**# Project Name:** Finding MQLS (Most qualified Leads) by analysis on website traffic and usage pattern

**Team Size - 3**

Anjali Mangal

Archana Soni

Christie Lafrance

**## Background**

The data dive decisions. Companies which are data driven organization make

Informed decisions rather than based on intuitions and those who are not using data are falling behind.

**DataGuru.com** - Company has a new business unit and investing in cloud services (subscription model) for few of its traditional licensed suits on premise products.

Business Unit has launched 3 of its data-cloud services

* Data Logging
* Data packaging
* SDK

In Beta /Trial phase to almost 100 Companies. The trial period is about 60 days.

While the trial is going on the engineering team has instrumented the click stream telemetry in the services. They have tried Google analytics /Mixpanel in previous launches. But for these new services they have chosen Fullstory [www.fullstory.com](http://www.fullstory.com).

The tool gives us info on pageviews, click paths, and heat maps as a proxy for user engagement,

But irrespective of chose of tool of collecting customer usage info all of the data is raw numbers and which does not tell any story or meaning full insights.

Each day usage data is captured in data exports which is a big chunk of information.

The team has put a plan together to answer most important business question which stakeholders need to resolve to make some strategic moves.

Business Unit hired a small team of data analysts and data scientists who will identify and following

**### 1. Most Active Customers based on** – (For this project scope we will work on top 10-15)

* Total number of sessions
* Average Total Session time
* Average Active Session time
* Average Nonactive Session time

**###2. Most Popular Services -** Engineering, product management and VPS wants to know the most used service during trial phase to decide the future investment areas w.r.t development, testing etc.

* Most Used Services based on PageUrl Hits and Events triggered.
* Most Used Services by top 15 active customers found in #1

**###3. Most Visited Pages Analysis -**

* Team will identify the pages of services having maximum hits and events overall
* Most Visited pages by our top 15 customers found in #1.

**### 4 Most Popular Browser Analysis –**

These are the web services, so it is very important to know which is the most popular choice of browser (Firefox/IE/chrome)

**### 5 Most Popular Devices Analysis –**

As these are web services so getting to know choice of device is also a key area like desktop vs mobile

**Data Analytics Team Plan and considerations:**

Team will use the Fullstory APIS to get the data exports of last 30 days.

What is data export - The Data Export pack provides these events, pageclicks,

Session info and much more to you in an easily readable JSON file.

This “JSON blurb” would represent a single row of data in an Excel sheet. Purple text represents the “name of the column” and the orange text is the value for that specific row. Therefore, a single session will be represented by up to hundreds of these JSON event blurbs, but they can all be easily tied back to a sequence of actions taken by a specific user in a specific session.



**APIS used using python are -**

https://help.fullstory.com/develop-rest/data-export-api

**LIST**

https://export.fullstory.com/api/v1/export/list

#### *GET*

https://export.fullstory.com/api/v1/export/get

**Code Structure and flow:** The Project includes following -

**Config.cfg ->** This file contains the Fullstory API Key and some other variables for parsing.

**data folder - >** This folder is a place holder for all CVSs generated. If the folder does not exist, then code takes care of creating it.

**imgs folder ->** All the graphs and plots are saved at this location

**Python Notebooks ->**

1. GetData\_fromAPI.ipynb - This file actually calls the API to collect all the data export of website traffic and usage in saves them in local csv.
2. CleanSensitiveInfo.ipynb - There are some company and customer sensitive info which needs to be masked, this notebook take care of that and saves results in another local csv.
3. Website\_Traffic\_Analysis.ipynb - This is the main file and does all data cleaning, columns and rows dropping, filtering, indexing, data munging, Sorting and all logical and flow up to creating graphs.

**Consideration and facts**

* Team has heavily used the Python /Pandas NumPy for data retrieval, munging and cleaning
* Team has also used Python statistics analysis for mean, average and a

algorithm of weighted score /ranking to rank top 15 customers

* Matplotlib and used seaborn for plotting and charting
* List comprehension/Lambda are favorite tools of senior data scientist in this team
* What new we have learnt during the Project:
  + Configures library to read and parse the variables from Config .cfg
  + Use of logger module to generate the log file for debugging purposes
  + Extensive use of lambda functions
  + Iterating graph creation via user defined function calls for plotting same graphs.
  + Converting regular dataframes to pivot tables for easier plotting.
  + Lots if added features w.r.t plots and graphs like palate colors, size fit in a given window etc.