

MDBK DATA ANALYSIS REPORT

DESCRIPTIVE & DIAGNOSTIC (partial) DATA ANALYSIS OF OFWAT FILES

Version

Document number	1
Revision number	1
Effective date	
Owner	
Approver	
Status	Draft

The following individuals contributed to the preparation of this document. For any clarification or questions on this document, please contact any of these individuals.

Name	Role/Title	Telephone Number	Email Address
Reuben Siwela	Analyst	+27677167716	reubensiwela@gmail.com

MDBK DATA ANALYSIS REPORT

Table of Contents

VERSION	1
TABLE OF CONTENTS	2
PURPOSE	3
Target Audience	3
Scope	3
Project Rationale	3
DATA SECTION	4
Data Sets	4
Data Issues & Management	6
METHODOLOGY SECTION	7
ANALYSIS	8
Trend Metrics	8
Households Connectivity	10
Revenue Metrics	12
Total Costs	16
Average Bill Size	19
Council Tax Collection Rate	20
% Defaulted Households	21
OUTCOMES	22
I. Observed Patterns	22
II. What to Draw out	22
SIGN-OFF	25

MDBK DATA ANALYSIS REPORT

PURPOSE

The purpose of the analysis was to perform data analysis on OFWAT files and achieve the following objectives

- identify costs and revenue outliers
- Identify main cost drivers
- Identify any associated correlations and potential insights

Target Audience

Primary: The MDBK management team

Secondary: MDBK targeted market

Scope

Analysis restricted to the 3 files that were provided

Project Rationale

MDBK aims to understand the OFWAT data and elicit insights and themes which can be explored either for thought leadership/white paper publications and or business development.

MDBK DATA ANALYSIS REPORT

DATA SECTION

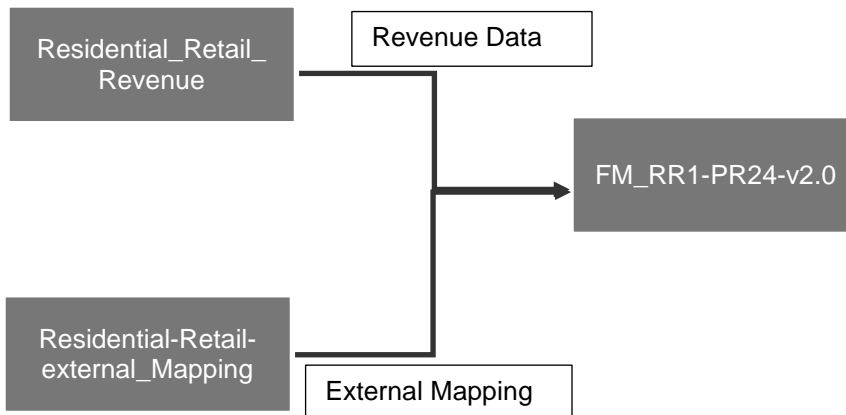
Data Sets

The analysis was based on three spreadsheets from OFWAT

The spreadsheets are

Residential_Retail_Revenue	Revenue model
FM_RR1-PR24-v2.0	Cost model
Residential-Retail-External Mapping	External data Mapping

How the Spreadsheets are Related:



Workbook	Category	Worksheets
Purpose	The Purpose of the Residential_Retail_Revenue Data is to calculate companies outrun historical revenue, which is then used in the calculation of the companies' average bill size in the FM_RR1 file	
Residential Retail Revenue	Input Data tabs	F_Inputs PR 14(Price Review Data) with data from 2011-12 and 2012-2013
		F_Input blind year: Blind year data for years 2013-14 and 2014-15
		F_Inputs cyclical: Historical company Annual Performance Report (APR) data with data from 2015-16 to 2021-2022
	The Transformed Data sheet	Merge Revenue
	Outputs	Revenue

The data in the three input tabs depicts the revenue generators shown in the 3 tables below. The revenue generator names are grouped into 3 periods reflecting the period they covered. The periods are FY12 to FY15, FY16 to FY20 and FY21 to FY22

MDBK DATA ANALYSIS REPORT

Reference	Description	Period
BR581_PR14	Water: Unmeasured - household	FY12 to FY15
BR881_PR14	Sewerage: Unmeasured - household	FY12 to FY15
BR582_PR14	Water: Measured - household	FY12 to FY15
BR882_PR14	Sewerage: Measured - household	FY12 to FY15

Reference	Description	Period
R3017TR	Unmeasured water only customer - Total revenue £m	FY16 to FY20
R3018TR	Measured water only customer - Total revenue £m	FY16 to FY20
R3019TR	Unmeasured wastewater only customer - Total revenue £m	FY16 to FY20
R3020TR	Measured wastewater only customer - Total revenue £m	FY16 to FY20
R3021TR	Unmeasured water and wastewater customer - Total revenue £m	FY16 to FY20
R3022TR	Measured water and wastewater customer - Total revenue £m	FY16 to FY20

Reference	Description	Period
B0072TRR	Residential revenue - Total residential revenue - Revenue	FY21 to FY22

Workbook	Category	Worksheets	Description
Its purpose is to Collate historical cost and cost driver information for residential retail			
The Feeder Model Residential Retail 1 file (FM_RR1-PR24v2.0)	Input	Item dictionary	Provides bon codes, labels, units, and calculation notes for the data in this dataset.
		F_Inputs	Pulls the raw historical costs and cost drivers data from Ofwat's database.
		F_inputs cyclical	Historical cost and drivers' data (2013-14 to 2021-22)
		F_inputs PR19	Raw historical and forecast data (up to 2024-25) from PR19 Business Plan Data Tables.
		External data	Provides external data for Equifax, migration, income score, council tax collection rate, revenue and CPIH.
	Processing Sheets	Input overrides	Overrides the data where required (e.g. mergers, PR14 blind year, random errors in data).
		F_inputs cyclical corrected	Combines inputs sheet and overrides.
		Depreciation	Calculates depreciation data.
		Recharges	Calculates net recharges (costs minus income).
	Output sheets	nom statafile	Collates the data and calculates the dependent and independent variables required for modelling. This is the master dataset to be used in Stata, costs in nominal terms.
		real statafile	Transforms data into real prices (2017-18 CPIH price base).

MDBK DATA ANALYSIS REPORT

Data Issues & Management

1. Excluded Revenue Generators

The following revenue generators pertaining to the two companies were excluded for comparative purposes

Company	Reference	Description	Period
SVT	BR986	Wholesale Total - Household	2018-19
DVW	BR986	Wholesale Total - Household	2018-19

2. There was no revenue generators data prior to the financial years 2015-16

3. Company merges

- Merge Revenue tab had 26 companies
- Nom Stata tab and Revenue Tab had the same 22 companies

Combination	Application	Notes
SST +CAM = SSC	Norm Stata Tab,Revenue Tab,Merge Revenue Tab	
VCE + VEA +VSE =AFW	Norm Stata Tab,Revenue Tab,Merge Revenue Tab	From FY12 to FY15
BWH +SWT = SWB	Norm Stata Tab,Revenue Tab,Merge Revenue Tab,External Tab	BHW & SWT were already added to give SWB figures
DVW +SVT +HDD =SVE	Norm Stata Tab,Revenue Tab,Merge Revenue Tab, External Tab	From FY19 to FY 22 only used SVE figures (does it already include the others?)

- Following the combinations 16 companies remained.

MDBK DATA ANALYSIS REPORT

METHODOLOGY SECTION

- Descriptive and Diagnostics analysis
- Focused on outliers, drivers, correlations, and any emerging patterns(themes)
- Tools used
 - Power BI
 - IBM Cloud
 - Excel

MDBK DATA ANALYSIS REPORT

ANALYSIS

- In total 16 companies were analyzed

TMS	YKY	SWB	SSC
NWT	SRN	WSX	BRL
SVE	WSH	AFW	SES
ANH	NES	SEW	PRT

- Key metrics analyzed are indicated below

Metric	Analysis Period
Revenue Trend -aggregated	FY14 to FY22
Revenue per company	FY12 to FY22
Revenue driver contribution	FY12 to FY22
Total Costs	FY14 to FY22
Household connected	FY14 to FY22
Household connected by driver	FY14 to FY22

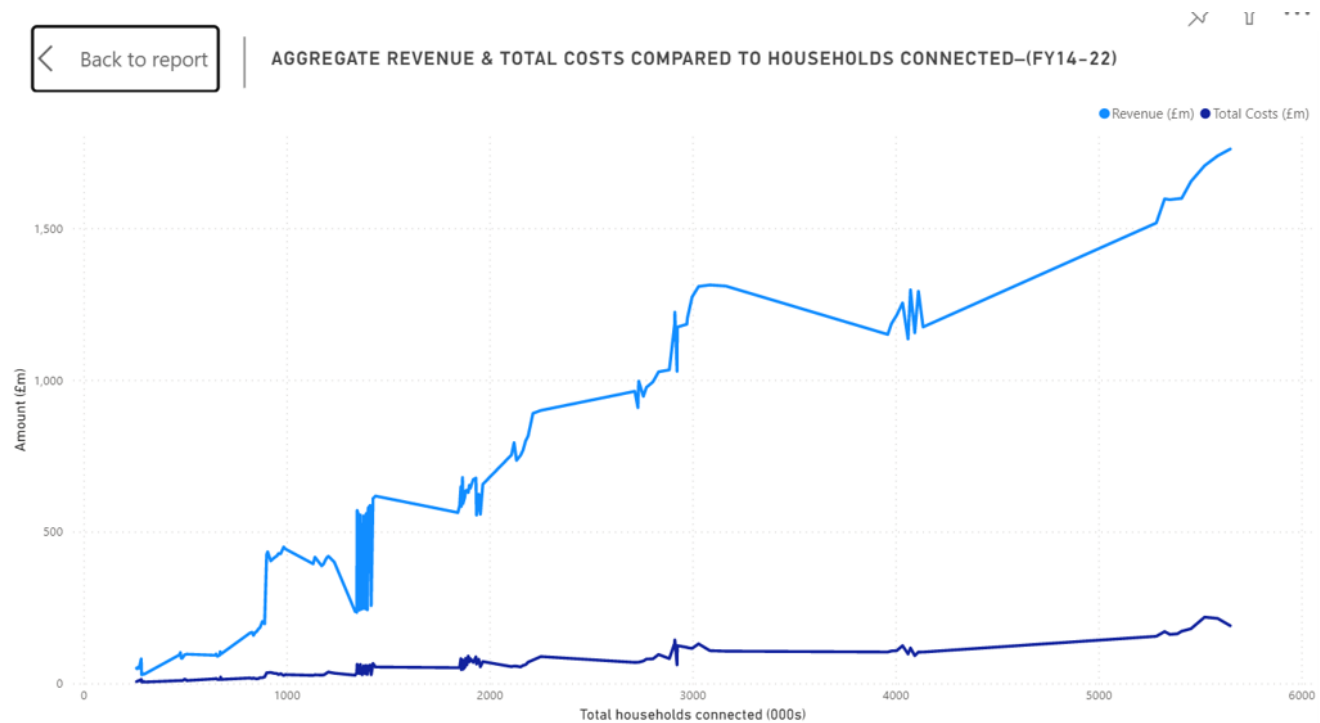
Trend Metrics



- An upward trend observed from FY14 to FY22.

MDBK DATA ANALYSIS REPORT

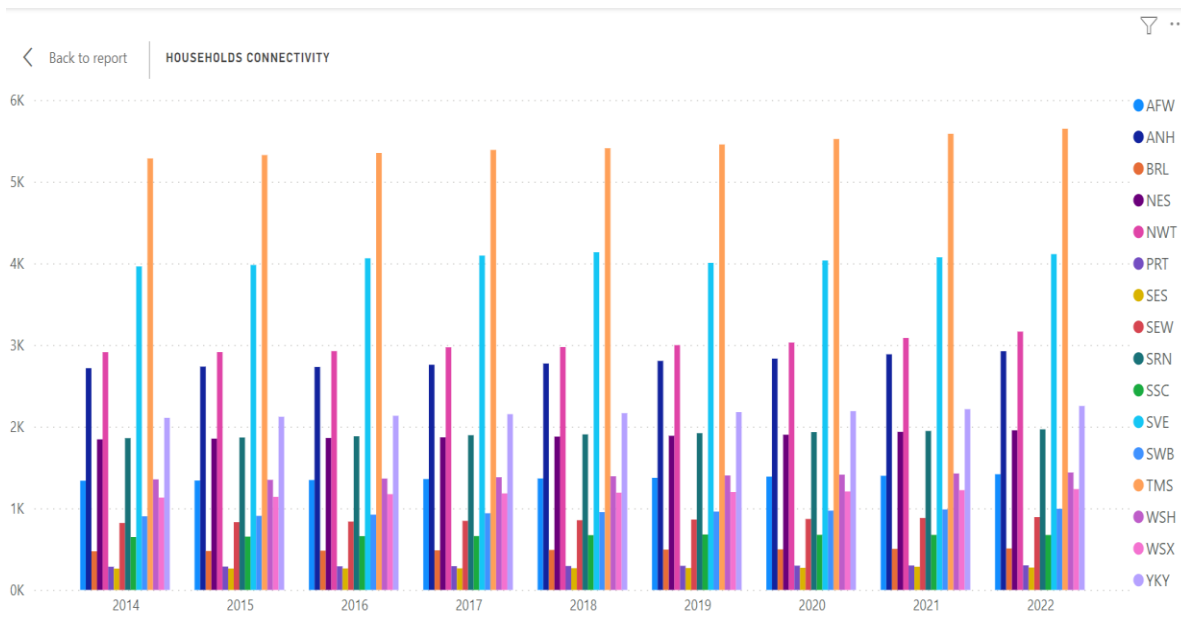
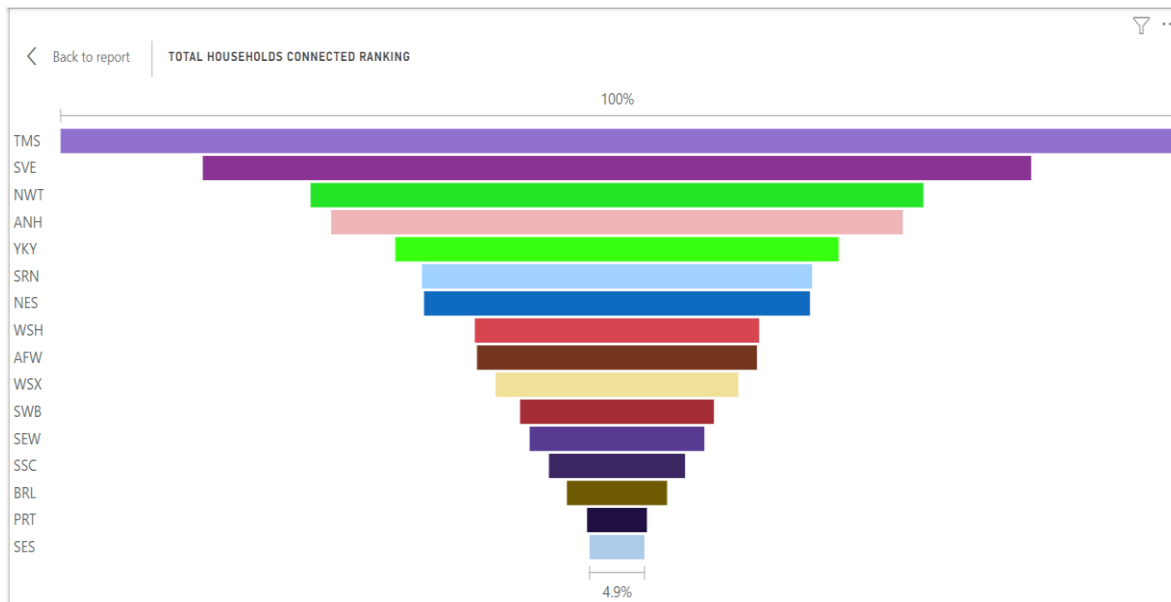
- Significant downward trend observed from FY15 with a lowest revenue recorded in FY16



- Aggregate Revenue has increased in line with Total Household Connected
- Revenue increases generally upward and more aggressive
- Total costs maintain an upward slop but within restricted regime and costs increases are subdued resulting in huge margins

MDBK DATA ANALYSIS REPORT

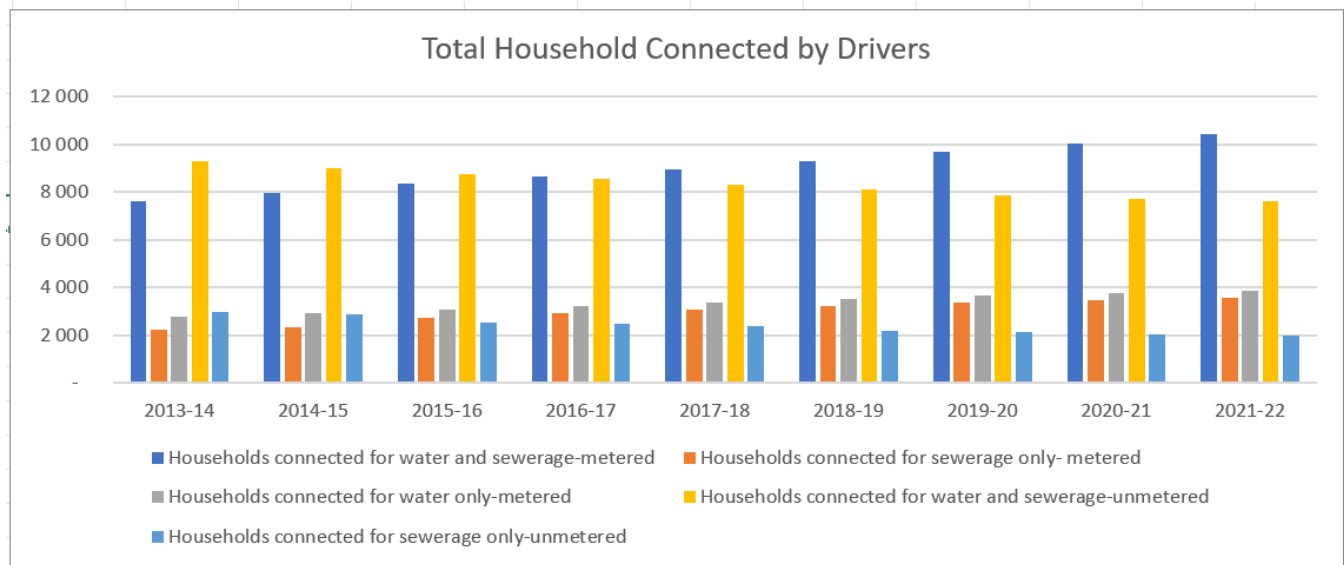
Households Connectivity



Top 5 Companies Connected	Bottom 5 Companies Connected
TMS	SES
SVE	PRT
NWT	BRL
ANH	SSC
YKY	SEW

MDBK DATA ANALYSIS REPORT

Total Household Connected by Connectivity drivers



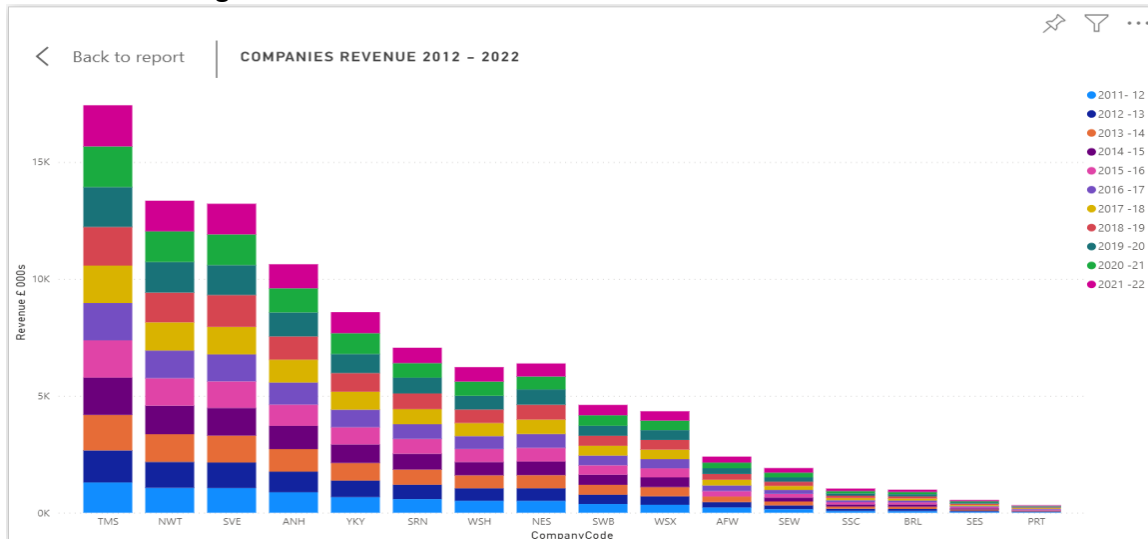
Household Connectivity as a Revenue Generators - Sorted by most Revenue generator

1	Household connected for water and sewerage - metered
2	Household connected for water and sewerage - unmetered
3	Household connected for water only- metered
4	Household connected for sewerage only- metered
5	Household connected for sewerage only- unmetered

MDBK DATA ANALYSIS REPORT

Revenue Metrics

Revenue Ranking



- The Ranking shows TMS as the leader in Revenue and an outlier
- TMS has maintained dominant revenue position through the years

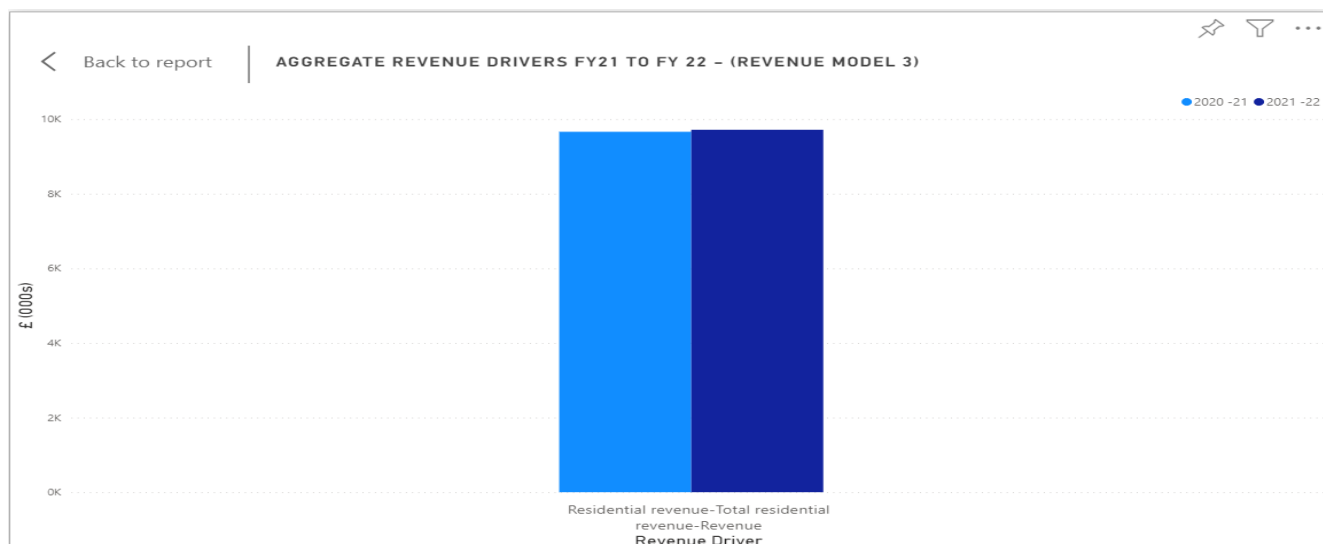
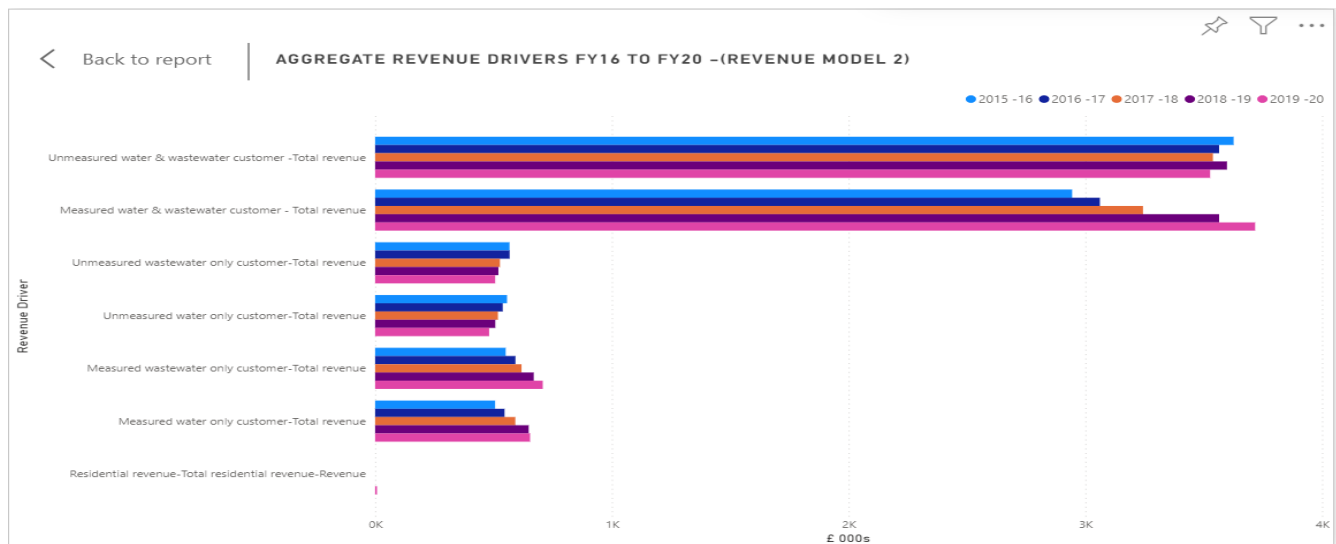
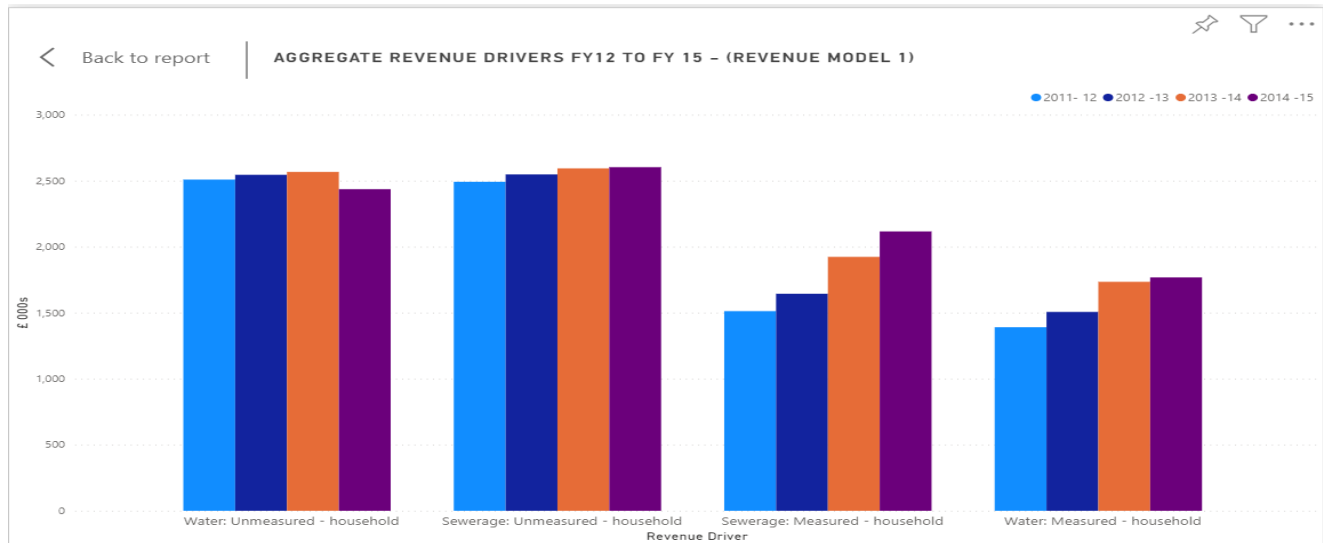
Revenue by Revenue Generating Driver

3 Revenue Generating models were identified based on the periods they covered.

The revenue generating models were used in

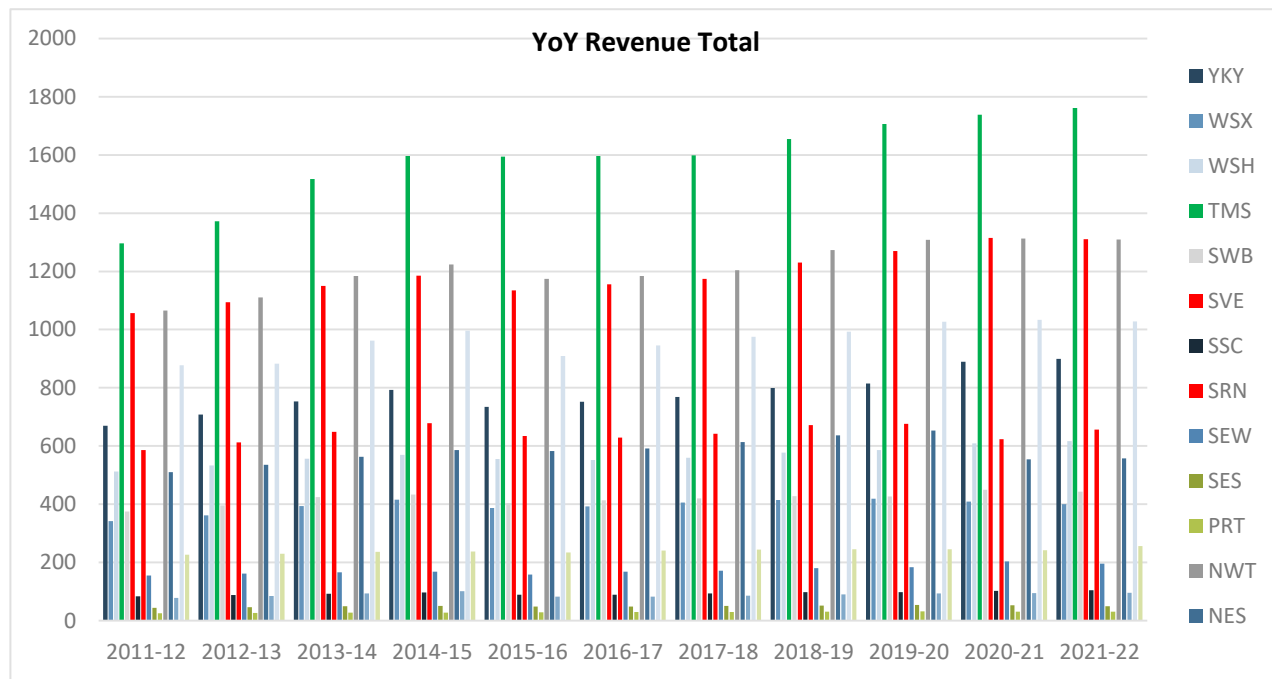
- FY12 to FY15
- FY 16 to FY20
- FY 21 to FY22

MDBK DATA ANALYSIS REPORT



MDBK DATA ANALYSIS REPORT

Revenue Regime Bands

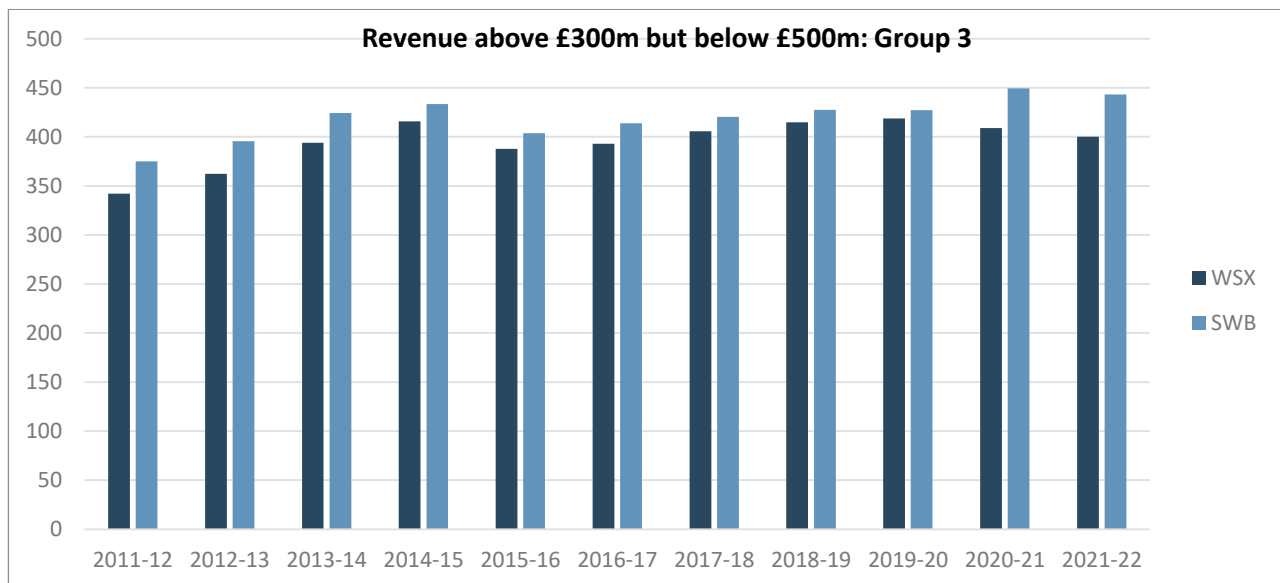
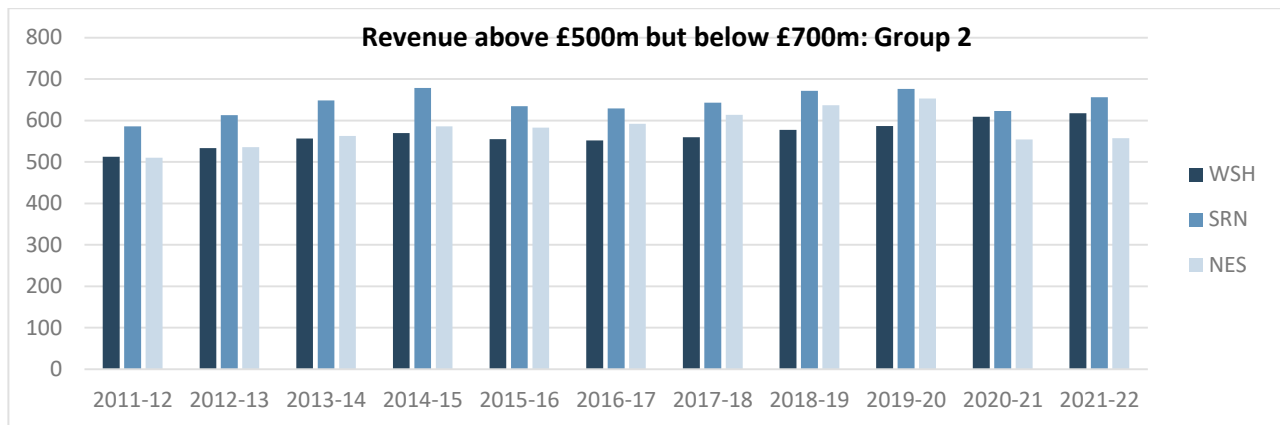
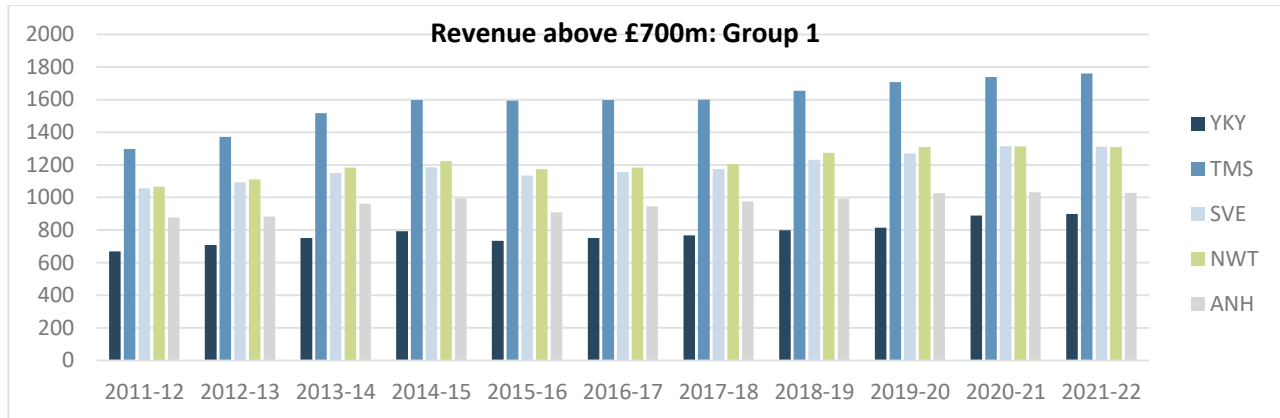


Revenue Band	Company	Count
Under 300m	SSC, SEW, SES, PRT, BRL, AFW	6
Above 300m & Under 500m	WSX, SWB	2
Above 500m & Under 700m	NES, WSH, SRN	3
Above 700	ANH, NWT, SVE, TMS, YKY	5

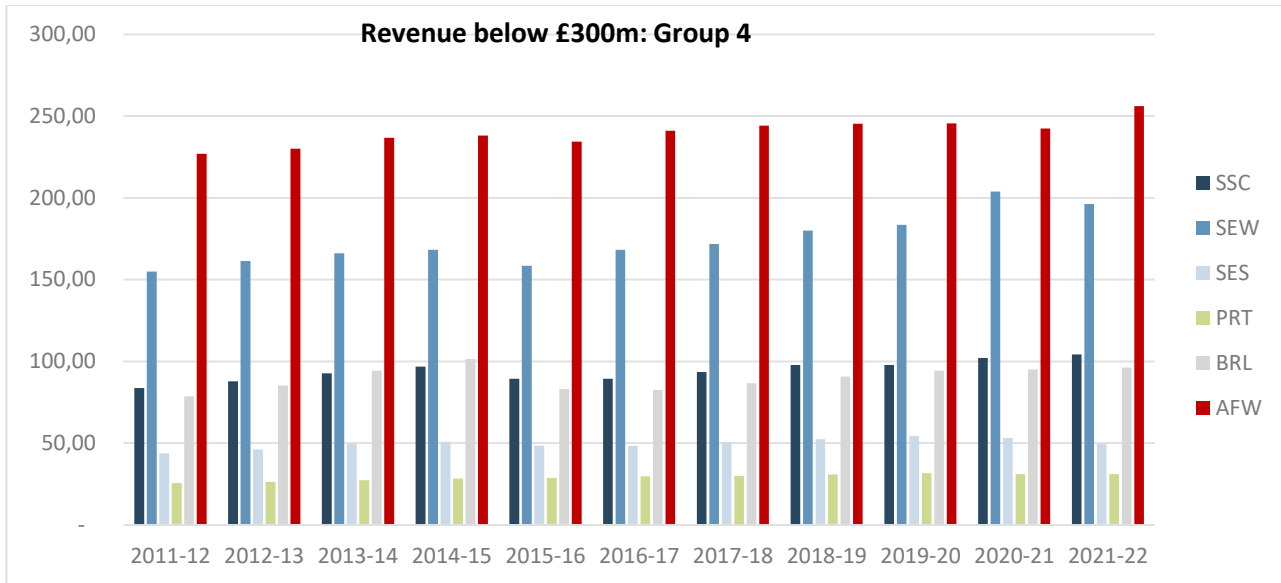
The grouping of households connected bands are shown in the table below

Population Band	Company	Count
Under 1m	BRL, SES, SEW, SWB, PRT, SSC	6
Above 1m but under 2m	AFW, NES, SRN, WSH, WSX	5
Above 2m but below 3m	ANH, YKY	2
3m and above	SVE, TMS, NWT	3

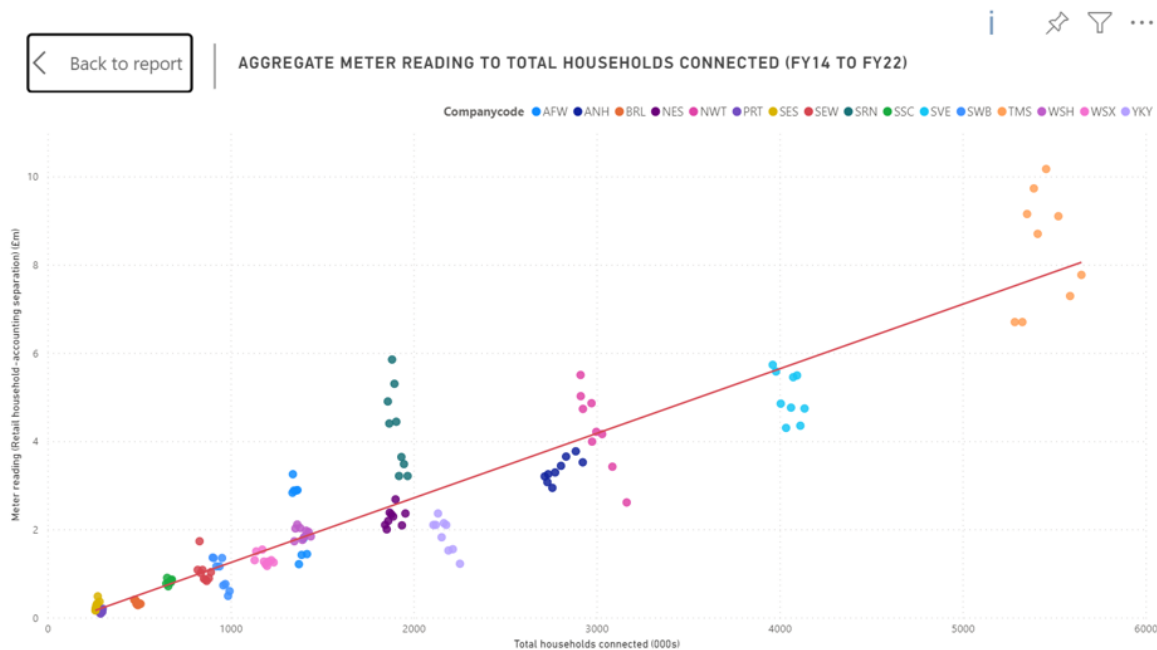
MDBK DATA ANALYSIS REPORT



MDBK DATA ANALYSIS REPORT

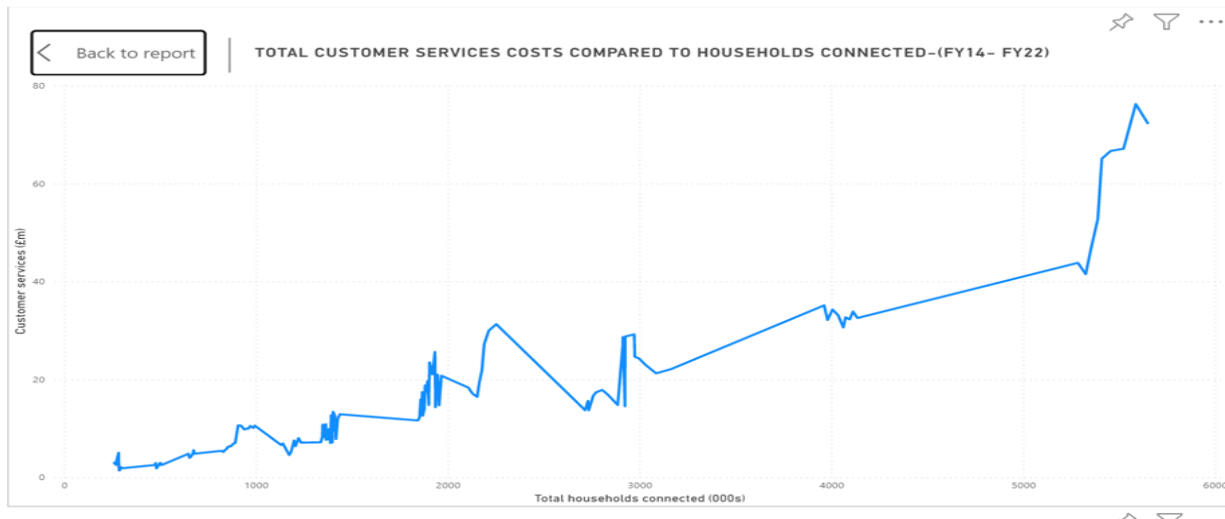
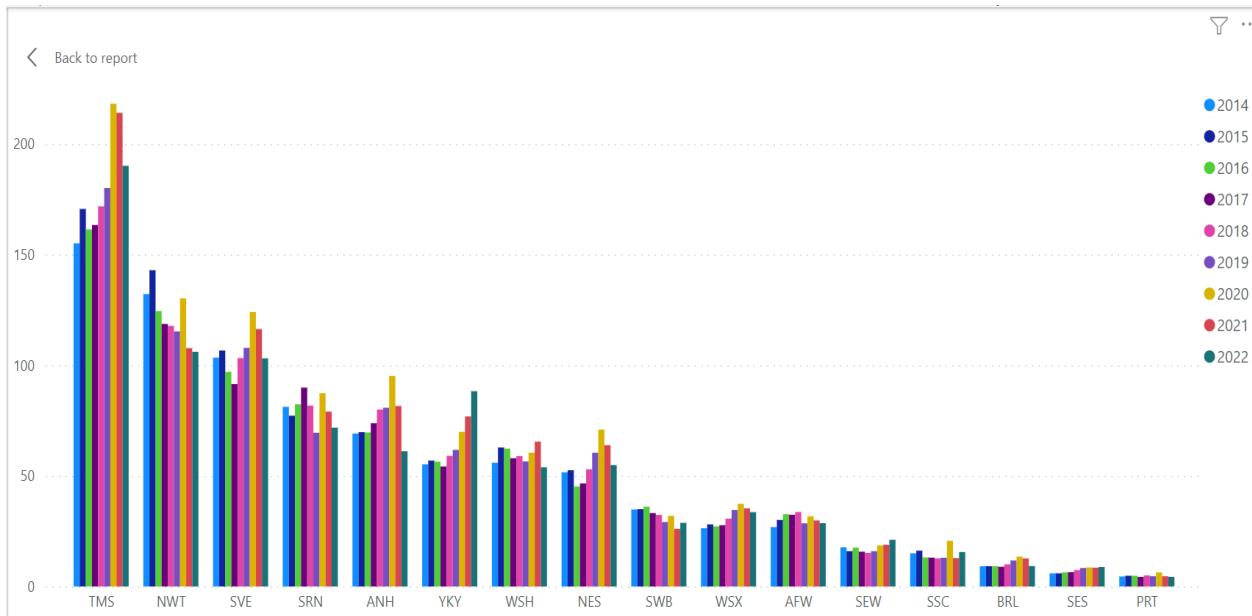


Total Costs



MDBK DATA ANALYSIS REPORT

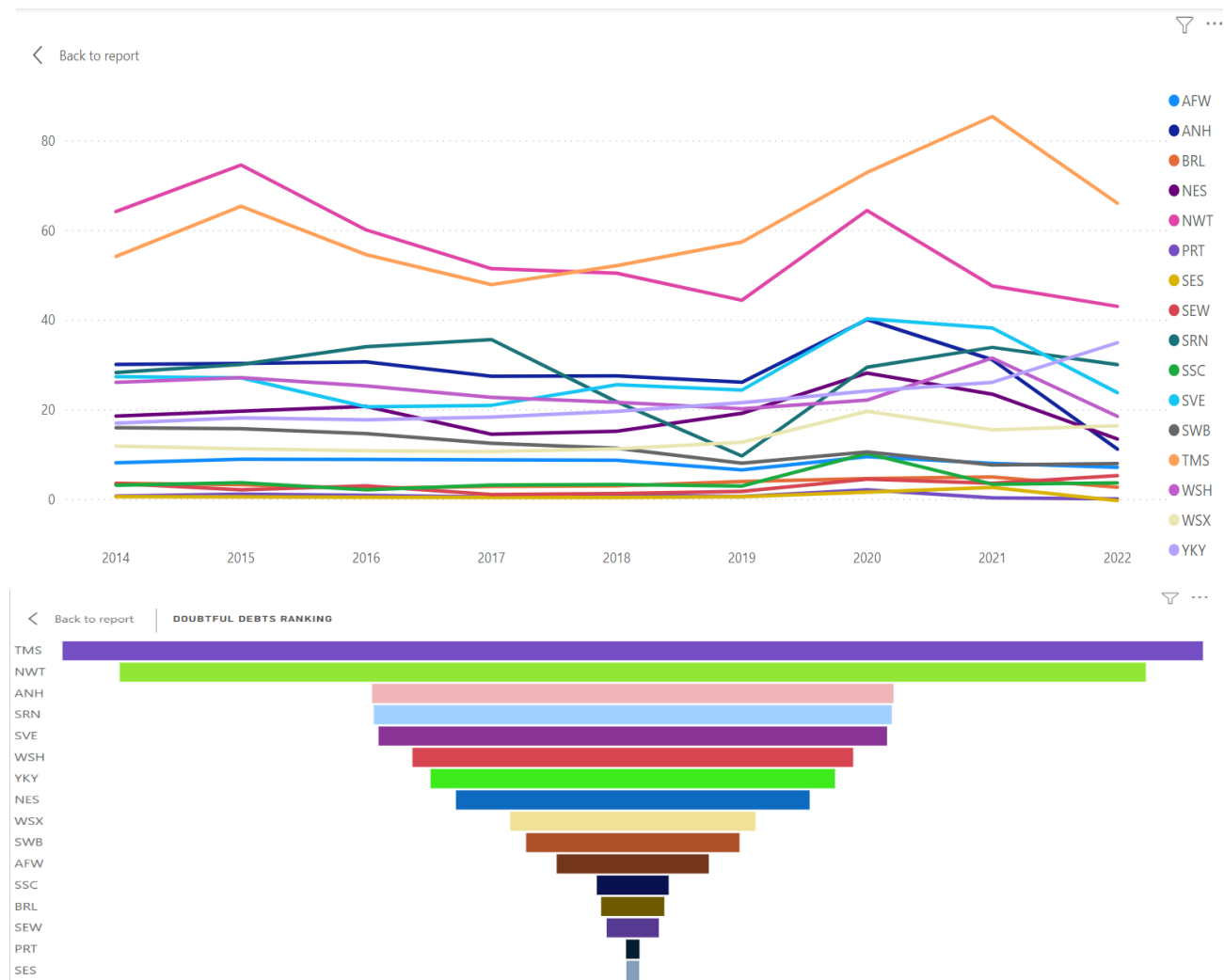
YoY Total Costs



MDBK DATA ANALYSIS REPORT

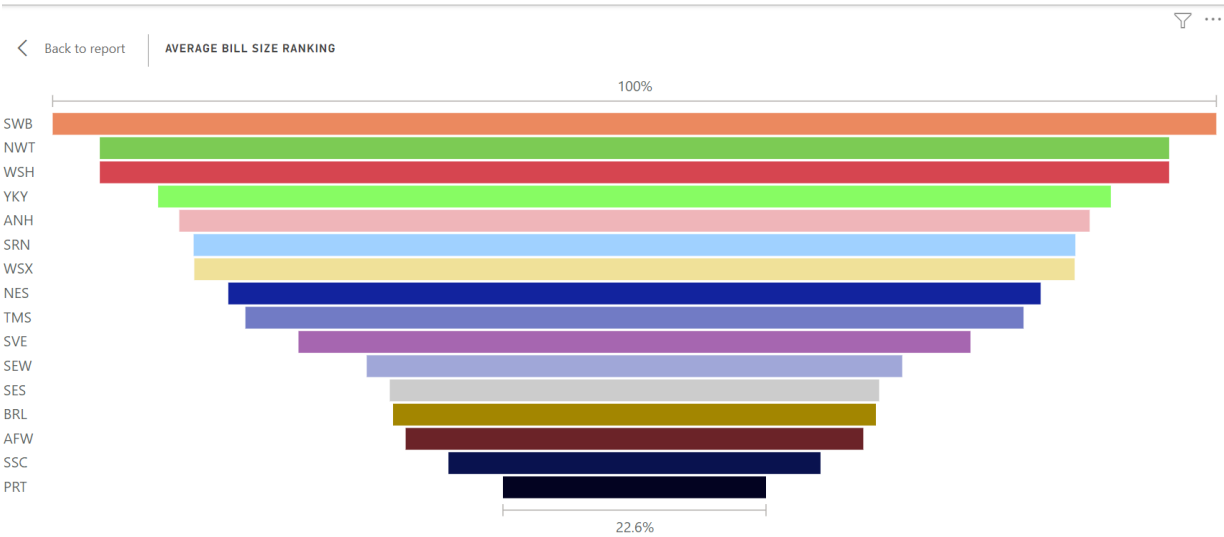
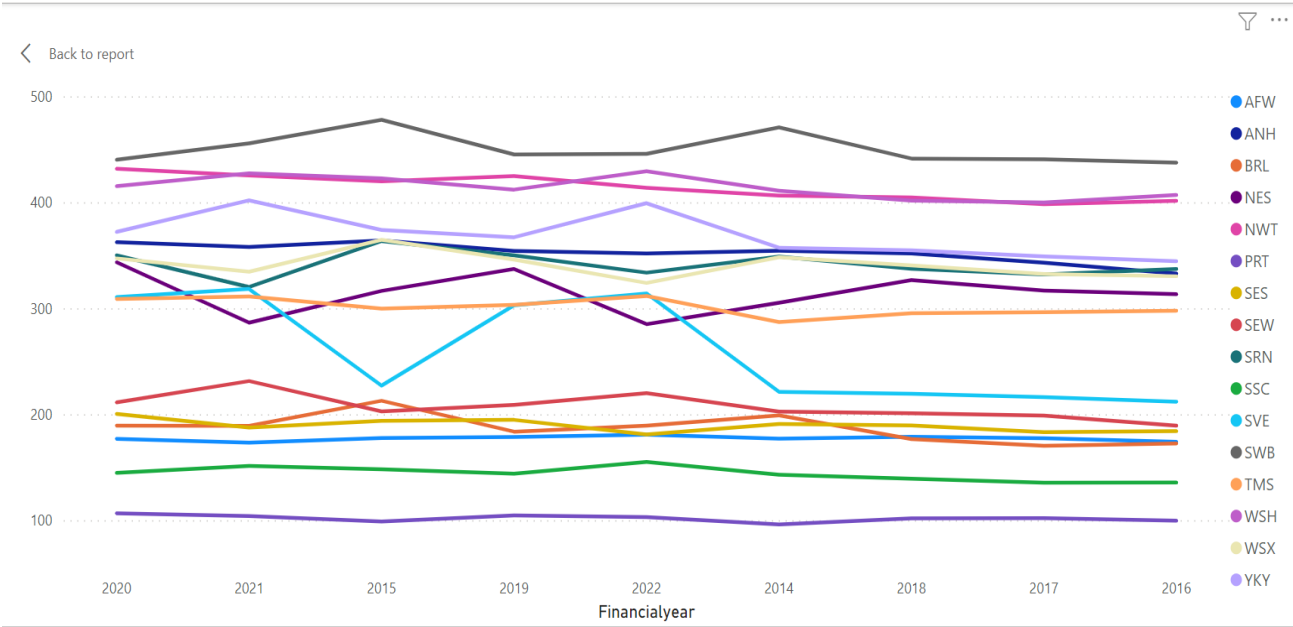
Doubtful Debts

Increase in Doubtful debts: *(Are these made up of IFRS 9 expected credit losses)*



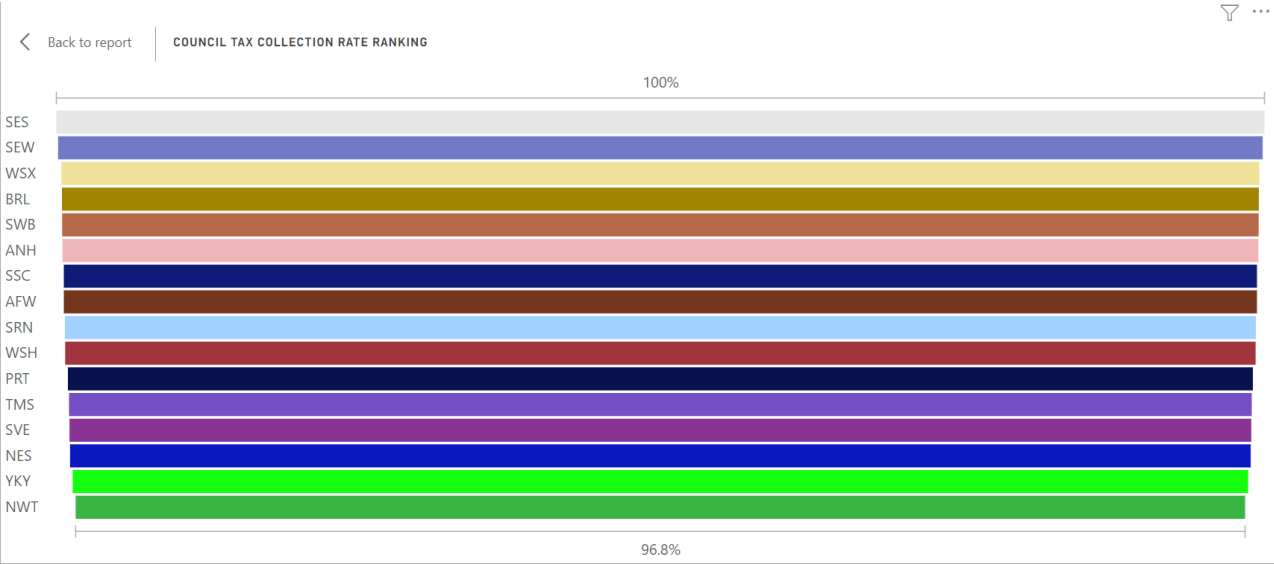
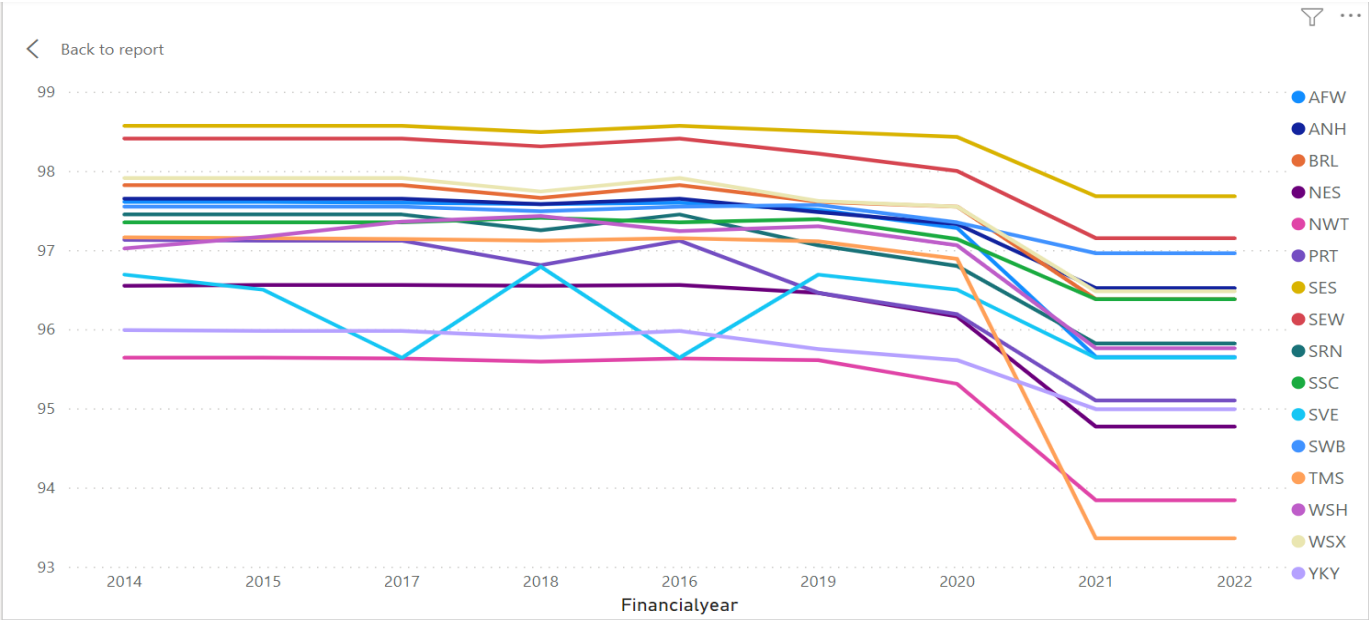
MDBK DATA ANALYSIS REPORT

Average Bill Size



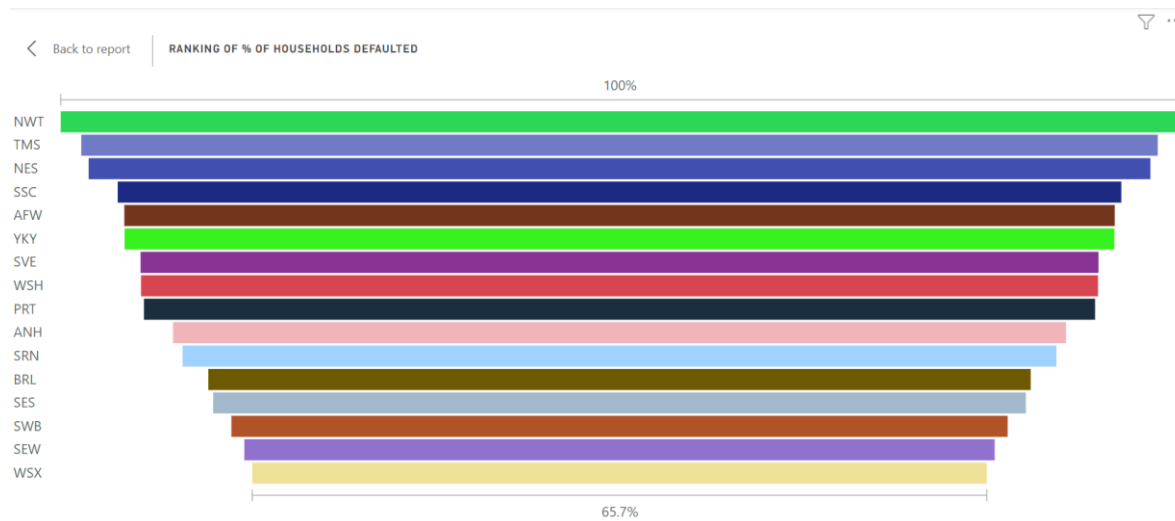
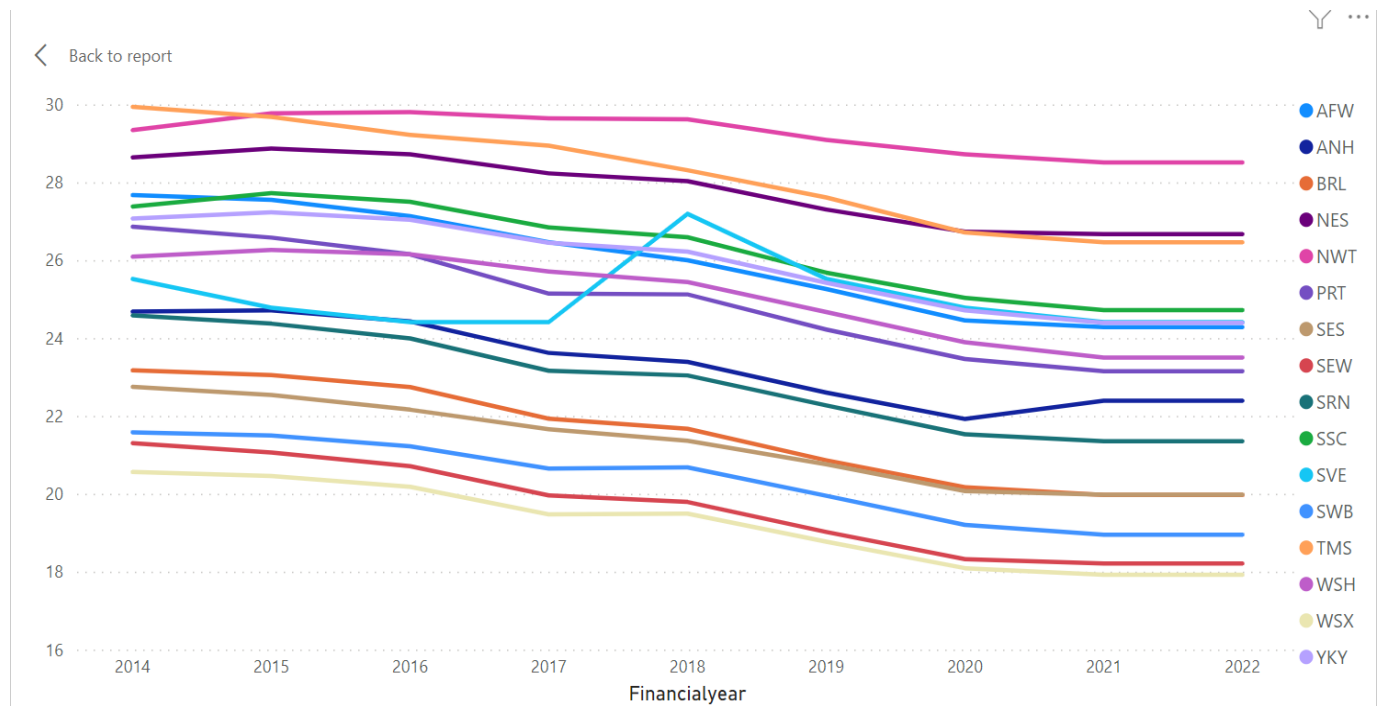
MDBK DATA ANALYSIS REPORT

Council Tax Collection Rate



MDBK DATA ANALYSIS REPORT

% Defaulted Households



MDBK DATA ANALYSIS REPORT

OUTCOMES

I. Observed Patterns

From the various visualizations and tabulations, below are some of the patterns we observed, and in section II that follows we provide some potential initiatives that can be extended from the observations

Item	Comment
Revenue	Ranking order generally consistent over the periods
Household Connected	Ranking order generally consistent over the periods
Revenue size	This was generally proportional to the household connected size
Total Costs	Doubtful Debts and customers services costs were main costs drivers under total costs.
Household Connected for Water and sewage-unmetered	This was the main connectivity driver from FY14 to FY16
Household Connected for Water and sewage-metered	This was the main connectivity driver from FY17 to FY22
% Defaulted household	The order ranking followed a different pattern to the general order ranking due to households connected and Revenue size
Doubtful Debts	The Ranking order generally followed the ranking order pattern of the Household Connected/Revenue size of the company.
Council Tax Collection Rate	The order ranking did not generally follow the Company size in terms of Household connected and Revenue earned.
Average bill	The order ranking for Average bill did not generally follow the Company size in terms of Household connected and Revenue earned.

II. What to Draw out

A. Doubtful Debts

Doubtful Debts Ranking

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
TMS	NWT	SVE	ANH	SRN	YKY	WSH	NES	WSX	SWB	AFW	SSC	BRL	SES	SEW	PRT

There was a notable increase in doubtful debts from which coincided with the introduction of IFRS 9 Accounting Standard on Financial Instruments which was effective from 1 January 2018. One of the

MDBK DATA ANALYSIS REPORT

major requirements of the standard is that companies with a debtor's book must make a predictive recognition of expected loss (which form the figure for doubtful debts). The standard replaced IAS39 was an incurred loss model meaning, doubtful debts were only recognized when a debtor started to show signs of financial distress. The new standard requires that expected losses be booked for all trade receivables meaning the companies now need to actively monitor and, in some instances, manage the credit standing of their clients to avoid a deterioration in credit quality which cause the company to end up booking more expected losses.

Problems

- If not properly modeled various risks arise
 - Increase in bad debts
 - Wrong expected loss amount booked hence misrepresenting the costs of doubtful debts
- Proactive Client Information Management
 - If no proactive client account management (gleaning insights from account behavior-payment and others), doubtful debts, and bad debts increase
 - Not proactively engaging with client data poses a missed opportunity to take remedial action before the client credit quality deteriorates.

Potential Opportunities

- Software as a Service (SaaS) - Online Expected Loss calculator
- Model Validation of the expected loss models if companies already have these in place
- Development of Predictive Credit Early Warning Systems (another SaaS)

B. Efficiency

Using average bill as proxy measure of efficiency

(It would have been good to have an industry average or benchmark metric)

The average bill ranking is as follows

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
SWB	WSH	NWT	YKY	ANH	WSX	SRN	NES	TMS	SVE	SEW	SES	BRL	AFW	SSC	PRT

This table is contrasted with the two tables below

The Revenue Ranking

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
TMS	NWT	SVE	ANH	YKY	SRN	WSH	NES	SWB	WSX	AFW	SEW	SSC	BRL	SES	PRT

Household Connected Ranking

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
TMS	SVE	NWT	ANH	YKY	NES	SRN	AFW	WSH	WSX	SWB	SEW	SSC	BRL	PRT	SES

MDBK DATA ANALYSIS REPORT

Comment

The Ranking pattern established by the Revenue and Household Connected Data is breached when it comes to the average bill ranking. The Revenue and Household connected order ranking is significantly breached suggesting efficiency and revenue collection capacities in the absence of other explanations.

Potential Opportunities

- Reconciling existing accounts
- Establish gaps in existing capacities to improve efficiency
- Revenue leakage or improvement solutions

Council Tax Collection Rate

This is another metric that can provide more information about the properties and collection efficiency depending on available information.

Council Tax Collection Rate order ranking

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
SES	SEW	SWB	WSX	BRL	AFW	ANH	SSC	SRN	WSH	SVE	PRT	YKY	NES	NWT	TMS

Whilst the levying of council tax is different in various areas, it also presents a validation business opportunity as it is affected by many factors including:
size, layout, character, location, change in use, value etc.

C. Bad Debts

Ranking of % Households Defaulted

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NWT	TMS	NES	SSC	AFW	YKY	SVE	PRT	WSH	ANH	SRN	BRL	SES	SWB	SEW	WSX

Comments

The order ranking shows that some of the big companies by Revenue and Household connected are also leaders in having a high number of defaulted clients. TMS and NWT top the Revenue ladder. SWB which tops the average Bill ranking and is in the top 3 on the Council Tax Collection Rate ladder, is in the bottom 3 in terms of having a subdued incident of defaulted clients.

Problem

If Bad Debts escalate and not managed properly, this can lead to more defaults resulting in liquidity challenges.

Potential Opportunities

- Bad Debt Management

MDBK DATA ANALYSIS REPORT

- Understand risk factors and unpack remedial action

Conclusion

Whilst we tried to suggest potential initiatives that can grow from the themes we observed in the data, there would be need for further information to ensure that our suggestions can indeed lead to initiatives that are worth following up on.

Sign-off

MDBK DATA ANALYSIS REPORT

Project Coordinator Date

Stakeholder Date

Stakeholder Date

Stakeholder Date

Stakeholder Date
