Case Study: Cross-Device User Path Analysis

Context/Objective:

An advertiser wants to understand their customers' journeys across multiple devices and touchpoints using cross-device event log-level data from Google Ads Data Hub. The analytics team has access to the aggregated cross-device path data from Google ADH and aims to analyze the conversion paths that generate a high return on ad spend (ROAS) to develop a predictive model for identifying potential high-value audiences in real-time.

Dataset: Path.csv with columns ['path', 'cost', 'sales']

Deliverables:

- Report: The report should cover standard ML processes such as data cleaning and preprocessing, exploratory data analysis, feature engineering, model selection, model evaluation, and interpretation. Additionally, it should demonstrate analytical thinking and research ability.
- Model: A predictive model that accurately identifies conversion paths that lead to high ROAS. Candidates are encouraged to simulate user path data to demonstrate how the model can be used for real-time scoring. Prefer Python programming language.
- Preferred submission: GitHub repository with clear instructions on how to read the documentation, set up the runtime environments, and experiment with the code.

Hints:

- Candidates should consider whether a classification or regression model is more appropriate for this problem and identify any assumptions they make.
- Candidates should evaluate whether all the touchpoints on the raw paths are meant to analyze (eg: remove meaningless devices categories or associated touchpoints).
- They should also consider how to encode or decode the path data for the predictive model.
- It's important to investigate whether the lengths and sequences of the paths matter.
- Candidates are welcome to brainstorm the next steps if the existing dataset is not enough for addressing the real-life business problem.