DATA STORYTELLING
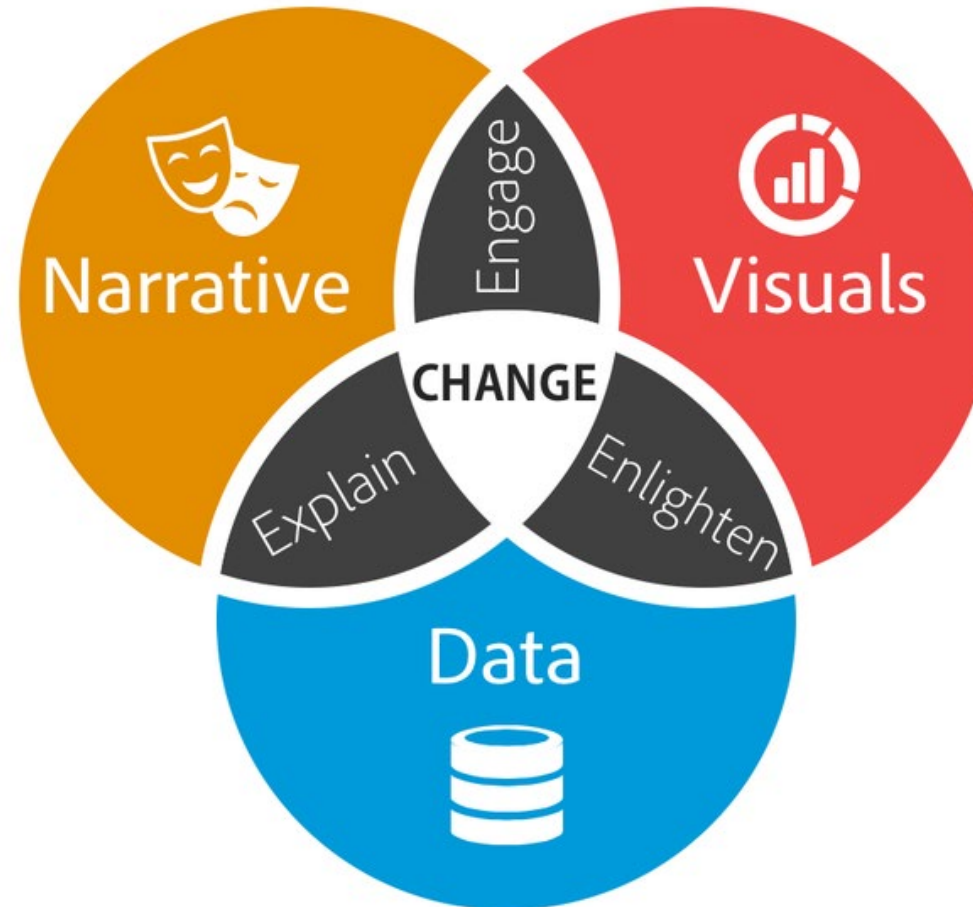
DR DANNY POO

BIG DATA ANALYTICS AND VISUALISATION

Combine Data with Visuals and Narrative

Data without a Narrative

AFTER | BEFORE

	A	B	C	D	E	F	G
1	Population (thousands)						
2	City	Country	2015	2020	2025	2050	2015-2050 Growth
3	Tokyo	Japan	36,090	36,370	36,400	36,400	0.9%
4	Mumbai	India	20,070	21,950	24,050	26,385	31.5%
5	Delhi	India	17,020	18,670	20,480	22,498	32.2%
6	Dhaka	Bangladesh	14,800	17,020	19,420	22,015	48.8%
7	São Paulo	Brasil	19,580	20,540	21,120	21,428	9.4%
8	Ciudad de México	México	19,480	20,190	20,700	21,009	7.8%
9	New York-Newark	US	19,440	19,970	20,370	20,628	6.1%
10	Kolkata	India	15,580	17,040	18,710	20,560	32.0%
11	Shanghai	China	15,790	17,210	18,470	19,412	22.9%
12	Karachi	Pakistan	13,050	14,860	16,920	19,095	46.3%
13	Kinshasa	D.R. Congo	9,050	11,310	13,880	16,762	85.2%
14	Lagos	Nigeria	10,570	12,400	14,130	15,796	49.4%
15	Total top 12		210,520	227,530	244,650	261,988	24.4%
16	Source: Nordpil World Database of Large Cities, 2016						

Data with a Narrative

AFTER

The world rapidly urbanizes
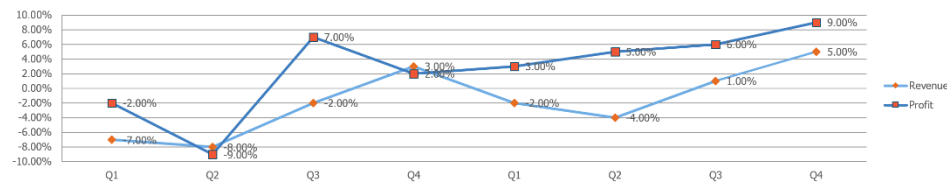
BY 2050, TOP 12 METROS ADD

51M PEOPLE

Data without a Narrative

BEFORE

Quarterly business review

Results
- Revenue 5% above target
- Profit 9% above target
- Profit and revenue performance driven by growth in the direct channel
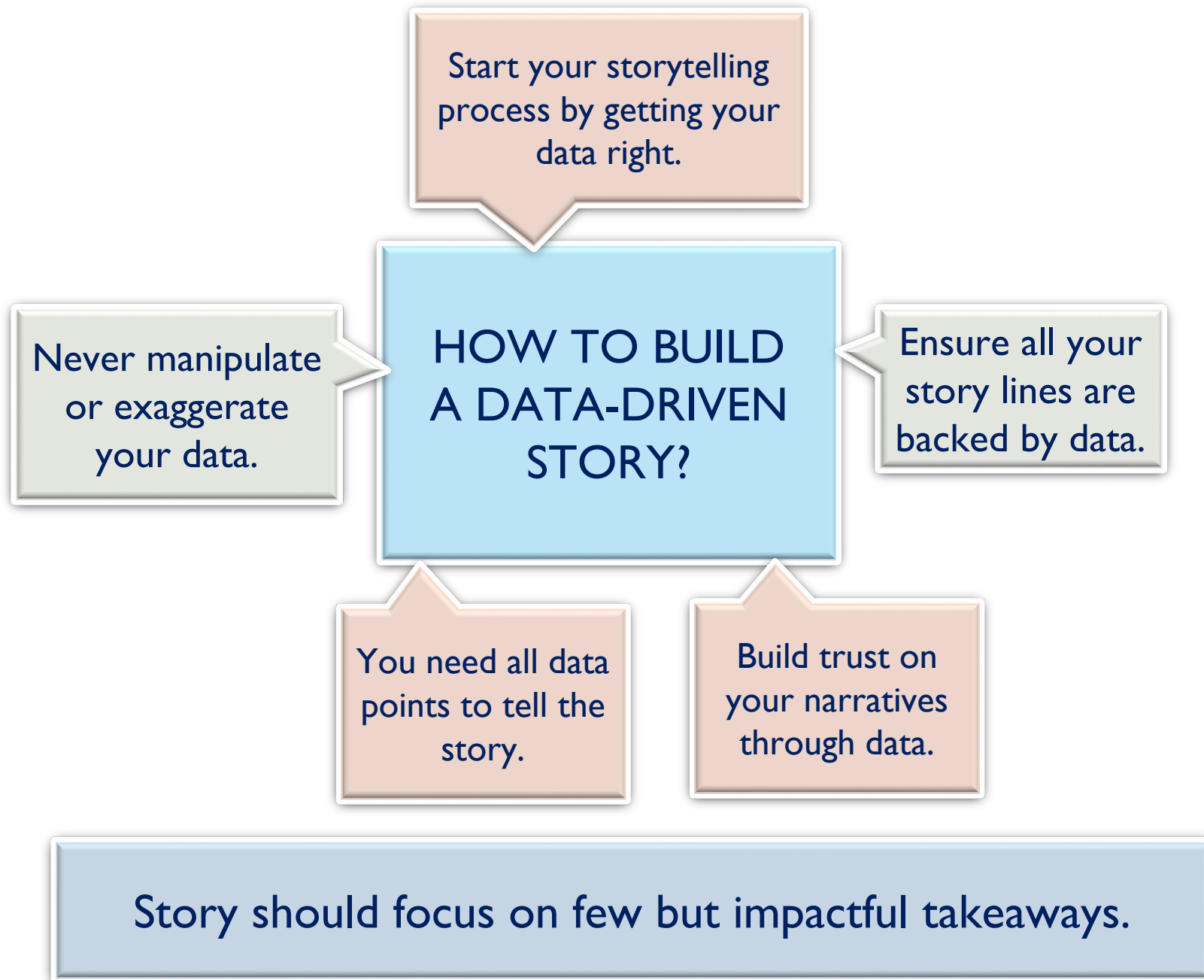- EMEA profits (-4% v target/-2%YoY) trend needs to be monitored
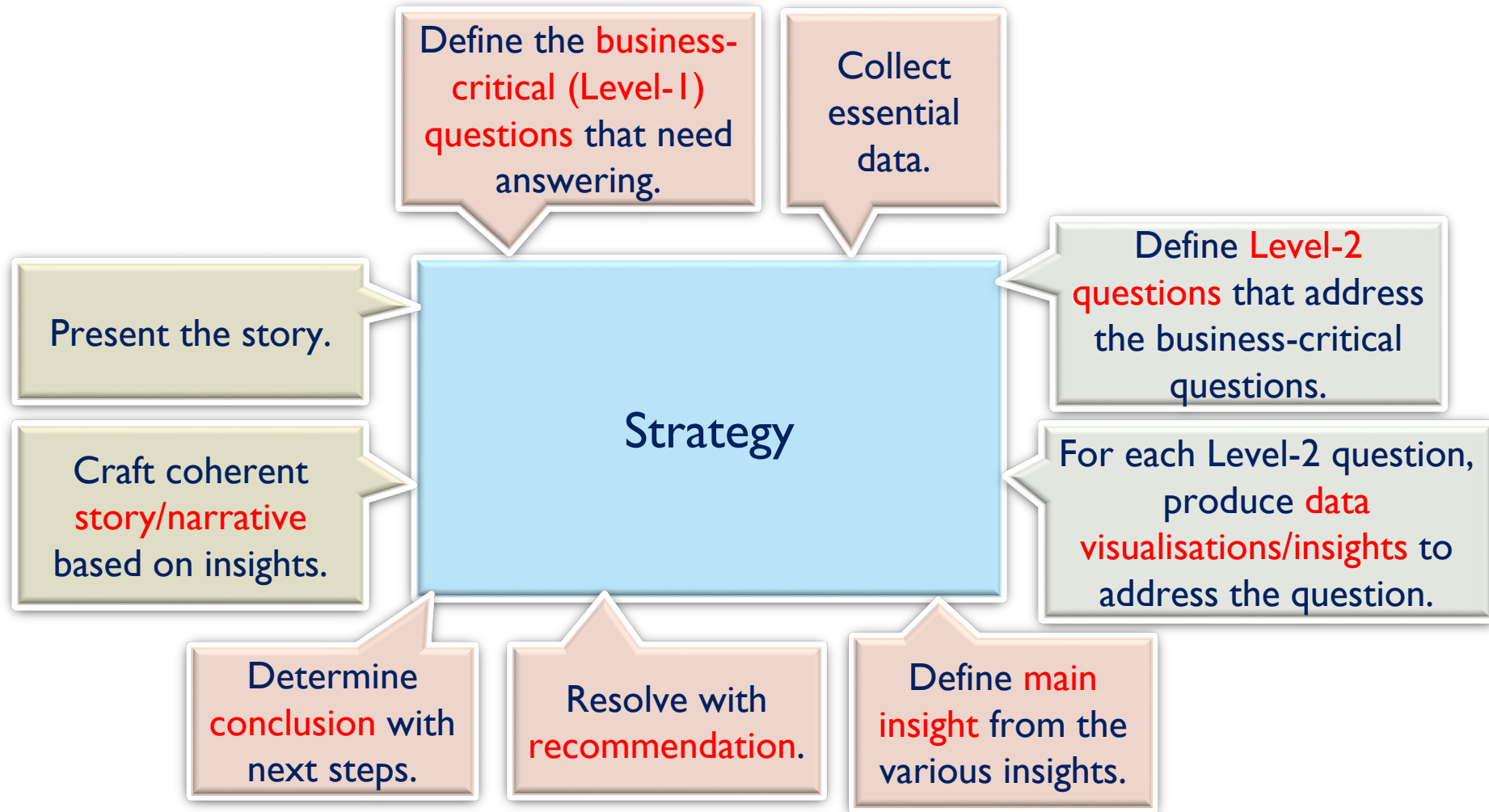
Data with a Narrative

AFTER

Q4 delivers excellent financial results, Forecast suggests challenges ahead

Revenue and profits exceed target

Economic indicators lag in 3 of 4 regions

Mixed performance on IT initiatives

Start your storytelling process by getting your data right.

HOW TO BUILD A DATA-DRIVEN STORY?

Never manipulate or exaggerate your data.

Ensure all your story lines are backed by data.

You need all data points to tell the story.

Build trust on your narratives through data.

Story should focus on few but impactful takeaways.

Strategy

- Define the business-critical (Level-1) questions that need answering.
- Collect essential data.
- Define Level-2 questions that address the business-critical questions.
- For each Level-2 question, produce data visualisations/insights to address the question.
- Define main insight from the various insights.
- Resolve with recommendation.
- Determine conclusion with next steps.
- Craft coherent story/narrative based on insights.
- Present the story.

Freytag's Pyramid & Data Stories

Climax = Main Insight

Rising Action = Supporting Facts

Resolution = Recommendation

Conclusion = Next Steps

Inciting Incident = *Deviation from expected*

Introduction = Set-up/Background

Character Growth = *Increase in audience awareness*
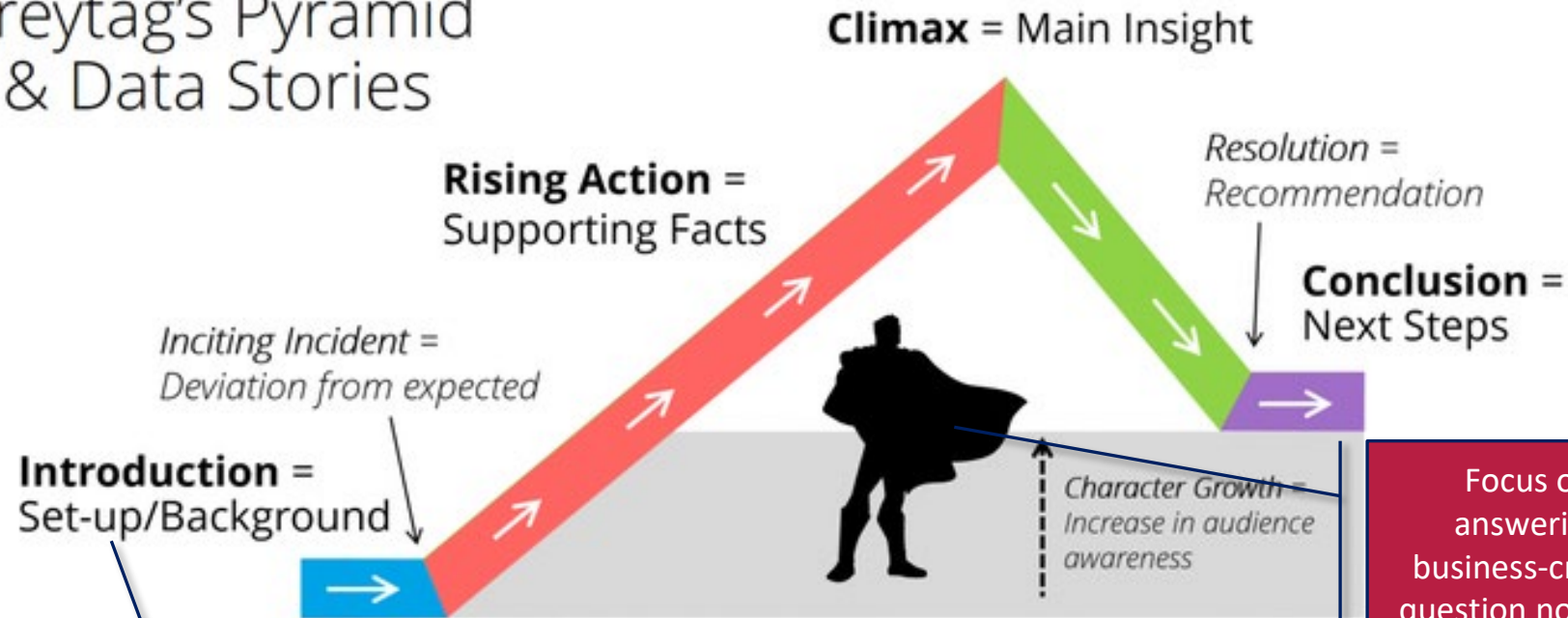
Beginning should focus on providing sufficient context (time period, data source, previous trends, etc.)

Focus on answering business-critical question not data as your character of story.

Start by providing background information, then highlight something interesting (Level-1 questions) that deviated from it, show supporting data points (Level-2 questions and data visualisations) until you reach the climax (main insight) and then discuss recommendation and possibly next steps.

Example: Housing Prices

Business-critical questions (Level-1 questions):

◦ What are the factors influencing housing price?

Level-2 questions:

◦ Does the availability of subway affect housing price?

  ◦ **Level-3 questions:**

    ◦ To what extend does the availability of subway affect housing price?

◦ Does location affect housing price?

◦ What is the relationship, if any, among the variables to housing price?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	url	id	Lng	Lat	Cid	tradeTime	DOM	followers	totalPrice	price	square	livingRoom	drawingRoom	kitchen	bathRoom	floor	buildingType	constructionTime	renovationCondition	buildingStructure	ladderRatio	elevator	fiveYearsProperty	subway	district	communityAverage
2	https://bj	1.01085E+11	116.47549	40.01952	1.111E+12	9/8/2016	1464	106	415	31680	131	2	1	1	1	ß 26	1	2005	3	6	0.217	1	0	1	7	56021
3	https://bj	1.01086E+11	116.45392	39.881534	1.111E+12	28/7/2016	903	126	575	43436	132.4	2	2	1	2	ß 22	1	2004	4	6	0.667	1	1	0	7	71539
4	https://bj	1.01086E+11	116.56198	39.877145	1.111E+12	11/12/2016	1271	48	1030	52021	198	3	2	1	3	ÖÐ 4	4	2005	3	6	0.5	1	0	0	7	48160
5	https://bj	1.01086E+11	116.45392	40.076114	1.111E+12	30/9/2016	965	138	297.5	22202	134	3	1	1	1	µ× 21	1	2008	1	6	0.273	1	0	0	6	51238
6	https://bj	1.01087E+11	116.42839	39.886229	1.111E+12	28/8/2016	927	286	392	48396	81	2	1	1	1	ÖÐ 6	4	1960	2	2	0.333	0	1	1	1	62588
7	https://bj	1.01087E+11	116.46628	39.991363	1.111E+12	22/7/2016	861	57	275.6	52000	53	1	0	1	1	ÖÐ 8	4	2005	3	6	0.333	1	1	0	7	67738
8	https://bj	1.01087E+11	116.48259	39.891991	1.111E+12	14/7/2016	851	167	275	37672	73	2	1	1	1	ß 6	4	1997	4	2	0.5	0	0	0	7	50112
9	https://bj	1.01087E+11	116.45392	39.881534	1.111E+12	7/9/2016	904	138	800	49521	161.6	3	2	1	2	ß 22	1	2004	4	6	0.667	1	1	0	7	71539

DATA STORYTELLING

DR DANNY POO

BIG DATA ANALYTICS AND VISUALISATION