Sapien Designs Web Scraping and Marketing Quality Tool Sc. Project

Research Development Support for Mike Jensen

Research & Business Development Center – Rexburg, Idaho (BYU-Idaho)

Version: 4/14/2021

**Background**

Sapien Designs is a digital marketing company that focuses on multimedia and SEO marketing. It was founded in 2012 in Huntington beach, California but now has offices in Salt Lake City, Rexburg, and Jackson. They develop and design brands for companies to promote business. Sapien Designs promise their product is helping businesses with advertisements, “Getting you found” and “Finding you Customers”. They also create media content for clients such as video and audio recordings, branding, design, website photography and animation.

In their goal to promote their clients Sapien Designs identifies what causes a company to be noticed via search engines. Once these attributes are found, they implement necessary steps within their clients business.

**Overview**

Sapien Designs would like to develop a tool that would scrap business related data from Google My Business and Apple Maps that is industry specific. With that scraped data they want to identify key indicators that are common in the “tri-pack”. Based on the key indicators, create a report that identifies companies that are established but not at peak advertising and what area they can improve in with Sapien Design’s services.

NOTE: The Research & Business Development Center (RBDC) and Brigham Young University - Idaho who participate in this study must sign a Confidentiality Agreement protecting Sapien Designs’ business interests, data and business practices.

**Specific Scope:**

**There seems to be several phases to this project, the following scope of the project will spill over into more than one semester**, there will need to be more discussion about the specific work that will be done on a per semester basis.

**Phase 1:**

**Create a web scraping tool that scrapes business and marketing data from Google My Business and Apple maps.** - This tool will be used to scrape the top responses from an industry related search. The tool will compile a data set that consists of key indicators, essential to SEO and digital marketing. **Below are questions and subsets of this portion of the project.**

* 1. Industries of interest: Medical (Optician, surgeon, Surgical Center, Medical Center, Plastic Surgeon, Orthopedic Surgeon, Chiropractors), Professional (Real Estate, Insurance Agency/Agent)
  2. Analyze what aspects of the data leads to a company being in the top three search results.
  3. Created a database structure to store data scraped data.
  4. Allow for personal selection of the industry that is being scraped, also allow for personal selection of location.
     1. Create this tool as an app for ease of user input
  5. Look into web scraping vs API’s

**Phase 2:**

**Identify key indicators that provide a stable platform for the business use in the “tri-pack”.** Some indicators have already been selected and identified by the client. The team will determine if there are any other indicators that should be added to the previous developed list. The team will also use these indicators to rank the quality of individual businesses’ ability to be searched and identify what can be approved. **Below are questions and subsets of this portion of the project.**

* 1. Identify which indicators are most impactful and create the ranking system off of those indicators

**Phase 3:**

**Create a machine learning algorithm that identifies or determines companies that are established but could use improvement in their SEO or digital marketing.** The responses from the algorithm will be used to create a report of potential customers to contact. **Below are questions and subsets of this portion of the project.**

* 1. Create an automated email response when new potential clients are found
  2. How should the report be written? (Dashboard, PDF, Datatable?)

**Team Makeup:**

Team will be composed of 3-5 Data Science students. They will be chosen for this team based upon their backgrounds and how well it will support this effort. Students will also have a say on whether or not they wish to participate on the project. This team will have an experienced (Technical) Project Manager and either work within the Statistical Consulting class or work under the direction of the RBDC with oversight by a BYUI Professor with expertise in this field.

**Target Audience:**

The targeted audience for this study is Mike Jensen from Sapien Designs.

**Client Requirements:**

To aid in the progress of the project we request the client or a representative of the company to engage with the student team. Frequency of the interaction will be based on the availability of the client. The interactions between the students and the client would be to provide updates, insight and to answer questions about the project. We also request the client to provide data as needed for the project.

**Deliverables**

1. Meet with students and client to kickoff this project
2. Provide informational updates to Mike Jensen and other leaders as desired throughout the semester or as needed for the purpose of clarifying questions about the data. Use these reviews to adjust the development effort as necessary to meet the company’s needs.
3. Provide a summary outbrief of the project by mid of the semester.
4. Provide the desire deliverables [dashboard, algorithms, report]
5. Other - As desired