

1. The Core Problem & Goal

- I need a centralized sales and marketing funnel that I can track to understand and engage some sense of business.
- The main issue, which is the most critical to me, is there is no unified view or source of truth that I can refer to at all times and feel confident about the data that I am looking at.
- The problem with Google Scripts is Google has made its rate limits quite stringent recently, making it impossible for me to use solely that for calling all the data in one, as it fails daily.
- I am unable to perform any actions, which is why I need another solution, not Google Script.
- The desired end state would be to have the final data that I can refer to when creating dashboards.
- Right now, all I need is the complete raw data ready to use any way I want.

2. Organization Context

- I am working on is a problem at my little organization.
- It has very few employees, but it is an ed-tech organization based out of India, and we offer international financial certifications as well as a few other programs.

3. The Sales Funnel to Be Tracked

This is more or less the sales funnel:

- **Top of the funnel (TOFU):**
 - I need to track the top funnel.
 - Top of the funnel is performance marketing.
 - I need to understand: How many leads I get from my top funnel when I run a marketing campaign.
- **Communication Layer (Out of Scope):**
 - Right next to that, what happens is those users get some communications from us via email or WhatsApp. This is not something that I am needing to track right now.
- **Middle of the funnel (MOFU):**
 - Then there is some middle engagement (webinar or something).
 - How many of those leads actually register for the next event (which is usually the webinar).
 - How many of those registered webinar leads actually attend the webinar.
- **Bottom of the funnel (BOFU):**
 - Finally, there is a bottom of the funnel which is the attended users are handed over to sales team.
 - Once the webinar is attended, the lead is sort of handed over to the sales team.
 - When the sales team gets those leads, they call them, they convert it.
 - This is regarding the bottom of the funnel.

4. Data Sources & Current Processing

A. TOFU: Performance Marketing

- For performance marketing or the top of the funnel, the data is captured in Google Sheets.
- Whenever a campaign is run, all the leads captured are saved in Google Sheets automatically.
- Performance Marketing Lead Capture - Make.com automation brings leads to google sheet. It is working fine - leads are coming as expected.
- This sheet has usually has columns like these -
 - Name
 - Email
 - Phone number
 - City
 - Question 1
 - utmSource
 - utmMedium
 - utmCamp
 - created date
- Performance Marketing Spend Data - this sheet is photo updated and made public it can be accessed directly.

B. MOFU: Zoom Events (Webinars & Boot Camps)

- For the middle of the funnel, we'll have Zoom webinars.
- Within Zoom webinars, there will be two events that will be captured:
 1. Webinar registration
 2. Webinar attendance
- What usually happens is I download a report for webinar registration as well as webinar attendance. The format for both these reports is different.
- The information that I usually capture is this -
 - Webinar Date
 - Webinar ID
 - Attended
 - User Name (Original Name)
 - First Name
 - Last Name
 - Email
 - Phone
 - Registration Time
 - Registration Source
 - Attendance Type
 - Join Time
 - Leave Time
 - Time in Session (minutes)
 - Country/Region Name
 - Webinar name
 - Webinar conductor
- Another event that we usually capture... would be boot camps. They are very similar to

Zoom webinars but they span over a course of two days.

- We also capture boot camp day, i.e., which day did the user attend? Was it day one or day two, or was it both days?

C. BOFU: Sales CRM

- After this, the leads are handed over to the sales team, and all of the data after that is available in our CRM and can be accessed via API calls.
- So we do not need to change anything; we can just use APIs to call the data and work from there.

5. Lead Attribution & Tagging

- At this stage, we also label each of our leads so that they can be tagged to a "source".
- There can be multiple sources, for example:
 - This webinar source
 - Bootcamp
 - A few others as well
- One thing you have to understand is Zoom webinar or boot camp is just one or two of the sources that I'm talking about. There will be a few other sources as well, like CCC (Commerce Career Club), another one would be eWebinar, Seminar, etc.
- I know exactly what the conversion ratio for each of the sources is, as the engagement matrix or the processes in between at the middle of the final stage are different for each of these.
- Users follow different journeys for different sources.
- Lead attribution is also very complex and follows very specific logics that I have built already.
- I will use mobile phone number as the primary key, not email. Email can be used complementarily to fetch unique users at times, but my primary key will always remain the phone number.

6. Critical System Architecture: The Two-Part System

- One very critical piece of information that I want you to keep in mind is that you should treat this as a two-part system:
- **Part 1: Top → Middle Funnel**
 - From top funnel to middle, which includes your performance marketing data, your Zoom webinars, and everything, all the events that happen in between.
 - Zoom Data Processing Flow... It will give me a much more clearer picture of the marketing funnel.
- **Part 2: Sales Funnel (Post-Assignment)**
 - Once the lead is assigned, that should be considered part of phase II.
 - So once the leads are handed over to sales team, this data is now being called via API into Google Sheets as per my current process, and then cleaned and manipulated.
- **Integration & Separation**
 - I am looking to build both these parts separately.
 - Attribution in Part 2 - the CRM stores assigned lead sources. Do not need to look

- up the leads for number from Part 1. Hence, treating Part One and Part Two separately.
- I have not thought of an integration between the two, but since these two systems will be independent, we can later figure out an integration.
 - Between the two, the integration will likely be based on phone number as a primary key, but that phone number will be normalized.
 - This will mean the user ID (generated by cleaning & normalizing phone number) in the second part will be the foreign key for the normalized phone number from Pt1.

7. Existing Assets & Reusable Logic

- I have a script already made for this that cleans the Zoom webinar file that it gets as input and outputs the clean version of the file with the required fields pre-filled.
- This data cleaning and enrichment step is done already.
- The system that I have created is live on Streamlit UI, but it is not a full-fledged system. So, we can use some components of that and build on top of it.
- I am sharing the Streamlit app script that I currently use for cleaning and enriching data. That you will be able to find the logics there.
- I am attaching my manual workflow steps to clean and calculate data that is needed for my sales and Funnels dashboard.
- I'm attaching the funnel and the appends that I use.

8. Technical Constraints & Manual Steps

- **Google Sheets Blocker:**
 - The problem with Google Scripts is Google has made its rate limits quite stringent recently, making it impossible for me to use solely that for calling all the data in one, as it fails daily.
- **CRM (In-House):**
 - The CRM that we use is an in-house one. We have developed it ourselves, so any plugin/play thing would not be available.
 - The API calls are easily accessible though.
 - The CRM API that I use doesn't necessarily need any complexity or rate limits. It works the way that I am using it right now without facing any problems.
- **CRM API Calls:**
 - These are two separate endpoints without any authentication complexity. It takes about a minute or so to fetch all the data.
 - Just running this API endpoint URL basically downloads a CSV file.
 - API calls usually return full dumps instead of data range filters.
- **Zoom API Constraint:**
 - The Zoom limitation is that I do not have access to create apps. Neither do the scopes are configured... we have a specific contract with Zoom that does not allow us to do that.
 - Zoom APIs do not work for our specific system. A bit of manual effort will be required.
- **Manual Processes (Acceptable):**
 - Lead Assignment Trigger - this is done manually. No intervention required.
 - The API calls might need manual trigger. I do not know... I have scheduled them to

trigger every day. That worked... before the Google rate limits hindered my work.

- Manual mapping is run once every month; it is not a regular thing that we do on a daily basis.
- This [Monthly Manual Mapping Workflow] does not need to be included in the system that we are building. This can be a final layer that we manually perform after all the raw data is ready.

9. Solution Requirements

- **Data Freshness:**

- The data updates don't need to be real-time in any way. I need everything to be up to date only till D-1.
- So if I am looking at the data on the 4th of November, I need all the data till the 3rd of November.
- For those purposes, I think syncing everything once every day in the morning or in the last night, and will work just fine!

- **Flexibility:**

- I'd like this to be a slightly dynamic and flexible. If I want to add new funnels or remove existing ones, I want to be able to do that.
- Some of the sources in this funnel mapping sheet have been deprecated, for example -ACCA_FinProClub, ACCA_Influencer, Which is why I insisted on a solution, which is flexible.

- **Infrastructure & Cost:**

- I do not have a hosting or infrastructure preference per se. I want a simple solution that is easy to maintain.
- It would be better if it is open source or free. I think the free tier should cover my use case if we decide to go that route.

- **Final Output Format:**

- I don't have a preference for that, but a single Google sheet for final output might not be possible due to Google Apps Script rate limits. I do not know if rate limits will be applicable for that thing.
- I am okay with having CSV files in the cloud storage or having a new database. Anything works as long as it is simple and easy to maintain, Without needing too much technical intervention at my end.