

Problem solving & effective questioning

Instructor: Ximena

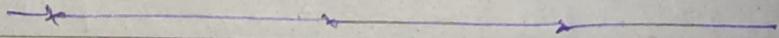
Google Finance Data Analyst

- > Effective & Ineffective questions
- > Learn how to ask great questions that lead to insights that can help you solve business problems.
- > Different types of data
- > Communicating strategies

Structured Thinking :-

The process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying the options.

- * Address vague & complex problem by breaking it down into smaller steps.



2/4

Story: Anywhere Craving Repair

They come to you to repair

Owner wanted to expand his business

Confused about choosing advertising mode

- * Think about target audience & where you can reach them. And cost

- > Zoom out and look at the whole situation not just its symptoms.
- > collaborate with stakeholders & understand their needs.

Ask: Main & stakeholders Agreed upon

Problem: Determine which advertising method is best for reaching Anywhere Gaming Repairs' target audience.

Prepare: Collecting data on different advertising methods.

Process: Clean data, eliminate inaccuracies, remove outliers.

Analyze: Need to find 2 things.

- ① Who's most likely to own a gaming system?
- ② Where these people are more likely to see the advertisement?

Discovery:-
① Ages 18-34 Target audience
② TV commercial & podcast

Since Anywhere Gaming had budget constraint manager recommended 'podcast' as its more cost-effective.

Share: Visualizations

Act:- Anywhere Gaming took action, place 30s podcast with local agency.

Ran for a month, visible growth in first week end of week 9, 88 new consumers.

Nikki: Education, Assessment, Research Manager

Noogler onboarding program:

The process to "onboard new hires".

Ask: If Nooglers are onboarding faster than
hires with old onboarding programs?
(used to lecture them)

What do you mean by onboarding faster?

Prepare: ① Prepare data by understanding the
target population
sample set? Experiment group?
control group data sources?

Should be clean and digestible to write proper
scripts for.

Process: Process the data & make sure its in the
① right format ② right columns
2 ③ right tables

Analyze: write script in SQL & R to correlate data
to control or experiment group to
understand changes.

Share: Report in such a way stakeholders can
understand. Prepare Reports, Dashboards and presentations
while keeping in mind the shareholders.

After: All's cool Continue (in this case)

Solve problems with Data,

- * Problems can be small, large, simple or complex and they are unique and require slightly different approach.
- ** However, the first step is always same:-
Understanding what kind of problem you're trying to solve.

* Six common Types of Problems:-

- ① Making Predictions :
Using data to make an informed decision about how things may be in the future.
- ② Categorizing things :
Assigning information to different groups or clusters based on common features.
- ③ Spotting something unusual :→
Identifying data that is different from the norm.
- ④ Identifying themes :→
Grouping categorized information into broader concepts.

⑤ Discovering Connections :-

Finding similar challenges faced by different entities and combining data and insights to address them.

⑥ Finding patterns:-

Using historical data to understand what happened in the past and therefore likely to happen again.



*problems in real world. Consider: Anywhere Cleaning Repair story

Problem: How to determine best advertising method for a target audience

Problem type:- Making prediction.

predicting which advertising method would work out best.

Now let's talk about categorizing things:-

Business wants to improve customer satisfaction levels.

DA reviews call recordings & assigns them into categories such as polite, rude, general, satisfaction and use this data to reward best performing call representative & coach low performers

Spotting something unusual in smart watch identifying 120 BPM of athletic women when her usual rest heart rate is 70 BPM. Eventually saving her life!

(4) Identifying themes:

finding categories for keywords and phrases in consumer service conversations.

But identifying themes goes even further by grouping each insight into a broader theme [coffee maker example]

(5) Discovering Connections →

Third party logistics, time waste for both companies when truck comes to pickup a shipment isn't ready.

By sharing data they can view each other's timeline & avoid future problems.

(6) Finding Patterns →

Oil & gas companies constantly working to keep machines running properly.

finding pattern in historical data, how & ~~when~~ machines were breaking, duration of maintenance, thereby reaching at solution.



Anmol:

"At its core his job is to connect eight with eight message at eight time"

Head of Flange

Advertiser
marketing
Analytics

Hypothesis

Testing, verify, go back to stakeholders (Marketing)

* SMART Questions →

The more questions you ask the more you'll learn about the data and the more powerful your insights will be at the end of the day.

* Leading Question: Leading you to a answer in a certain way.

Ex. This sandwich is good, isn't it?
No room for expressing your opinion

* Close ended question: Yes/No answer
Won't lead to valuable insights

* Do you like chocolate or vanilla?
What are they talking about ice cream, pudding, coffee?
You might like vanilla ice cream but chocolate in coffee, or hate both flavours.

This question is too vague & complex.

* Knowing the difference b/w effective & ineffective questions is essential for your career as a data analyst.

After all DA process starts with "Ask" phase so its important that we ask right questions!

Effective Questions follow the SMART methodology

S : Specific

M : measurable

A : Action oriented

R : Relevant

T : Time-bound.

- ⑤ Specific \rightarrow Simple, significant and focused on a single topic or few closely related ideas.

Rather than asking

Are kids getting enough physical activity these days?

Ask:

what % of kids achieve recommended 60 min physical activity atleast 5 days a week.

(m)

Measurable \rightarrow

measurable questions can be quantified & assessed.

Why did a recent video go viral? X

How many time video was shared? ✓

↑

This question helps us arrive a concrete number.

(A)

Action Oriented \rightarrow Encourage change.

How do we get customer to recycle our packaging?
What design feature will make our packaging easier to recycle?

X

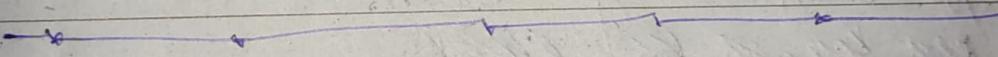
① Relevant Questions:

Relevant questions matter, are important and have significance to the problem you're trying to solve.

- Pine barren tree from frogs started disappearing,
why does it matter? X
- What environmental changes made frogs
disappear? ✓

② Time bound : Specify time to be studied.

Suppose we need to study data b/w 1984-2004.
This limits our range of possibilities and enables data analysts to focus on relevant data.



While crafting question, fairness is very import to be considered.

Fairness means ensuring that your question don't create or reinforce bias.

Ex. sandwich is good, isn't it?

Leads us to predefined answer & creates bias.

> Not assuming things before asking question

> Clear & straight forward wording

Learning portfolio manager

Evan: Data opens doors:

"One of the coolest jobs in the world."

Roles
only 3

"There are so many different paths for someone
who is interested in data" ^{careers}

Data Analyst: SQL, spreadsheets, databases,
BI team member creating dashboards,

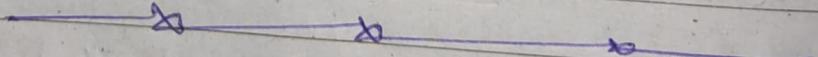
Data Engineer: Turn raw to actionable pipeline
data

Data Scientist: Turn this into a model or
statistical inference.

"* You don't have to know the outset, which path
you want to go down."

~~best to go so fast~~

Try them all | See what you like



Understanding the power of data

~~DATA~~ Data Piranha



Data - Inspired



Exploring different data sources
to find out what they have in
common.

Uses data to help cool data centres ~~over~~ ^{over} 40%.

Use data to hire Google's people operation team.
& get them started

Algorithm →

A process or set of rules to be followed for a specific task.

Ed:

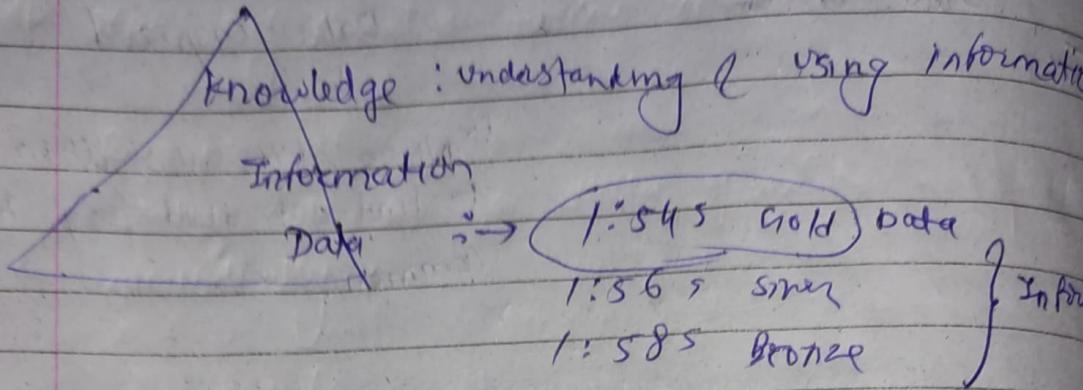
Just having data isn't enough. Data in itself provides little value. We have to do something meaningful with it.

Jack Dorsey: "Every single action that we do in this world is triggering of some amount of data, most of that data is meaningless until someone adds ~~meaning~~ to it"

~~meaning~~
Some interpretation

Data is simple: facts collected together, values that describe something

Individual data points become useful when they're selected & collected



Data becomes information by comparing with other data

Limitation of Dg:

- limited data
- Biased data

wrong

Interpretation Example: Coke replaced by new-Coke turned out to be a flop!

Mars orbiter loss \$125M

LOL → Navigation team: SI / metric system
Engineers : Pound / Engineering units
To calculate force

* Create & Barzel: sales jumped 40% in 1 year

* Pepsico: Hired Analytics team after the Coke blonde

* Qualitative vs Quantitative

Quantitative: Specific and objective measures of numerical facts.

Then what?

charts

How many?

graphs

How often?

Qualitative: Subjective or explanatory measures of qualities and characteristics. (Data that can't be put in nos)

Why?'s

(Hair & color)

Helps us understand why numbers are the way they are

Adds context to a problem.

Story: Ice-cream shop

negative reviews, start asking questions:

- How many 5-star reviews?
- What's average rating?
- Sample keyword?

17

"frustrated"

These questions generate numerical answers.

- Why are customers unsatisfied?

Running out of flavor

- What can we do to change that?

Extra supply

These questions lead to qualitative answers.

* The Big Reveal : sharing your findings

Report :

Static collection of data given to stakeholders periodically.

Dashboard :

Monitors like incoming data

Reports

Pros:

- > High-level historical data
- > Easy to design
- > Pre-clean and sorted data

Cons:

- > Continual maintenance
- > less visually appealing
- > static

Dashboard

Pros:

- > Dynamic, automatic & interactive
- > more stakeholder access
- > low maintenance

Cons:

- > labor-intensive design
- > can be confusing

If at any point base table breaks they need a lot of maintenance

Pivot tables

A data summarization tool that is used in data processing. Pivot tables are used to summarize, sort, reorganize, group, count, total or average data stored in a database.

→ To create Pivot Table →

Insert > Pivot table

Then add rows, columns, values & filters,

→ Data versus Metric

[Metric:]

single, quantifiable type of data that can be used for measurement

Data starts off as raw facts and it becomes a single type of data when we organize them into individual metric

ROI : Return on investment

$$ROI = \frac{\text{net profit over a period of time}}{\text{cost of investment}}$$

Metric can be used to help calculate customer retention rates, or a company's ability to keep its customers over time.

Metric goals

A measurable goal set by company & evaluated using metrics.

monthly sales goal, etc.

* Mathematical Thinking

Looking at a problem and logically breaking it down step-by-step to see if there are patterns in your data.

And use that to analyze your problem.

Which tool to use is

> Size of dataset

Small data

: Specific

: Short time period

: Day-to-day decision



Spreadsheet

Big Data

: Large & less specific

: Long time period

: Big decisions



SQl

Story: Hospital: over / under use of beds



or get rid

wants to provide bed who needs them & doesn't want to use resources on empty beds.

Bed occupancy rate =

$$\frac{\text{Total # of inpatient days for a given period} \times 100}{(\text{Available beds} \times \# \text{ of days in period})}$$

Considering hospitals generate lots of data the choice should be to go with SQl.

* V-words's

Volume : The amount of data

Variety : The different kinds of data

Velocity : How fast data can be processed

Veracity : Quality & reliability of data.

Big data often needs to be broken into smaller pieces in order to be organized and analyzed effectively for decision making

However

Small data is usually already a manageable size for analysis.

More Spreadsheet Basics

* Spreadsheet will be the first tool you'll go for

They are the unsung hero of the data world.

You should definitely learn Spreadsheet,

"Add data in a sheet"

"Set up formula's in one tab"

"Let it perform on other tabs"

Spreadsheets can do basic as well as complex calculations automatically.

Common math fn: → sum, min, max, count, average



① Load the data into spreadsheet

② Organize data

> put it into pivot table

> sort & filter (first step)

> Calculations on data

> Formulae

> Functions



Step-by-step in spreadsheets →

- > New spreadsheet
- > Give it a title, relevant
- > move it to a new folder

Open data sources:-

- World Bank
- World Health Organization
- U.S. Census Bureau
- Google Public Data Explorer

Currently > WHO : population of Latin & Caribbean countries
2010-2019.

03/01/23
Week 3
2/4

Formulas in Spreadsheet →

Formula. is a set of instructions that performs a specific calculation

Operator : symbol that names the type of operation or calculation to be performed.

$$\boxed{= 3 - 2} \quad \boxed{= 31982 - 17795} \quad \boxed{\leftarrow}$$

Cell reference: single / range of cell in a worksheet that can be used in a formula.

Range: two or more cells.

Cell references update Automatically

Page No.	
Date	

You can copy paste the formula & it'll update automatically.



Errors:

#DIV/0! (denom zero/

↓
= IFERROR(B4/A4, "Not Applicable")

#ERROR! (sheets only)

formula can't be interpreted as input (also known as a parsing error).

sum(B2:B6 C2:C6)

↓ comma missing (delimiter)

#N/A : data not found (404)

commonly occurs with VLOOKUP

IFNA(sum(→), 'Not Found')

#NAME? : formula/function name isn't understood

VLOOKUP(→)

#NUM! : calculation can't be performed

DATETIME(B2, C2, "M") (B2>C2)

∴ Error (interchange of values)

#DIV/0! #N/A!

#ERROR! #NUM!

#NAME? #REF!

Page No.	
Date	

#VALUE! : General error, problem with formula / referenced cells.

Date IF("John Melty", $\frac{1}{1}$, "11/20/16", "m")

impossible to calculate

will raise #VALUE! error

#REF! : Reference cell invalid / deleted.

should use function instead of referencing
absolute

→ → → → →

* functions: A preset command that automatically performs a specific process or task using the data

Data to text →

Text(B2, "mmmm")

extracts month from B2
short date

Absolute Referencing: \$A\$2

Mixed Referencing: \$A2, A\$2

Relative Referencing: A2

[F4] To toggle betw. Relative & Absolute

Albert Einstein :

"IF I were given one hour to save the planet,
I would spend 59 minutes defining the problem
and one minute resolving it"

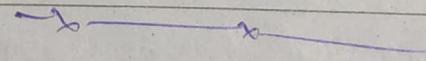
You must not jump into analysis before defining a problem.

first of all

→ Problem Domain:

Specific area of analysis that encompasses every activity affecting or affected by the problem.

* A big part of a DA's job is to develop a structured approach and use critical thinking to find the best solution.



Scope of work & structured thinking

* Structured thinking is the process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying the options.

Simple words → A list of what to deliver, timeline for major tasks and activities & checkpoints

Statement of Work & Scope of Work

Are [DIFFERENT]

Scope of Work

more commonly asked to q
Junior DA Manager

Page No.

Date

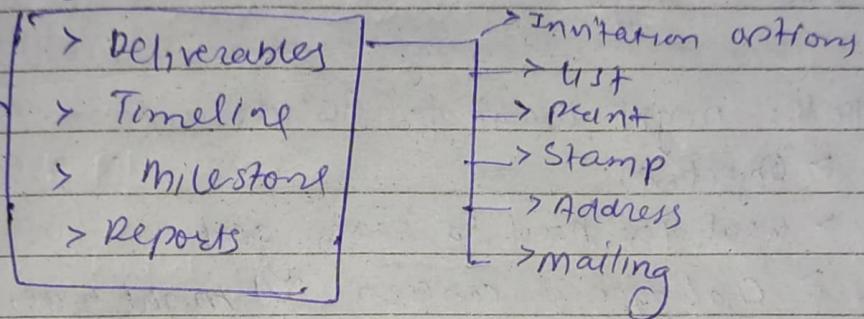
Scope of Work : (SOW)

An agreed upon outline of the work you're going to perform on a project.

Example : A wedding planner, focused on invitations

SOW

To keep
on
track



Scope of work is simple but a powerful tool

Any confusions, contradictions & questions can be easily answered.

who
what
where
when
why
how

Stating Objective →

Context:

The condition in which something exists or happens.

Number don't mean much without context.

Ed: "

Descriptive

what happened? Diagnostic

why it happened?

Predictive

will it happen?

Prescriptive

How can we make it happen

*Everyone approaches uniquely!