**Team Name**- Team Zeit

**Team Members**- Ashok Anand B, Vaishnavi Bihare, Snehal Ilawe, Shruti Kulkarni.

**Dataset**- <https://www.kaggle.com/c/ga-customer-revenue-prediction/data> (Competition Dataset)

**Problem being addressed**- Marketing teams often face challenges in making appropriate investments in promotional strategies, as only a small percentage of customers produce most of the revenue. This dataset is useful to demonstrate the business impact that thorough data analysis can have.

In this competition, the challenge is to analyze a Google Merchandise Store (also known as GStore, where Google swag is sold) customer dataset to predict revenue per customer. The outcome of this project will be to use the data available to get insights on consumer spending habits and additionally, to discover a more efficient and actionable operational change and better use of marketing budgets for those companies who choose to use data analysis on top of Google Analytics data.

Proposal - Build a model that can predict the probability of returning customers. Secondly, predict the revenue from the transactions for those customers who returned using regression.

**Data Description**-

a) No of rows and columns-

There are 12 original columns of which 4 of the columns have a json object that can be converted:

1. device: 16 new columns (includes, text, numeric, binary and categorical data types)
2. geoNetwork: 11 new columns
3. totals: 6 new columns
4. trafficSource: 14 new columns (includes, text, numeric, binary and categorical data types)

Thus a total of 55 columns and 903653 rows

b) We are predicting the probability of returning customers- classification and the total transactional revenue --Regression

c) Datatype-

**fullVisitorIdv** - A unique identifier for each user of the Google Merchandise Store. (Numeric)

**channelGrouping** - The channel via which the user came to the Store. (Categorical)

**date** - The date on which the user visited the Store. (Numeric)

**device** - The specifications for the device used to access the Store.

device.browser: (String)

device.browserSize: (String)

device\_browserVersion (String)

device\_deviceCategory (String)

device\_flashVersion (String)

device\_isMobile (Binary)

device\_language (String)

device\_mobileDeviceBranding (String)

device\_mobileDeviceInfo (String)

device\_mobileDeviceMarketingName (String)

device\_mobileDeviceModel (String)

device\_mobileInputSelector(String)

device.operatingSystem(String)

device\_operatingSystemVersion (String)

device\_screenColors (String)

device\_screenResolution(String)

device\_deviceCategory(String)

**geoNetwork** - This section contains information about the geography of the user.

geoNetwork\_city(String)

geoNetwork\_cityId

geoNetwork\_continent (categorical)

geoNetwork\_country (String)

geoNetwork\_latitude(String)

geoNetwork\_longitude(String)

geoNetwork\_metro(String)

geoNetwork\_networkDomain(String)

geoNetwork\_networkLocation(String)

geoNetwork\_region(String)

geoNetwork\_subContinent (String)

**sessionId** - A unique identifier for this visit to the store. (Categorical)

**socialEngagementType** - Engagement type, either "Socially Engaged" or "Not Socially Engaged".(Categorical)

**totals** - This section contains aggregate values across the session.

Totals\_bounces (numeric)

Totals\_hits (numeric)

totals\_newVisits (numeric)

Totals\_pageviews (numeric)

totals\_transactionRevenue (numeric)

Totals\_visits (numeric)

**trafficSource** - This section contains information about the Traffic Source from which the session originated.

trafficSource\_adContent

trafficSource\_adwordsClickInfo.adNetworkType (categorical)

trafficSource\_adwordsClickInfo.criteriaParameters

trafficSource\_adwordsClickInfo.gclId(binary)

trafficSource\_adwordsClickInfo.isVideoAd

trafficSource\_adwordsClickInfo.page

trafficSource\_adwordsClickInfo.slot

trafficSource\_campaign (categorical)

trafficSource\_campaignCode (numeric)

trafficSource\_isTrueDirect (binary)

trafficSource\_keyword

trafficSource\_medium

trafficSource\_referralPath

trafficSource\_source

**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.(Numeric)

**visitNumber** - The session number for this user. If this is the first session, then this is set to 1.(Numeric)

**visitStartTime** - The timestamp (expressed as POSIX time).