Machine Learning Engineer / Data Scientist / Data Science Intern

At CloudTara Technologies, we are a team of skilled problem-solvers with a deep technology background and want to drive companies to be technology and data-centric through the use of modern technology and actionable insights thereby bringing meaningful value, efficiency and competitive advantage to the business.

Job Description

CloudTara Technologies is hiring for Machine Learning Engineers / Data Scientists (for full-time applicants) and Data Science Interns (for part-time/interns) to help us businesses provide next-gen digital experience to our consumers by building revolutionary, cloud-based, scalable and reliable products using Agile-Scrum methodology.

Problem Statement

An advertiser would like to estimate the performance of his campaign in the future.

For the given data, please try to address the below questions:

1. Perform exploratory data analysis and specify your findings with respect to any trends or outliers based on raw or efficiency metrics
2. Identify correlation in performance of the ads by time metrics if any (Ex: day of the week, day etc.)
3. Given the past performance data, predict the performance (Impressions, Clicks, Conversions, Revenue) of an ad between March 1st and March 15th. Use the data in predict.csv for your predictions and submit your response as shown in the file sample_submssion.csv)

Resources

The required data dumps for the assignment can be found here:

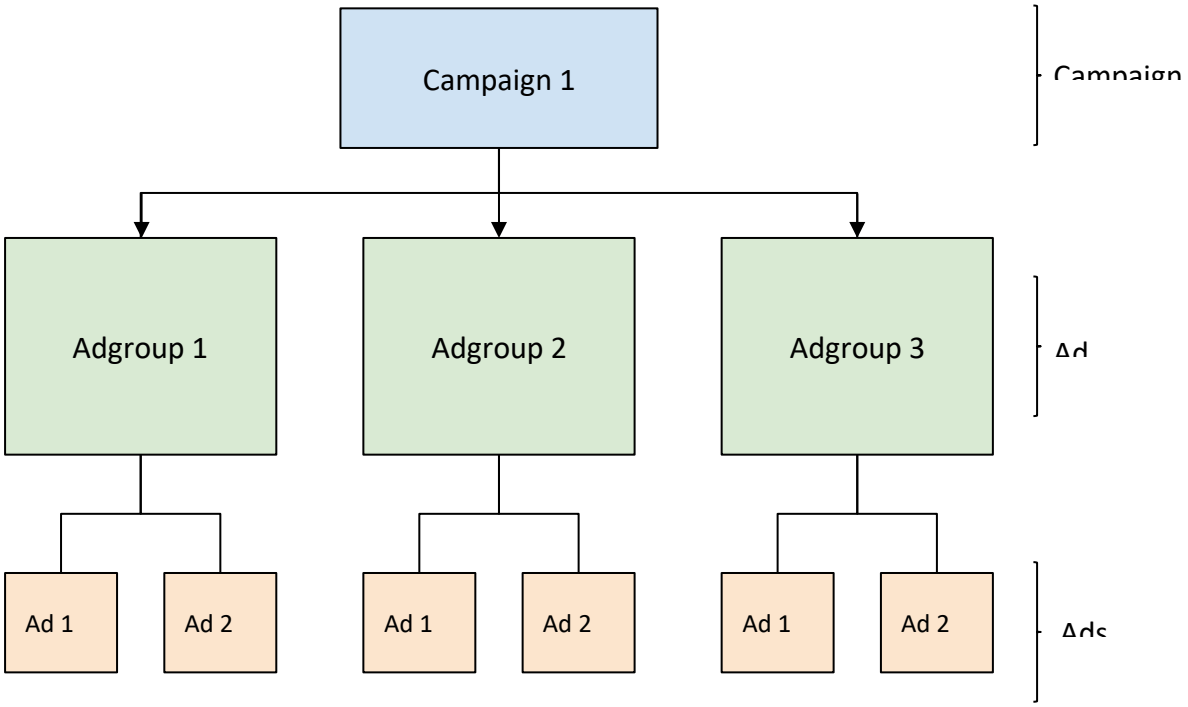
Filename	Comments	Link
data.csv	Raw metrics from 01-Aug-2020 to 28th Feb-2021.	Drive > Resources/data.csv
predict.csv	Your model needs to predict for these data points.	Drive > Resources/predict.csv
sample_submission.csv	Contains column format for the submission of assignment - submission.csv file. Your model needs to predict for columns (impressions, clicks, conversions, revenue) using input data points from predict.csv	Drive > Resources/sample_submission.csv

CLOUDTARA TECHNOLOGIES

About the Data

The sample data dump shared above is the ad performance for the dates Feb 2021.

Here is a brief hierarchy for relationships in adtech systems:



One Campaign can have one or more Ad Groups. One Ad group can have one or more ads. In the given dataset - we have only 1 campaign; 4 Ad Groups and multiple ads.
Here is a brief description of all performance columns in the data set:

Raw metrics:
Impressions - Number of times the ad was shown
Clicks - Number of times the ad clicked shown
Cost - Amount spent to show ad
Conversions - Number of transactions received (higher the better)
Revenue - Total value of transactions received (higher the better)

Efficiency metrics:
Needs to be calculated and are based on raw metrics
CTR - Clicks / Impression (higher the better - used to evaluate if the users find the ad relevant)
CPC - Cost / Click (lower the better - used to evaluate if the cost for getting a click)
CPA - Cost / Conversion (lower the better - used to evaluate if the cost for getting a conversion)
ROI - Revenue/Cost (higher the better - used to evaluate the effectiveness of the advertising budget spent)

How to Submit Assignment

Keep the subject line of your email as “YOUR NAME - Assignment” (eg: “Shubhang Y - Assignment”).

Note:

- Please share the following information in the submission mailer:
    - Full-time: Current/Last Job Role, Company Name, Last drawn CTC, Notice Period / Joining Date Availability
    - Interns: Stream/branch of education, Current Semester, Joining Date Availability
- All submissions should include the project code (as a zip or hosted API link) along with a clear readme/setup file
- Being able to summarise your findings visually is important for us.
    - Please use appropriate charts/graphs.
    - Please write your thought process and findings below the visualization as comments in the Juypter notebook
- Your submission.csv will be evaluated using RMSE and R^2 metrics.
- The Jupyter notebook that you share with us, should fetch the data from the shared CSV file to the notebook and should be fully executable without any errors.
- Other reference material

We are expecting the following attachments in your submissions:

1. Resume
2. Jupyter notebook (.ipynb)
3. Attach **submission.csv** file. Your model needs to predict for columns (impressions, clicks, conversions, revenue) using input data points from predict.csv. We have also shared a sample sample_submission.csv format so that you can share appropriately.
4. Requirements.txt file
5. Text file containing sample input and expected response

Email your submissions to sadanand@cloudtara.com.

The last date for submission is June 12, 2021 (the date differs based on your batch, refer to the message which was sent to you). Feel free to reach out to us on the same email if you have any questions related to the assignment or submission date.

CLOUDTARA TECHNOLOGIES