Please tackle the 2 prompts below.

1. Data Manipulation

Leads interact with several different channels before becoming a qualified opportunity with pipeline. This can include touchpoints through marketing efforts such as events and webinars or touchpoints through sales outreach such as direct phone calls. The sales and marketing teams are looking to understand which touchpoints and channels are sourcing the most pipeline.

The team wants to build a simple attribution system to tackle this problem. The attribution system has the following properties and assumptions.

- ➔ Attribution window = 90 days *(i.e., only include touchpoints that happened up to 90 days prior to the opportunity creation date)*
- ➔ Attribution model = first touch
  - ◆ The first touchpoint that happened within the 90 days before opportunity creation gets full credit for sourcing the opportunity. Let's walk through a hypothetical to see how this works.
  - Opportunity A has a pipeline of $1M
  - There were 4 touchpoints in the 90 day period prior to opportunity creation
  - Event A was the first of the 4 touchpoints
  - So for Opportunity A, Event A sourced $1M in pipeline and other 3 touchpoints sourced $0 in pipeline

We have provided 4 .csv files as part of this exercise. Below are the files and their respective schema.

- ➔ Marketing data
  - ◆ Marketing touchpoint id
  - ◆ Channel name
  - ◆ Contact id
  - ◆ Marketing touchpoint date
- ➔ Sales outreach data
  - ◆ Sales touchpoint id
  - ◆ Channel name
  - ◆ Contact id
  - ◆ Sales touchpoint date

- ➔ Contact data
  - ◆ Contact id
  - ◆ Account id
- ➔ Opportunity data
  - ◆ Opportunity id
  - ◆ Account id
  - ◆ Pipeline amount
  - ◆ Opportunity created date
  - ◆ Sales segment

For this deliverable, please do the following:

- ➔ Transform provided data to build a data table with all the touchpoints, the opportunity they interacted with and amount of pipeline they sourced
  - ◆ You can assume that interactions on an account are related to opportunities on that account ➔ Answer the following questions:
  - ◆ Which channel sourced the most pipeline? How does this look by sales segment?
  - ◆ What information do you need to know to understand the ROI (return on investment) of each channel?
  - ◆ How did you structure your data table and why? What do you think are the important output dimensions?
  - ◆ This table is an important input into other data and business systems. What kind of data validations and checks would you implement to make sure that downstream stakeholders have confidence in the insights they are generating from this?

Output includes the final data table, code used to generate the final data table and a short write-up answering the questions above.

2. Systems Design

How would you think about architecting a robust attribution system? A 'robust attribution system' should, at minimum, be able to do the following:

- ➔ Ingest, transform and map interaction and opportunity data from several sources
- ➔ Apply several different attribution models with unique logic

- ➔ Pipe outputs into an easily accessible and usable system for end stakeholder consumption. End stakeholders include executive leadership, sales, marketing and finance.

Please answer the following questions:

- ➔ What data platforms will you use and what will your data stack look like? Why? ➔ Please identify risks and vulnerabilities in your system ➔ How will you maintain and scale this system?