Chemicals In Cosmetics

This project explores harmful and potentially harmful chemicals found in cosmetic products according to information reported to the California Safe Cosmetics Program (CSCP) in the California Department of Public Health.

Data source: https://www.kaggle.com/datasets/zarinhelena/chemicals-in-cosmetics?select=chemicals-in-cosmetics.csv

Business Issue

- How can we give consumers easy access to how many chemicals are found in their favorite cosmetic products?

- Do certain companies have a higher amount of harmful chemicals per product than others?

- Which companies have made the most effort to remove harmful chemicals from cosmetic products?

Exploratory Data Analysis

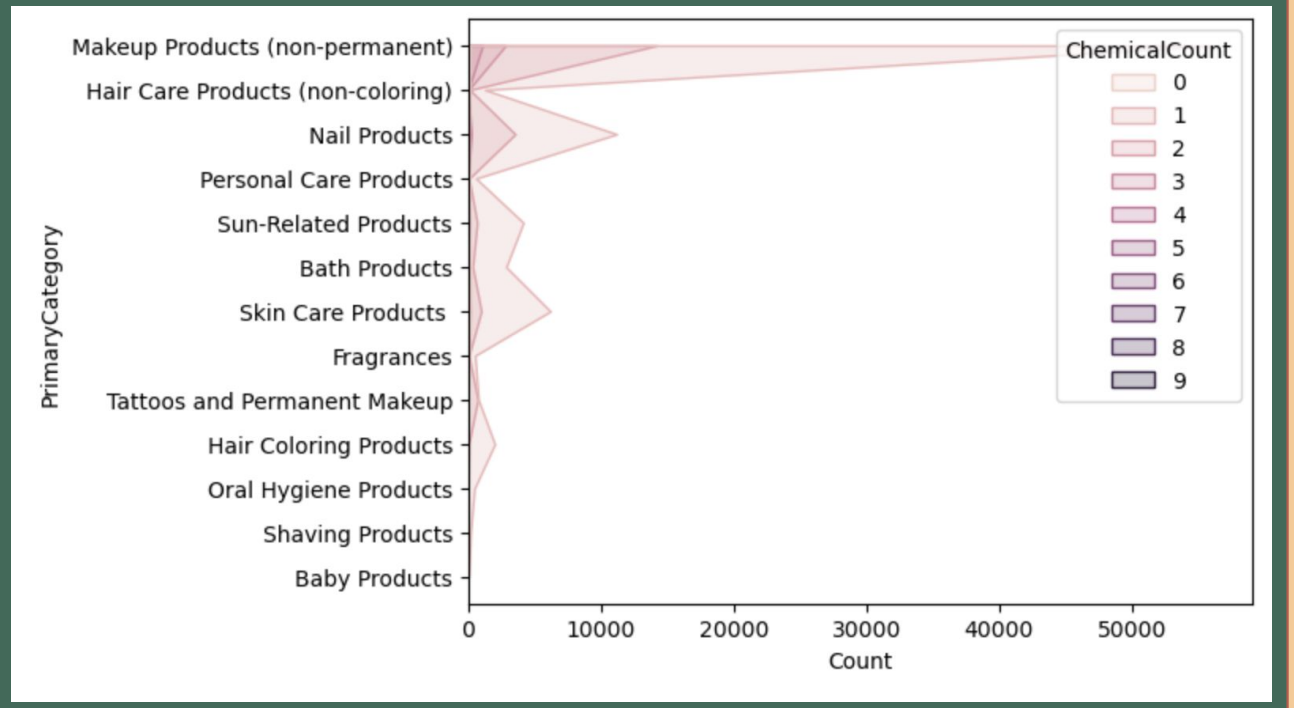
- 123 unique Chemical Names
- 13 Primary Categories
- 601 unique Company Names
- 2683 unique Brand Names

	CDPHId	ProductName	CSFId	CSF	CompanyId	CompanyName	BrandName	PrimaryCategoryId	PrimaryCategory	SubCategoryId	...	CasNumber
0	2	ULTRA COLOR RICH EXTRA PLUMP LIPSTICK-ALL SHADES	NaN	NaN	4	New Avon LLC	AVON	44	Makeup Products (non-permanent)	53	...	13463-67-7

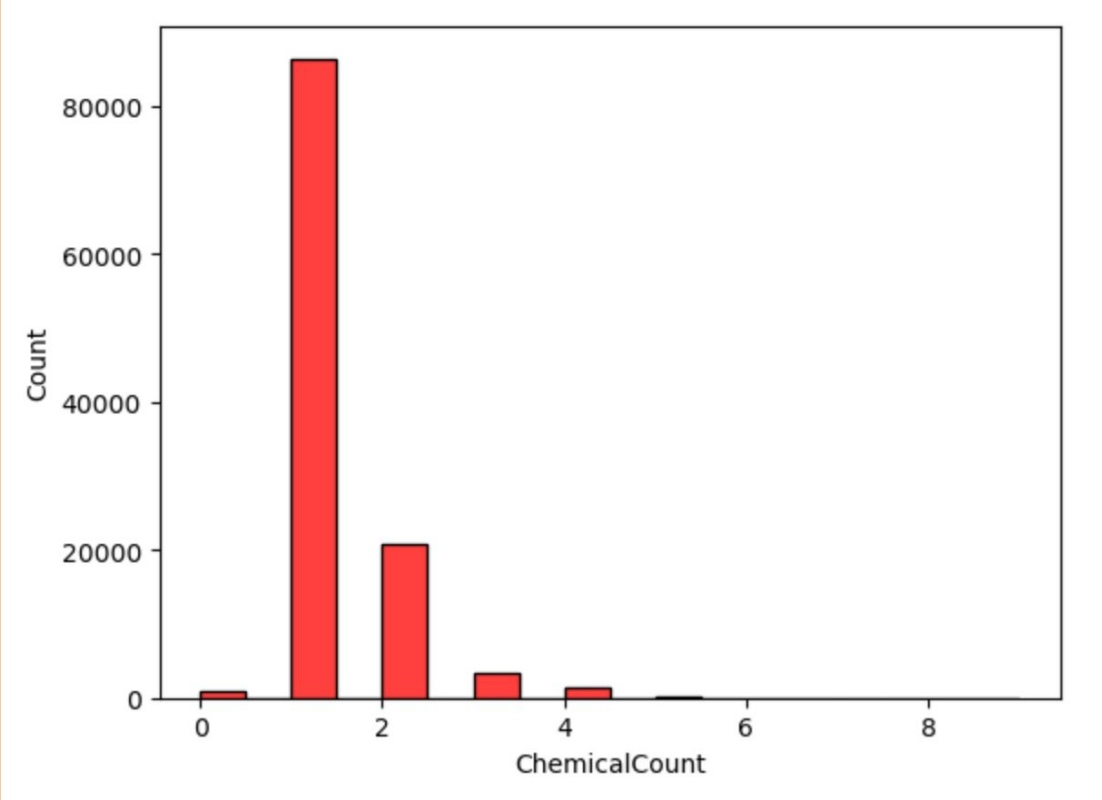
ChemicalName	InitialDateReported	MostRecentDateReported	DiscontinuedDate	ChemicalCreatedAt	ChemicalUpdatedAt	ChemicalDateRemoved	ChemicalCount
Titanium dioxide	06/17/2009	08/28/2013	02/01/2011	07/09/2009	07/09/2009	NaN	1

Exploratory Data Analysis

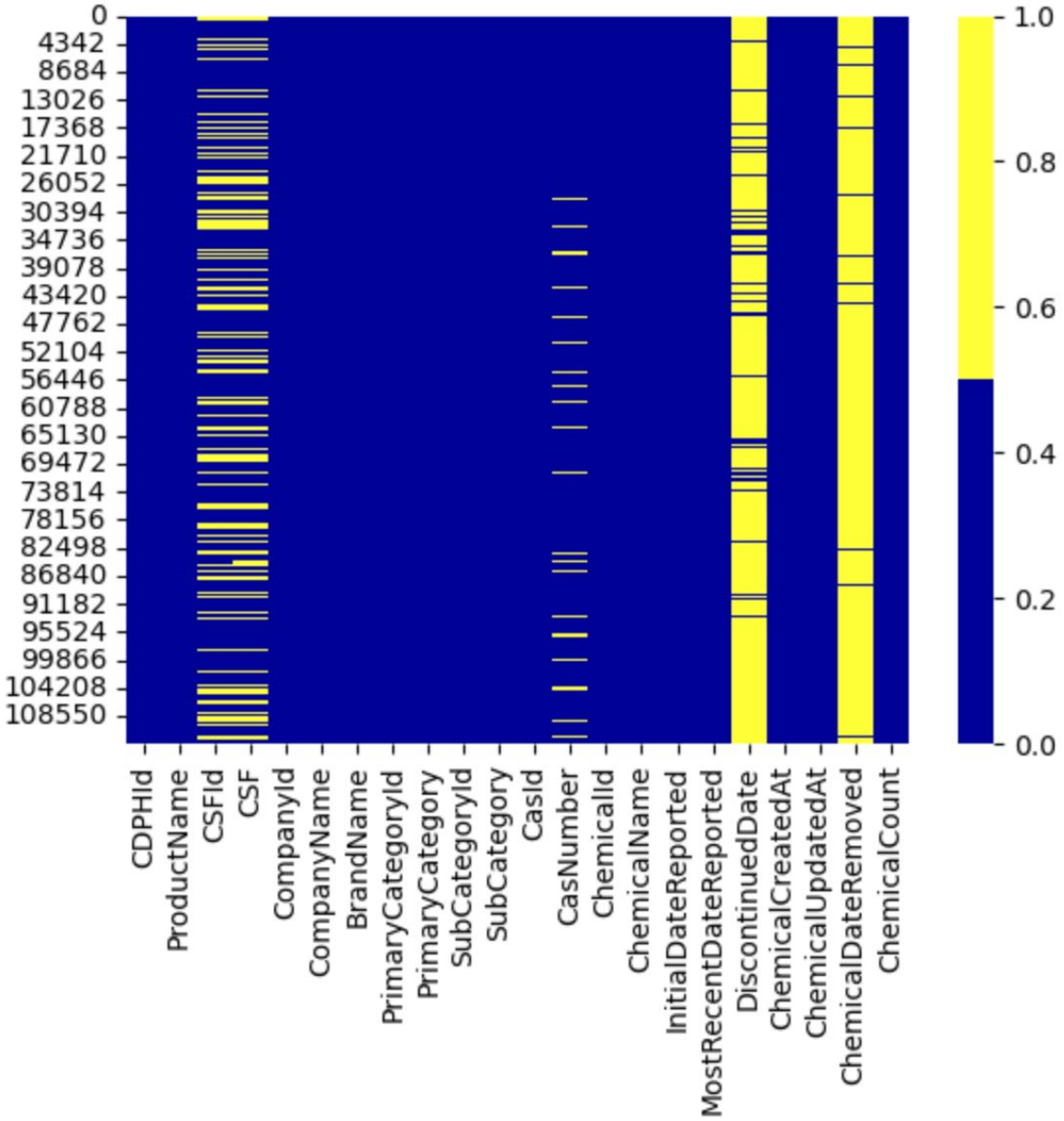
Majority of products are Makeup Products



Majority of products have a chemical count of 1



Data Cleaning



Missing Values

The following columns contained missing data:

- Dropped Columns:
  - CSFId (30%)
  - CSF (30%)
    - color, scent, or flavor

- Filled Null Values with “MISSING”
  - CasNumber (6%)
  - DiscontinuedDate (89%)
  - ChemicalDateRemoved (97%)

Data Cleaning

Duplicates

```
df.shape
```

```
(112870, 20)
```

```
df.duplicated().value_counts()
```

```
False    63215
True     49655
dtype: int64
```

```
df = df.drop_duplicates()
df.shape
```

```
(63215, 20)
```

Inconsistent Data

- Changed all columns containing string values to lowercase, then checked again for duplicates.
  - Product Name
  - Company Name
  - Brand Name
  - Chemical Name