

# G2 Data Engineering Project Overview

## I. Objective

The purpose is to create a tool/dashboard that provides a user with information about selected companies on G2. Visualize several KPIs and comparisons of the selected company to its competitors to show where the selected company stands within that category.

## II. High Level Requirements

Requirement	Description
Filter by the name of companies	A drop-down bar that prompts the user to pick a company
Compare the ratings of the selected company to its top 10 competitors	A number that shows the average rating of the top 10 competitors
Compare the ratings of the selected company to the average ratings of the categories the selected company is in	A graph that shows the average rating of the companies within that category and the rating of the selected company
Show number of reviews and the average rating (number of stars) of the vendor of choice	Highest level of information that gives a general overview of the company

## III. Metrics

To address the aforementioned technical requirements, the following list of metrics must be extracted:

- Number of reviews
- Average number of stars
  - By the companies in the same category
  - By the top 10 competitors
  - By industry

Top 10 competitors provided by G2 can give users a better idea on where the company stands among its competitors in technical terms instead of relying solely on the number of stars.

#### IV. Data Sources

The dataset is obtained from this [link](#) provided by [Seattle Data Guy](#) which he obtained through the Web Scraper IDE on Bright Data for G2.com.

#### V. Infrastructure

##### A. Snowflake:

- Process raw, scraped data from Bright Data to structured tables
- Query layers of processed tables to get desired metrics

##### B. Tableau:

- Turn queried tables into interactive dashboards

#### VI. Objects

##### A. vendor\_rating:

Parameter	Data Type	Description
vendor_name	varchar	Name of the vendor
num_reviews	int	Total number of reviews of the vendor
num_stars	double	Average number of stars/ratings of the vendor
category_list	list[varchar]	List of categories the specific vendor is classified as

##### B. vendor\_competitor\_rating:

Parameter	Data Type	Description
vendor_name	varchar	Name of the vendor
competitor_name	varchar	Name of the competitor of the

		vendor
competitor_num_review	int	Total number of reviews of the competitor
competitor_num_stars	double	Average number of stars/ratings of the competitor

C. vendor\_category\_comparison:

Parameter	Data Type	Description
vendor_name	varchar	Name of the vendor
product_category	varchar	Name of the product category
num_of_stars	double	Average number of stars of the vendor
num_stars_category	double	Average number of stars/ratings of the companies in the category