**Introduction**

I would like to build an attribution model to better understand the ROI of different ad campaigns and ad channels my employer currently has running.

**Problem**

In many cases digital marketing takes the form of a multi-touch process. An advertiser, particularly a company looking to advertise its services or products to acquire customers or users often utilises multiple advertising channels to both create brand awareness and drive conversions to its offerings. As a result of this, in many cases a user may interact with multiple adverts from a single company before converting into a paying customer. While this process creates an essential funnel where the user goes from learning about the company’s services or products and developing a deeper affinity to its service offerings to converting into a paying customer, it also creates a measurement problem for the company’s marketing department. Namely, it raises issues regarding how to correctly attribute credit to an advert for its contribution to getting a user to convert to a customer. Two problems arise as a result of this attribution problem; firstly the company will not have an accurate understanding of the return on investment on its existing advertisement campaigns, secondly, the inability to accurately and reliably estimate the contribution of each ad to the final conversion reduces the company’s ability to effectively allocate its budget to the channels that can best help the company acquire more conversions at a lower cost.

Solutions addressing this problem have generally taken two forms; i.) rule based approaches- examples of these include scenarios where full credit is given to the first or advert a customer clicked before converting (first or last click attribution), equal credit is assigned to each touch point (linear attribution), or ‘more credit is assigned to touch points closer to the final conversion (time decay attribution) ,and ii.) algorithmic attribution- this includes instances where custom models are built to ascertain the importance/contribution of each touch based through the analysis of existing customer data to understand different customer journeys leading to a conversion.

While Google Analytics offers solutions to attribution modeling this its existing models, data collected on GA is aggregated and anonymized, making it less impactful for companies operating with larger digital marketing budgets and needing a more in-depth analysis of customer journeys

**Data**

I will be using internal data, some clean up will be required. The data shows timestamps upon which a given ad was clicked by a user, the url which the user interacted with, the referrer channel and medium and whether or not the user became an opportunity. According to the company’s definition a user becomes a lead the moment they start filling in information on our signup form, specifically when they enter their business details. When this happens the user is then assigned with a unique user id. Prior to this, a user can be identified using its anonymous\_id. However, the user may be assigned with a different anonymous id if s/he continues the signup process with a different device.

**EDA**

I will be looking at the volume of clicks from each channel, grouped by month/date or year. I would also like to understand the most prevalent sequence of clicks leading to a conversion and the channel most likely to be the first and the last click before a conversion

**Deliverables**

* Cleaning up the data from the source
* Creating visualisations and other analysis to understand the distribution of clicks across different ad channels
* Creating an attribution modelto better understand the ROI of different marketing campaigns