Task 6: Sample Data

To populate our database we opted to obtain real world commerce data focussing on the following categories:

- Laptops
- Desktop Computers
- Headphones & Earbuds
- Keyboards & Computer Mice
- Monitors
- Smartphones
- Tablets

The data was obtained via querying SerpApi, a trusted interface for scraping search engine results. The API was queried with the following parameters:

```
{
      "engine": "google_shopping",
      "q": <search_string>
}
```

Passing "google_shopping" as the value for the "engine" parameters instructs the API to return scraped results from the Google Shopping engine.
The value passed for the "q" parameter is the search string that will be inputted into the engine.

The following algorithm was followed to obtain similar products from different retailers:

```
initialize a list of categories
for each category in categories:
      query the API with q = category, save first n products of response
      for each of the n products in response
            query the API with q = product title
            record responses in data structure
dump all responses to a structured JSON file
```

The database was populated based on extracted data from the generated JSON file via a python script. For reference, the scripts can be found in our repository here.