THE ANALYST'S AI PROMPT HANDBOOK

Platform-specific prompts for the questions analysts ask every day

Excel / Sheets

Write a VLOOKUP (or XLOOKUP) formula

Match data across two tables without remembering the syntax.

PROMPT

```
I have two sheets in Excel:

- Sheet1 has columns: Order ID, Customer ID, Order Date, Amount, Status
- Sheet2 has columns: Customer ID, Customer Name, Email, Segment, Region

Write the XLOOKUP formula to bring Customer Name into Sheet1, placed in
column E.
Also bring in Segment (column F) and Region (column G) from Sheet2 using
the same lookup key.

Additionally:
1. Give me the VLOOKUP equivalent for all three fields and explain the key
differences
   between VLOOKUP and XLOOKUP (return direction, error handling, default
behaviour).
2. Show me how to handle the case where a Customer ID in Sheet1 doesn't
exist in Sheet2
   — I want it to return "Unknown" instead of an error.
3. Explain whether I should lock the lookup array with $ signs and why it
matters
   when copying formulas down a column.
4. If I were using Google Sheets instead of Excel, what would change?
```

Tags: formula · lookup · XLOOKUP · beginner

THE ANALYST'S AI PROMPT HANDBOOK

Build a dynamic pivot-style summary

Create a SUMIFS or COUNTIFS formula with multiple conditions.

PROMPT

```
I need a formula that sums column D (Revenue) where:

- Column A = "North" region
- Column B = "Q2"
- Column C is not blank

Write the SUMIFS formula for Excel. Then extend it with the following:

1. Make it fully dynamic so I can change the region and quarter from
dropdown
   cells (B1 for region, B2 for quarter). The formula should reference
those
   cells instead of hardcoded strings.
2. Add a COUNTIFS version that counts the number of qualifying rows
instead of
   summing revenue — I want this to sit next to the SUMIFS result.
3. Show me how to write a version where I can pass multiple regions at
once
   (e.g., "North" and "East" both selected), using SUMPRODUCT or another
approach.
4. Explain why SUMIFS is better than a filtered pivot table for this use
case,
   and when I'd choose a pivot table instead.
```

Tags: formula · SUMIFS · dynamic · multi-condition

Clean messy date formats

Fix inconsistent date strings into a usable format.

PROMPT

```
My Excel column (column A) has dates in mixed formats:
"Jan 5 2024", "1/5/2024", "2024-01-05", "05-Jan-24", and some blanks.

I need them all standardised as YYYY-MM-DD in column B. Please:

1. Write a formula approach using TEXT, DATEVALUE, and IF/IFERROR
combinations
   to handle each format variant. Explain what each part of the formula
does.
2. Also write a Power Query (Get & Transform) approach that handles all
formats
   automatically — I have Excel 365 so Power Query is available.
3. Tell me how to identify which rows still have unrecognised formats
after
   applying the formula, so I can fix them manually.
4. Explain the ambiguity risk: if a date says "01/05/2024", how does Excel
   decide whether it's January 5th or May 1st? How do I control this?
5. What's the safest way to store dates in Excel to avoid this problem in
future?
```

Tags: data cleaning · dates · text functions · Power Query

Conditional formatting for exceptions

Highlight rows that breach a threshold or match a condition.

PROMPT

```
I want conditional formatting in Excel that highlights an entire row red
when the value in column E (Actual) is more than 20% above the value in
column F (Budget). My data runs from row 2 to row 500.

Please provide:
1. The exact conditional formatting formula to use, and walk me through
   exactly where to apply it (New Rule > Use a formula, selection range,
etc.).
2. Explain why the formula uses $ signs on some references but not others
—
   this always confuses me.
3. Add a second rule: highlight the row amber (orange) if Actual is
between
   10% and 20% above Budget. Explain how to set rule priority so red and
   amber don't conflict.
4. A third rule: highlight green if Actual is more than 10% BELOW Budget
   (i.e., favourable variance).
5. How do I copy these conditional formatting rules to a new sheet or
range
   without breaking the references?
```

Tags: conditional formatting · exceptions · variance · rules

Debug a broken formula

Paste a formula that returns an error and get a fix.

PROMPT

```
This Excel formula returns a #VALUE! error:

=IFERROR(INDEX(Sheet2!B:B,MATCH(1,(Sheet2!A:A=A2)*(Sheet2!C:C="Active"),0)),"Not
found")

Please:
1. Explain what this formula is trying to do in plain English.
2. Diagnose exactly why it's returning #VALUE! — is it the array multiplication,
   the MATCH, the INDEX, or something else?
3. Give me the corrected version. If it needs to be entered as an array formula
   (Ctrl+Shift+Enter in older Excel), tell me.
4. Show me the XLOOKUP equivalent that avoids the array formula complexity.
5. Suggest a way to test the formula step by step using F9 in the formula bar,
   so I can debug similar issues myself in future.
6. Are there any performance concerns with using full column references (B:B)
   in INDEX/MATCH on large datasets?
```

Tags: debugging · INDEX-MATCH · errors · array formulas

Write a macro / VBA snippet

Automate a repetitive task in Excel with VBA.

PROMPT

```
Write an Excel VBA macro that does the following:

1. Loops through every sheet in the workbook (excluding any sheet named
"Summary")
2. On each sheet, deletes any row where column A is blank
3. Also deletes any row where column B contains the word "DELETE" (case-
insensitive)
4. After cleaning, auto-fits all columns on each sheet
5. Saves the file when done

Requirements:
- Add detailed comments explaining every block of code
- Loop rows from bottom to top (explain why this matters for deletion)
- Show me where to paste the macro (which module) and how to run it
  (keyboard shortcut, Developer tab, etc.)
- Add basic error handling so it doesn't crash if a sheet is protected
- How do I assign this macro to a button on the sheet so non-technical
  colleagues can run it?
```

Tags: VBA · automation · macro · loops · error handling

THE ANALYST'S AI PROMPT HANDBOOK

SQL

Write a cohort retention query

Calculate week-over-week or month-over-month retention.

PROMPT

```
Write a SQL query to calculate weekly cohort retention.

Table: user_events
Columns: user_id (STRING), event_date (DATE), event_type (STRING)

A user is "retained" in week N if they had any event in that calendar
week,
regardless of event_type. Define week 0 as the user's first-ever activity
week.

The output should show:
- cohort_week: the week the user was first seen (ISO week format)
- retention_week: weeks since first seen (0, 1, 2, ... up to 12)
- cohort_size: total users in the cohort
- retained_users: users still active in that retention week
- retention_rate: retained_users / cohort_size as a percentage

Additional requirements:
1. Write this in BigQuery SQL syntax using DATE_TRUNC and WITH clauses.
2. Also show me how the query changes for monthly cohorts.
3. Explain the self-join or window function approach you used, and why.
4. How would I modify this to track retention on a specific event type
only
   (e.g., only "purchase" events count as retention)?
```

Tags: retention · cohorts · BigQuery · analytics

Find duplicates in a table

Identify and optionally remove duplicate rows.

PROMPT

```
I have a PostgreSQL table called customers with columns:
id (SERIAL), email (VARCHAR), name (VARCHAR), created_at (TIMESTAMP),
source (VARCHAR)

Step 1 — Identify:
Write a query to find all rows where email appears more than once.
Show: email, count of duplicates, and the list of IDs sharing that email.

Step 2 — Preview before deleting:
Write a SELECT query that shows exactly which rows would be deleted —
keeping only the earliest created_at record per email, removing the rest.
I want to review before I delete anything.

Step 3 — Delete:
Write the DELETE query (using a CTE or subquery) that executes the dedup
logic.
Make it safe: wrap it in a transaction so I can ROLLBACK if something
looks wrong.

Step 4 — Prevent future duplicates:
Write the ALTER TABLE statement to add a UNIQUE constraint on email.
What happens to existing duplicates when I try to add this constraint?
How do I add it as DEFERRABLE if needed?
```

Tags: deduplication · data quality · DELETE · PostgreSQL · constraints

Running totals and window functions

Add cumulative sum, rank, or row number to a result set.

PROMPT

```
I have a sales table with columns: sale_date (DATE), region (VARCHAR),
revenue (DECIMAL).
Write SQL (Snowflake syntax) to add these calculated columns to each row:

1. cumulative_revenue: running total of revenue by region, ordered by
sale_date
2. daily_rank: rank of each day within its region by revenue (1 = highest,
ties share rank)
3. rolling_7d_avg: 7-day rolling average revenue per region (including
current day)
4. pct_of_region_total: what percentage of the region's all-time revenue
this row represents

For each window function:
- Write the full OVER() clause with PARTITION BY and ORDER BY explained
- Explain the ROWS vs RANGE frame clause difference, and which applies
here
- Show me what changes if I want DENSE_RANK instead of RANK for #2
- For the rolling average, what happens on the first 6 days when there
  aren't 7 prior rows yet? How do I control that behaviour?
```

Tags: window functions · running total · Snowflake · analytics

Optimise a slow query

Diagnose and fix a query that's timing out or running slow.

PROMPT

```
This query is taking 40+ seconds in PostgreSQL:

[PASTE YOUR QUERY HERE]

Context:
- Table: orders (~50M rows), columns: id, customer_id, created_at, status, amount
- Table: customers (~2M rows), columns: id, email, region, tier
- No indexes currently exist beyond primary keys
- Running on PostgreSQL 15, RDS instance type db.r6g.2xlarge

Please:
1. Identify the most likely bottlenecks based on common patterns in slow queries.
2. Recommend specific indexes to add — write the exact CREATE INDEX statements,
   including whether to use btree, hash, or partial indexes for each case.
3. Show me the rewritten query using best practices (avoid SELECT *, push filters
   early, avoid functions on indexed columns in WHERE, etc.)
4. Explain what EXPLAIN ANALYZE output to look for: what does "Seq Scan" vs
   "Index Scan" mean? What are alarming row estimate mismatches?
5. Are there any query-level settings (work_mem, parallel workers) that could
   help without schema changes?
```

Tags: performance · indexing · EXPLAIN ANALYZE · PostgreSQL

Pivot rows to columns

Reshape long-format data into a wide crosstab.

PROMPT

```
I have a table with columns:
- user_id (INT)
- month (VARCHAR, values: 'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun')
- revenue (DECIMAL)

Goal: pivot so each month becomes its own column:
user_id | Jan | Feb | Mar | Apr | May | Jun

Please provide:
1. BigQuery SQL using the PIVOT operator — with the full syntax explained.
2. A CASE WHEN version for databases that don't support PIVOT
   (e.g., MySQL, older PostgreSQL) — show the full query.
3. How to handle months that may have no data for some users
   (should show 0 or NULL — let me control which).
4. How to make this dynamic if the number of months isn't fixed —
   i.e., the month values come from the data itself, not hardcoded.
   (BigQuery dynamic PIVOT or a stored procedure approach.)
5. How to reverse this — i.e., unpivot a wide table back to long format
   using UNPIVOT or CROSS JOIN UNNEST.
```

Tags: pivot · reshape · crosstab · BigQuery · dynamic SQL

Build a funnel conversion query

Count users at each step of a multi-step funnel.

PROMPT

```
I have an events table:
- user_id (STRING)
- event_name (VARCHAR)
- event_timestamp (TIMESTAMP)
- session_id (STRING)

Funnel steps (in order):
'page_view' → 'signup_start' → 'signup_complete' → 'first_purchase'

Write SQL (Redshift syntax) that:
1. Counts unique users who reached each step
2. Shows the conversion rate from the previous step (step-over-step %)
3. Shows the overall conversion rate from step 1 to step 4
4. Enforces strict ordering: a user counts at step N only if they
   completed step N-1 first (earlier timestamp)
5. Allows a configurable time window: users must complete all steps
   within 7 days of their first page_view

Bonus:
- Show how to segment this funnel by a user property (e.g., traffic source
  or device type) if I JOIN to a users table.
- What changes if I want to allow users to revisit steps (non-strict
funnel)?
```

Tags: funnel · conversion · product analytics · Redshift

THE ANALYST'S AI PROMPT HANDBOOK

Python / R

Clean and reshape a DataFrame

Fix messy CSVs, rename columns, handle nulls, change dtypes.

PROMPT

```
I have a pandas DataFrame loaded from a CSV with these issues:

- Column names have spaces, mixed case, and special characters (e.g.,
"Total Revenue ($)")
- The 'date' column is a string in format "DD/MM/YYYY" with some entries
as "DD-MM-YYYY"
- The 'revenue' column has "$1,234.56" strings mixed with plain floats and
NaN
- ~5% of rows have nulls in the 'region' column
- The 'customer_id' column has leading/trailing whitespace in some entries
- There's a 'status' column with values like "Active", "active", "ACTIVE"
(inconsistent case)

Write Python code to:
1. Standardise all column names to snake_case (handle spaces, brackets,
special chars)
2. Parse dates flexibly, handling both DD/MM/YYYY and DD-MM-YYYY
3. Clean revenue to float (strip "$", ",", handle NaN gracefully)
4. Fill region nulls with "Unknown"
5. Strip whitespace from customer_id and standardise status to title case
6. Print before/after dtypes and a .head(5) comparison
7. Add assertions at the end to validate the cleaning worked correctly

Also: should I do this cleaning in pandas or in the SQL layer before
loading?
When is each approach better?
```

Tags: pandas · data cleaning · dtypes · validation

Merge multiple CSVs into one

Combine files from a folder with consistent or mismatched schemas.

PROMPT

```
I have a folder /data/monthly_reports/ with 24 CSV files,
each named like "2023_01.csv", "2023_02.csv", ..., "2024_12.csv".

Write Python code using pandas and pathlib to:
1. Read all CSVs into one DataFrame with pd.concat (not a loop of appends)
2. Add a 'source_file' column showing the original filename for each row
3. Parse the filename to extract 'year' (int) and 'month' (int) columns
4. Handle schema mismatches: if some files have extra or missing columns,
   align them gracefully (fill missing columns with NaN, don't crash)
5. Report which files were loaded and how many rows each contributed
6. Export the combined DataFrame as /data/combined.parquet
   (explain why Parquet instead of CSV for downstream use)

Edge cases to handle:
- What if one of the files is empty (0 rows)?
- What if a file has encoding issues (latin-1 vs UTF-8)?
- How do I make the glob pattern future-proof if new years are added?
```

Tags: pandas · pathlib · file I/O · parquet · automation

Run a quick regression / model

Fit a linear or logistic regression and interpret the output.

PROMPT

```
I have a pandas DataFrame with columns:
- marketing_spend (float): weekly ad spend in USD
- seasonality_index (float): 0–1 normalised seasonal factor
- competitor_price (float): average competitor price that week
- promotions_active (int): 1 if a promotion ran that week, else 0
- sales (float): weekly units sold — this is the target variable

Write Python using scikit-learn to:
1. Split into train/test sets (80/20, random_state=42), stratify if applicable
2. Fit a linear regression, print R², RMSE, and MAE on both train and test
3. Print a coefficient table showing: feature name, coefficient, and
   a plain-English interpretation of each (e.g., "each $1k increase in
   marketing_spend is associated with X more units sold")
4. Plot predicted vs actual values with a 45-degree reference line
5. Plot a residuals chart to check for patterns

Then explain:
- What does a high train R² but low test R² indicate?
- Should I scale the features? What difference does it make for linear regression?
- When would I use Ridge or Lasso instead of plain linear regression?
```

Tags: scikit-learn · regression · modelling · interpretation

Automate an Excel report with openpyxl

Generate formatted Excel files from Python without manual work.

PROMPT

```
Write Python using openpyxl to create an Excel report from a pandas
DataFrame called df with columns: Region, Product, Revenue, Units, Month.

The Excel output should:
1. Have a bold header row with dark navy background (#1F2D5A) and white
text,
   font size 11, all caps
2. Auto-fit column widths (write the helper function to calculate this
correctly)
3. Add a totals row at the bottom for Revenue and Units with SUM formulas
   (use actual Excel formulas, not hardcoded Python values)
4. Format the Revenue column as currency ($#,##0.00) and Units as integer
5. Add alternating row shading (white / light grey) for readability
6. Add a bar chart of Revenue by Region on a separate "Charts" sheet,
   with a chart title, axis labels, and no gridlines
7. Freeze the top row so headers are always visible
8. Save as "weekly_report.xlsx"

Also: how do I add a drop-down validation list to the Region column using
openpyxl's DataValidation?
```

Tags: openpyxl · automation · Excel · charts · formatting

Scrape a table from a webpage

Extract structured data from a public web page.

PROMPT

```
Write Python code to scrape the first HTML table from this URL:
[PASTE URL]

Requirements:
1. Use requests with a proper User-Agent header and a timeout of 10 seconds
2. Use BeautifulSoup to parse the HTML and find the first <table> element
3. Parse it into a pandas DataFrame, including handling multi-row headers
   (if the table uses <thead> with multiple rows)
4. Clean column headers: strip whitespace, lowercase, replace spaces with underscores
5. Save as a CSV to ./output/scraped_table.csv
6. Handle errors gracefully:
   - Connection timeout or network error → print a clear message, exit cleanly
   - 403 Forbidden → suggest adding headers or using a delay
   - Table not found on page → explain what to do next
7. Add comments explaining each step

Bonus: how would this code change if the table is loaded dynamically by
JavaScript (i.e., not in the raw HTML)? What tool would I use instead?
```

Tags: web scraping · BeautifulSoup · requests · error handling

Visualise data with matplotlib / seaborn

Create a chart that's presentation-ready, not just exploratory.

PROMPT

```
I have a pandas DataFrame with columns:
- month (Jan–Dec as strings, in order)
- channel (values: "Organic", "Paid", "Referral", "Direct")
- revenue (float)

Write Python using seaborn and matplotlib to create a grouped bar chart
showing revenue by channel for each month.

The chart must be presentation-ready:
1. Professional colour palette — not the default seaborn colours.
   Suggest and use a palette that works well in both print and on screen.
2. Value labels above each bar (e.g., "$12.4K") — write the label
formatting helper
3. Clean grid: horizontal gridlines only, no top or right spines (despine)
4. Legend placed outside the plot area (top-right, no box border)
5. X-axis shows month abbreviations, properly ordered Jan→Dec
6. Chart title, x-axis label, and y-axis label (with $ and K formatting)
7. Figure size suitable for a 16:9 slide
8. Export as "revenue_chart.png" at 300 DPI for print quality

Also: when would I use seaborn vs plotly for this type of chart?
What's the main trade-off?
```

Tags: matplotlib · seaborn · visualisation · presentation · export

Power BI Tools

Write a Looker / LookML measure

Define a custom metric, ratio, or filtered measure in LookML.

PROMPT

```
I'm writing LookML for a Looker view called "orders".
The view has these existing dimensions: id, customer_id, created_date,
status (values: 'completed', 'refunded', 'cancelled'), revenue,
item_count.

Write LookML measures for:
1. total_revenue: SUM of revenue across all orders
2. average_order_value: total_revenue divided by count of orders,
   formatted as currency with 2 decimal places
3. completed_revenue: revenue from orders where status = 'completed' only
   (use a filtered measure, not a separate dimension)
4. refund_rate: percentage of orders with status = 'refunded'
5. 30-day rolling revenue: if this is achievable in LookML, show how —
   and if not, explain the limitation and the workaround

For each measure:
- Include label, description, and value_format_name or value_format
- Explain when to use measure type: sum vs type: number for ratios
- Flag any measures that may cause fanout issues on joined views
```

Tags: Looker · LookML · measures · filtered metrics

Write a Tableau calculated field

Build a LOD expression, table calc, or conditional formula.

PROMPT

```
I'm working in Tableau Desktop (version 2023.3+) with an Orders data
source.
Fields: Order Date, Region, Customer ID, Revenue, Category, Profit.

Write calculated fields for:
1. Revenue % of Total: each row's revenue as a % of the grand total.
   Use a FIXED LOD expression. Explain why FIXED and not INCLUDE or
EXCLUDE.
2. Running Revenue by Month: cumulative revenue month-by-month using
   a RUNNING_SUM table calculation. Explain: what does "compute using"
mean
   and which setting do I choose for this to work correctly in a bar
chart?
3. Customer Value Tier: classify customers as "High" (>$10k total spend),
   "Mid" ($2k–$10k), or "Low" (<$2k) using a FIXED LOD + CASE WHEN.
4. Profit Margin %: profit divided by revenue, formatted as a percentage.
   What happens when revenue is zero — how do I handle divide-by-zero?

For each field: explain where to drag it in the view, and when to use
LOD expressions vs table calculations vs row-level calculations.
```

Tags: Tableau · LOD · table calc · FIXED · calculated fields

Write a Power BI DAX measure

Calculate a KPI, ratio, or time-intelligence metric in DAX.

PROMPT

```
I'm in Power BI Desktop. My data model has:
- Fact table: Sales (Date [DATE], ProductID, CustomerID, Revenue, Units,
Cost)
- Dimension: Calendar (Date, Year, Month, MonthName, Quarter, WeekNum)
  — marked as a Date Table
- Dimension: Products (ProductID, Category, SubCategory, UnitPrice)
- Dimension: Customers (CustomerID, Region, Segment, Tier)

Write DAX measures for:
1. [Total Revenue]: sum of Revenue
2. [Revenue MTD]: month-to-date revenue using time intelligence
3. [Revenue SPLY]: revenue for the same period last year
4. [YoY Growth %]: year-on-year percentage change with a DIVIDE for safety
5. [Gross Margin %]: (Revenue - Cost) / Revenue, formatted as a percentage
6. [Active Customers]: count of distinct customers with at least one sale
   in the selected period

For each measure:
- Write the full DAX with comments
- Explain any CALCULATE, FILTER, or ALL modifiers used
- Flag which measures require the Calendar table to be marked as a date
table
- Show how to format the measure correctly in the Modeling pane
```

Tags: Power BI · DAX · time intelligence · CALCULATE · measures

Set up a Metabase / Superset filter

Configure cross-filter logic or dashboard variables.

PROMPT

```
I'm using Metabase (version 0.47+) with a dashboard containing 4 charts:
1. A bar chart: Revenue by Month
2. A line chart: Orders over Time
3. A table: Top 10 Products by Revenue
4. A KPI tile: Total Active Customers

All 4 charts are based on the same "orders" table in our PostgreSQL
warehouse.
The table has columns: order_id, region, product_category, order_date,
revenue.

Walk me through setting up:
1. A "Region" dropdown filter that simultaneously updates all 4 charts —
   including which filter type to choose (Field Filter vs. Category), why,
   and how to map it to each question's field.
2. A date range filter (start date / end date) that also applies to all
charts.
3. Common mistakes to watch out for — e.g., filter mapping to the wrong
field
   type, charts not updating, questions using native queries vs. GUI
editor.
4. How to set a default value for the Region filter so the dashboard loads
   showing "All Regions" by default.
5. Can filters in Metabase cross-filter between charts (i.e., clicking a
bar
   filters other charts)? If not, what's the workaround?
```

Tags: Metabase · filters · dashboards · cross-filtering

Design a KPI dashboard layout

Get a structure and metric hierarchy for an exec-facing dashboard.

PROMPT

```
I'm building a weekly executive dashboard for a B2C e-commerce company.
Tool: [Tableau / Power BI / Looker — choose the most common approach].
Audience: C-suite and VP-level, reviewed every Monday morning.
Available metrics: Revenue, Orders, AOV (Average Order Value), New vs Returning
customer split, Conversion Rate, Cart Abandonment Rate, Top 5 SKUs by revenue,
Return Rate, Marketing Spend, CAC (Customer Acquisition Cost).

Design the dashboard with:
1. A layout recommendation: what goes at the top (hero KPIs), middle (trend charts),
   and bottom (detail tables / breakdowns). Explain the visual hierarchy logic.
2. For each of the 10+ metrics: what chart type to use (KPI tile, line, bar,
   table, etc.) and why. Which metrics need WoW / MoM / YoY comparisons shown?
3. Calculations for non-trivial metrics: AOV, Conversion Rate, CAC.
   What are the common definition disputes around these — and how to document
   the agreed definition in the dashboard?
4. What to leave out: which metrics look important but add noise at the exec level?
5. Colour and alert conventions: when should a number be red vs green?
   (Revenue down = bad, Return Rate down = good — how do I handle directional logic?)
```

Tags: dashboard design · KPIs · executive reporting · layout

THE ANALYST'S AI PROMPT HANDBOOK

Slides / Decks

Structure a business case deck

Get a slide-by-slide outline for a persuasive business case.

PROMPT

```
I need to build a business case presentation recommending [YOUR
INITIATIVE]
to senior leadership. The ask is [BUDGET / HEADCOUNT / APPROVAL].

Write a detailed slide-by-slide outline following this structure:
- Slide title
- 3-4 bullet points of what to cover on this slide
- The ONE key message this slide must land (the "headline")
- Suggested chart or visual type (if applicable)

The narrative arc should follow: Problem → Insight → Options Considered →
Recommended Approach → Financial Case → Risks & Mitigations → The Ask.

Additional requirements:
1. Keep it to 8-10 slides. Audience is time-poor and sceptical — no fluff.
2. Include a "Pre-mortem" slide (what could go wrong) — explain why this
builds
   credibility rather than undermining the case.
3. Suggest what to put in the appendix vs the main deck.
4. What's the single most common reason business cases get rejected at
this level,
   and how should the deck pre-empt it?
5. How do I tailor the opening slide differently for a CFO vs a COO
audience?
```

Tags: storytelling · business case · structure · exec communication

Rewrite bullets into executive language

Transform verbose analyst bullets into punchy exec-ready copy.

PROMPT

```
Rewrite these slide bullets into tighter, exec-friendly language.

Rules for each rewritten bullet:
- Maximum 12 words
- Lead with the insight or outcome, not the method or process
- Use active voice ("Revenue grew" not "Revenue was observed to grow")
- Quantify wherever possible — if a number is missing, flag it with
[NUMBER]
- Avoid jargon: no "leveraged", "utilised", "synergised", "deep-dived"

Here are the current bullets:
[PASTE YOUR BULLETS HERE]

After rewriting, also:
1. Flag any bullets that are pure observations with no "so what" —
   and for each one, suggest what the implication or recommendation should
be.
2. Identify any bullets that are actually two separate points crammed
together
   and split them.
3. Rank the rewritten bullets by impact for this specific audience
   (most important first).
4. Suggest a single headline sentence (max 15 words) that could sit at the
   top of the slide and summarise all the bullets.
```

Tags: copywriting · exec communication · clarity · bullets

THE ANALYST'S AI PROMPT HANDBOOK

Write the 'so what' for an analysis

Turn data findings into a clear, actionable recommendation.

PROMPT

```
Here is my analysis finding:
[PASTE YOUR FINDING — e.g., "Cohort retention drops from 45% to 18% between week 1
and week 2 for users who never completed the onboarding checklist."]

Write 3 versions of the "so what" statement:

Version 1 — For a data / analytics audience:
Include the statistical basis, caveats, and what further analysis is needed.

Version 2 — For a business stakeholder (e.g., Product Manager, Marketing Director):
Focus on business implications, what decision this should inform, and
what would happen if we don't act.

Version 3 — For a C-suite exec:
Max 3 sentences. Lead with the business impact in dollar or percentage terms.
End with a clear, time-bound recommendation.

For each version, answer:
- What does this mean?
- Why does it matter right now?
- What should we do about it, and by when?

Also: what additional data would make this finding more compelling or defensible?
```

Tags: insight writing · recommendation · communication · storytelling

Choose the right chart type

Get a recommendation for how to visualise a specific data scenario.

PROMPT

```
I want to visualise the following data:
[DESCRIBE YOUR DATA — e.g., "Monthly revenue for 5 product categories over 24 months,
I want to show both the trend and the composition."]

Please:
1. Recommend the best chart type for this scenario and explain why it's the
   right choice (what visual question it answers most clearly).
2. List 2–3 alternative chart types with their trade-offs:
   - What each one does better
   - What each one obscures or makes harder to read
   - When you'd choose each one instead
3. Sketch the structure of the recommended chart:
   - What goes on the X and Y axes
   - How the data needs to be structured (long vs wide format)
   - What colours, labels, and annotations to include
4. What are the most common mistakes people make with this chart type?
   (e.g., using 3D bar charts, pie charts with too many slices, dual-axis charts
   that mislead)
5. Does the answer change if this chart is for a live dashboard vs a static slide?
```

Tags: data viz · chart selection · design · best practices

Write slide commentary / speaker notes

Turn a data slide into a clear narrative for the presenter.

PROMPT

```
Here is what's on my slide:

Title: [YOUR SLIDE TITLE]
Chart type: [DESCRIBE THE CHART — e.g., "Line chart showing weekly active users
Jan–Dec 2024, with a dip in March and a sharp rise from September"]
Key numbers: [PASTE YOUR KEY DATA POINTS]
Audience: [WHO WILL SEE THIS]

Write speaker notes for this slide targeting 2–3 minutes of talking time.
Structure them as:

Opening hook (15 seconds): a question or statement that frames why this slide matters.
What the data shows (45 seconds): walk through the chart, guide attention to the
key features (highs, lows, inflections).
The key insight (30 seconds): the ONE thing the audience must remember.
Caveats (20 seconds): what this data doesn't tell us / known limitations.
Transition (10 seconds): a bridge sentence leading naturally to the next slide.

Format: write in natural spoken language — contractions, short sentences, pauses
indicated by "..." — not formal written prose. It should sound like a person
talking, not reading.

Also flag any data points that are likely to get challenged with a suggested response.
```

Tags: speaker notes · presentation · narrative · delivery

THE ANALYST'S AI PROMPT HANDBOOK

General Analysis

Sanity-check my analysis

Get a structured critique of your methodology and conclusions.

PROMPT

```
I'm about to share the following analysis with stakeholders. Before I do,
critique it rigorously:

[PASTE YOUR ANALYSIS OR DESCRIBE YOUR APPROACH AND CONCLUSION IN DETAIL]

I want you to act as a senior data scientist reviewing my work. Look
specifically for:

1. Logical gaps: are the conclusions I've drawn actually supported by the
data shown?
2. Confounding factors: what other variables could explain the pattern I'm
seeing,
   that I haven't controlled for?
3. Selection bias: how was the data filtered or sampled, and could that
skew results?
4. Correlation vs causation errors: am I implying causation where I only
have
   correlation evidence?
5. Sample size and statistical significance: are the numbers large enough
to
   support the claims, or am I reading signal in noise?
6. Time period issues: is my date range cherry-picked, too short, or
affected
   by a one-off event?
7. Definition inconsistencies: are all metrics defined consistently
throughout?

After the critique:
- Rate the overall robustness of the analysis (High / Medium / Low) with
reasoning.
- Suggest the 3 most impactful additions that would strengthen the
conclusion.
- Identify the 2 questions most likely to be asked by a sceptical
stakeholder
  and suggest how to answer them.
```

Tags: QA · methodology · critical thinking · peer review

Define a metric from scratch

Get the formula, edge cases, and caveats for a new business metric.

PROMPT

```
I need to define a new metric for: [METRIC NAME — e.g., "Customer Health Score",
"Product-Led Growth Score", "Supplier On-Time Delivery Rate"].

Context: [DESCRIBE YOUR BUSINESS AND WHAT THIS METRIC SHOULD MEASURE]

Provide a complete metric definition document covering:

1. Official definition: a precise, unambiguous one-sentence definition
   (the kind that goes in a data dictionary)
2. Formula: the exact calculation with variable names defined
3. Data inputs required: which tables, columns, and systems this depends on,
   and potential data quality issues with each
4. Granularity and grain: what is one row of this metric?
   (per user? per day? per user per day?)
5. Edge cases and how to handle them:
   - Null or missing values in inputs
   - New users / zero denominators
   - Negative values (are they valid? what do they mean?)
6. Segmentation: how should this metric be sliced?
   (by region, product, cohort, etc.)
7. Benchmarks and targets: what are typical ranges in [your industry]?
8. Common misreadings: how do people typically misinterpret this metric,
   and what guardrail metrics should sit alongside it?
9. Common gaming risks: how might teams optimise for this metric in ways
   that don't reflect real improvement?
```

Tags: metric design · KPIs · definitions · data dictionary

Write a data request for engineering

Draft a clear spec for a data pipeline or reporting table.

PROMPT

```
Write a data request document for the data engineering team.

I need a new table in our data warehouse with the following requirements:

- Source tables/systems: [LIST YOUR SOURCES — e.g., "Salesforce
opportunities table,
  Stripe payments table, internal CRM via daily CSV export"]
- Grain: one row per [USER / ORDER / DAY / CUSTOMER-PRODUCT COMBINATION]
- Approximate row volume: [ROWS PER DAY / TOTAL ROWS]
- Key output fields: [LIST THE FIELDS YOU NEED]
- Most complex transformation: [DESCRIBE YOUR HARDEST CALCULATION]
- Refresh cadence: daily / SLA: data must be available by 6am UTC
- Consumers: [WHO WILL USE THIS — BI tool, downstream model, spreadsheet?]

Format the document with these sections:
1. Purpose & Business Context (2–3 sentences on why this table is needed)
2. Source Tables (with relevant columns and join keys identified)
3. Output Schema (field name, data type, description, nullable Y/N)
4. Transformation Logic (pseudocode or plain English for each calculated
field)
5. Testing Criteria (how the engineering team should validate correctness
—
   row counts, null checks, referential integrity checks, etc.)
6. Open Questions (flag decisions that need engineering input)
7. Priority & Timeline (suggest how to frame urgency without being
demanding)
```

Tags: data engineering · spec · PRD · collaboration · warehouse

Interpret a statistical output

Understand what a p-value, confidence interval, or model output means.

PROMPT

```
I ran a [t-test / linear regression / chi-square / ANOVA / logistic regression]
and got this output:

[PASTE YOUR STATISTICAL OUTPUT HERE]

Please explain:

1. What each number in the output means in plain English — not textbook definitions,
   but what it means for my specific analysis.
2. Was the result statistically significant? What threshold should I use (p < 0.05
   is the default, but is it always right here)?
3. What does "statistically significant" actually mean in practice — and what does it
   NOT mean? (Common misconceptions around p-values.)
4. What is the practical / business significance of this result?
   Is the effect size large enough to matter, even if p < 0.05?
5. What assumptions does this test make, and based on my data description,
   should I be concerned about any of them?
   (Normality, equal variance, independence, sample size, etc.)
6. If the assumptions aren't fully met, what's the alternative test I should use?
7. How should I report this finding to a non-statistical audience — in one sentence?
```

Tags: statistics · p-value · interpretation · significance · assumptions

Design an A/B test

Calculate sample size, duration, and success criteria for an experiment.

PROMPT

```
I want to run an A/B test on: [FEATURE / CHANGE — e.g., "a new checkout flow",
"a subject line variant in our weekly email", "a new onboarding screen"].

Context:
- Primary metric: [e.g., conversion rate, click-through rate, revenue per user]
- Current baseline value: [e.g., conversion rate = 3.2%]
- Minimum detectable effect (MDE) I care about: [e.g., 0.5 percentage point lift]
- Statistical power: 80%, significance level (alpha): 5% (two-tailed)
- Daily eligible users: [e.g., 8,000 users per day]
- Number of variants: Control + 1 Treatment (or specify more)

Please:
1. Calculate the required sample size per variant, showing the formula used.
2. Estimate the test duration in days given the daily user volume.
3. Define the primary metric and 2–3 guardrail metrics (things that shouldn't
   get worse — e.g., don't improve conversion by harming AOV).
4. Write the statistical hypotheses (H0 and H1) in plain English.
5. Flag the top 5 mistakes analysts make when running A/B tests:
   (peeking early, multiple testing, novelty effects, SUTVA violations, etc.)
6. When should I use a one-tailed vs two-tailed test here, and why?
7. How do I analyse the results once the test concludes — what's the exact
   decision rule for "ship it", "kill it", or "run longer"?
```

Tags: A/B testing · experimentation · sample size · statistics

Anticipate stakeholder questions

Prepare for the tough questions your data will get challenged with.

PROMPT

```
I'm presenting the following analysis to [AUDIENCE — e.g., "the CFO and finance
leadership team", "the product team in our weekly review", "the board"]:

[DESCRIBE YOUR ANALYSIS AND MAIN CONCLUSION IN 3–4 SENTENCES]

Generate the 10 hardest questions they're likely to ask — especially questions
that could undermine my conclusion, expose a methodology flaw, or request
context I haven't provided.

For each question:
1. Write the question as they would actually phrase it (not politely softened)
2. Rate the severity if I can't answer it well: High / Medium / Low
3. Suggest a confident, prepared response (2–3 sentences)
4. Indicate if I need additional data or analysis to answer it properly

At the end:
- Flag the 2 questions I absolutely must have a pre-prepared backup slide for
- Suggest the one thing I should add to my analysis before presenting
  that would neutralise the most dangerous line of questioning
- What's the single "concession" I should proactively make
  (acknowledging a limitation before being asked) to build credibility?
```

Tags: presentation prep · stakeholder management · QA · pre-mortem

THE ANALYST'S AI PROMPT HANDBOOK

Write a data story narrative

Turn a set of findings into a coherent written narrative.

PROMPT

```
I have the following analysis findings:
[PASTE 3–5 KEY DATA POINTS OR FINDINGS — e.g.,
"1. Revenue grew 18% YoY in Q3 but only 4% in Q4.
 2. The drop is entirely driven by the Enterprise segment (-12% QoQ).
 3. SMB and Mid-Market both grew QoQ.
 4. Enterprise churned 3 large accounts in October, representing $2.1M
ARR.
 5. No new Enterprise logos closed in Q4 despite a strong pipeline at the
start."]

Write a short narrative (4–5 paragraphs) that weaves these into a coherent
story
using the Situation → Complication → Resolution structure.

Requirements:
1. Lead with the most important finding, not the chronology of how I found
it.
2. Audience: business stakeholder, non-technical, time-poor.
3. Use the active voice. Write like a journalist, not an academic.
4. Avoid analyst hedging language ("it appears that", "this may suggest")
—
   state findings directly while flagging genuine uncertainty where it
exists.
5. End with a clear recommended action or decision, not an open question.

After the narrative, also:
- Write a single subject line (10 words max) I could use to send this as
an email.
- Suggest one chart that would most effectively accompany this narrative.
```

Tags: storytelling · writing · communication · narrative · SCR

These prompts are starting points, not scripts. The more context you give — your actual table names, column headers, business definitions, and specific goal — the better the output will be. Treat each prompt as a template to personalise.