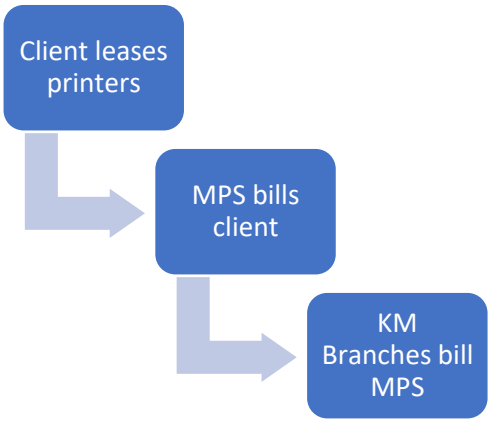
Data Handling Process

By Tufail Ismail

This presentation showcases the meticulous process I undertook to develop comprehensive reports and interactive dashboards tailored specifically for each of KMSA's MPS clients on a monthly basis. It is important to note that the data presented in this demonstration is mock data and for illustrative purposes only.

For a little context, KMSA would lease printers to clients which would include support and service. Support was provided by MPS staff, which I managed, and my team would administer service from KMSA branches and dealers around SA. For the service and support delivered by KMSA, MPS would charge the clients for the prints they made, and since each client had 100 plus devices spanning across the country, it made most sense for MPS to total up all printer usage, invoice the client and then have each branch invoice MPS. Thus, MPS was compelled to supply clients and branches with statistics and analysis.



I split this presentation is split into three parts.

The first part is based on a report that I produce monthly between the 25th and the last day of each month. It entails collecting printer usage data from several sources, importing it into a spreadsheet database, then summarizing the data to create a billing report. I would then use the data collected to fill in a template that would enable me to upload the data into SAP.

In the second part we look at data that is collected from the print tracking software that records each user's print, copy and scan transactions. It includes user and printer data with metrics that we will use for analysis. This data comes in the form of a csv or by extracting the data off a SQL database.

In the third part, I highlight important reports that were instrumental in investigating and identifying the root causes of various issues. These reports played a crucial role in conducting inquiries and gaining an understanding of the problems at hand.

Part 1

The process begins by collecting meter readings off the printers. We used a sniffing tool that scans the IP addresses of network printers and provides us with serial numbers and meter readings. This report included extra details and serial numbers with a prefix that was not listed on our records when captured, therefore to use a VLOOKUP to import the data into my spreadsheet, I would need to clean the data. This was possible with the “find and replace function” in excel. I would also need to filter out data where the dates weren’t current and convert the Total, Color and BW fields to numbers.

Before

Date	Re*istere* na*e	Serial Number	IP Address / DNS Host Name	Total	Color Total	Black Total	MAC Address	Model Name
03/02/2020 09:18:03	*ouw*ar*ia *RC	A5C2**11*7659	1*2.26.40.1*8	283,410	123,959	159,451	00:20:6*:98:E6:7*	KONICA MIN
03/08/2020 00:32:28	*ouw*ar*ia *RC 01	A5C1**1*16927	1*2.26.41.20*	127,902	44	127,858	00:20:6*:A9:7C:D8	KONICA MIN
04/20/2020 04:33:47	*ouw*ar*ia *RC 03	A5C1**1**6*77	1*2.2*.4.49	260,662	39,092	221,694	00:20:6*:8C:D4:74	KONICA MIN
04/20/2020 04:33:48	*RC *ouw*ar*ia	D**3***556	1*2.26.41.209	869,351	282,673	586,678	00:20:6*:67:7A:*E	KONICA MIN
04/20/2020 04:33:20	Icon	D**2*17173	1*2.26.214.18	355,712	4,247	351,465	00:20:6*:69:92:07	KONICA MIN
04/20/2020 04:33:45	C220 Port E*iza*eth2	D**3*1*562	1*2.26.214.80	541,751	37,448	504,303	00:20:6*:6D:3*:53	KONICA MIN
12/18/2019 16:44:40	Po**e* – Inte*rate* Care R*P	D**3**977*	1*2.2*.5.96	324,416	15,508	308,908	00:20:6*:6C:32:*2	KONICA MIN
04/20/2020 04:33:45	Po**e* *ouw*ar*ia	D**3**3665	1*2.2*.5.100	731,710	175,551	556,159	00:20:6*:6A:*E:DE	KONICA MIN
04/20/2020 04:33:29	Po**e* – Ca** Centre *ouw*ar*ia	D**3**37*8	1*2.2*.5.101	791,684	69,439	722,245	00:20:6*:6A:68:*0	KONICA MIN
04/20/2020 04:33:47	ScriptFar* *ouw*ar*ia	A5C1**1*16581	1*2.2*.*.5*	230,509	92,282	138,227	00:20:6*:A9:61:2*	KONICA MIN
04/20/2020 04:33:36	RAF1 A*S Site Centurion	M**1***361	1*2.2*.*.51	378,529	38,993	339,536	00:20:6*:84:A0:*8	KONICA MIN
04/20/2020 04:33:42	Centurion He*los Key Accounts R*P	D**3*1*533	1*2.2*.25.221	174,499	9,715	164,784	00:20:6*:6D:3C:A9	KONICA MIN

After

Date	Regis*ere*Na*e	SerialNumber	BlackTotal	ColorTotal	Total
04/20/2020 04:33:16	A*A *rea**en* Suppor*	*0*3013793	194,02	16,766	210,786
04/20/2020 04:33:16	*ea*** In*e**ence Uni*	*0*3013767	415,496	129,772	545,268
04/20/2020 04:33:16	*ega*	*0*301*0**	194,017	75,172	269,168
04/20/2020 04:33:16	*e**ea*** *raining	*0*301**39	249,915	38,371	288,286
04/20/2020 04:33:17	***	*0*301**16	230,265	40,851	271,116
04/20/2020 04:33:17	*usiness *eve*op*en*	*0*301377*	499,019	186,115	685,134
04/20/2020 04:33:17	A*A	*0*3010*38	200,548	48,79	249,338

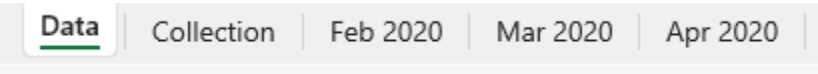
Next I would import that data into my usage spreadsheet.

My “usage” spreadsheet was comprised of a data sheet, that listed all devices, their IP addresses, Location, Model, Serial number, and BW and Color usage for every month.

I used a “collection” sheet to paste the values collected from the sniffing report, I don’t like to import data from external sources as it can be a problem if that document changes or moves from its directory.

And finally a summary sheet that summed up usage for the month and auto updated values for billing. I used color to separate and distinguish the different areas, and conditional formatting to identify outliers.

“Usage Spreadsheet” tabs



“Data sheet”

				Jan-20		Feb-20		Mar-20		Apr-20	
IP Address	Location	Model	Ser. No.	B&W	Color	B&W	Color	B&W	Color	B&W	Color
1**.**.*10.**9	K*e*ks*o**-*e*icove* *ui**in*	C35	A121**1**91*5	149837	5452	149837	5452	149837	5452	149837	5452
1**.**.*09.*10	*TA- *onitas-A*ca*ia 0	C224e	A5C***1111652	360825	5144	364722	5158	368666	5182	370168	5283
1**.**.*09.1*	*o*okwane -Checke*s Cente*	C35	1**1**915*	347840	4193	353442	4201	359082	4215	360763	4288
1**.**.*0*.11	O** *utua*- Sc*a**e*	C35	1**1**8315	94426	42173	94426	42173	94426	42173	94426	42173
1**.**.*10.133	San*a*- Sc*a**e*	C35	1**1**6*67	24093	41273	24093	41273	24093	41273	24093	41273
1**.**.*14.3*	*TA -**C, Ne**ank **aza	C3350	A*Y***1**1523	7687	7481	7687	7481	7687	7481	7687	7481
1**.**.*09.1**	-Ki**e**y	C3350	A*Y***1**3676	92113	1342	93615	1349	94403	1349	94932	1375
1**.**.*10.99	Ne*s**uit-unit Squa*e	C3350	A*Y***1**3713	87238	1906	89080	1906	90972	1906	91421	1906

“Summary Sheet”

		For the Month		Feb	2020
	Printer Name	Serial Number	BW	COLOUR	TOTAL
JHB	*o*it*s-***se 2 1 *o*t*	**89*21***161	3 443	0	3 443
	*o*it*s-***se 2 1	**89*21******	63 225	0	63 225
	*e*s *ew	**2K*21**8368	20 997	1 006	22 003
	V**-**i*	**89*21****89	4 100	0	4 100
	**ock * 1F	**89*21***119	3 304	0	3 304
	Co**o**te Sc*e*es-***se 1 * Sout*	**89*21***132	2 198	0	2 198
	*os*e*-***se 2 -1 Sout*	**89*21***13*	10 745	0	10 745
	**i* *oo*	**89*21**32*6	3 537	0	3 537
	*e*ios-***se 1 1 *o*t*	**89*21***28*	1 994	0	1 994
	*o*it*s-***se 2 1 Sout*	**89*21***1*2	11 625	0	11 625
	Fe**e**t*-**ock F 1st f*oo* *o*t*	**89*21***323	2 694	0	2 694
	y*o-**ock * * E*st	**R**21*38328	13 812	2 709	16 521
	*e*ios T*e **ee*s	D**3*1**33	843	29	872
VRG	***i*ist**tio* V**-V**	D**3*1**12	3 472	0	3 472
	I*te*ve*tio* V**-V**	D**3*1**21	2 359	9	2 368
	ioo* V**-V**	D**3*1**2*	26 586	7 736	34 322
	#*/*	***1***1*3	27 403	4 471	31 874
	C**tu*i** V**-V**	**89*21***118	1 484	0	1 484
	C*ie*t Se*vices V**-V**	**89*21****93	4 918	0	4 918
	Qu**ity *ssu***ce V**-V**	**89*21***129	641	0	641
KZN	C*** Ce*te* 1 KZ*-KZ*	D**3*1**1*	15 638	4 215	19 853
	ioo* KZ*-KZ*	D**3*1**28	7 232	255	7 487
	C*O *e*t. KZ*-KZ*	D**3*1****	119 814	25 435	145 249
	W**ki*s*e*t KZ*-KZ*	D**3*1****	119 814	25 435	145 249
	C*** ce*t*e 2 KZ*-KZ*	D**3*1**3*	43 522	8 074	51 596
	C*** ce*t*e 3 KZ*-KZ*	***1***3*9	11 300	1 916	13 216
CTN	**O-*ou*ev*** 1 F **ock	D**2*1*228	2 498	167	2 665
	Fu** ***-*ou*ev*** * * **ock	J**1***21*	80	20	100
	Fu** ***-*ou*ev*** * * **ock	J**1***138	2 984	556	3 540
	Q*-*ou*ev*** * * **ock	D**2*1*232	0	0	0
	S**es & ***keti**-*ou*ev*** * * **ock	D**2*1*23*	0	0	0
	IT-*ou*ev*** * * **ock	D**2*1*182	1 450	303	1 753

"Summary Sheet Continued"

Location	Model	BW	Col	BW Price	Col Price
B*N	***8	12 106	2 758	R100,00	R200,00
***P	*33*1	636	281	R200,00	R100,00
	***8	32 963	26 894	R100,00	R300,00
	*6*8	62 719	33 011	R300,00	R100,00
*P*IA	***8	80 846	30 772	R100,00	R50,00
G*J	*4*8E	10 749	-	R200,00	R100,00
	***8	5 407	1 632	R100,00	R200,00
KIM	***8	15 259	6 496	R50,00	R100,00
***IA	*33*1	1 196	1 085	R50,00	R100,00
	***8	98 726	51 486	R20,00	R40,00
	*6*8	32 554	24 690	R30,00	R60,00
U*N	*4*8E	3 285	-	R70,00	R150,00
	***8	5 036	3 295	R60,00	R125,00
E*S	*4*8E	3 434	-	R500,00	R250,00
	*33*1	8 337	1 961	R300,00	R500,00
	***8	11 121	3 458	R100,00	R50,00
P*Z	*4*8E	12 730	-	R50,00	R50,00
	***8	21 699	25 863	R250,00	R125,00
				R2 580,00	R2 600,00
					R5 180,00

By using formulas to summarize and sum up totals, I was able to eliminate errors and the need to re calculate every month. All I would need to do is, duplicate the previous month's summary sheet, rename the sheet and adjust the dates, and make two changes to the formula in the BW and Color fields to reference the new month's data.

I used a SUMIF here as a VLOOKUP range would be unnecessarily large.

=SUMIF(Data!$G:G;D8;Data!CB:CB)-SUMIF(Data!$G:G;D8;Data!BZ:BZ)

G refers to the serial number on the data sheet, D is the serial number in the summary sheet, and CB:CB refers to the data that needs to be imported from the data sheet. It subtracts the current month results from the previous months and displays the usage for the current month.

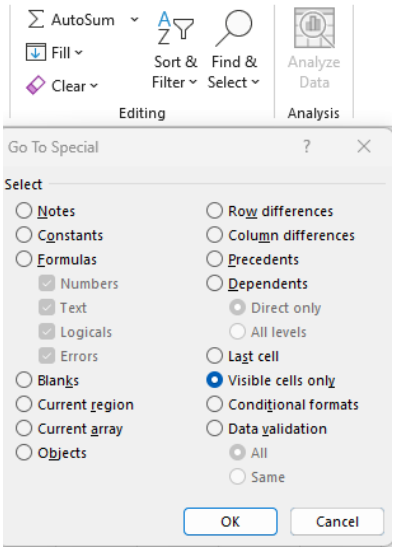
By setting up my data this way and using conditional formatting, it was easy for me to identify outliers, incorrect values or any data that was not fitting.

Once this spreadsheet was completed correctly, I would move on to transferring the collected data into a template that we used to upload the data to SAP for all the internal billing to commence. And thereafter provide this billing data to the client.

"Template"

Company	Contact	Telephone	Mode	Serial Number	Meter Reading Type	Opening	Closing F
INTERNA* ACC * *O**INGS (PT	R******AB B**	011 ****1 *000	bizhub	**43*14534	Black & White Copies	711128	
INTERNA* ACC * *O**INGS (PT	R******AB B**	011 ****1 *000	bizhub	**43*14534	Colour Copies	37437	
INTERNA* ACC * *O**INGS (PT	R******AB B**	011 ****1 *000	bizhub	**43*14515	Black & White Copies	344969	
INTERNA* ACC * *O**INGS (PT	R******AB B**	011 ****1 *000	bizhub	**43*14515	Colour Copies	42343	
INTERNA* ACC * *O**INGS (PT	R******AB B**	011 ****1 *000	bizhub	**43*145*4	Black & White Copies	244259	
INTERNA* ACC * *O**INGS (PT	R******AB B**	011 ****1 *000	bizhub	**43*145*4	Colour Copies	35623	
INTERNA* ACC * *O**INGS (PT	F**E	021 50* 1171	bizhub	*9*21**4*9*	Black & White Copies	103995	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*1*41*16581	Black & White Copies	133804	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*1*41*16581	Colour Copies	91661	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*2*411*7659	Black & White Copies	159451	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*2*411*7659	Colour Copies	123959	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*1*41**6477	Black & White Copies	216854	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*1*41**6477	Colour Copies	38975	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*1*41*16927	Black & White Copies	127858	
INTERNA* ACC * *O**INGS (PT	**ISO **OLO	011 ****1 *000	bizhub	*1*41*16927	Colour Copies	44	

The problem here is that BW and color fields are listed on a separate line, whereas my spreadsheet has the BW and color fields in the same row.



To import the data here quickly I would copy the data from my data sheet and paste the values on a sperate sheet in template file. I would then filter the template by the "Meter Reading Type" column, select "Black & White Copies", put in the formula for a VLOOKUP, copy the formula, highlight the range to paste the formula in, but before pasting in the formula, I would go the "Find & Select" option on the ribbon menu, select the "Go To Special Option" and select "Visible cells only". This would prevent the formula from being pasted into hidden fields.

I would repeat this for "Colour Copies" and perform a few checks.

In some cases, meter reading data would be imported into SAP automatically via the network, or by technicians visiting the device. This would bypass our collection of data and create incorrect stats when reconciling. Therefore I would make sure I check that the data I am providing is correct and overrides other methods, and matches up with what I am billing the customer.

This concludes my process for managing and handling data for billing.

Part 2

As a respect for confidentiality, and not having access to the actual data, I will use simulated data from www.mockaroo.com to replicate the data and carry out the data handling process.

The print tracking application would provide us with a log of all transactions. The fields that it tracked are as follows:

- Device – name
- User- name
- Job title
- Date
- Time
- B/W print (A4)- Pages- count
- B/W copy (A4)- Pages- count
- Scan- Pages- count
- Color print (A4)- Pages- count
- Color copy (A4)- Pages- count
- B/W pages
- Color pages
- Total- Pages- count

With some tracking applications, you have the ability to generate reports through their interface, but customization is limited. Therefore, I would import the data into excel and create a pivot table to identify the top users, printers and departments and explore how the printers are being used.

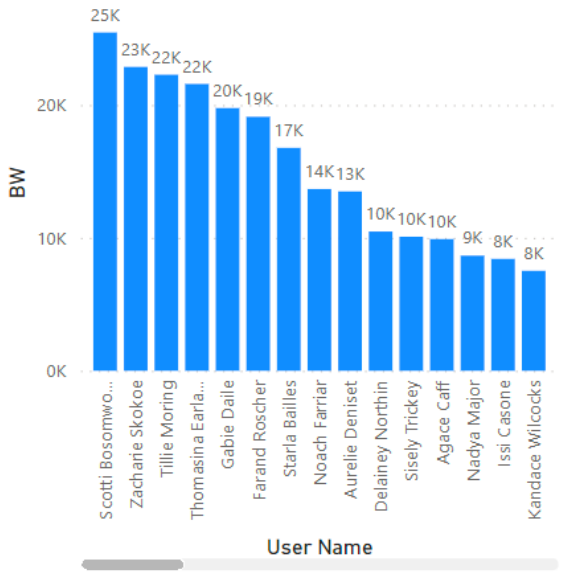
User Name	B/W print	B/W copy	Color print	Color copy	Scans
Abigale Scarsbrick	249	32	300	15	1653
Agace Caff	4954	17	800	116	2310
Aldridge Stokey	216	0	137	1	189
Alvis Beswick	898	55	1817	21	2770
Annelise Vaney	168	0	0	0	39
Arlene Pobjay	0	0	0	0	1
Aurelie Deniset	6747	198	4	4	11988
Austin Murrhardt	0	23	0	1	0
Avery Oxborrow	1766	1	2703	61	1227
Bernie Slamaker	3	0	0	0	0
Bibbye Hatchell	796	1	618	2	392
Blakelee Rambadt	1671	0	2245	180	1762
Blinnie Thornally	64	16	0	4	164
Bob Riccardi	51	0	0	0	16
Brittan Roser	1286	4	80	0	322
Bronny Rowlson	2500	284	1313	178	1450

We eventually switched to using Power BI. Below, you will find insightful snapshots displaying the top users at *** Client and visualize the usage patterns of the devices. These snapshots provide a clear understanding of the data..

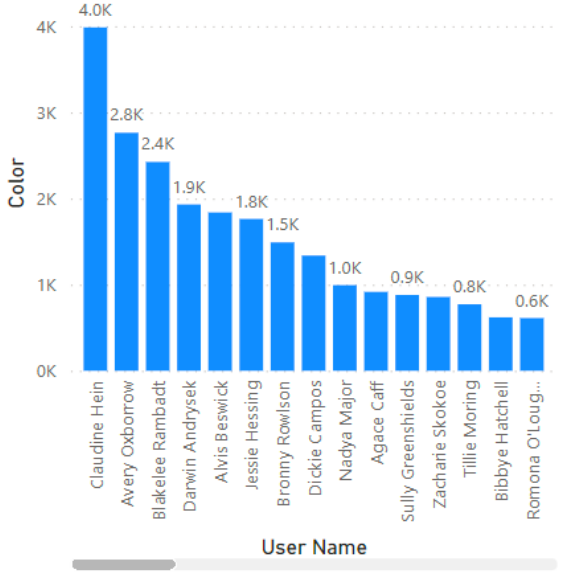
Top Users at *** Client

69	324k	28K	166K
Users	BW	Color	Scans

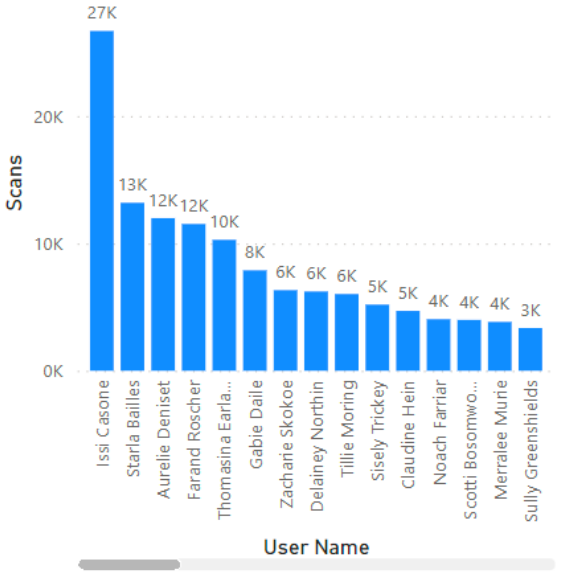
Top Users BW



Top Users Color



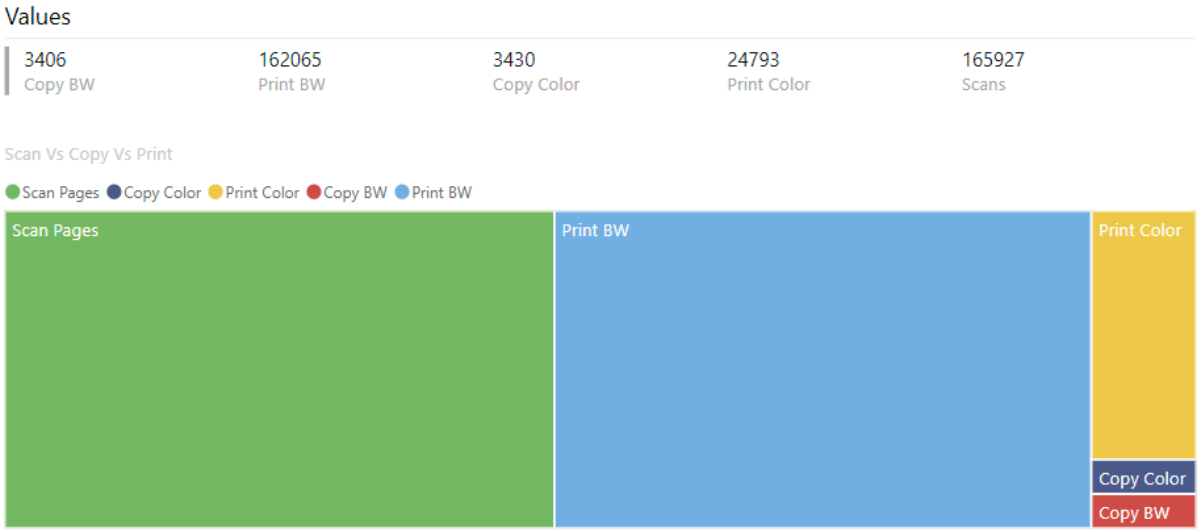
Top Users Scan

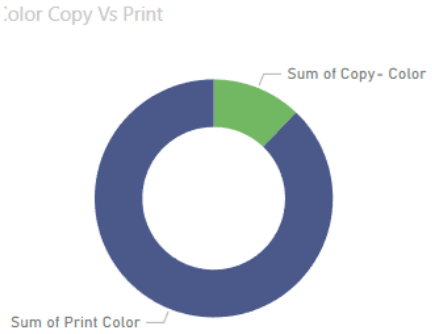


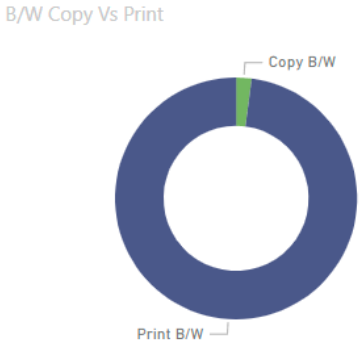
How are the printers being used at *** Client?

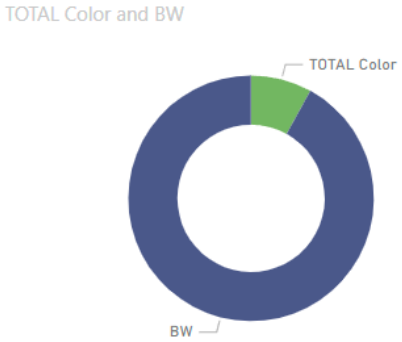
A careful examination reveals that the devices are primarily used for printing and scanning purposes. Black and white printing is notably more frequent than color printing, highlighting the client's preference for cost-effective printing options and their effective expenditure control.

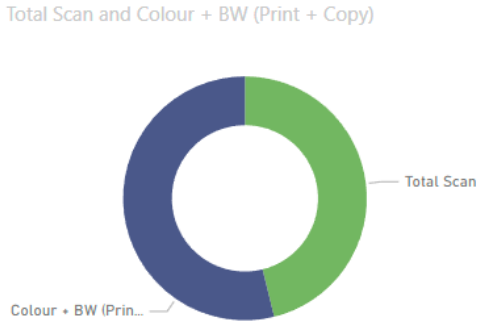
Moreover, based on the data, there is a potential need for a robust bulk scanning or OCR solution. Given the client's scanning frequency and volume, implementing such a solution would significantly improve their document management and processing capabilities.











With revisions to the summary, the results can be presented to the client and our sales team.

This brings us to the conclusion of my data presentation process.

Part 3

Here, I'd like to highlight a few noteworthy situations that I'm proud of since they demonstrate how my analytical and data management skills have been applied.

Case 1:

Prior to taking over managing this client, our department acknowledged that this specific client was not yielding any profits throughout the duration of it's contract. The reasons behind this were unknown, but on a monthly basis, my colleague would report losses ranging from a few hundred rands to a few thousand rands in some months.

Upon taking over this client, I initiated an investigation and made a significant discovery. There was a device situated in a remote location, operating under a sub-client code, that was not being billed. The client had been directly interacting with the nearby branch, where the branch technician would submit meter readings, and the costs were being internally billed. Remarkably, this device's existence had remained unknown for years.

Unaware of this situation, I meticulously sifted through years of data. Finally, I successfully identified this device and managed to recover approximately R200,000 from the client.

Case 2:

We encountered a similar situation with another client managed by my colleague, where we consistently incurred losses every month. Despite balancing all reconciliations, we still fell short without understanding the underlying cause.

Taking charge of this client, I headed an investigation to uncover why we were experiencing financial losses despite the client having minimal overhead expenses. Through extensive research and careful examination of internal and external billing reports and invoices, I eventually made a significant breakthrough. It was revealed that a specific set of devices at this client had incorrect VAT codes assigned, leading to double VAT charges. The VAT was erroneously included in the cost per copy as well as being charged separately on each invoice.

While I cannot claim credit for the entire recovery, I initiated an investigation that prompted our finance team to review the VAT codes assigned to all internal devices. As a result, we successfully recovered close to a million rand from several accounts in double VAT charges.