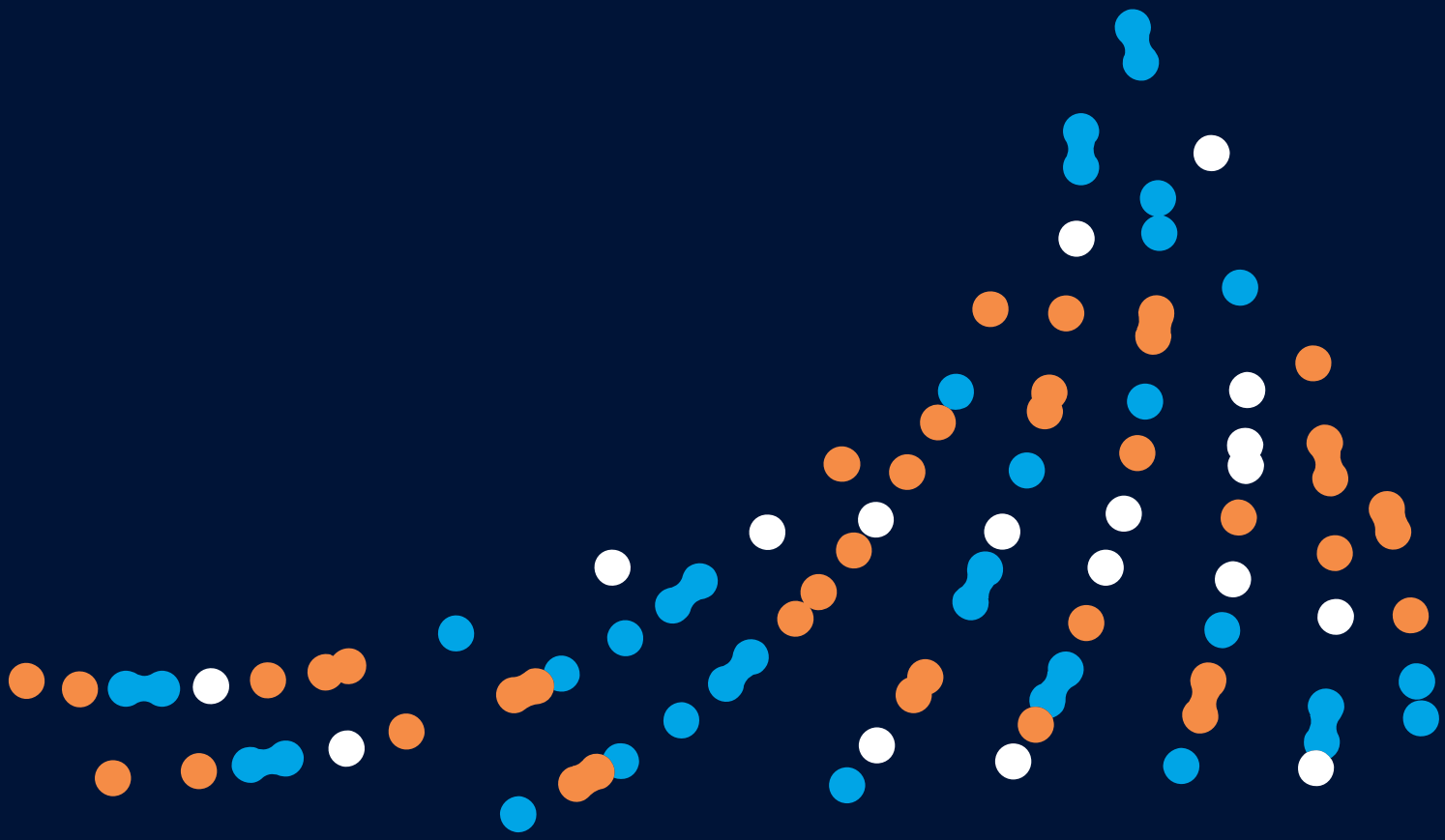


ETL Project

Two Sunnys

Report

Sunmin Lee
Thidar Swe Tin

21 April 2019

Contents

Project Description

Web scraping of products and reviews from 2 sources (Sephora and Strawberrynet) for the following categories:

Skin care categories:

- Eye & lip
- Cleansers
- Masks
- Moisturizers / Treatments
- Sun Care

Steps:

1. Scrape all the product urls from the category page
2. Scrape all the product details from the product urls
3. Store raw data from scraped URLs into MongoDB
4. Retrieve data from MongoDB and transform the data using panda dataframes
5. Load the transformations back into MySQL
6. Create views in MySQL for both sites to look at
    - Top 100 products by ratings
    - Top 10 brands by ratings

Extract

Data Source 1 – Sephora (Sunmin Lee)

Step 1: Scrape all the product urls from the category page

URL of webpage to be scraped

- eye_url = 'https://www.sephora.com/shop/eye-treatment-dark-circle-treatment?pageSize=300'
- lip_url = 'https://www.sephora.com/shop/lip-treatments?pageSize=300'
- cleanser1_url = 'https://www.sephora.com/shop/cleanser?pageSize=300&currentPage=1'
- cleanser2_url = 'https://www.sephora.com/shop/cleanser?pageSize=300&currentPage=2'
- mask1_url = 'https://www.sephora.com/shop/face-mask?pageSize=300&currentPage=1'
- mask2_url = 'https://www.sephora.com/shop/face-mask?pageSize=300&currentPage=2'
- moist1_url = 'https://www.sephora.com/shop/moisturizing-cream-oils-mists?pageSize=300&currentPage=1'
- moist2_url = 'https://www.sephora.com/shop/moisturizing-cream-oils-mists?pageSize=300&currentPage=2'
- moist3_url = 'https://www.sephora.com/shop/moisturizing-cream-oils-mists?pageSize=300&currentPage=3'
- treat1_url = 'https://www.sephora.com/shop/facial-treatments?pageSize=300&currentPage=1'
- treat2_url = 'https://www.sephora.com/shop/facial-treatments?pageSize=300&currentPage=2'
- suncare_url = 'https://www.sephora.com/shop/sunscreen-sun-protection?pageSize=300'

*Products found [Total: **2852**]*

- Eye & lip [**381**]
- Cleansers [**517**]
- Masks [**338**]
- Moisturizers / Treatments [**1398**]
- Sun Care [**218**]

Step 2: Scrape all the product details from the product urls

Took a total of **7 hours** to scrape all the **2852** product details.

Step 3: Store raw data from scraped URLs into MongoDB

Refer to Appendices: MongoDB Database Structure (Sephora) for more details.

Data Source 2 – Strawberrynet (Thidar Swe Tin)

Step 1: Scrape all the product urls from the category page

URL of webpage to be scraped

urls = [{"eyes_lips":"/en-us/skincare/?groupid=3&sort=popularity"},
{"cleansers":"/en-us/skincare/cleansers/t/?groupid=4&sort=popularity"},
{"masks":"/en-us/skincare/masks/t/?groupid=4&sort=popularity"},
{"moisturisers":"/en-us/skincare/moisturizers-treatments/t/?groupid=4&sort=popularity"},
{"sun_care":"/en-us/skincare/?groupid=7&sort=popularity"}]

*Products found [Total: **3247**]*

- Eye & lip [**641**]
- Cleansers [**444**]
- Masks [**521**]
- Moisturizers / Treatments [**1228**]
- Sun Care [**413**]

Step 2: Scrape all the product details from the product urls

Took a total of **8.5 hours** to scrape all the **3247** products details and reviews.

Step 3: Store raw data from scraped URLs into MongoDB

Refer to Appendices: MongoDB Database Structure (Strawberrynet) for more details.

Transform

Data Source 1 – Sephora (Sunmin Lee)

Step 4: Retrieve data from MongoDB and transform the data using pandas dataframes

1. Export the collections in JSON format from MongoDB and import into Pandas dataframe

```
In [45]: # Export the collections in JSON format from MongoDB and import into Pandas Dataframe
json_moisttreat = '../project2sunnys_SL/sephora_db_MongoDB/col_moisttreat.json'
moisttreat_df = pd.read_json(json_moisttreat, lines=True)
# Insert category name
moisttreat_df.insert(1, 'category_name', 'moisturisers', True)
moisttreat_df.head()
```

Out[45]:

	_id	category_name	avg_stars	brand_name	product_name	product_price	product_url	review_num
0	{'$oid': '5cbab5bfe26c8d611aa0fc8b'}	moisturisers	4 stars	Drunk Elephant	Protini Polypeptide Moisturizer	$68.00	https://www.sephora.com/product/protini-tm-pol...	1K reviews
1	{'$oid': '5cbab5bfe26c8d611aa0fc8c'}	moisturisers	4 stars	La Mer	Crème de la Mer	$175.00	https://www.sephora.com/product/creme-de-la-me...	495 reviews
2	{'$oid': '5cbab5bfe26c8d611aa0fc8d'}	moisturisers	4 stars	IT Cosmetics	CC+ Cream with SPF 50+	$39.00	https://www.sephora.com/product/your-skin-but-...	2K reviews
3	{'$oid': '5cbab5bfe26c8d611aa0fc8e'}	moisturisers	4 stars	Tatcha	The Water Cream	$68.00	https://www.sephora.com/product/the-water-crea...	1K reviews
4	{'$oid': '5cbab5bfe26c8d611aa0fc8f'}	moisturisers	4 stars	SK-II	Facial Treatment Essence	179.00(214.00 value)	https://www.sephora.com/product/facial-treatme...	790 reviews

```
In [33]: len(moisttreat_df)
```

Out[33]: 1398

2. Combine all product details into one dataframe

```
In [47]: # Combine all product details into one dataframe
prd_detail_df = cleanser_df.append(eyelip_df)
prd_detail_df1 = prd_detail_df.append(mask_df)
prd_detail_df2 = prd_detail_df1.append(moisttreat_df)
prd_detail_df3 = prd_detail_df2.append(suncare_df)
len(prd_detail_df3)
```

Out[47]: 2852

```
In [49]: # Show all product details in one dataframe
prd_detail_df3.head()
```

Out[49]:

	_id	category_name	avg_stars	brand_name	product_name	product_price	product_url	review_num
0	{'$oid': '5cba94c2e26c8d611aa0f934'}	cleansers	4.5 stars	Drunk Elephant	T.L.C. Framboos™ Glycolic Night Serum	$90.00	https://www.sephora.com/product/t-l-c-framboos...	1K reviews
1	{'$oid': '5cba94c2e26c8d611aa0f935'}	cleansers	4.5 stars	Drunk Elephant	T.L.C Sukari Babyfacial Mask	$80.00	https://www.sephora.com/product/t-l-c-sukari-b...	1K reviews
2	{'$oid': '5cba94c2e26c8d611aa0f936'}	cleansers	4 stars	SK-II	Facial Treatment Essence	179.00(214.00 value)	https://www.sephora.com/product/facial-treatme...	790 reviews
3	{'$oid': '5cba94c2e26c8d611aa0f937'}	cleansers	4.5 stars	Fresh	Soy Face Cleanser	$38.00	https://www.sephora.com/product/soy-face-clean...	6K reviews
4	{'$oid': '5cba94c2e26c8d611aa0f938'}	cleansers	4.5 stars	The Ordinary	Glycolic Acid 7% Toning Solution	$8.70	https://www.sephora.com/product/the-ordinary-d...	433 reviews

3. Create a filtered dataframe with specific columns

In [50]:
```
# Create a filtered dataframe with specific columns
prd_detail_df4 = prd_detail_df3[['category_name', 'avg_stars', 'brand_name', 'product_name', 'product_price', 'product_
prd_detail_df4.head()
```

Out[50]:

	category_name	avg_stars	brand_name	product_name	product_price	product_url	review_num
0	cleansers	4.5 stars	Drunk Elephant	T.L.C. Framboos™ Glycolic Night Serum	$90.00	https://www.sephora.com/product/t-l-c-framboos...	1K reviews
1	cleansers	4.5 stars	Drunk Elephant	T.L.C Sukari Babyfacial Mask	$80.00	https://www.sephora.com/product/t-l-c-sukari-b...	1K reviews
2	cleansers	4 stars	SK-II	Facial Treatment Essence	179.00(214.00 value)	https://www.sephora.com/product/facial-treatme...	790 reviews
3	cleansers	4.5 stars	Fresh	Soy Face Cleanser	$38.00	https://www.sephora.com/product/soy-face-clean...	6K reviews
4	cleansers	4.5 stars	The Ordinary	Glycolic Acid 7% Toning Solution	$8.70	https://www.sephora.com/product/the-ordinary-d...	433 reviews

4. Clean the data by dropping duplicates / Clean the data by removing units

In [51]:
```
# Clean the data by dropping duplicates
prd_detail_df4.drop_duplicates('product_url', inplace=True)
len(prd_detail_df4)
```

Out[51]: 2525

In [56]:
```
# Clean the data by removing units
prd_detail_df4['avg_stars'] = prd_detail_df4['avg_stars'].str.replace('stars', '')
prd_detail_df4['review_num'] = prd_detail_df4['review_num'].str.replace('reviews', '')
prd_detail_df4['review_num'] = prd_detail_df4['review_num'].str.replace('K', '000')
prd_detail_df4['product_price'] = prd_detail_df4['product_price'].str.replace('$', '')
prd_detail_df4.head()
```

Out[56]:

	category_name	avg_stars	brand_name	product_name	product_price	product_url	review_num
0	cleansers	4.5	Drunk Elephant	T.L.C. Framboos™ Glycolic Night Serum	90.00	https://www.sephora.com/product/t-l-c-framboos...	1000
1	cleansers	4.5	Drunk Elephant	T.L.C Sukari Babyfacial Mask	80.00	https://www.sephora.com/product/t-l-c-sukari-b...	1000
2	cleansers	4	SK-II	Facial Treatment Essence	179.00(214.00 value)	https://www.sephora.com/product/facial-treatme...	790
3	cleansers	4.5	Fresh	Soy Face Cleanser	38.00	https://www.sephora.com/product/soy-face-clean...	6000
4	cleansers	4.5	The Ordinary	Glycolic Acid 7% Toning Solution	8.70	https://www.sephora.com/product/the-ordinary-d...	433

Data Source 2 – Strawberrynet (Thidar Swe Tin)

Step 4: Retrieve data from MongoDB and transform the data using pandas dataframes

1. Get all items from product details collection in MongoDB


```
Out[362]: [{'eyes_lips': [{'product_url': 'https://www.strawberrynet.com/en-us/skincare/elizabeth-arden/
          488/',
              'brand_name': 'Elizabeth Arden',
              'product_name': 'Eight Hour Lipcare Stick ',
              'product_price': '9.50',
              'review_num': 169,
              'avg_stars': 4.437869822485207,
              'review_details': [[{'review_date': '11/4/2019',
                 'review_header': 'lip moisture',
                 'review_comment': 'excellent for the lips. give good feeling.',
                 'stars_num': 5},
                {'review_date': '7/4/2019',
                 'review_header': 'excellent',
                 'review_comment': 'Very smooth and keeps your lips hydrated',
                 'stars_num': 5},
```


2. Store categories in DataFrame (category_df)

	category_name
0	eyes_lips
1	cleansers
2	masks
3	moisturisers
4	sun_care

3. Store unique brand names in DataFrame (brand_df)

	brand_name	ds_id
0	Elizabeth Arden	1
1	Shiseido	1
2	Clinique	1
3	L'Occitane	1
4	Clarins	1

4. Store products in DataFrame (product_df)

	category_name	product_url	brand_name	product_name	product_price	review_num	avg_stars	ds_id
0	eyes_lips	https://www.strawberrynet.com/en-us/skincare/e...	Elizabeth Arden	Eight Hour Lipcare Stick	9.50	169	4.44	1
1	eyes_lips	https://www.strawberrynet.com/en-us/skincare/s...	Shiseido	Benefiance WrinkleResist24 Intensive Eye Conto...	61.50	0	0.00	1
2	eyes_lips	https://www.strawberrynet.com/en-us/skincare/c...	Clinique	All About Eye Serum De-Puffing Eye Massage	34.50	0	0.00	1
3	eyes_lips	https://www.strawberrynet.com/en-us/skincare/l...	L'Occitane	Shea Butter Lip Balm Stick	11.00	0	0.00	1
4	eyes_lips	https://www.strawberrynet.com/en-us/skincare/c...	Clarins	Hydra-Essentiel Moisture Replenishing Lip Balm	21.50	0	0.00	1

5. Get all the products that has reviews (review_df)

	index	category_name	product_url	brand_name	product_name	product_price	review_num	avg_stars	ds_id
0	0	eyes_lips	https://www.strawberrynet.com/en-us/skincare/e...	Elizabeth Arden	Eight Hour Lipcare Stick	9.50	169	4.44	1
1	9	eyes_lips	https://www.strawberrynet.com/en-us/skincare/c...	Clarins	Eye Contour Gel	32.50	41	4.68	1
2	15	eyes_lips	https://www.strawberrynet.com/en-us/skincare/d...	Darphin	Wrinkle Corrective Eye Contour Cream	55.00	8	3.75	1
3	17	eyes_lips	https://www.strawberrynet.com/en-us/skincare/c...	Clarins	Instant Eye Make Up Remover	28.00	41	5.05	1
4	20	eyes_lips	https://www.strawberrynet.com/en-us/skincare/d...	Decleor	Prolagene Lift Lift & Firm Eye Care	42.00	1	5.00	1

6. Retrieve reviews and store reviews in dataframe (review_df)

	review_date	review_header	review_comment	stars_num	product_id
0	11/4/2019	lip moisture	excellent for the lips. give good feeling.	5	1
1	7/4/2019	excellent	Very smooth and keeps your lips hydrated	5	1
2	2/4/2019	lip stick	very good make the lips smooth.	5	1
3	17/03/2019	love it	Love the way it feels and lasts long!	5	1
4	9/3/2019	good product	Is a lip that I like very much, It helps to ke...	5	1

7. Drop duplicates, invalid dates and convert to date time (review_df)

	review_date	review_header	review_comment	stars_num	product_id
0	2019-04-11	lip moisture	excellent for the lips. give good feeling.	5	1
1	2019-04-07	excellent	Very smooth and keeps your lips hydrated	5	1
2	2019-04-02	lip stick	very good make the lips smooth.	5	1
3	2019-03-17	love it	Love the way it feels and lasts long!	5	1
4	2019-03-09	good product	Is a lip that I like very much, It helps to ke...	5	1

8. Replace brand names with brand id (product_df)

	category_name	product_url	brand_id	product_name	product_price	review_num	avg_stars	ds_id
0	eyes_lips	https://www.strawberrynet.com/en-us/skincare/e...	1	Eight Hour Lipcare Stick	9.50	169	4.44	1
1	eyes_lips	https://www.strawberrynet.com/en-us/skincare/s...	2	Benefiance WrinkleResist24 Intensive Eye Conto...	61.50	0	0.00	1
2	eyes_lips	https://www.strawberrynet.com/en-us/skincare/c...	3	All About Eye Serum De-Puffing Eye Massage	34.50	0	0.00	1
3	eyes_lips	https://www.strawberrynet.com/en-us/skincare/l...	4	Shea Butter Lip Balm Stick	11.00	0	0.00	1
4	eyes_lips	https://www.strawberrynet.com/en-us/skincare/c...	5	Hydra-Essentiel Moisture Replenishing Lip Balm	21.50	0	0.00	1

9. Replace category names with category id (product_df)

	category_id	product_url	brand_id	product_name	product_price	review_num	avg_stars	ds_id
0	1	https://www.strawberrynet.com/en-us/skincare/e...	1	Eight Hour Lipcare Stick	9.50	169	4.44	1
1	1	https://www.strawberrynet.com/en-us/skincare/s...	2	Benefiance WrinkleResist24 Intensive Eye Conto...	61.50	0	0.00	1
2	1	https://www.strawberrynet.com/en-us/skincare/c...	3	All About Eye Serum De-Puffing Eye Massage	34.50	0	0.00	1
3	1	https://www.strawberrynet.com/en-us/skincare/l...	4	Shea Butter Lip Balm Stick	11.00	0	0.00	1
4	1	https://www.strawberrynet.com/en-us/skincare/c...	5	Hydra-Essentiel Moisture Replenishing Lip Balm	21.50	0	0.00	1

10. Clean product price (product_df)

	category_id	product_url	brand_id	product_name	price	review_num	avg_stars	ds_id
0	1	https://www.strawberrynet.com/en-us/skincare/e...	1	Eight Hour Lipcare Stick	9.5	169	4.44	1
1	1	https://www.strawberrynet.com/en-us/skincare/s...	2	Benefiance WrinkleResist24 Intensive Eye Conto...	61.5	0	0.00	1
2	1	https://www.strawberrynet.com/en-us/skincare/c...	3	All About Eye Serum De-Puffing Eye Massage	34.5	0	0.00	1
3	1	https://www.strawberrynet.com/en-us/skincare/l...	4	Shea Butter Lip Balm Stick	11.0	0	0.00	1
4	1	https://www.strawberrynet.com/en-us/skincare/c...	5	Hydra-Essentiel Moisture Replenishing Lip Balm	21.5	0	0.00	1

Load

Step 5: Load the transformations back into MySQL

Load the 4 dataframes into MySQL (.to_sql)

- category_df
- brand_df
- product_df
- review_df

Refer to Appendices: MySQL Database Structure for more details.

```
In [115]: # Connect to local MySQL database
          connection = 'root:password@localhost/skincare_db'
          #engine = create_engine(f'mysql:///{connection}')
          engine = create_engine(f'mysql://{connection}?charset=utf8mb4')

In [116]: # Confirm tables
          engine.table_names()

Out[116]: ['brand', 'category', 'data_source', 'product', 'review']

In [ ]:   # Load dataframe into database
          prd_detail_df4.to_sql(name='product', con=engine, if_exists='append', index=False)
```

Step 6: Create views in MySQL for both sites to look at

- Top 100 products by ratings

category_name	brand_name	product_name	price
cleansers	Yves Saint Laurent	Top Secrets Toning & Cleansing Micellar Water	40.5
eyes_lips	Payot	Techni Regard - Anti-Wrinkles Smoothing Care (...	43.5
moisturisers	L'Occitane	Shea Ultra Rich Comforting Cream - Dry to Very...	40.5
moisturisers	Clarins	Hydra-Essentiel Moisturizes & Quenches Silky Cr...	46
sun_care	EltaMD	UV Facial Moisturizing Facial Sunscreen SPF 30 -...	22
masks	Biotherm	Pure.Fect Skin 2 in1 Pore Mask (Normal to Oily S...	30.5
moisturisers	Guinot	Hydra Sensitive Face Cream	47.5
moisturisers	Aveda	Botanical Kinetics Hydrating Lotion	50.5
moisturisers	Ahava	Time To Hydrate Active Moisture Gel Cream	38
moisturisers	Aveda	Botanical Kinetics Hydrating Lotion	50.5
cleansers	Darphin	Cleansing Foam Gel with Water Lily	32.5
cleansers	Payot	Les Demaquillantes Efface' Cils Douceur Instant ...	25.5
eyes_lips	Aesop	Rosehip Seed Lip Cream	19.5
moisturisers	Clarins	Gentle Day Cream	61.5

- Top 10 brands by ratings

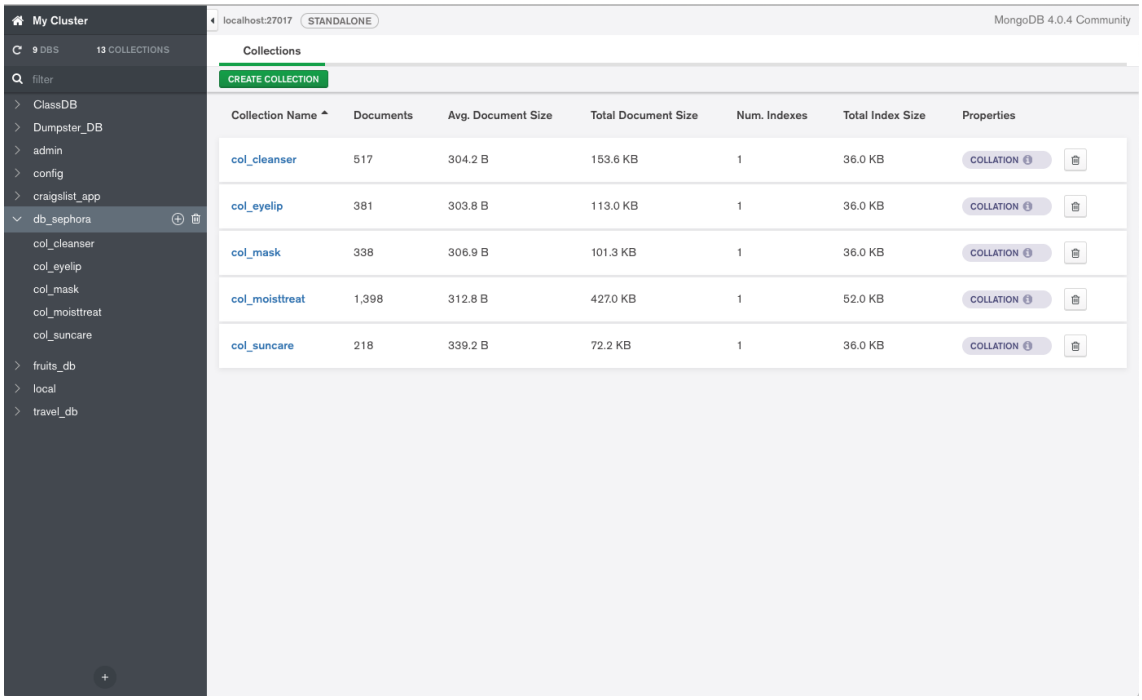
brand_name	avg_rating
Innisfree	4
Joey New York	3.75
Orico London	3.605
Prevage	2.5883333333333334
Jane Iredale	2.36
Aesop	2.158666666666667
EltaMD	2.12
Coryse Salome	2.0957142857142856
My Beauty Diary	2.0714285714285716

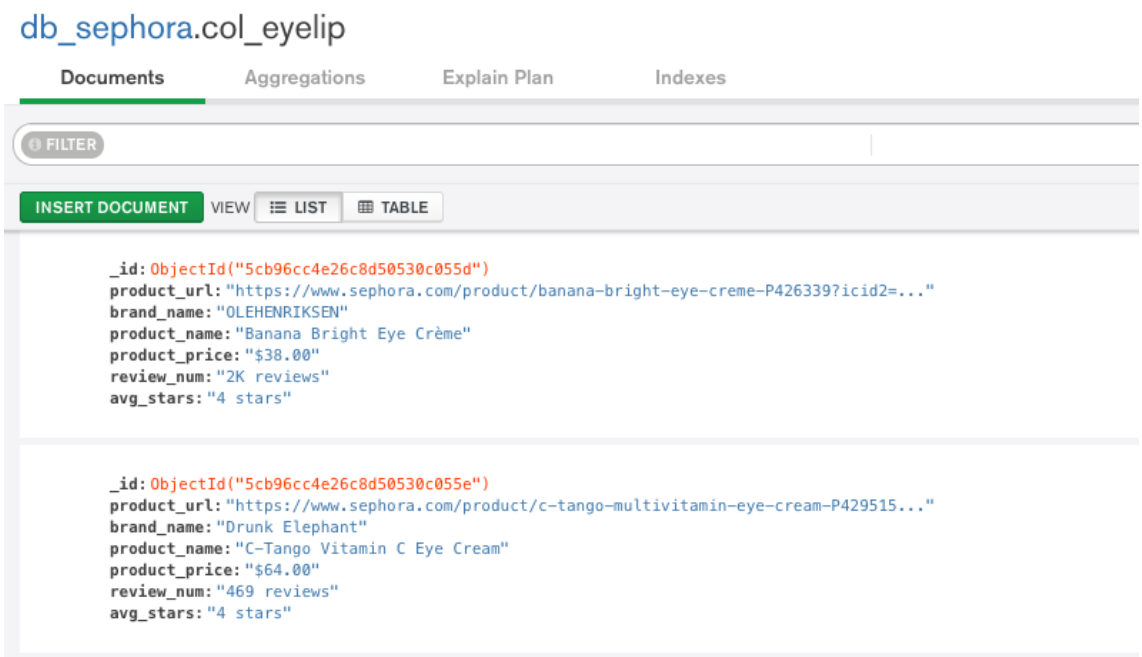
- Refer to Appendices: MySQL Database Script for more details.

Appendices

MongoDB Database

Database Structure (Sephora)





Database Structure (Strawberrynet)

category

	_id ObjectId	eyes_lips Array	cleansers Array
1	5cb888ea6896a81d7065099a	[] 641 elements	[] 444 elements

masks Array	moisturisers Array	sun_care Array
[] 521 elements	[] 1228 elements	[] 413 elements

sbn_db.category

DOCUMENTS 1 TOTAL SIZE 235.4KB AVG. SIZE 235.4KB

Documents Aggregations Explain Plan Indexes

FILTER OPTI

INSERT DOCUMENT VIEW LIST TABLE Displaying


```
_id: ObjectId("5cb888ea6896a81d7065099a")
eyes_lips: Array
  0: "/en-us/skincare/elizabeth-arden/eight-hour-lipcare-stick/13488/"
  1: "/en-us/skincare/shiseido/benefiance-wrinkleresist24-intensive/140188/"
  2: "/en-us/skincare/clinique/all-about-eye-serum-de-puffing/104889/"
  3: "/en-us/skincare/l-occitane/shea-butter-lip-balm-stick/42067/"
  4: "/en-us/skincare/clarins/hydra-essentiel-moisture-replenishing/94452/"
  5: "/en-us/skincare/clarins/extra-firming-eye-lift-perfecting/167004/"
  6: "/en-us/skincare/estee-lauder/advanced-night-repair-eye-synchronized/17..."
```


sbn_db.product_details

DOCUMENTS 5 TOTAL SIZE 1.7MB AVG. SIZE 354.6KB

Documents Aggregations Explain Plan Indexes

FILTER OPTI

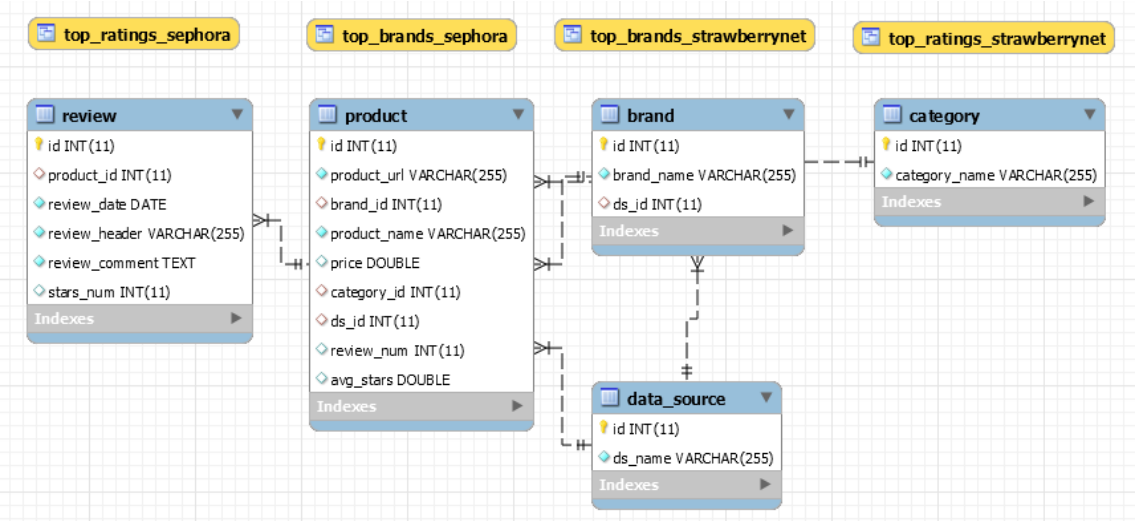
INSERT DOCUMENT VIEW LIST TABLE Displaying


```
_id: ObjectId("5cb89c3d6896a81d7065099b")
eyes_lips: Array
  0: Object
    product_url: "https://www.strawberrynet.com/en-us/skincare/elizabeth-arden/eight-hou..."
    brand_name: "Elizabeth Arden"
    product_name: "Eight Hour Lipcare Stick "
    product_price: "9.50"
    review_num: 169
    avg_stars: 4.437869822485207
    review_details: Array
      0: Array
        0: Object
          review_date: "11/4/2019"
          review_header: "lip moisture"
          review_comment: "excellent for the lips. give good feeling."
          stars_num: 5
        1: Object
          review_date: "7/4/2019"
          review_header: "excellent"
          review_comment: "Very smooth and keeps your lips hydrated"
          stars_num: 5
        2: Object
          review_date: "2/4/2019"
          review_header: "lip stick"
          review_comment: "very good make the lips smooth."
          stars_num: 5
```

MySQL Database

Database Structure

- **5 tables and 4 views**



Database Script

```
CREATE DATABASE skincare_db;

USE skincare_db;


CREATE TABLE IF NOT EXISTS category (
   id INT AUTO_INCREMENT,
   category_name VARCHAR(255) NOT NULL,
   PRIMARY KEY (id)
);


CREATE TABLE IF NOT EXISTS data_source (
   id INT AUTO_INCREMENT,
   ds_name VARCHAR(255) NOT NULL,
```

```
    PRIMARY KEY (id)
);

CREATE TABLE IF NOT EXISTS brand (
    id INT AUTO_INCREMENT,
    brand_name VARCHAR(255) NOT NULL,
    ds_id INT,
    PRIMARY KEY (id),
    FOREIGN KEY (ds_id) REFERENCES data_source(id)
);

CREATE TABLE IF NOT EXISTS product (
    id INT AUTO_INCREMENT,
    product_url VARCHAR(255) NOT NULL,
    brand_id INT,
    product_name VARCHAR(255) NOT NULL,
    price DOUBLE,
    category_id INT,
    ds_id INT,
    review_num INT,
    avg_stars DOUBLE,
    PRIMARY KEY (id),
    FOREIGN KEY (brand_id) REFERENCES brand(id),
    FOREIGN KEY (category_id) REFERENCES category(id),
    FOREIGN KEY (ds_id) REFERENCES data_source(id)
);
```

```
CREATE TABLE IF NOT EXISTS review (
   id INT AUTO_INCREMENT,
   product_id INT,
   review_date DATE NOT NULL,
   review_header VARCHAR(255) NOT NULL,
   review_comment TEXT NOT NULL,
   stars_num INT,
   PRIMARY KEY (id),
         FOREIGN KEY (product_id) REFERENCES product(id)
);

INSERT INTO data_source (ds_name) VALUES ('Strawberrynet');
INSERT INTO data_source (ds_name) VALUES ('Sephora');
INSERT INTO data_source (ds_name) VALUES ('Both');

CREATE VIEW top_ratings_strawberrynet AS SELECT c.category_name, b.brand_name,
p.product_name, p.price, p.avg_stars
                  FROM product p
    JOIN brand b
    ON (p.brand_id = b.id)
    JOIN category c
    ON (p.category_id = c.id)
    JOIN data_source ds
    ON (p.ds_id = ds.id)
    WHERE ds.id = 1 ORDER BY p.avg_stars DESC, c.category_name ASC LIMIT 100;
CREATE VIEW top_ratings_sephora AS SELECT c.category_name, b.brand_name, p.product_name,
p.price, p.avg_stars
                  FROM product p
```

```
    JOIN brand b
    ON (p.brand_id = b.id)
    JOIN category c
    ON (p.category_id = c.id)
    JOIN data_source ds
    ON (p.ds_id = ds.id)
    WHERE ds.id = 2 ORDER BY p.avg_stars DESC, c.category_name ASC LIMIT 100;
CREATE VIEW top_brands_strawberrynet AS SELECT b.brand_name, avg(p.avg_stars) as
avg_rating
                FROM product p
    JOIN brand b
    ON (p.brand_id = b.id)
    WHERE p.ds_id = 1 GROUP BY b.brand_name ORDER BY avg_rating DESC LIMIT 10;
CREATE VIEW top_brands_sephora AS SELECT b.brand_name, avg(p.avg_stars) as avg_rating
                FROM product p
    JOIN brand b
    ON (p.brand_id = b.id)
    WHERE p.ds_id = 2 GROUP BY b.brand_name ORDER BY avg_rating DESC LIMIT 10;
```