IMPACT Model

I	ID the questions
M	Master the data
P	Perform test plan
A	Address and refine results
C	Communicate insights
T	Track outcomes
Identify	➔understand the problem that needs addressing
	Attributes = Audience, Scope, Use
Master	➔what data is available & will that data help address the problem
	Need to know about the data:
	how to access
	availability
	reliability
	frequency of updates
	time line of data coverage
Perform Test Plan	➔think of the right approach to the data to be able to answer the question
	ID relationship between response/ dependent and predictor/ explanatory/ independent variables
	8 approaches
	● Classification "sorting into predefined categories"
	● Regression "a number"
	● Similarity Matching "similar individuals"
	● Clustering "finding natural groups"
	●Co-occurrence grouping "associations based on transactions"
	● Profiling "the typical"
	● Link Prediction "relationship between two data"
	● Data Reduction "Reduces to most critical"
Address and refine	➔Data analysis is iterative

IMPACT Model (cont)

	slice, dice, and manipulate the data
Commun-icate	➔insights are formed by decision makers and are communicated
	executive summaries, static reports, digital dashboards, and data visualizations
Track outcomes	➔can we predict future outcomes
	then you can test how accurate the predictions were

Business Environment and Big Data

Big Data	
4 V's	Volume (size)
	Velocity (speed)
	Variety (types)
	Veracity (quality)
	Impact of data
Auditing	Audits must embrace technology.
	Technology = better quality, transparency, accuracy in audit.
	gathering data ➔ rationale behind data queries
	expands auditors' capabilities in fraud detection
	automating compliance-monitoring activities
Management Accounting	(most similar to analytics)
	Job of MA:
	are asked questions by management
	find data to address those questions
	analyze the data
	report the results to management
Financial Reporting	
Financial Statement Analysis	
Tax	

Relational Database

Benefits of (3NF) Relational DB	
	Completeness
	No redundancy
	Business rules enforcement (IC)
	Communication and integration of business processes
ETL process	extract, transform, load
	doing what to the data?
Extract	
	1. Determining the purpose and scope
	2. Obtaining
Transform	
	3. Validating for completeness and integrity
	4. Cleaning
Load	
	5. Loading the data
	DO a VICy Clean Load
Step 1 Determining	the purpose, to solve, what problem
	reliability, usefulness
	nature, timing, and extent
Step 2 Obtaining	How obtain?
	standard data request form??
	Where data?
	What specific data?
	what tools needed
ID what **you** need to Extract	Where is **your** info
	tables
	attributes
	relations between the 2
Step 3 Validating	ensure extracted data = complete & integrity
4 steps after extraction	1. Compare number of records in OG and extracted

Relational Database (cont)

	2. Compare descriptive statistics
	3. Validate Date/Time fields
	4. Compare string limits
Step 4	Cleaning
4 steps to clean	Remove headings or subtotals
	Clean leading zeroes and (NPC's) nonprintable characters
	Format negative numbers
	Correct inconsistencies (US, U.S., United States, States)
Common quality issues	Dates
	Numbers
	International characters and encoding
	Languages and measures
	Human error
Step 5	Loading
	if previous steps were done correctly then no loading necessary

Extract = Determining & Obtaining / 1 & 2
Transform = Validating & Cleaning / 3 & 4
Load = Load / 5

4 quads

	Declarative	Exploratory
Qual	Quad 1	Quad 3
Quan	Quad 2	Quad 4

Distributions

Normal Distribution	ANY mean and ANY standard deviation
	The mean, median, and mode are all equal
	Half the data falls below the mean, half above

Distributions (cont)

	SAT scores, IQ scores, heights and weights of newborn babies
Standard Normal Distribution	special ☆
not typical for data-driven quantitative data	A mean of exactly 0
	A standard deviation of exactly 1
	the mean is 0, the median and mode are also 0
Poisson Distribution ⊘	the probability of a specific number of events happening in a fixed time period

Data Governance and Ethics

Institute of Business Ethics

6 Questions

Does the company....	How does the company use data?
	...send a privacy notices?
	...assess the risks for the customer?
	...have safeguards for the risks of data misuse?
	... have the appropriate tools to manage the risks of data misuse?
	...conduct appropriate due diligence when sharing with or acquiring data from third parties?

🏴‍☠️ LMAO this is America, we have no data protection
$ This country is ruled by the amount of profit that share holders can extract
🏛 No data rights for our own data
#bigbrother4profit

7 goals CH1

Developed Analytics Mindset	Recognize when and how data analytics can address business questions.
Data Scrubbing and Data Preparation	Comprehend the process needed to clean and prepare the data before analysis.

7 goals CH1 (cont)

Data Quality	Recognize what is meant by data quality, be it completeness, reliability, or validity.
Descriptive Data Analysis	Perform basic analysis to understand the quality of the underlying data and its ability to address the business question.
Data Analysis through Data Manipulation	Demonstrate the ability to sort, rearrange, merge, and reconfigure data in a manner that allows enhanced analysis.
Statistical Data Analysis Competency	Identify and implement an approach that will use statistical data analysis to draw conclusions and make recommendations on a timely basis.
Data Visualization and Data Reporting	Report results of analysis in an accessible way to each varied decision maker and their specific needs.

Main Types of Data Analytics

Descriptive (past)	summarize existing data
	what has happened
Diagnostic (current)	explore the data
	why something has happened the way it has
Predictive (future)	used to generate a model
	what is likely to happen
Prescriptive (current/ for future)	identify the best possible options given constraints
	more advanced AI
	optimize current processes
Descriptive	
Summary statistics	mean, median, standard deviation, ect

Main Types of Data Analytics (cont)

Data reduction or filtering	*IFIF (Identify attribute, Filter, Interpret, Follow up)*
Diagnostic	
Benefits of Diagnostic	reduced external audit fees, reduced audit delay, lower material weaknesses, restatements
Profiling →	Characterizes the **typical** behavior
	IDSIF (ID, Determine, Set boundaries, Interpret, Follow-up
How profile?	*Z-score*
	Box Whisker
	interquartile range (IQR)
Clustering (finding natural groups) →	**Divides** individuals into groups that share common underlying characteristics
Hypothesis Testing	proof if P&C are meaningful
Similarity matching →	Identifies similar individuals based on data already known about them
Cooccurrence grouping	Discovers associations between individuals based on transactions they are both involved in
Predictive	
	target = attribute or value to evaluate
	class = assigned category (to record for event)
Regression (number)	Estimates or predicts a numerical value for a variable using a statistical model
IV/DFF/IP/EF	Identify the variables, Determine the functional form, Identify parameters, Evaluate fit

Main Types of Data Analytics (cont)

Classification (sorting into predefined categories)	Assigns each unit into a small set of categories or classes
Link Prediction (relationship)	Predicts a relationship between two data items
Prescriptive	
Decision support systems	Rule-based systems that gather data and recommend actions
Artificial intelligence	Learning models that adapt to new data over time to make recommendations
Benford's law	The law states that in many naturally occurring collections of numbers, the significant leading digit is likely to be small.
overfitting	complex models
underfitting	simple models

significance level = alpha
t-test = p-value < alpha = statistically significant else 👎

Graphs Charts and Tables OH MY!

Why Pictures?	
statistics	alone can be misleading
visualizations	visualizations
Purpose	
Declarative	presenting findings
Exploratory	discovering insights
Qualitative	categorical
Quantitative	numerical
Data Types	
Qualitative Data (categorical):	
Nominal	only count and group
Ordinal	⌄ AND rank
Quantitative Data (numerical):	
Interval (no 0)	⌄ and measure differences
Ratio	⌄ has a meaningful zero
The right chart	
QUAL	

Graphs Charts and Tables OH MY! (cont)	
Bar/Column C.	compares proportions of categories
Pie chart C.	parts of a whole
Stacked bar C.	shows proportion AND allows comparison
Word cloud	used for text data
QUAN	
Line C. (continuous)	trends over time
Box and whisker P.	quartiles, medians, and outliers
Scatter P.	correlation between two variables or a trend line
Filled geographic map	data ranges across geography
Refining	
readability	How much data do you need to show
	Should outliers be displayed or removed?
make differences look dramatic	What scale should be used?
	Do you need reference points to make the scale meaningful?
distract from data	When should you use multiple colors?
Reports	
get to the point	
I	Explain what was being researched
M	Overview of the data source and what data was included
P	Describe the analytical approach used
A	Present the results of the analysis
C	Communicate the insights and what they mean
T	Describe what outcomes will be tracked going forward

Data Types	
Discrete Data	whole numbers
	number of students in a class
Interval Data	the differences between values are meaningful and equal
	Fahrenheit temperature
	time on a clock
Ratio data	requires a true zero
	0 pounds = no weight
	0 dollars = no money → no problems