Summary:

Scope of work: What you need to know

As a data analyst, it's hard to overstate the importance of an SOW document. A well-defined SOW keeps you, your team, and everyone involved with a project on the same page. It ensures that all contributors, sponsors, and stakeholders share the same understanding of the relevant details.

Why do you need an SOW?

The point of data analysis projects is to complete business tasks that are useful to the stakeholders. Creating an SOW helps to make sure that everyone involved, from analysts and engineers to managers and stakeholders, shares the understanding of what those business goals are, and the plan for accomplishing them.

Clarifying requirements and setting expectations are two of the most important parts of a project. Recall the first phase of the Data Analysis Process—asking questions.

What is a good SOW?

There's no standard format for an SOW. They may differ significantly from one organization to another, or from project to project. However, they all have a few foundational pieces of content in common.

- Deliverables: What work is being done, and what things are being created as a result of this project? When the project is complete, what are you expected to deliver to the stakeholders? Be specific here. Will you collect data for this project? How much, or for how long?
- Milestones: This is closely related to your timeline. What are the major milestones for progress in your project? How do you know when a given part of the project is considered complete?
- Timeline: Your timeline will be closely tied to the milestones you create for your project. The timeline is a way of mapping expectations for how long each step of the process should take. The timeline should be specific enough to help all involved decide if a project is on schedule. When will the deliverables be completed? How long do you expect the project will take to complete? If all goes as planned, how long do you expect each component of the project will take? When can we expect to reach each milestone?
- Reports: Good SOWs also set boundaries for how and when you'll give status updates to stakeholders. How will you communicate progress with stakeholders

and sponsors, and how often? Will progress be reported weekly? Monthly? When milestones are completed? What information will status reports contain?

At a minimum, any SOW should answer all the relevant questions in the above areas. Note that these areas may differ depending on the project. But at their core, the SOW document should always serve the same purpose by containing information that is specific, relevant, and accurate. If something changes in the project, your SOW should reflect those changes.

Context can turn raw data into meaningful information. It is very important for data analysts to contextualize their data. This means giving the data perspective by defining it. To do this, you need to identify:

- Who: The person or organization that created, collected, and/or funded the data collection
- What: The things in the world that data could have an impact on
- Where: The origin of the data
- When: The time when the data was created or collected
- Why: The motivation behind the creation or collection
- How: The method used to create or collect it

First I understood the goal of the analysis and then figured out what steps we should take in order to achieve the result as quickly and efficiently as possible. I asked questions like:

> What is the first thing we should do after getting the data? it cleaning the data

> Do we have a parameter to differentiate which video goes in list and which doesn't ? if not create one!

> We should have some spare time to make up for blunders if any occured.

Data Analysis Project

Data Analyst: *SirFarkade*

Client/Sponsor:
Youtube

Purpose:

To clean the dataset and then find perform Exploratory Data Analysis

Scope / Major Project Activities:

What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.

Activity	Description
Data Cleaning	We'll figure out what columns we need for the analysis, delete the rest and then assign them proper data types, delete null rows and value missing rows.
Parameter Selection	Selecting best possible parameter which will help us indicate why the video is trending and will other type of video with same stats trend or not
Analysis	Calculate the agreed upon stats for each video
Predictions	Predict whether the target video will land in the trending list or not

This project does not include:

Specify the things that this project isn't responsible for doing (out of scope). For instance, "this project does not involve a summation of 2019 data analysis"

- This project does not include live statistics about the data
- It won't work if the dataset used is much older
- Does not considered whether videos are promoted or not

Deliverables:

A specific list of things that your project will deliver.

Deliverable	Description/ Details
Analysis	*Calculating stats of videos that went into the trending list*
Class wise	*It works on class basic and isn't same for all*
Predictions	*Predict whether the target video can go or not*

Schedule Overview / Major Milestones:

The expected schedule for the project. This can be defined by milestones (e.g. "all data is cleaned and processed"), periods of time ("Week 1 / Week 2"), or other ways based on the needs of the project.

Milestone	Expected Completion Date	Description/Details
Data Cleaning Report	*15/01/2023*	*Basic data cleaning*
Parameter Selection	*30/01/2023*	*With at least 2 group meeting we should decide upon which parameter to use for calculating the insights after some research*
Analysis	*20/02/2023*	*Calculating the parameter and testing for predictions, if the predictions are off by a long shot iterate the selection process*
Wait Period	*05/03/2023*	*Iteration period*
Final Report	*5/03/2023*	*Predictions on the target*

		video

Learning Log: Define problems and ask questions with data

Date: 04/01/2023	**Course/topic:** Course 2: Ask Questions to Make Data-Driven Decisions
	Learning Log: Define problems and ask questions with data
Summarize your findings	Create a summary of key information you think a stakeholder would need to know about this project. In this case, your stakeholder could be a member of the executive team, like a project manager. Here are some questions to help you get started: ● What is the problem? ● Can it be solved with data? If so, what data? ● Where is this data? Does it exist, or do you need to collect it? ● Are you using private data that someone will need to give you access to, or publicly available data? ● Who are the relevant sponsors and stakeholders for this project? Who is involved, and how? ● What are the boundaries for your project? What do you consider "in-scope?" What do you consider "out-of-scope?" ● Is there any other information you think is relevant to the project? ● Is there any information you need or questions you need answered before you can begin?
Reflection:	Write 5-7 sentences (100-140 words) summarizing the key questions, the data available, and the answers or insights you have gained so far
Questions and responses:	*Here the objective of the project is to identify which videos are likely to go in the trending list. For this the datasets are available however this project requires current data, so its always better to create our own data set using some web scraping tools or creating a python script for us. The data is publicly available and does not require permission to be used. This project treats promoted and organic videos the same way, as we are yet to find a solution for that. Information like artist history could prove useful in determining whether it'll end up trending or not. We can begin as soon as executives give us the go signal.*