

Think Green Supplies – Marketing Data Scientist Interview Round 2 Practicum

Question: Can we predict if a visitor, in a session, will add item(s) to the cart?

Our company, Think Green Supplies, sells various merchandise to businesses & consumers via one sales channel, our ecommerce store (website). Using obfuscated Google Analytics data on each visitor session, you will build a model that predicts whether each session will result in the visitor performing an action to add an item to the cart.

Goal: It is your job to predict for each visitor session, whether it will result in addedToCart (1) or not (0).

Evaluation: Submissions are evaluated on AUC, which is a number between 0 and 1, where 0 means 100% of the predictions are wrong and 1 means 100% of the predictions are correct. Randomly guessing should give a score of 0.5.

The Data:

Training: <https://drive.google.com/file/d/1LqO1PtN-NU8lmwLVIEWD4hgJBiePu5TI/view>

Test: <https://drive.google.com/file/d/17BdSbCPI5p96GWeoU52s7BjmWayLWfy8/view>

Submission File Format: The file should contain a header and have the following format:

```
sessionId,addedToCart
1234123412341234,0
2341234123412341,1
3412341234123412,1
4123412341234123,0
```

What is the table schema?

- fullVisitorId: The unique visitor ID.
- visitStartTime: The timestamp (expressed as POSIX time).
- date: The date of the session.
- deviceCategory: The type of device (Mobile, Tablet, Desktop).
- isMobile: If the user is on a mobile device, this value is true, otherwise false.
- operatingSystem: The operating system of the device (e.g., "Macintosh" or "Windows").
- browser: The browser used (e.g., "Chrome" or "Firefox").
- country: The country from which sessions originated, based on IP address.
- city: Users' city, derived from their IP addresses or Geographical IDs

- trafficSource: The source of the traffic source. Could be the name of the search engine, the referring hostname, or a value of the utm_source URL parameter.
- trafficMedium: The medium of the traffic source. Could be "organic", "cpc", "referral", or the value of the utm_medium URL parameter.
- trafficCampaign: The campaign value. Usually set by the utm_campaign URL parameter.
- isFirstVisit: 1 if this is the first visit made by the visitor, otherwise 0.
- isBounce: 1 if the visitor went to only one webpage and did not come back, otherwise 0.
- totalVisits: The total number of visits by the visitor across sessions.
- totalHits: The total number of interactions across sessions.
- totalPageviews: The total number of page views across sessions.
- totalTimeOnSite: The total time on site.
- totalTransactions: The total number of interactions across sessions.
- productPagesViewed: The number of product pages viewed in the session.
- addToCart: If the visitor added an item to the cart or made further downstream actions (checkout, transaction), otherwise 0 (if the visitor only browsed but never added an item to the cart).