Capstone Project | BI Data Science Learning Path|

**Abstract**

**Project 1**

**Website Traffic Source Analysis**

Nowadays use of Internet is a part of nearly every day for private and occupational purposes by its users. In the virtual environment of Internet there exists a lot of Internet websites. Website content itself is very important for users. To find websites it is very necessary to use some tools of internet marketing. Traffic Source in the simplest form can be defined as the origin through which people found the site. Every session or visit to any website has origin or source of the hit. Whenever someone visits the website through desktop, mobile or any Internet-enabled device, the Web and Google Analytics track its source. Understanding traffic sources helps derive the understanding of the sources which are driving the most and least traffic to the website.

**Tools / Skills Used**

1. MySQL Server
2. Traffic Source Analysis using UTM parameters
3. Conversion Rate
4. Joins
5. A\B Testing

**Introduction to the project**

The given dataset is a company’s dataset which wants to know more about their customers and the sales of their company. In this project, Traffic source analysis (UTM Parameters) and A\B Testing are used to ensure the site is performing well and delivering conversions. Website Traffic Analysis tools were used to identify the trends, prioritize developing patterns, and streamline for the site, weather the visitors are searching the site organically or landing on the page via social media, paid searches, or any other means, and which are the best customers of the company.

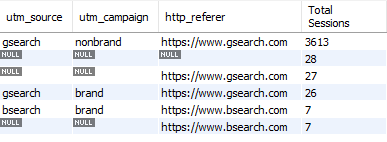
**Implementation**

**Problem Statement 1:** To understand from where the bulk of the website sessions are coming from (specifically breakdown of UTM source, campaign, and referring domain)

**Codes:**

****

**Output**

****

**Conclusion:**

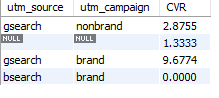
gsearch, nonbrand, <https://www.gsearch.com> is getting the maximum number of sessions.

**Problem Statement 2:** Calculate Conversion Rate from session to order (to manage bids based on CVR)

**Codes:**

****

**Output:**

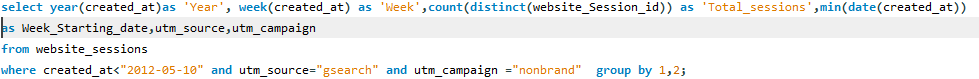
****

**Conclusion:**

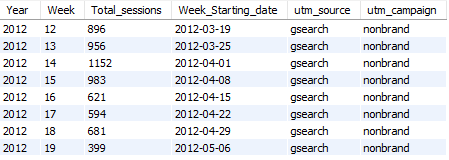
The CVR for gsearch nonbrand id the highest and we are overbidding for it..

**Problem Statement 3:** To find gsearch non brand trended session, volume by week.

**Codes:**

****

**Output:**

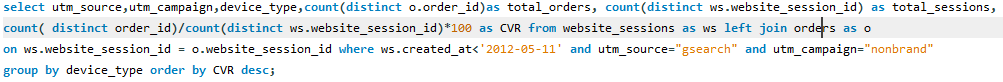
****

**Conclusion:**

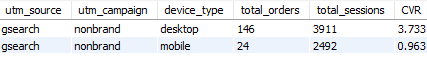
After above conclusion the company bid down for gsearch nonbrand but after that the sessions per week first increased then decreased at a very high rate.

**Problem Statement 4:** To calculate Conversion Rates from session to order by device type (to increase the session volume)

**Codes:**

****

**Output:**

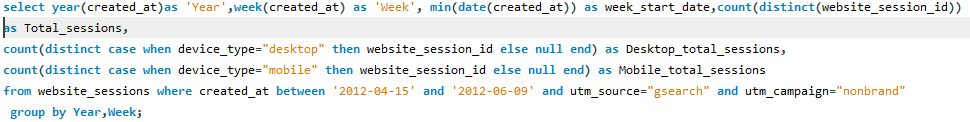
****

**Conclusion:**

The CVR for desktop device type is more so the company should bid more for desktop to increase the volume of sessions

**Problem Statement 5:** To figure out weekly trends by device type to see the impact on volume after the company raised the bid for gsearch nonbrand desktop.

**Codes:**

****

**Output:**

****

**Conclusion:**

The CVR for gsearch nonbrand desktop is high so the volume for sessions has definitely increased after bid up for desktop.