

Project Group 1

Project Title: Customer Revenue Prediction

Abhishek Kumar, UB#50291056
Narendra Badam, UB#50289385

Description: We would like to participate in the Kaggle competition hosted by Google to analyze the Google Merchandise store customer dataset and predict the revenue per customer. The dataset is provided by Google and its parameters are:

- *fullVisitorId* - A unique identifier for each user of the Google Merchandise Store.
- *channelGrouping* - The channel via which the user came to the Store.
- *date* - The date on which the user visited the Store.
- *device* - The specifications for the device used to access the Store.
- *geoNetwork* - This section contains information about the geography of the user.
- *sessionId* - A unique identifier for this visit to the store.
- *socialEngagementType* - Engagement type, either "Socially Engaged" or "Not Socially Engaged".
- *totals* - This section contains aggregate values across the session.
- *trafficSource* - This section contains information about the Traffic Source from which the session originated.
- *visitId* - An identifier for this session. This is part of the value usually stored as the `_utmb` cookie. This is only unique to the user. For a completely unique ID, you should use a combination of *fullVisitorId* and *visitId*.
- *visitNumber* - The session number for this user. If this is the first session, then this is set to 1.
- *visitStartTime* - The timestamp (expressed as POSIX time).

We are planning to use exploratory data analysis using Numpy, Pandas and Matplotlib. If time permits we would like to use time series analysis to predict revenue per customer and also try to identify traits of revenue generating customers.

Hopefully, the outcome will be more actionable operational changes and a better use of marketing budgets for those companies who choose to use data analysis on top of Google Analytics data.