

# Google Analytics Customer Revenue Prediction

Predict how much GStore customers will spend

## Solution proposal

The goal of the project is to predict how much customers will spend at Gstore, an online retail store. The original dataset comprises of 12 attributes, namely fullVisitorId, channelGrouping, date, device, geoNetwork, sessionId, socialEngagementType, totals, trafficSource, visitId, visitNumber and visitStartTime. Each record in the dataset corresponds to a single visit to the store. Our task is to predict the value of the natural log of the sum of all transactions per user.

$$Y_{\text{user}} = \sum \text{transaction}_{\text{user}i}$$

$$\text{Target}_{\text{user}} = \ln(y_{\text{user}} + 1)$$

The first step toward the solution is data preprocessing. Some attributes have sub-columns, so they need to be separated and only relevant attribute columns are used. Then we are going to use **regression** to build the model that we will use for the predictions.

There are two separate files for the dataset, the train.csv and test.csv. We will use the train.csv data to build the model and the test.csv data to evaluate the performance of the resulting model. We have sampled 200 instances from each of the two files and they can be accessed through the following link: [https://drive.google.com/open?id=1Qs3fBiu6swWejf27\\_ijHkh18QxTzQ7NU](https://drive.google.com/open?id=1Qs3fBiu6swWejf27_ijHkh18QxTzQ7NU). The complete dataset will be made available once we finish the preprocessing step.