Data Requirements

# for **Lead Generation**

# for **Catalog Enrichment Service**

## **Product Catalog**

* List of all available products that are being offered on the platform, along with the relevant information about each product such as mentioned below:
  + title
  + brand
  + complete category hierarchy
  + quantitative information about the product such as weight, volume, price, size etc.
  + Inventory and Discount that is being offered at different stores
* Product catalog data is dynamic, it’s attributes might change with different store/location and time. We need to have access to the latest catalog dump that is being used for POS machines, along with all the attributes as mentioned in the previous point.

## **Clickstream Data**

**query logs:**

* complete list of queries searched by users
* for each query we need
  + total search count:
    - # of times the query was searched by any user
  + clickthrough rate:
    - (# of times a query is searched and led to a click)/(total search count of that query)
  + add to cart rate:
    - (# of times a query is searched and led to a product being added to cart)/(total search count of that query)
  + order rate
    - (# of times a query is searched and led to a product being ordered)/(total search count of that query)
  + query to product mapping
    - list of all the products that were interacted with for the given query
    - for each product:
      * separate count of click, add to cart, and order

**session data:**

* for each session of a particular user on the search platform, we need the below data:
  + user metadata
    - gender, device type, session id, location (latitude, longitude) etc.,
  + transactions made by user such as
    - query search, page view, add to cart, order
    - all these transactions when done in a same session should be tagged with same session id

## **Synonym List** (optional)

This data enables the Spell check module to identify synonyms for search terms for example:

* **magenta** 🡪 **purple**
* **pillow** 🡪 **cushion** etc.