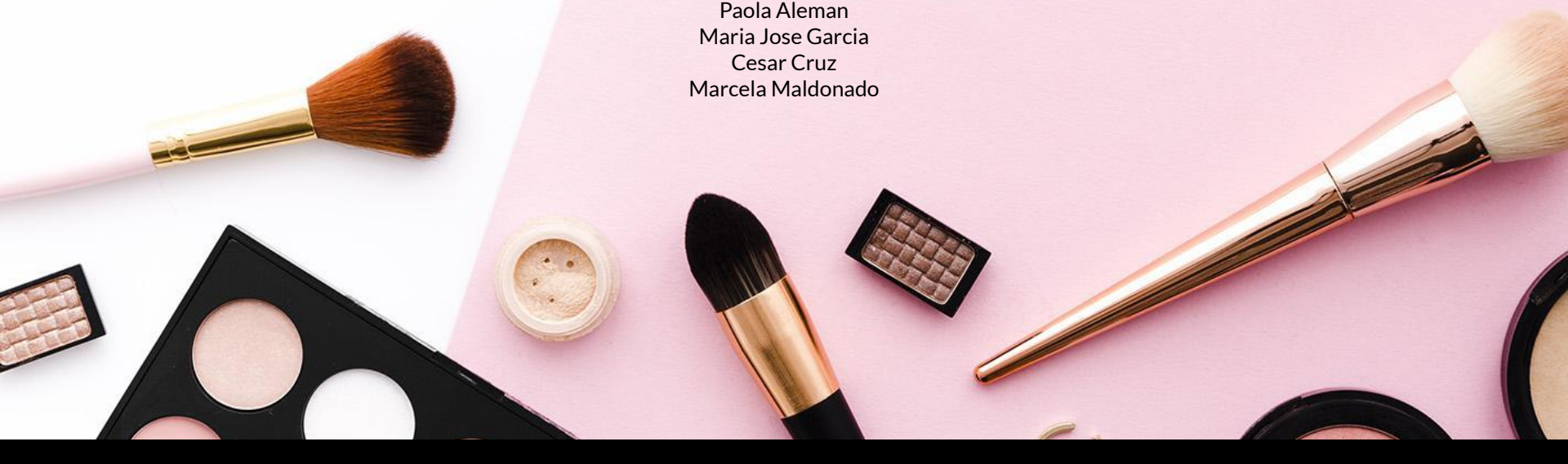
MAY 2023

# CHEMICALS IN COSMETICS

## Data Analytics: PROJECT 3

Paola Aleman
Maria Jose Garcia
Cesar Cruz
Marcela Maldonado



# INTRODUCTION

The data that we have chosen to work with has been reported to the California State Cosmetics Program (CSCP) in the California Department of Public Health (CDPH).

The **purpose of the CSPC** is to collect information of products that contain any ingredients known or suspected to cause: cancer, birth defects, developmental or reproductive harm. There is a list of “reportable” ingredients in which the California Safe Cosmetics Act requires all the manufacturers, packers, and or distributors to submit if:

- the company has an annual aggregate sale of cosmetic products of 1 million dollars or more and
- has sold cosmetic products in California on or after January 2007.

California Department of PublicHealth

Github Link

# INDEX

1

## Data Set

Explains some of the variables included

2

## Project Development

Describes workflow and issues encountered

3

## HTML

The user friendly web page created

4

## Graphics

**1.**All companies reported
**2.**Top 10 companies reported
**3.**Primary categories by count

5

## Conclusion

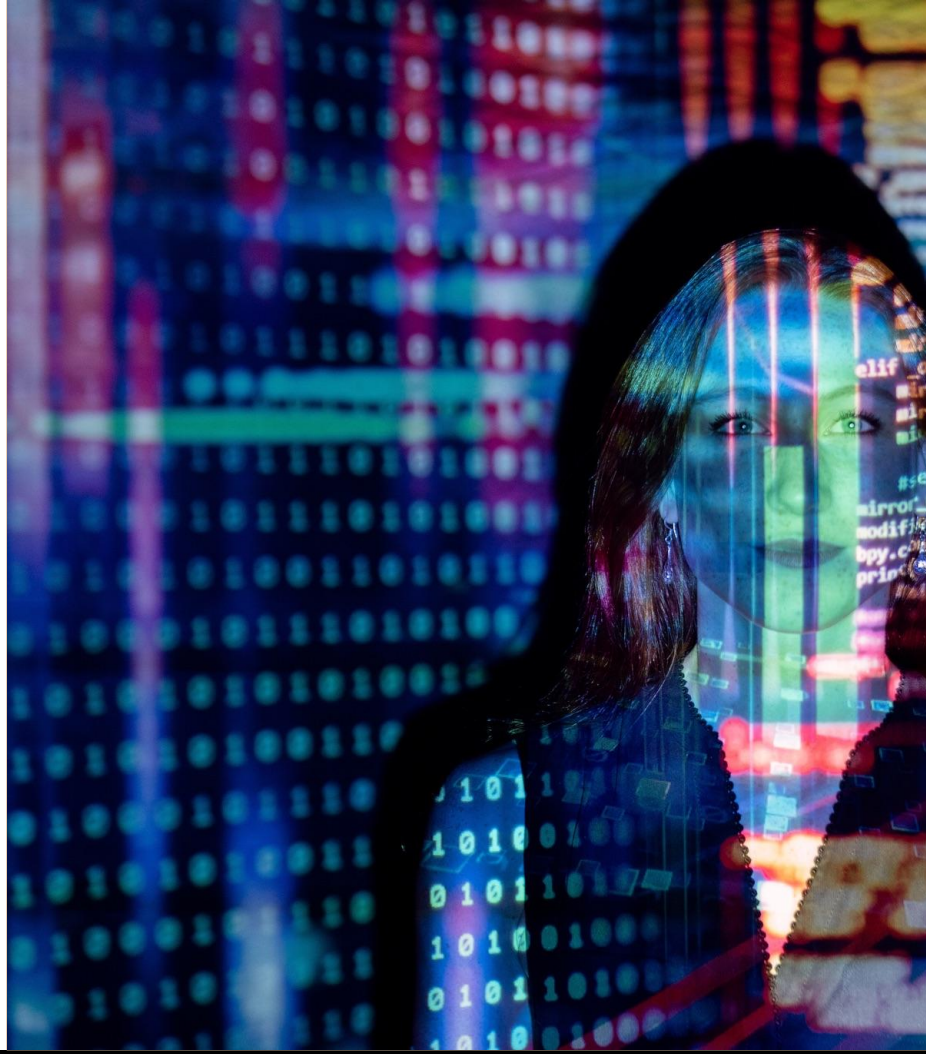
Data summary

# DATA SET

The data set includes 22 columns and 114,635 rows. To mention a few:

- Primary product categories
- Subcategories
- Company names reported
- Chemical names reported
- Product Names
- Dates Reported

We had out data in a csv file, which was uploaded to MongoDB. Using a flask we accessed this information for our  html.



# Project Development

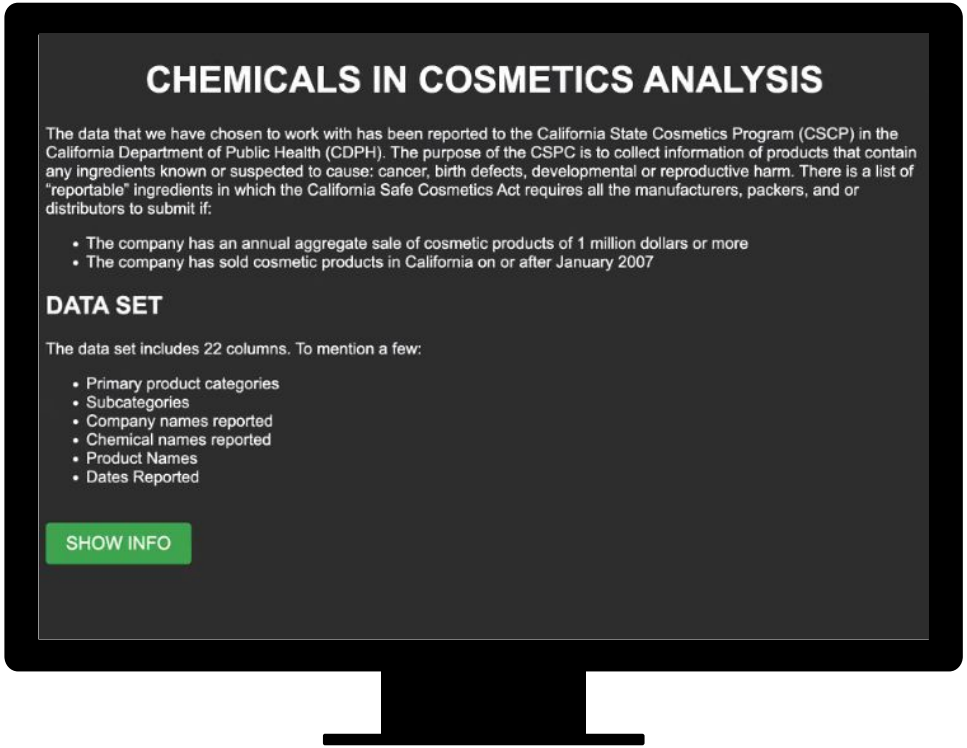
**Day 1:** Found the database we wanted to work with and worked in our 1 page proposal.
**Day 2:** Cleaned Data using Python and uploaded our dataset to MongoDB.
**Day 3:** The flask,index and API were created. We also had an issue with CORS, since all of us had Mac we lost a whole day with this issue.
Solutions we tried to solve the CORS problem:

- Using different computers
- Uploading our dataset to Life Server

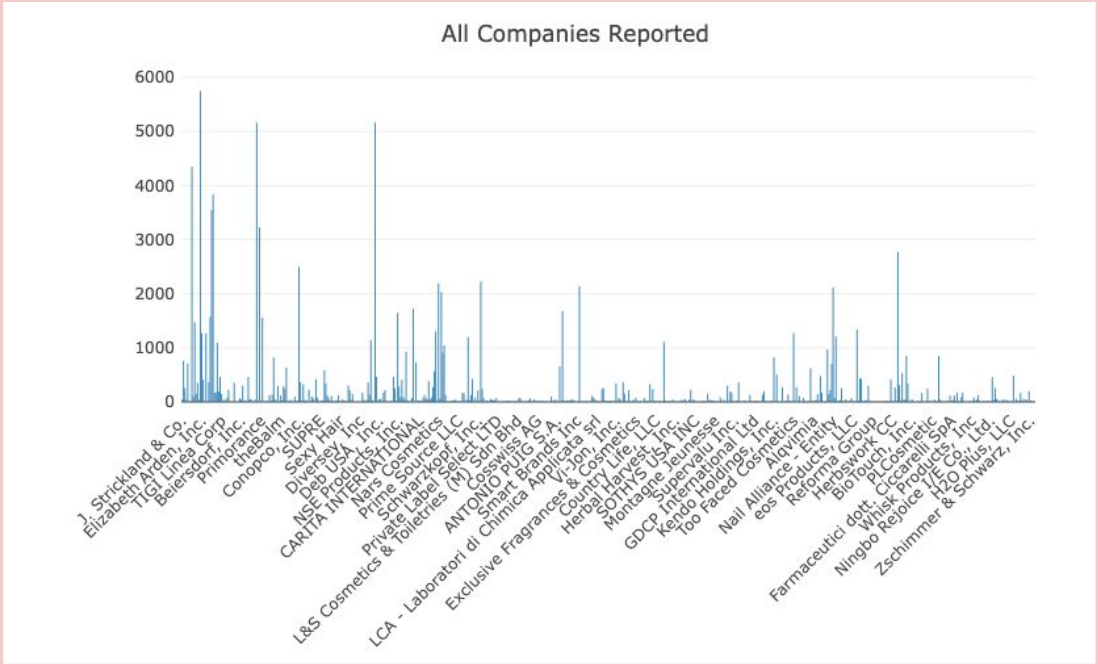
**Day 4:** Found a solution to our problem. We created the static and template folders. In the app.py we assigned the index to the render template and we used the localhost/index address to pull our dataset and overcome the CORS issue.
**Day 5:** We started pulling the information from the flask so we could show the graphs and making the html look a little nicer.
**Day 6:** Got the powerpoint presentation ready and final details.

# HTML

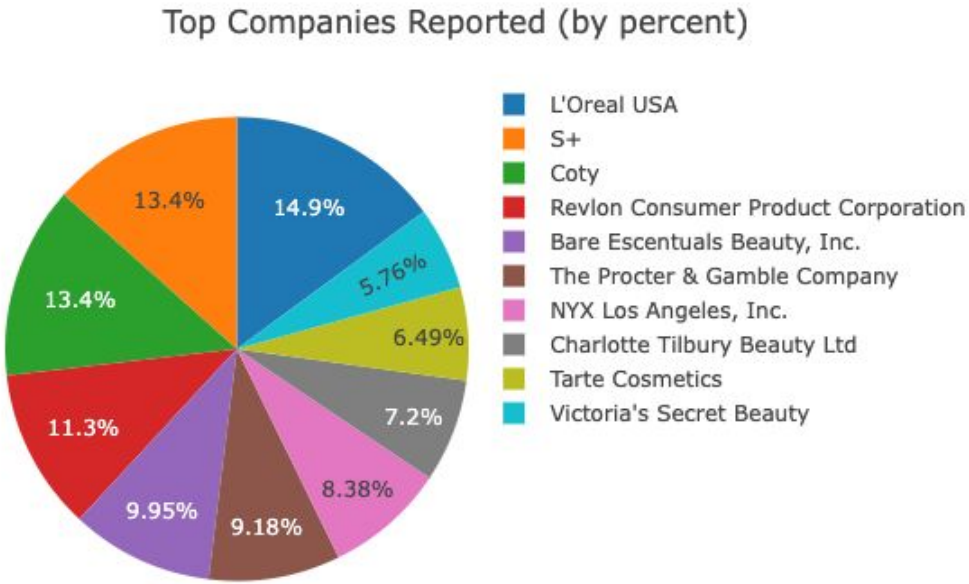
User driven interaction we included is a button which shows the info



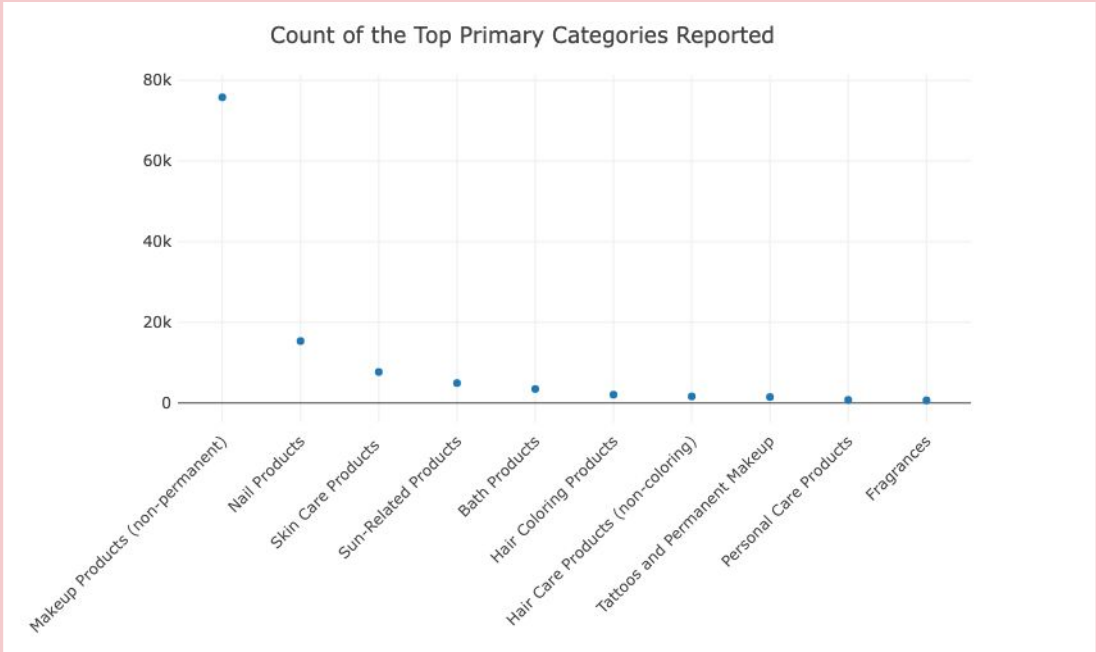
# GRAPHICS




# All Companies Reported




# Top 10 Companies Reported




# Primary Categories Reported by Count

# Conclusions

- Project
  - ➔ All the team members own a Mac so we had a lot of trouble working with the flask.
  - ➔ We also came to the conclusion that working with Java is not very easy and needs practice to dominate and feel more confident .
  - ➔ It is complicated working with a big database (100MG) .

# Conclusions

- Database
  - ➔ Out of the 44 companies reported, the top 3 were L'Oreal Paris USA, S+ and Coty.
  - ➔ Out of the 13 primary categories, the top 3 were make up products, nail products and skin care products; however the amount of makeup products reported is very high compared to the other ones (with more than 70k+).
  - ➔ In the scatter plot chart, most of the categories reported (8) have less than 10k products reported.

# THANKS!

Do you have any questions?