22. Available at: <https://stumejournals.com/journals/i4/2018/1/20> [Accessed 2 Aug, 2024].
- BBC Bitesize. (n.d.). Testing - Advanced programming techniques - AQA - GCSE Computer Science Revision - AQA. [online] Available at: <https://www.bbc.co.uk/bitesize/guides/zb33rwx/revision/8>.
- Castellanos, A et al. (2017). Understanding Benefits and Limitations of Unstructured Data Collection for Repurposing Organizational Data. *EuroSymposium on Systems Analysis and Design*  
[https://www.researchgate.net/publication/319341556\\_Understanding\\_Benefits\\_and\\_Limitations\\_of\\_Unstructured\\_Data\\_Collection\\_for\\_Repurposing\\_Organizational\\_Data](https://www.researchgate.net/publication/319341556_Understanding_Benefits_and_Limitations_of_Unstructured_Data_Collection_for_Repurposing_Organizational_Data)
- DuBois, J.M. (2008). Hidden Data for Research Ethicists: An Introduction to the Concept and a Series of Papers. *Journal of Empirical Research on Human Research Ethics*, 3(3), pp.3–5.  
 doi:<https://doi.org/10.1525/jer.2008.3.3.3>.
- Engineering Statistics Handbook (2019). What are outliers in the data? [online] Nist.gov. Available at: <https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>.
- GeeksforGeeks. (2023). Handling Inconsistent Data. [online] Available at: <https://www.geeksforgeeks.org/handling-inconsistent-data/>.

- Kolcz, A et al. (2003). Data duplication: an imbalance problem?. ResearchGate.  
[https://www.researchgate.net/publication/228547112\\_Data\\_duplication\\_an\\_imbalance\\_problem](https://www.researchgate.net/publication/228547112_Data_duplication_an_imbalance_problem)
- Kotu, V & Deshpande, B. (2019). Unstructured Data. Data Science.  
<https://www.sciencedirect.com/topics/mathematics/unstructured-data>
- Kwon, Y., Lemieux, M., McTavish, J. and Wathen, N. (2015). Identifying and removing duplicate records from systematic review searches. *Journal of the Medical Library Association : JMLA*, 103(4), pp.184–188. doi:<https://doi.org/10.3163/1536-5050.103.4.004>.
- Master's in Data Science. (n.d.). How to Deal with Missing Data. [online] Available at:  
<https://www.mastersindatascience.org/learning/how-to-deal-with-missing-data/#:~:text=When%20dealing%20with%20missing%20data.>
- Pham, T. M. Pandis, N. White, I. R. (2024). Missing data: Issues, concepts, methods. *Seminars in Orthodontics* 30 (1): 37:44  
<https://www.sciencedirect.com/science/article/pii/S1073874624000082#:~:text=Missing%20data%20are%20data%20that,which%20missing%20data%20can%20arise.>
- Vailshery, L.S. (2022). BI & analytics software market value worldwide 2019-2024. [online] Statista. Available at: <https://www.statista.com/statistics/590054/worldwide-business-analytics-software-vendor-market/>.
- Wiemer, H., Drowatzky, L. and Ihlenfeldt, S. (2019). Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model. *Applied Sciences*, 9(12), p.2407.