

## SI 507 Final Project Proposal

Yang Yang 8922 1006

The program is designed to show the average ratings for different cosmetic catalogs and analyze the relationship of different kinds of customers and products provided by Estee Lauder USA website.

**Data Sources:** HTML + API, Challenge Point: ~~10~~ 8

\*Estee Lauder: *crawling and scraping multiple pages (haven't used before)*, Score: 8  
Products (>100) provided by the site belongs to three catalogs (skincare, makeup and fragrance). Each catalog has at least 10 sub catalogs (face, eye, lip ...). Parsing the page of each sub catalog to get the basic information about products (name, price, rating, review counts). Detail information about reviews (customer name, region, age, skin type...) is obtained through the page of each product.

~~\*Google Place: web API (have used before), Score: 2~~

~~For displaying stores for each state on the map, street information is parsed through the Estee Lauder website and the latitude-longitude data are obtained by google place.~~

**Database:** 4 tables

\*Catalog: Name, Id

\*Product: Id, Name, CatalogId, Type, Price, Rating, ReviewNumber

\*Review: Name, Region, Age, SkinType, ProductId, Rating, ReviewContent

\*Store: Name, StreetAddress, City, State

**Presentation:** 4 3 different graphs + 1 table

*Plotly: selecting graphs by interactive commands and viewing graphs through plotly*

\*Bar Plot: showing the average rating for each catalog, X: Catalog Name, Y: Rating

<catalog>: X: skincare, makeup, fragrance

<sub catalog> <catalog name>: X: face, eye, lip...

\*Line Plot: for a given product, showing the relationship btw customer and rating

X: Customer Skin Type/Customer Age, Y: Rating

\*Dist Plot: for a given product, showing the distribution of different customer

X: Customer Skin Type/ Age/ Region, Y: Number of Customer

~~\*Map Plot: for a given state, showing the location of stores~~

*Flask App: displaying selected data using HTML tables*

\*Table: for a given catalog, showing products