

Question 1: the **goal** of the analysis: what question do you want to answer? Why? (100-200 words)

- a. What are the best products for the user based on their skin type and personal preferences?

Skincare is an essential part of many people's daily routine. Considering the major effect that it can have on the skin, it's important that one uses products that are suitable for their skin type to prevent irritation and damage. However, with the fast-growing popularity of the skincare industry and the large amount of products that it releases onto the market every year, it can be difficult to keep track of all the different skincare items and their purposes. The goal of our application, therefore, is to present the user with an overview of products that best fit their needs based on their input. In order to do so, the user is presented with a list of questions. The results of this questionnaire form the base of the application as they are used to retrieve related data from a set of ontologies and datasets. A clear overview of recommended skincare products for the user's skin will be the resulting output. It will inform the user about products' properties which, for instance, include the type of product (moisturizers, cleansers, etc.), the skin type it's suited for, and its brand, price, ranking, and ingredients.

Question 2: the **stakeholders** of the analysis: for whom is the analysis intended? How does the analysis satisfy the needs of the stakeholders? (100-200 words)

The analysis is intended for skincare companies and people with skin issues. These stakeholders both want the best skincare products. This application can help with their needs by retrieving information, such as information on specific ingredients, products, and price range.

The first stakeholders are skincare companies. Skincare companies want to find products with a good price that will give them maximum profit, however, they also want to give their clients products with high-quality ingredients. This application will give companies an idea of which products would be worth investing in and which products are worth selling to their customers. It is also a huge benefit for the companies because customers will see their products when they are using our application. This is a good way to promote the products they are selling and it will give them a good position in the skincare market. It will also keep companies updated with the current development within the skincare industry. As our application will not be outdated, because it keeps track of the newly developed skincare on the market.

The second stakeholders are people with skin issues. As for people with skin issues, they want to find products that are just right for their skin type. Since skincare products are constantly being improved, it is hard for consumers to stay up to date with the best skincare products. This application will provide these people with an idea on what products to buy depending on their preferences. It will give them recommendations depending on their skin type (oily, dry, normal) which will prepare a list of different brands of the product, the ingredient list, rankings of the product, pricing, and so on.

Question 3: the **design** of the analysis and visualizations: which steps will your analysis have, and why do they make sense? For instance, showing restaurants on a map may be more intuitive than showing a list (maybe show both?). (150-300 words)

The aim of the design is to lay out the important features that have been selected for the customer, in order to display the essential information to the customer in an efficient manner. As a result, the visualization of the data has been decided to be displayed through a table containing the following features.

- b. appropriate skin type or disease for the product (example image)
- c. product name
- d. main ingredients of the product
- e. usage and effects of the product
- f. side effects of the product
- g. manufacturer and company of the product
- h. customer rating of the product
- i. price of the product

The features above have been selected because they are asked and answered in the list of questions that the customer receives. In other words, the results will produce an ontology, which will contain the appropriate information to fill in the data for the questioned features. A table is capable of organizing and arranging complicated data by categorizing them into rows and columns. The arrangement of having certain types of information contained in a specific row or column will allow the customer to retrieve the information that is needed quickly and easily. A visual example of the table can be seen in Table 1. Different visual representations, such as a list would be less suitable, in comparison to a table, due to the fact that it would not be able to organize lengthy pieces of data. Moreover, lists do not contain images, thus when presenting data such as skin type will have to be presented in words. Therefore, the customer most likely will have to search up an example of the skin type to check whether it is the correct skin type that they have. Therefore, to effectively communicate the relevant information to the customers, in this situation a table would be the most appropriate choice.


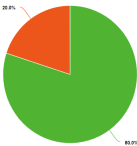


Product Name	Skin type	Ingredients	Effects	Side effects	Company	Rating	Price
tea tree cleanser		tea tree	reduce acne	dries skin	l'oreal		10.93
coral reef cleanser		coral reef, sea salt	cleanse pores	burning	ganiay		6.20

Table 1: Example Image of Data Layout Design

Question 4: Identify 2 ontologies that you can reuse, and motivate your choice. In some cases, it might be difficult to find an ontology that is closely related to the analysis domain. Then you can use more general ontologies and vocabularies, such as FOAF, DBO, VOCAB.ORG, SCHEMA.ORG, etc.

The first ontology we want to use is called the Schema.org¹ ontology. Schema.org contains vocabulary that can be used in applications that publish, discover, and integrate different kinds of data. This ontology has different classes, which include products and services. It is especially useful to online stores with a great diversity of products. We can use this ontology to describe the relevant information about skincare products. By using this we can show the different products with prices, reviews, and quality of the product. Customers will also be able to find related products and see images of each skincare product. Using this ontology will make our application user-friendly because they do not have to search again to find similar products.

The second ontology we want to use is called Dbo². This ontology can be used to add more information about the skincare products. By using this ontology, the customer will be able to search for products based on ingredients, product type, and existing skincare brands. By combining these two ontologies we will be able to make an application that can help users to find skincare products based on their preferences.

¹ <https://schema.org/Product>

² [Ontology \(DBO\) - DBpedia Association](#)

Question 5: a description of at least **2 external sources** of data that will be used in your analysis (i.e. for creating the instances). Motivate your choice of these 2 sources of data. At least one of these must be an **external SPARQL endpoint**. The other dataset **should not be in RDF**. (100-200 words). If you don't find any useful dataset related to the goal of your analysis (we doubt that), then you might have to think about changing your analysis domain.

To create instances to use in the analysis, external sources of data have to be extracted. The first source is the skincare products clean dataset³ contains more than 1,000 skincare products, which contain data about the product name, type ingredients, prices and URLs. The second data set, the cosmetics dataset⁴ provides almost 1,500 products that include the product name, brand, rank, combination, ingredients, skin preference and price. These data sets will be capable of providing instances for the classes or subclasses within the ontology. In other words, it will further advance the quality of the recommendation system for the appropriate product and to satisfy the customer by being able to provide further information about the products, if needed.

DBPedia⁵ is a database consisting of subsets of datasets in RDF format that offers various information when it comes to skincare such as pages on different brands⁶, and on skincare as a subject itself⁷. All this information is useful to our application since it provides a good starting point when searching for suitable and, therefore, recommended products for the user. Because DBPedia has a SPARQL endpoint⁸, it will enable the application to go through a wider range of products and make searching for instances based on the given query more efficient. In addition to its easily navigable database, DBPedia is a piece of open-source software. This means it may be used and redistributed without requiring permission. For these reasons, the use of DBPedia is ideal as it can considerably improve the application's searching mechanism, and the quality and size of the output.

³ <https://www.kaggle.com/eward96/skincare-products-clean-dataset>

⁴ <https://www.kaggle.com/kingabzpro/cosmetics-datasets>

⁵ <https://www.dbpedia.org/>

⁶ <https://dbpedia.org/page/Clinique>

⁷ https://dbpedia.org/page/Category:Skin_care

⁸ <https://dbpedia.org/sparql>