Entertainment Company

(Extraction of Sales Data from PDF Files)

Automated the process to read the Sales data from PDF documents and loaded the data into a database, for developing automated Power BI reports

AUTOMATED SALES REPORTING FOR A THEATRICAL PRODUCTION COMPANY

ABOUT THE CLIENT

Client is an entertainment company and one of the largest **theatrical production companies** in the world

SITUATION

- The client receives sales data in the form of non-standardized PDF files from induvial locations shared by its business planning team and network partners. The client was processing the sales data received though manual processes to create a consolidated Sales Report on Excel.
- Merilytics partnered with the client to **develop a process to automatically read the sales data from varied PDF reports and store in a database,** that is then connected to automated Power BI dashboards

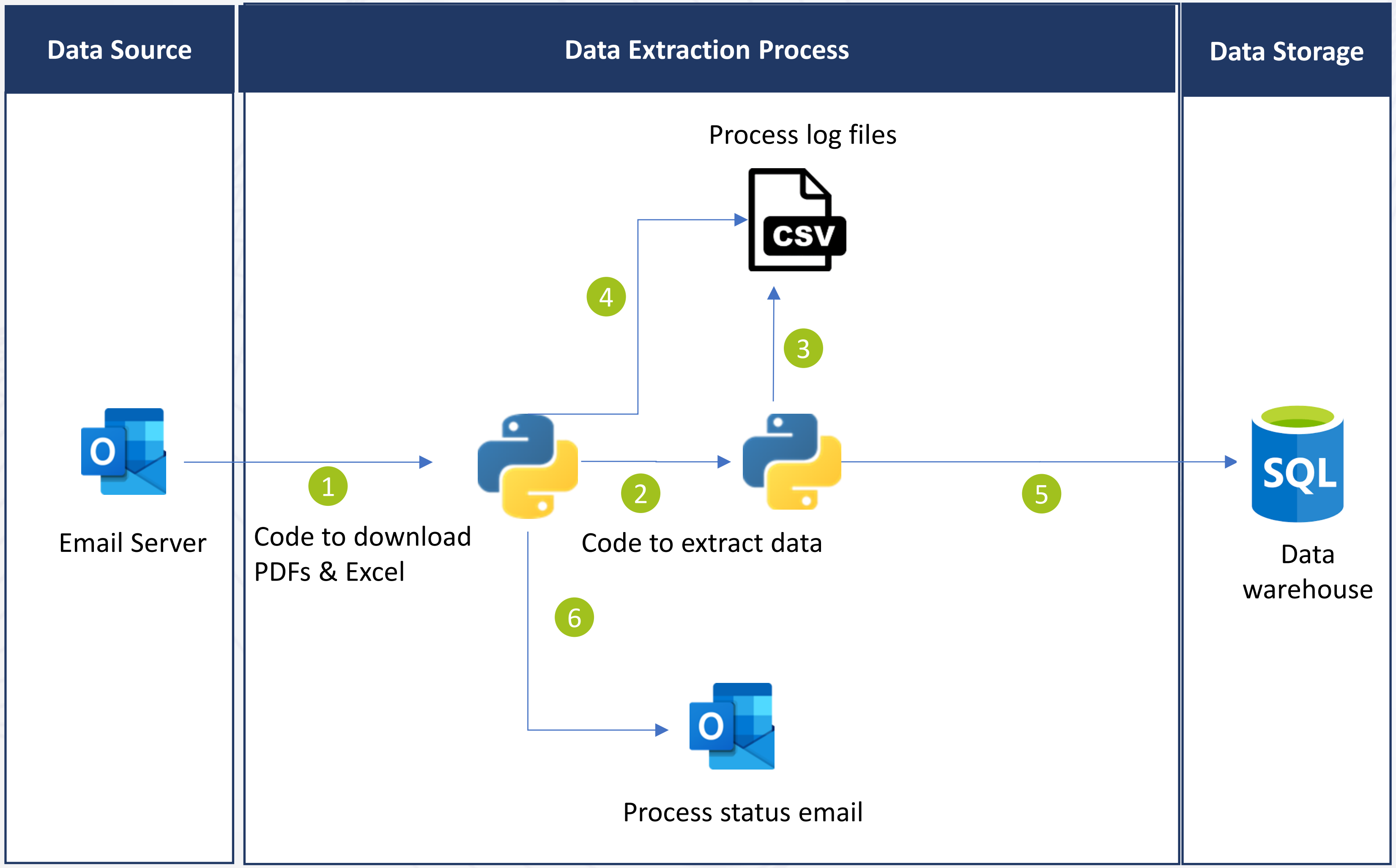
VALUE ADDITION

- Developed **automated process to store the Sales Reports** received from business planning team and network partners in PDF formats in a central location on a local system
- **Parsed the PDF files and extracted data from the Sales Reports** using **Python packages such as Camelot and Tabula**, and consolidated/stored the data on an on-premise Data Warehouse, by leveraging attributes present in the email body and the PDF files
- **Set up/deployed automated Power BI reports** to run off the consolidated data in the Data Warehouse

IMPACT

- Provided **visibility into the historical sales** at show, event, price tier etc. in a centralized visualization layer
- Reduced scope for human intervention led to **increased accuracy of the Sales Reports**, and the automation helped to **track the sales performance at a daily level**

DATA EXTRACTION PROCESS FLOW

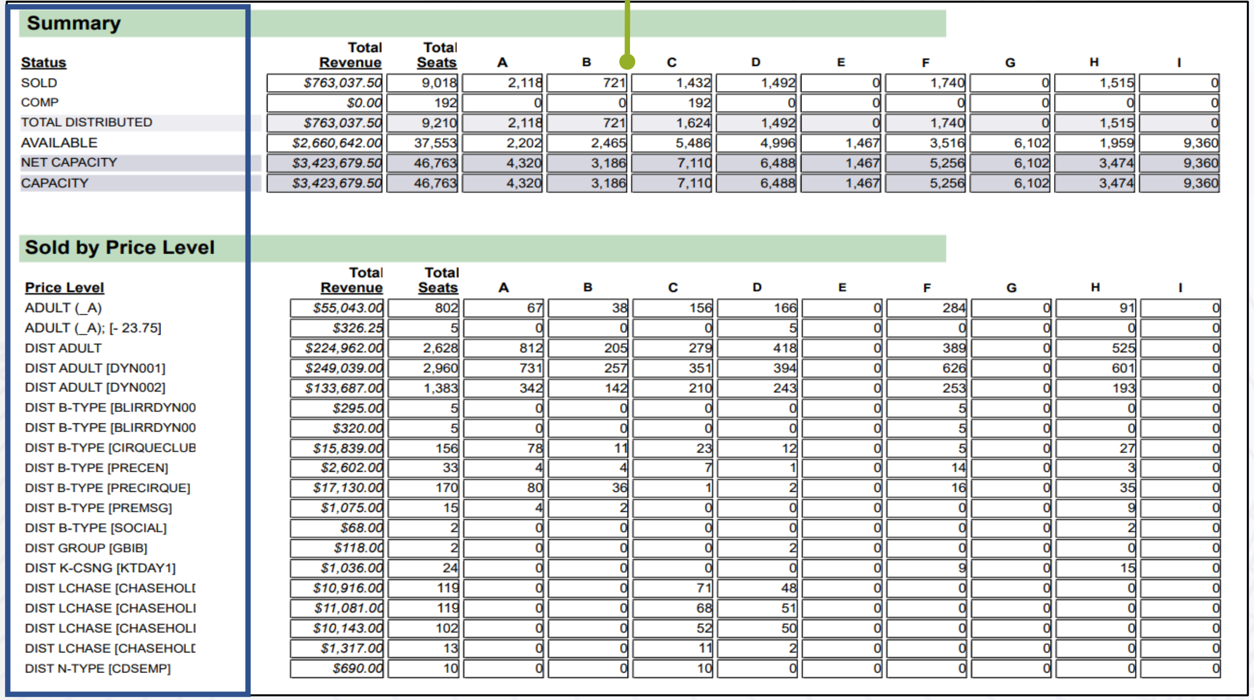


1. All the PDFs were received on an email account. The Python script was reading these emails and identifying the relevant emails using email subject and right attachments and downloading the PDFs in local system.
2. The downloaded PDFs were parsed using Python packages such as Camelot and Tabula to extract the data
3. Status of each extraction process was stored in process log files
4. The process log file is used to generate information for daily process status notifications
5. The aggregated and processes data was loaded in a data warehouse
6. Based on the status of each run associated with each PDFs, daily status notifications were generated using Python.

PDF TEMPLATES (1/2)

ILLUSTRATIVE

Sales data available for shows, events & price tiers in multiple PDF formats

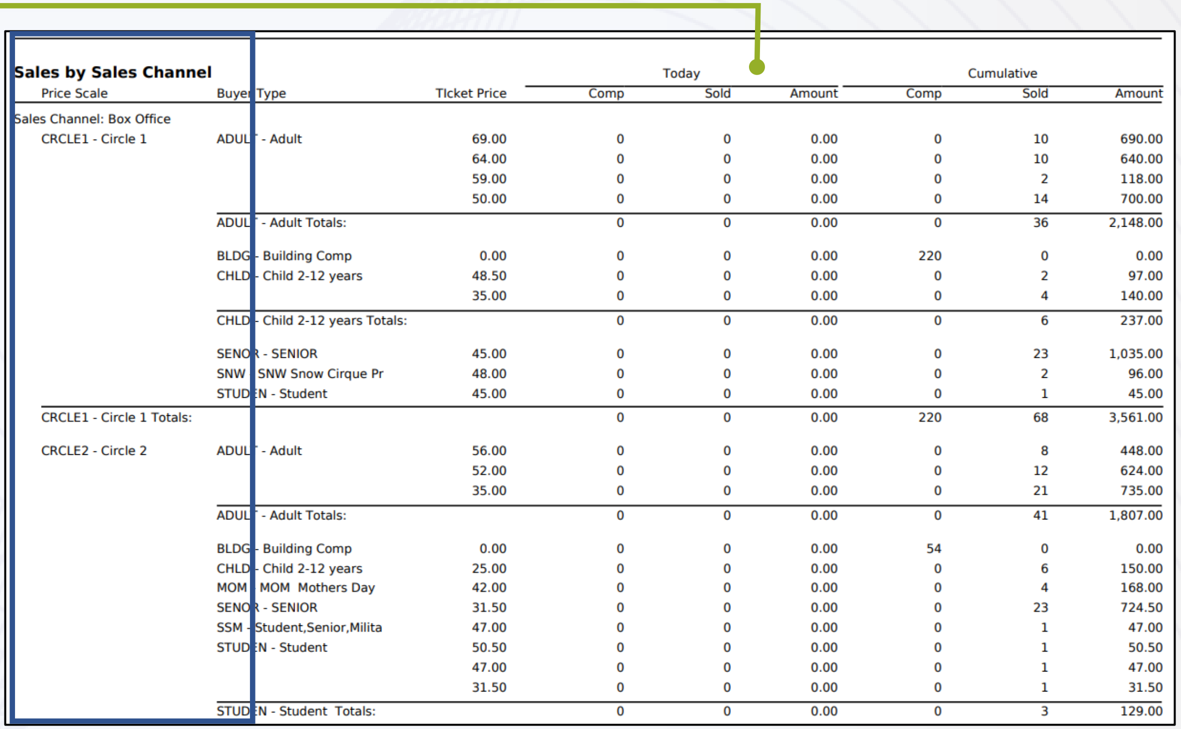


Summary

Status	Total Revenue	Total Seats	A	B	C	D	E	F	G	H	I
SOLD	$763,037.50	9,018	2,118	721	1,432	1,492	0	1,740	0	1,515	0
COMP	$0.00	192	0	0	192	0	0	0	0	0	0
TOTAL DISTRIBUTED	$763,037.50	9,210	2,118	721	1,624	1,492	0	1,740	0	1,515	0
AVAILABLE	$2,660,642.00	37,553	2,202	2,465	5,486	4,996	1,467	3,516	6,102	1,959	9,360
NET CAPACITY	$3,423,679.50	46,763	4,320	3,186	7,110	6,488	1,467	5,256	6,102	3,474	9,360
CAPACITY	$3,423,679.50	46,763	4,320	3,186	7,110	6,488	1,467	5,256	6,102	3,474	9,360

Sold by Price Level

Price Level	Total Revenue	Total Seats	A	B	C	D	E	F	G	H	I
ADULT (_A)	$55,043.00	802	67	38	156	166	0	284	0	91	0
ADULT (_A); [- 23.75]	$326.25	5	0	0	0	5	0	0	0	0	0
DIST ADULT	$224,962.00	2,628	812	205	279	418	0	389	0	525	0
DIST ADULT [DYN001]	$249,039.00	2,960	731	257	351	394	0	626	0	601	0
DIST ADULT [DYN002]	$133,687.00	1,383	342	142	210	243	0	253	0	193	0
DIST B-TYPE [BLIRRDYN00	$295.00	5	0	0	0	0	0	5	0	0	0
DIST B-TYPE [BLIRRDYN00	$320.00	5	0	0	0	0	0	5	0	0	0
DIST B-TYPE [CIRQUECLUB	$15,839.00	156	78	11	23	12	0	5	0	27	0
DIST B-TYPE [PRECEN]	$2,602.00	33	4	4	7	1	0	14	0	3	0
DIST B-TYPE [PRECIRQUE]	$17,130.00	170	80	36	1	2	0	16	0	35	0
DIST B-TYPE [PREMSG]	$1,075.00	15	4	2	0	0	0	0	0	9	0
DIST B-TYPE [SOCIAL]	$68.00	2	0	0	0	0	0	0	0	2	0
DIST GROUP [GBIB]	$118.00	2	0	0	0	2	0	0	0	0	0
DIST K-CSNG [KTDAY1]	$1,036.00	24	0	0	0	0	0	9	0	15	0
DIST LCHASE [CHASEHOLI	$10,916.00	119	0	0	71	48	0	0	0	0	0
DIST LCHASE [CHASEHOLI	$11,081.00	119	0	0	68	51	0	0	0	0	0
DIST LCHASE [CHASEHOLI	$10,143.00	102	0	0	52	50	0	0	0	0	0
DIST LCHASE [CHASEHOLI	$1,317.00	13	0	0	11	2	0	0	0	0	0
DIST N-TYPE [CDSEMP]	$690.00	10	0	0	10	0	0	0	0	0	0



Sales by Sales Channel

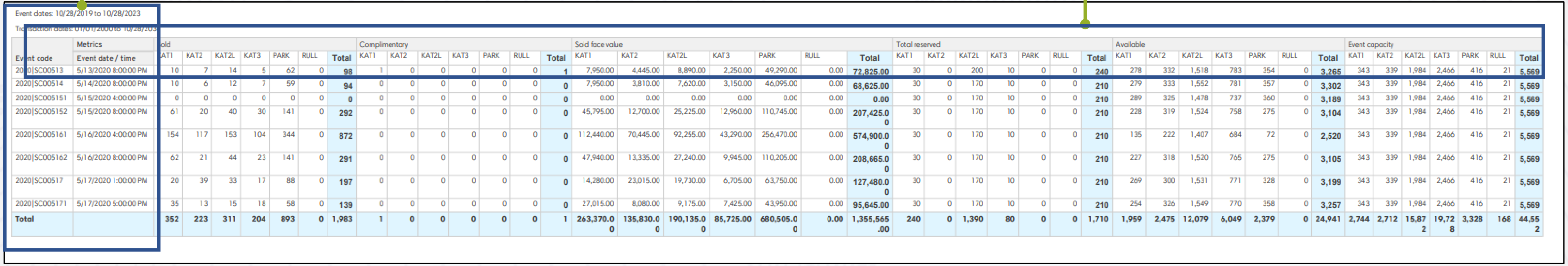
			Today			Cumulative		
Price Scale	Buyer Type	TIcket Price	Comp	Sold	Amount	Comp	Sold	Amount
Sales Channel: Box Office								
CRCLE1 - Circle 1	ADUL - Adult	69.00	0	0	0.00	0	10	690.00
		64.00	0	0	0.00	0	10	640.00
		59.00	0	0	0.00	0	2	118.00
		50.00	0	0	0.00	0	14	700.00
	ADUL - Adult Totals:		0	0	0.00	0	36	2,148.00
	BLDG - Building Comp	0.00	0	0	0.00	220	0	0.00
	CHLD - Child 2-12 years	48.50	0	0	0.00	0	2	97.00
		35.00	0	0	0.00	0	4	140.00
	CHLD - Child 2-12 years Totals:		0	0	0.00	0	6	237.00
	SENO R - SENIOR	45.00	0	0	0.00	0	23	1,035.00
	SNW SNW Snow Cirque Pr	48.00	0	0	0.00	0	2	96.00
	STUD N - Student	45.00	0	0	0.00	0	1	45.00
CRCLE1 - Circle 1 Totals:			0	0	0.00	220	68	3,561.00
CRCLE2 - Circle 2	ADUL - Adult	56.00	0	0	0.00	0	8	448.00
		52.00	0	0	0.00	0	12	624.00
		35.00	0	0	0.00	0	21	735.00
	ADUL - Adult Totals:		0	0	0.00	0	41	1,807.00
	BLDG - Building Comp	0.00	0	0	0.00	54	0	0.00
	CHLD - Child 2-12 years	25.00	0	0	0.00	0	6	150.00
	MOM MOM Mothers Day	42.00	0	0	0.00	0	4	168.00
	SENO R - SENIOR	31.50	0	0	0.00	0	23	724.50
	SSM -Student,Senior,Milita	47.00	0	0	0.00	0	1	47.00
	STUD N - Student	50.50	0	0	0.00	0	1	50.50
		47.00	0	0	0.00	0	1	47.00
		31.50	0	0	0.00	0	1	31.50
	STUD N - Student Totals:		0	0	0.00	0	3	129.00

PDF TEMPLATES (2/2)

ILLUSTRATIVE

Sales data available for multiple events

Sales data available for multiple price tiers



Event dates: 10/28/2019 to 10/28/2023

Transaction dates: 01/01/2000 to 10/28/203

	Metrics	Sold							Complimentary						
Event code	Event date / time	KAT1	KAT2	KAT2L	KAT3	PARK	RULL	Total	KAT1	KAT2	KAT2L	KAT3	PARK	RULL	Total
2020|SC00513	5/13/2020 8:00:00 PM	10	7	14	5	62	0	98	1	0	0	0	0	0	1
2020|SC00514	5/14/2020 8:00:00 PM	10	6	12	7	59	0	94	0	0	0	0	0	0	0
2020|SC005151	5/15/2020 4:00:00 PM	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2020|SC005152	5/15/2020 8:00:00 PM	61	20	40	30	141	0	292	0	0	0	0	0	0	0
2020|SC005161	5/16/2020 4:00:00 PM	154	117	153	104	344	0	872	0	0	0	0	0	0	0
2020|SC005162	5/16/2020 8:00:00 PM	62	21	44	23	141	0	291	0	0	0	0	0	0	0
2020|SC00517	5/17/2020 1:00:00 PM	20	39	33	17	88	0	197	0	0	0	0	0	0	0
2020|SC005171	5/17/2020 5:00:00 PM	35	13	15	18	58	0	139	0	0	0	0	0	0	0
Total		352	223	311	204	893	0	1,983	1	0	0	0	0	0	1

	Metrics	Sold face value						
Event code	Event date / time	KAT1	KAT2	KAT2L	KAT3	PARK	RULL	Total
2020|SC00513	5/13/2020 8:00:00 PM	7,950.00	4,445.00	8,890.00	2,250.00	49,290.00	0.00	72,825.00
2020|SC00514	5/14/2020 8:00:00 PM	7,950.00	3,810.00	7,620.00	3,150.00	46,095.00	0.00	68,625.00
2020|SC005151	5/15/2020 4:00:00 PM	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2020|SC005152	5/15/2020 8:00:00 PM	45,795.00	12,700.00	25,225.00	12,960.00	110,745.00	0.00	207,425.00
2020|SC005161	5/16/2020 4:00:00 PM	112,440.00	70,445.00	92,255.00	43,290.00	256,470.00	0.00	574,900.00
2020|SC005162	5/16/2020 8:00:00 PM	47,940.00	13,335.00	27,240.00	9,945.00	110,205.00	0.00	208,665.00
2020|SC00517	5/17/2020 1:00:00 PM	14,280.00	23,015.00	19,730.00	6,705.00	63,750.00	0.00	127,480.00
2020|SC005171	5/17/2020 5:00:00 PM	27,015.00	8,080.00	9,175.00	7,425.00	43,950.00	0.00	95,645.00
Total		263,370.00	135,830.00	190,135.00	85,725.00	680,505.00	0.00	1,355,565.00

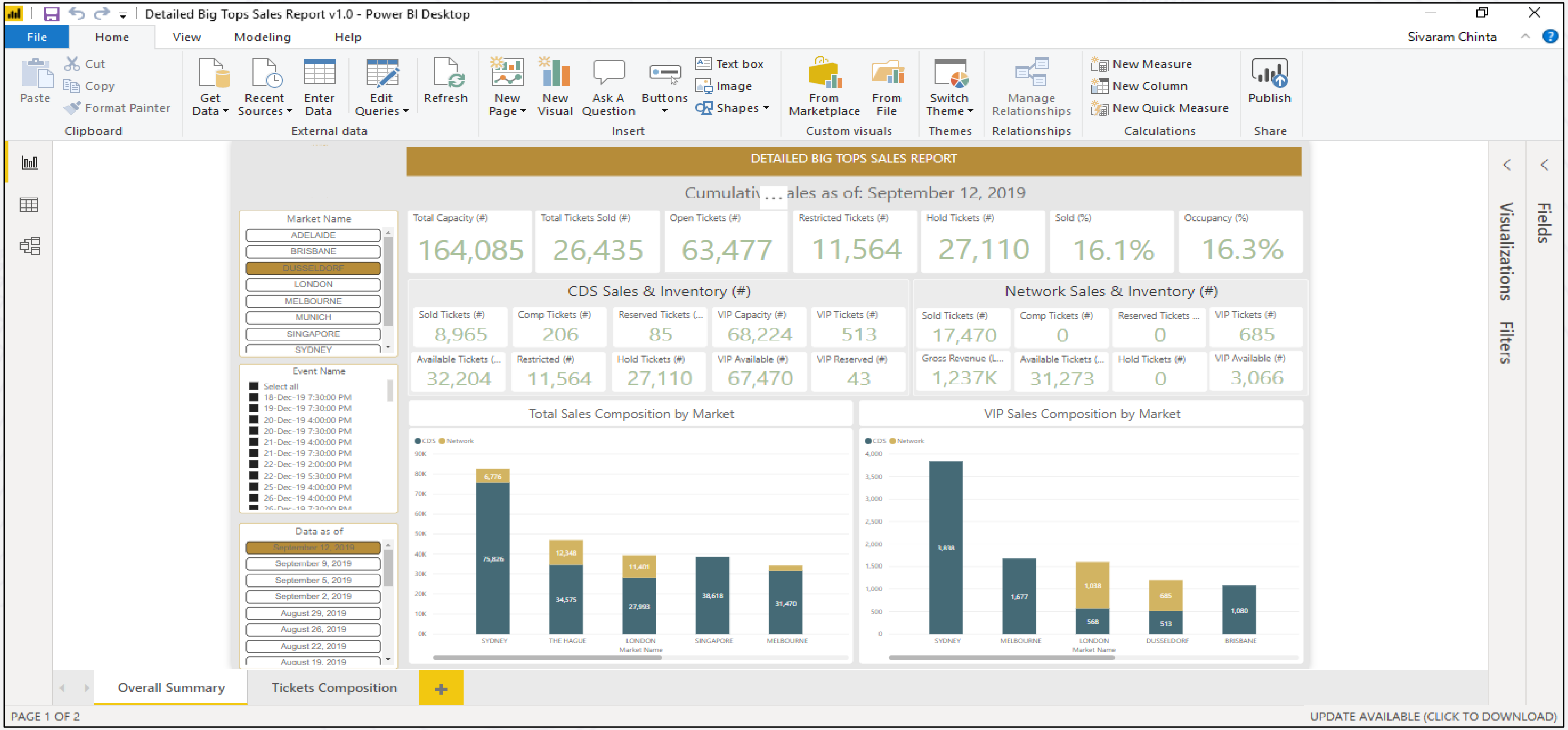
	Metrics	Total reserved							Available						
Event code	Event date / time	KAT1	KAT2	KAT2L	KAT3	PARK	RULL	Total	KAT1	KAT2	KAT2L	KAT3	PARK	RULL	Total
2020|SC00513	5/13/2020 8:00:00 PM	30	0	200	10	0	0	240	278	332	1,518	783	354	0	3,265
2020|SC00514	5/14/2020 8:00:00 PM	30	0	170	10	0	0	210	279	333	1,552	781	357	0	3,302
2020|SC005151	5/15/2020 4:00:00 PM	30	0	170	10	0	0	210	289	325	1,478	737	360	0	3,189
2020|SC005152	5/15/2020 8:00:00 PM	30	0	170	10	0	0	210	228	319	1,524	758	275	0	3,104
2020|SC005161	5/16/2020 4:00:00 PM	30	0	170	10	0	0	210	135	222	1,407	684	72	0	2,520
2020|SC005162	5/16/2020 8:00:00 PM	30	0	170	10	0	0	210	227	318	1,520	765	275	0	3,105
2020|SC00517	5/17/2020 1:00:00 PM	30	0	170	10	0	0	210	269	300	1,531	771	328	0	3,199
2020|SC005171	5/17/2020 5:00:00 PM	30	0	170	10	0	0	210	254	326	1,549	770	358	0	3,257
Total		240	0	1,390	80	0	0	1,710	1,959	2,475	12,079	6,049	2,379	0	24,941

	Metrics	Event capacity						
Event code	Event date / time	KAT1	KAT2	KAT2L	KAT3	PARK	RULL	Total
2020|SC00513	5/13/2020 8:00:00 PM	343	339	1,984	2,466	416	21	5,569
2020|SC00514	5/14/2020 8:00:00 PM	343	339	1,984	2,466	416	21	5,569
2020|SC005151	5/15/2020 4:00:00 PM	343	339	1,984	2,466	416	21	5,569
2020|SC005152	5/15/2020 8:00:00 PM	343	339	1,984	2,466	416	21	5,569
2020|SC005161	5/16/2020 4:00:00 PM	343	339	1,984	2,466	416	21	5,569
2020|SC005162	5/16/2020 8:00:00 PM	343	339	1,984	2,466	416	21	5,569
2020|SC00517	5/17/2020 1:00:00 PM	343	339	1,984	2,466	416	21	5,569
2020|SC005171	5/17/2020 5:00:00 PM	343	339	1,984	2,466	416	21	5,569
Total		2,744	2,712	15,872	19,728	3,328	168	44,552

PROCESS FLOW ARCHITECTURE

merilytics
AN ACCORDION COMPANY

ILLUSTRATIVE



DETAILED BIG TOPS SALES REPORT

Cumulativ… ales as of: September 12, 2019

Market Name
- ADELAIDE
- BRISBANE
- DUSSELDORF
- LONDON
- MELBOURNE
- MUNICH
- SINGAPORE
- SYDNEY

Event Name
- ■ Select all
- ■ 18-Dec-19 7:30:00 PM
- ■ 19-Dec-19 7:30:00 PM
- ■ 20-Dec-19 4:00:00 PM
- ■ 20-Dec-19 7:30:00 PM
- ■ 21-Dec-19 4:00:00 PM
- ■ 21-Dec-19 7:30:00 PM
- ■ 22-Dec-19 2:00:00 PM
- ■ 22-Dec-19 5:30:00 PM
- ■ 25-Dec-19 4:00:00 PM
- ■ 26-Dec-19 4:00:00 PM
- ■ 26-Dec-19 7:30:00 PM

Data as of
- September 12, 2019
- September 9, 2019
- September 5, 2019
- September 2, 2019
- August 29, 2019
- August 26, 2019
- August 22, 2019
- August 19, 2019

Total Capacity (#)	Total Tickets Sold (#)	Open Tickets (#)	Restricted Tickets (#)	Hold Tickets (#)	Sold (%)	Occupancy (%)
164,085	26,435	63,477	11,564	27,110	16.1%	16.3%

CDS Sales & Inventory (#)

Sold Tickets (#)	Comp Tickets (#)	Reserved Tickets (…	VIP Capacity (#)	VIP Tickets (#)
8,965	206	85	68,224	513
Available Tickets (…	**Restricted (#)**	**Hold Tickets (#)**	**VIP Available (#)**	**VIP Reserved (#)**
32,204	11,564	27,110	67,470	43

Network Sales & Inventory (#)

Sold Tickets (#)	Comp Tickets (#)	Reserved Tickets …	VIP Tickets (#)
17,470	0	0	685
Gross Revenue (L…	**Available Tickets (…**	**Hold Tickets (#)**	**VIP Available (#)**
1,237K	31,273	0	3,066

Total Sales Composition by Market

VIP Sales Composition by Market