



SOCIAL LISTENING & ANALYTICS REPORT

Consumer Stress Research & Product Insights on Skincare Category

FOR 2020 UNILEVER HACKATHON

| April 2020
| Group: 207



CONTEXT

1. RESEARCH TEAM 2. ANALYTICAL REPORT

- 2.1 STUDY OVERVIEW
- 2.2 STRESSOR CATEGORY UNDERSTANDING
- 2.3 KEY MOMENTS DECODING
- 2.4 RESULTS RECAP & SUMMARY

3. BUSINESS PROPOSAL 4. APPENDIX

- APPENDIX A: Data Collection and Mining Techniques
- APPENDIX B: Chinese Verbatim of Quotes

1. RESEARCH TEAM



Research Team

Group number: 207

Pu Shi

Academic Background

University of Edinburgh-Master in Business Analytics
Shanghai University of Finance and Economics- Bachelor in Investment

Internship/Project Experience

Ipsos social intelligence & information technology intern; Roland Berger PTA
Assist in projects for BMW, GM, Chery, etc.

Skills

Statistics, machine learning and programming, Fluent English
Python, SQL, Microsoft Office

In This Research Project:

Web scraping, text mining and k-means clustering in Python, technical file editing, part of the key moments decoding, quantitative analysis, data visualization

Siyi Wang

Academic Background

University College London-Master in Social Research Methods
Shanghai University of Finance and Economics- Bachelor in Taxation

Internship/Project Experience

Ipsos social intelligence & analytics Intern; EY Parthenon PTA; Aon Hewitt McLagan Intern
Assist in projects for Cetaphil, Nestle, Coke, Huawei, HSBC, etc.

Skills

Quantitative & Qualitative Methods, Fluent English
Stata, Tableau, PPT, & Excel

In This Research Project:

Proposal and determination of research methods and keywords for data collection, data visualization, construction of deck framework, qualitative analysis, slides production & formatting

2. ANALYTICAL REPORT

2.1 STUDY OVERVIEW



Research Background

In recent years, **young consumers in China** may face many **urban aggressors**, such as air pollution, psychological stress and workload. These stressors are likely to influence their lifestyle and consumer behaviors. For example, the problem of hair loss is now more talked about by young people, and the category of shampoo and hair care products has therefore gained new attention.

Another potential category influenced by the increasing stress on urban young people is **skincare**, which is the category of interest of this report. Through **capturing and analyzing online data** of consumer online behaviors such Weibo posts, online searches, product choice, preference or comments on e-commerce channels, we can understand more about their demands of anti-stress skincare products to address these problems. Thus, novel personal skincare produce business plan can be prepared.

This project will explore how urban aggressors affects consumer and come up with anti-stress new skincare product business plan.

Research Objectives

BUSINESS OBJECTIVE

- PROVIDE INSIGHT FOR A POTENTIAL NEW SKINCARE PRODUCT
- FIND PROPER AND EFFICIENT COMMUNICATION CHANNEL
- BUILD A WINNING POSITION IN MARKET

RESEARCH OBJECTIVE

- Identify Major Source of Stress and Related Skin Problems of Interest in China (through social intelligence)
 - Identify the most discussed stressors in China
 - Reveal the relationship between each source of stress and the major skin problems triggered by this stressor
- Find Consumer Demands, Preference, Motivations and Barriers to Related Skincare Products (through Text mining & social intelligence)
 - Understanding the key triggers & barriers to enter each sub-category of skincare
 - Understanding the areas of opportunity

Methodology



Weibo Data Collection

General seed words



Initial scrape

Extend key words set

Large scale scrape



Preprocessing

- Remove noises
- Remove duplicates
- Unify date formats
- Transform strings and numerical formats



Text Mining

- **Word Cut**
 - Cut the post contents
 - Word frequency
- **Word Clustering**
 - Cluster high frequency words with k-means
 - Match the clusters to stress sources
- **Visualization**
 - Word cloud
 - Social landscape



Qualitative Analysis

- Top Stressor Deep Dive
- Moment & Typical Scenario Decoding
- Consumer Triggers and Barriers

2. ANALYTICAL REPORT

2.2 STRESSOR CATEGORY UNDERSTANDING



Step 1: Identifying Stressors : What are people talking about stress and skincare on social media?

FIRST/PILOT WAVE OF DATA COLLECTION

DATA COLLECTION SCOPE:

Period tracked: Mar 10th, 2020-

April 10th, 2020

Sources tracked: Weibo

Key words tracked: Stress (e.g.

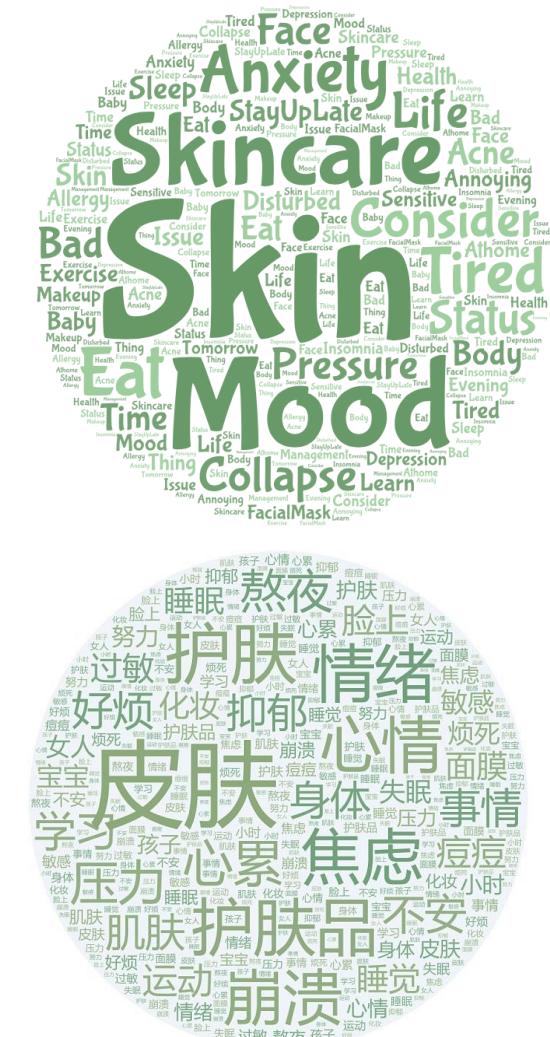
压力, 焦虑) & Skin (皮肤, 护肤)

TOTAL COLLECTED DATA VOLUME:

Total collected data volume: About 11145 data before data processing

Total UGC* data volume:

9450 posts



Through text analysis of the collected data, the word frequency is obtained after cutting the words. These words are divided into five categories, one of which is the pressure source. Through these pressure source related words, potential stressors can be identified.

Top 15 Words Of Each Category

NO.	Skin Problems	Mood	Stressor	Solution	Products
1	Skin/皮肤	Mood/心情	Stay up late/熬夜	Skincare/护肤	Facial Mask/面膜
2	Face/脸	Mood/情绪	Time/时间	Sleep/睡	Essence/精华
3	Body/身体	Anxiety/焦虑	Study/学习	Sport/运动	Product/产品
4	Skin/肌肤	Breakdown/崩溃	Baby/宝宝	Video/视频	Sunscreen/防晒
5	Allergy/过敏	Exhausted/心累	Thing/事	Weibo/微博	Cream/面霜
6	Acne/痘	Pressure/压力	Insomnia/失眠	Sleep/睡眠	Repair/修复
7	Sensitive/敏感	Sense/感觉	Lady/女人	Wash face/洗脸	Essential Oil/精油
8	Skin Disease/皮肤病	Status/状态	Work hard/努力	Exercise/健身	Eye cream/眼霜
9	Dry/干燥	Upset/不安	Child/孩子	Maintenance/保养	Lotion/乳
10	Mask/口罩	Depressed/抑郁	Age/岁	Keep early hours/早睡早起	Hormone/激素
11	Worse/变差	Annoying/好烦	Go to work/上班	Routine/规律	Cleanser/洗面奶
12	Acne Breakouts/长痘	Nill /不想	Money/钱	Read/看书	Lipstick/口红
13	Wrinkle/皱纹	Annoying/烦死	Season change/换季	Cure/治疗	Toner/水乳
14	Bump/闭口	Annoying/烦	Mom/妈妈	Clean/清洁	Lotion/霜
15	Dark Circles/黑眼圈	Depression/抑郁症	Epidemic/疫情	Rest/休息	Vitamin/维生素

*UGC: user generated contents ; non-UGC are contents generated by brands, KOL, promotions, etc.

Step 1: Identifying Stressors

Based the results of the word-cut and word cloud below, the stressors can be classified to four categories: work, study, environment and family. Then a second wave of data collection on these categories will be conducted.



Step 2: Top Stressors and Related Skin Problems

SECOND WAVE OF DATA
COLLECTION

DATA COLLECTION SCOPE:

Period tracked: Mar 10th, 2020-
April 10th, 2020

Sources tracked: Weibo

Key words tracked: 4 Categories of
Stress (工作, 学习, 环境, 家庭) &
Skin Problems (eg. 黑头, 痘痘)

TOTAL COLLECTED DATA VOLUME:

Total collected data volume:

About 27,506 data before data
processing

Total UGC* data volume:

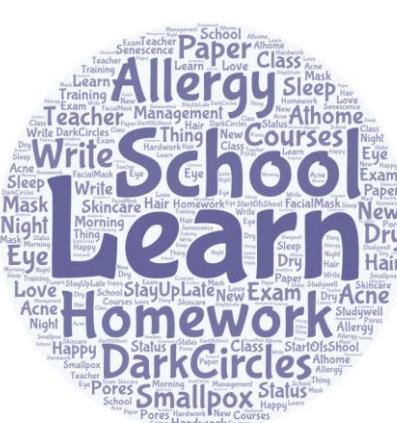
22,625 posts



Work



Study



Family



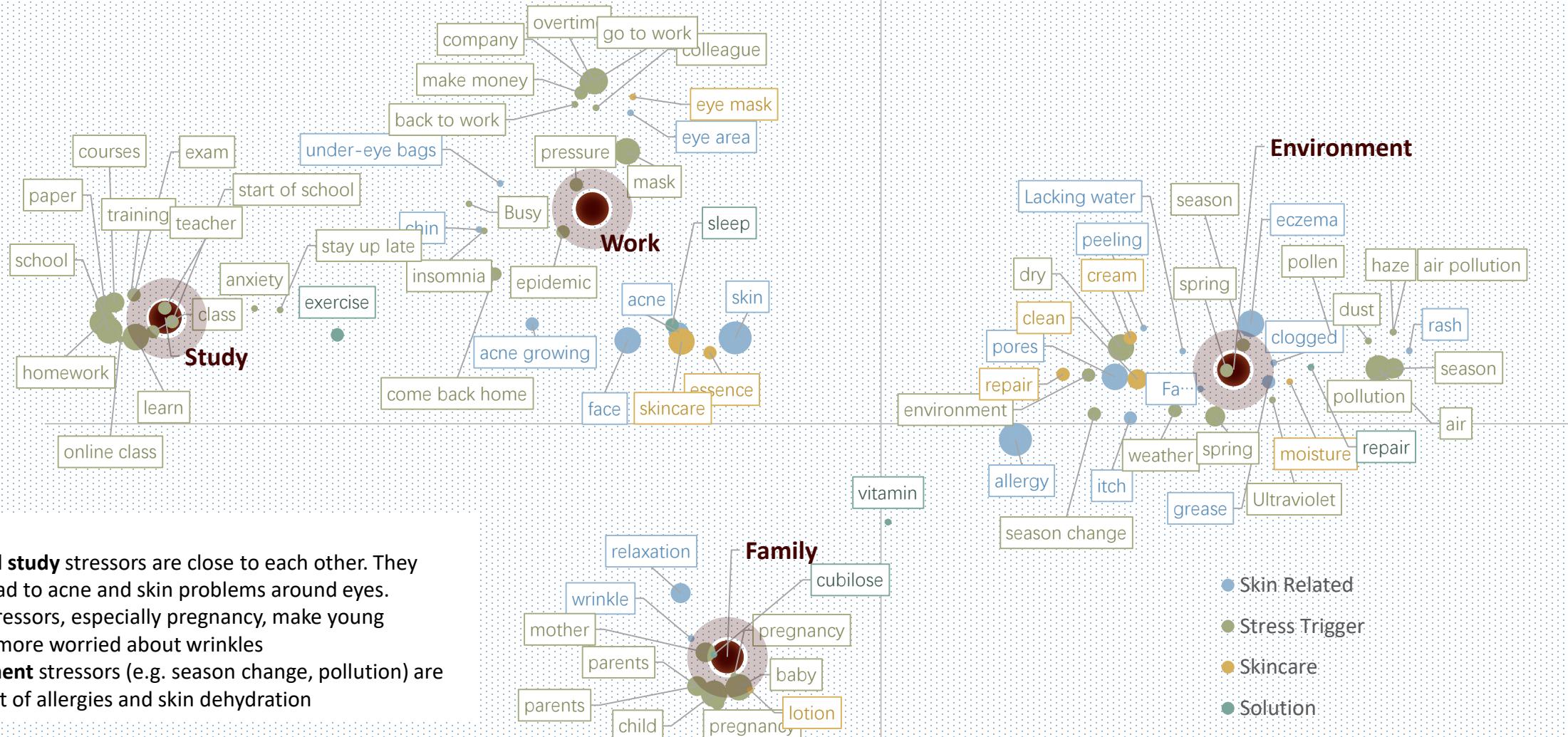
Environment



*UGC: user generated contents ; non-UGC are contents generated by brands, KOL, promotions, etc.

Step 2: Top Stressors and Related Skin Problems

Cluster Analysis: Social Landscape Map



Comments:

- **Work and study** stressors are close to each other. They mainly lead to acne and skin problems around eyes.
- **Family** stressors, especially pregnancy, make young mothers more worried about wrinkles
- **Environment** stressors (e.g. season change, pollution) are the culprit of allergies and skin dehydration

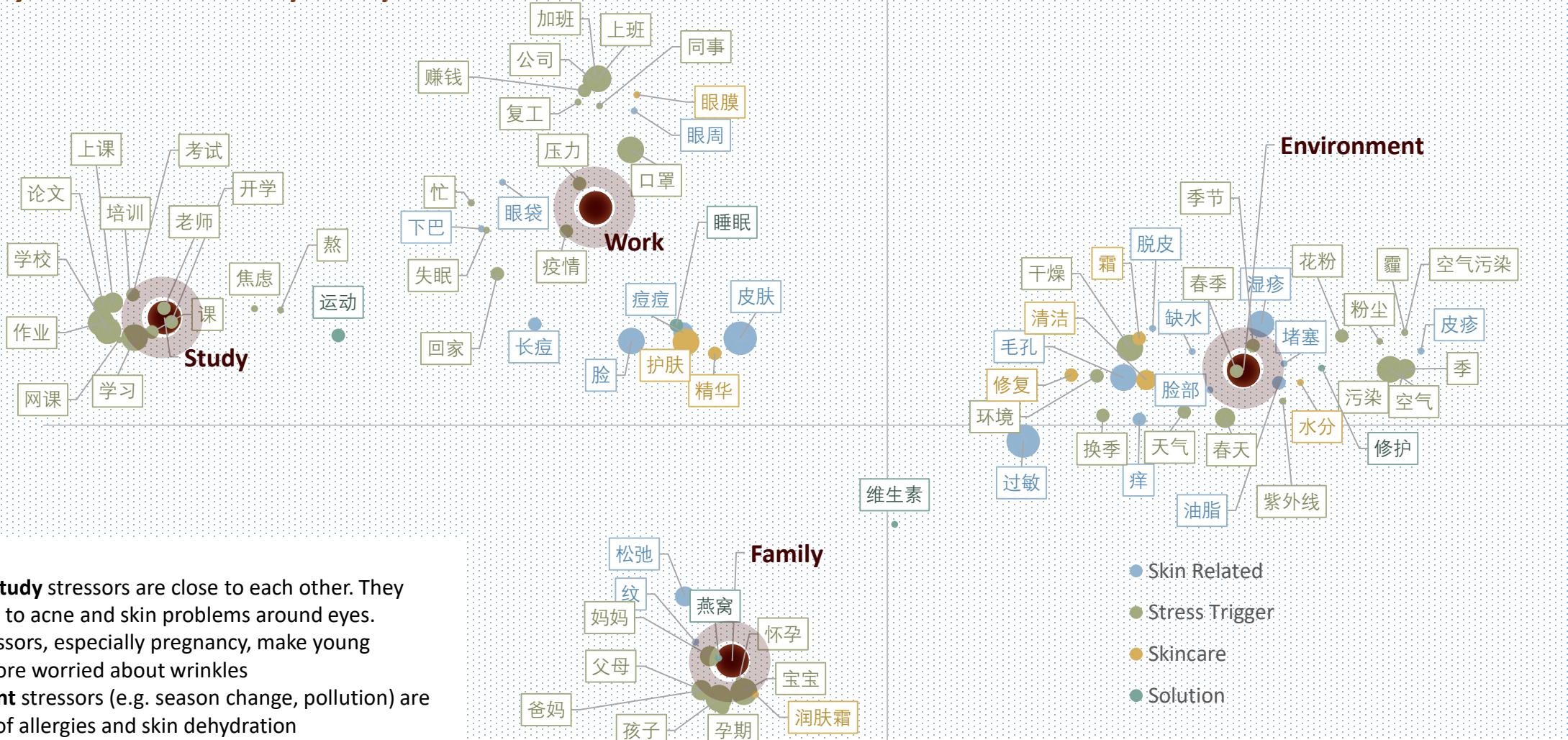
Note: The analysis was based on k-means clustering technique, and the details of the analysis can be found in APPENDIX A: TEXT MINING

The words are translated from Chinese so there may be some duplicates, but they are different in Chinese; see the Chinese version in the next page

Data Source: Weibo Search

Step 2: Top Stressors and Related Skin Problems

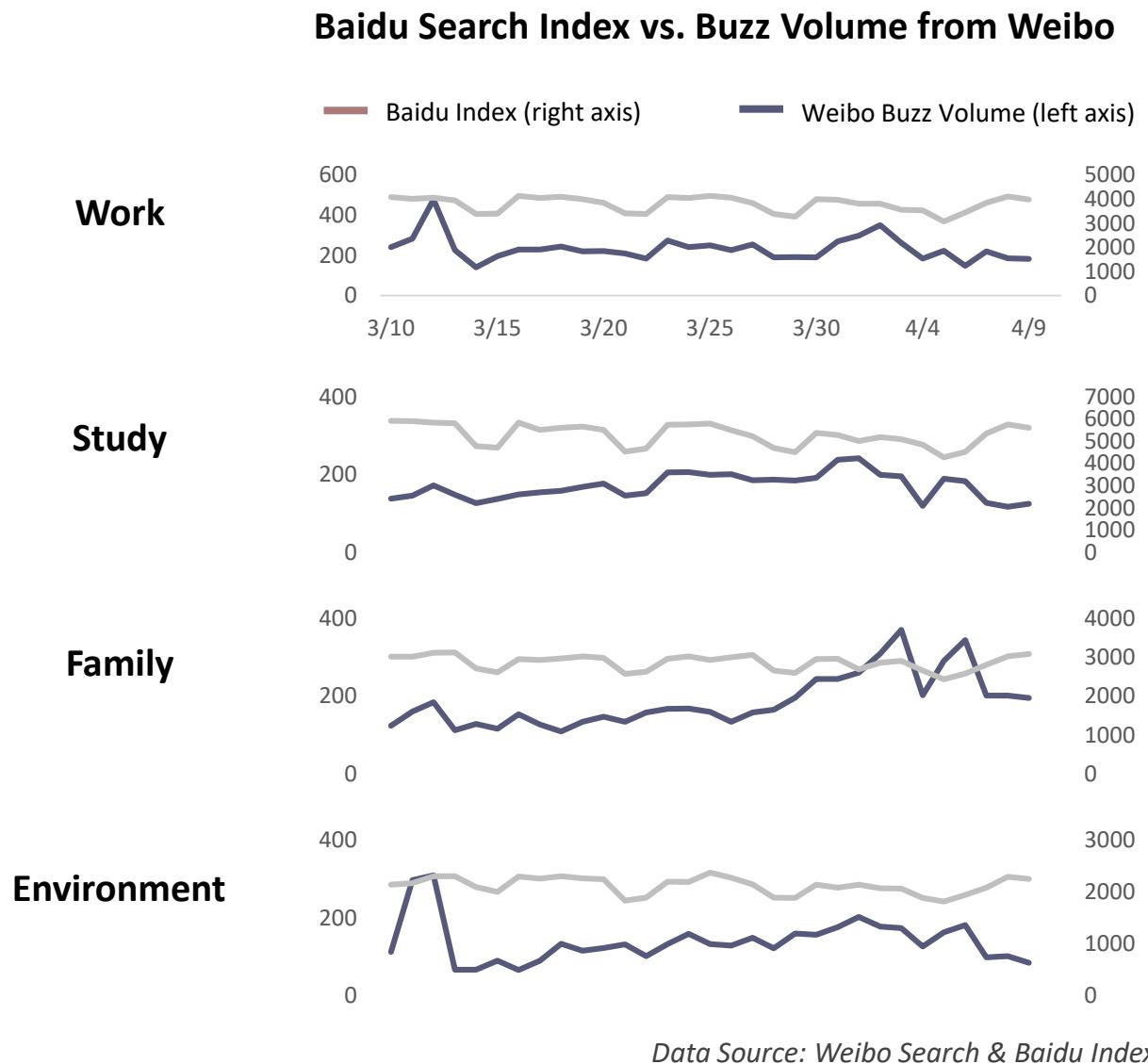
Cluster Analysis: Social Landscape Map



Comments:

- Work and study stressors are close to each other. They mainly lead to acne and skin problems around eyes.
- Family stressors, especially pregnancy, make young mothers more worried about wrinkles.
- Environment stressors (e.g. season change, pollution) are the culprit of allergies and skin dehydration.

Step 2: Top Stressors and Related Skin Problems



- Key words based scraping may return biased results because of the selection of words, so the data obtained from Weibo is compared to Baidu Search Index to verify the robustness of our analysis...
- It can be noticed that the scraped buzz volumes from **Work, study and environment** categories follows the similar trend with Baidu Search Index. However, result from the **family** category has a lower probability to be convincing...
- Work and study are similar based on the clustering analysis (Page 12), but work has higher buzz volume. So work category, will be analyzed in detail in next section. Environment category will be elaborated as well because it appears far away from other three categories in the social landscape.
- Our choice of deep dive category is also based on the potential business opportunities.

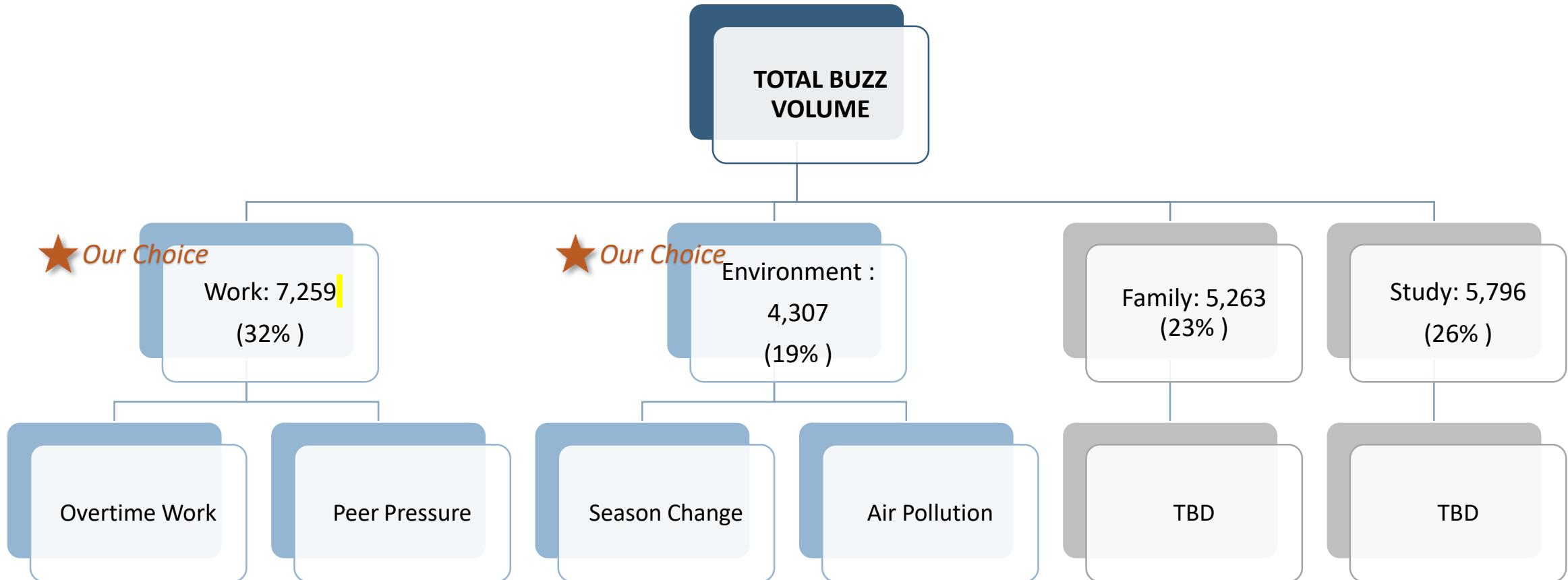
2. ANALYTICAL REPORT

2.3 KEY MOMENTS DECODING



Overall Topic Funnel

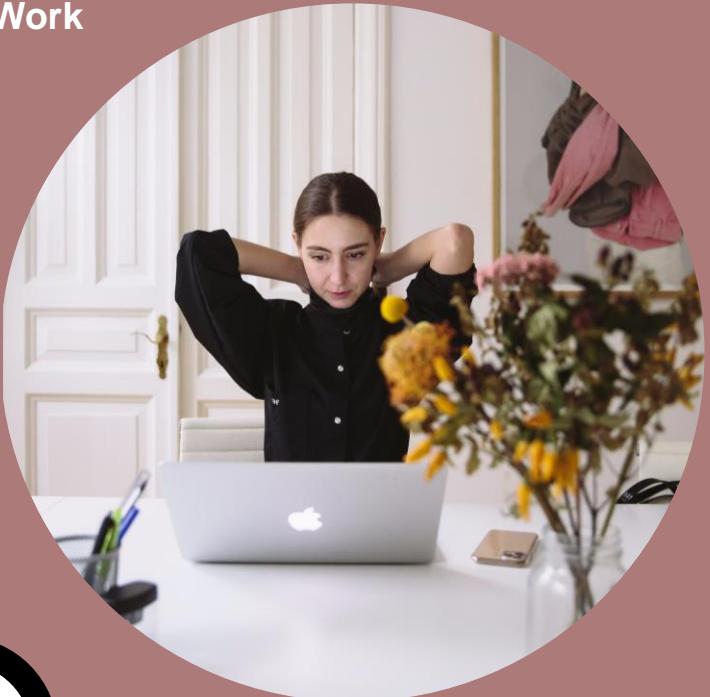
Some representative moments from work and environment categories will be analyzed in this section. Analysis for family and study categories will not appear in this report but can be conducted in the future.



Workload Stress Group Snapshot



Overtime Work



Workplace Relationships



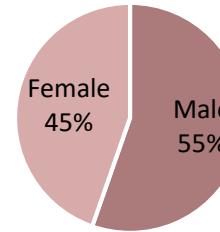
Promotion



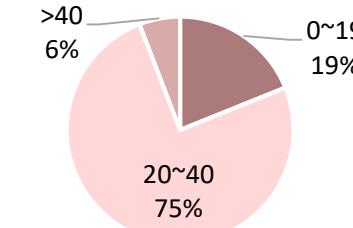
Peer Pressure

Demographics For “Work Stress” Group

Gender Distribution



Age Distribution

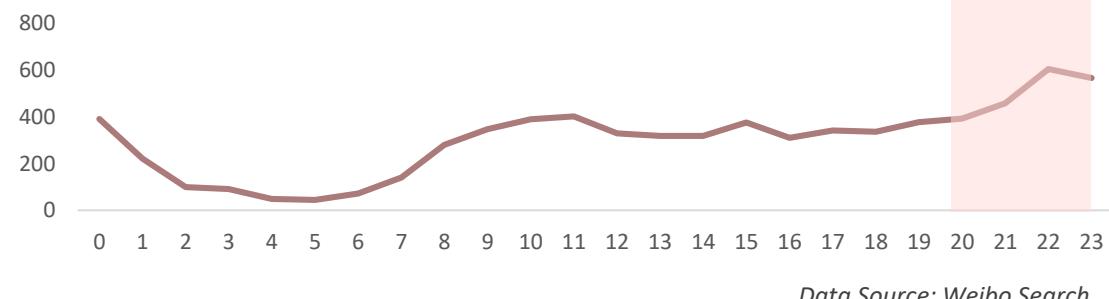


Searching City Rank

1	Beijing
2	Shanghai
3	Chengdu
4	Guangzhou
5	Shenzhen

Data Source: Baidu Index

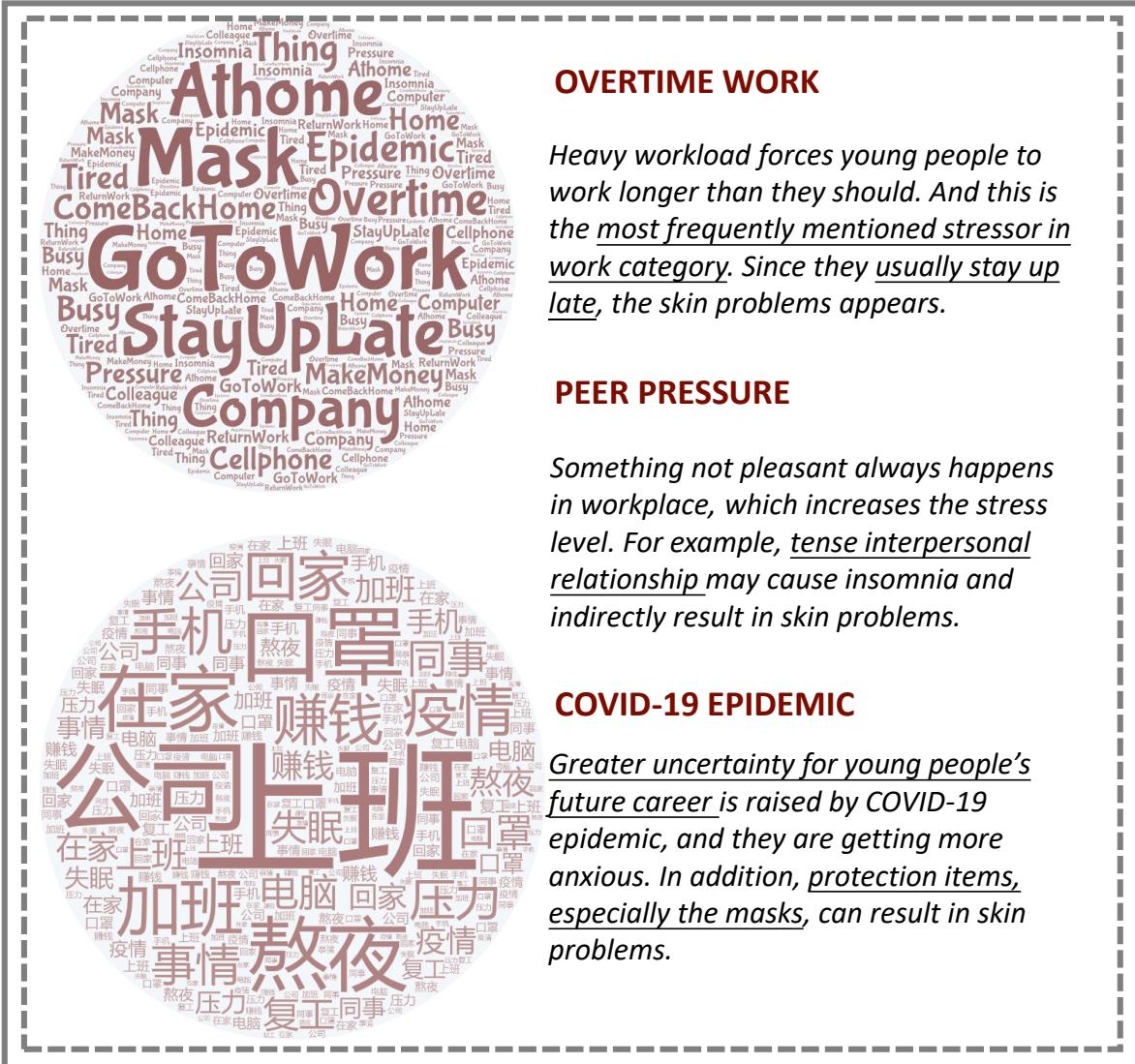
Number of Weibo Posts over Time



*Note: tags on the left are based on the qualitative analyze of the moment

- *Gender*
Males seems more likely to complain about the workload than females which is an interesting thing to research
- *Target Age & City*
The people being analyzed are mainly from 20 to 40 years in Tier 1 or Tier 2 cities
- *Posting Time*
contents are posted from 8pm until midnight, which means those stressed from work are very likely return home/ have leisure time late in the night.

Stressor Deep Dive – Work: specific stress source and skin problems



Stressor Deep Dive – Work: Skincare products and other solutions for work related skin problem

SKINCARE PRODUCTS

ESSENCE

- Why this product?

Wearing masks all day after returning to work brings acne, and good essence can provide overall protection for skin.

@Zenmeburangrenfaxianwo

My acnes are out of control after wearing masks working all day . Could anyone provide some advices about an essence for sensitive skins?

EYE MASKS

- Why this product?

People always keep working at home until midnight during lockdown . Such an unhealthy lifestyle leads to dark circles, and eye masks is a suitable product to deal with this.

@Zhangxiaoichi

My husband was shocked when he saw his black circles in the mirror after working for a long time. Then I helped him doing some "maintenance" by eye masks.

OTHER SOLUTIONS

GO TO BED EARLIER

- Why this solution?

Many skin problems are caused by overwhelming work, and going to bed earlier is a good way for one to recover from toil.

@Gaovaxuan

Working is annoying... But tiredness forced me to go to bed at around 9pm, thus my skin status is better than before...

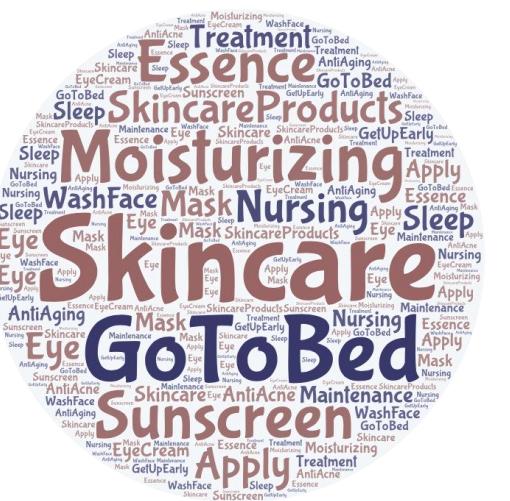
DOING SPORTS

- Why this solution?

Doing sports is a good way to keep health. And one's health status is highly correlated with his skin status.

@MISSwuditu

COVID-19 brought a boring 2-month vocation... The only thing I keep doing in this period is sports. Thanks to exercise, my skin is pretty fair now!



Skincare Products
Other Solutions

Typical Scenario Deep Dive: Moment 1

Typical Scenario #1: Overtime Work

There is no clear boundary between “onsite” and “off work” in a Chinese workplace due to the fierce market competition. And employees seldom reject the overtime working demand because they have loans to be paid and families to be raised. And extra workload leads to unhealthy lifestyle which performs clearly on bad skin status.

- **Stress Influence:** Lack of rest, Screen radiation, insomnia.
- **Skin Problem:** Sagging skin, Acnes, dark circles, eye bags, puffy eyes
- **Mentioned Solutions:** Autologous fat filling, traditional Chinese medicine, nutritious food, eye masks, essence
- **Interested Skincare Products:** Essence is a good choice to protect skin when staying up late, facial and eye masks are also mentioned. In addition, customers care about side effects such as allergies especially in spring.
- **Occasion of awareness:** bad skin status in the mirror before going to work; Feeling itch or pain on the face after working whole day with masks
- **Potential Barriers to Purchase Skincare Products:**
 - Not sure if their own situation applies or worried about allergies
 - Do not know how to choose skincare products on the market
 - Lack of time focusing on skincare



“

@Nlugulushiruhua

Overtime working makes me poor... Recently I noticed my sagging skin so I spent 20,000 yuan doing autologous fat filling, after which I even felt younger! Well...Beauty is a good thing.

@Zenmeburangrenfaxianwo

My acnes are out of control after wearing masks working all day . Could anyone provide some advices about an essence for sensitive skins?

”

Typical Scenario Deep Dive: Moment 2

Typical Scenario #2: Peer Pressure

Workplace is not always peaceful and no one can get along well with colleagues or boss all the time. Some unpleasant social experience can increase the one's stress level which can indirectly cause skin problems. Moreover, some peer activities even lead to skin problems directly as mentioned in Weibo posts.

- **Stress Influence:** Low mood, stress, interpersonal sensitivity, insomnia
- **Skin Problem:** allergies, acnes
- **Mentioned Solutions:** Relaxing after work, chatting with friends, complaining on social media
- **Interested Skincare Products:** Few people annoyed by interpersonal relationship choose to use skincare products, they tend to deal with the emotional problem but ignore the skin problem to some degree.
- **Occasion of awareness:** Rash after alcohol intake, told by colleagues that the skin status is unsatisfying, acnes realized from mirror
- **Potential Barriers to Purchase Skincare Products:**
 - When one has bad mood, he ignores or may not realize the skin problem, thus never focuses on skincare products
 - The interpersonal relationship can be improved by applying other methods
 - Do not know how to choose skincare products on the market



“

@langligelangerligelang

On a company party, I was asked whether I could drink liquor, and I told them I was allergic to alcohol. After drinking I always got rash on my face and neck. At that moment, I felt the social distance between me and my colleagues...

@Buxiangjiabandeyouzinigeningmeng

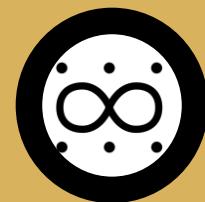
Thanks to my stupid colleague, I have to work more than 14 hours recently...It's really annoying that acnes grow on my face crazily!

”

Environment Stress Group Snapshot



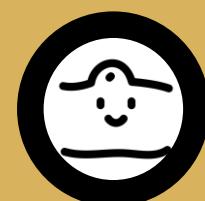
Season Change



Air Pollution



Allergy

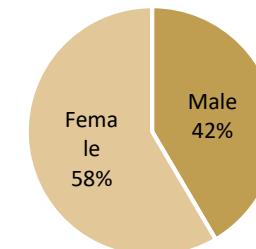


Eczema

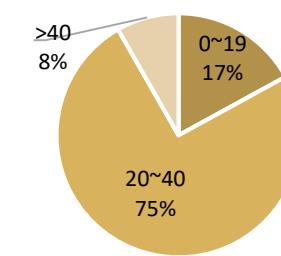
*Note: tags on the left are based on the qualitative analyze of the moment

Demographics For “Environment Stress ” Group

Gender Distribution



Age Distribution

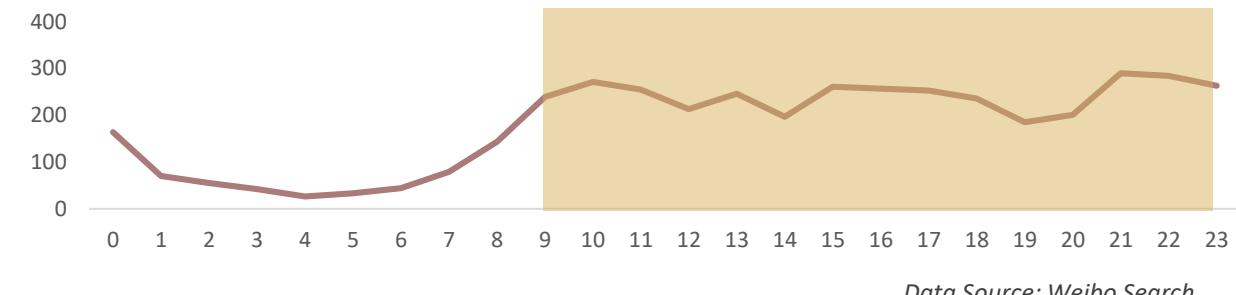


Searching City Rank

1	Beijing
2	Shanghai
3	Chengdu
4	Guangzhou
5	Shenzhen

Data Source: Baidu Index

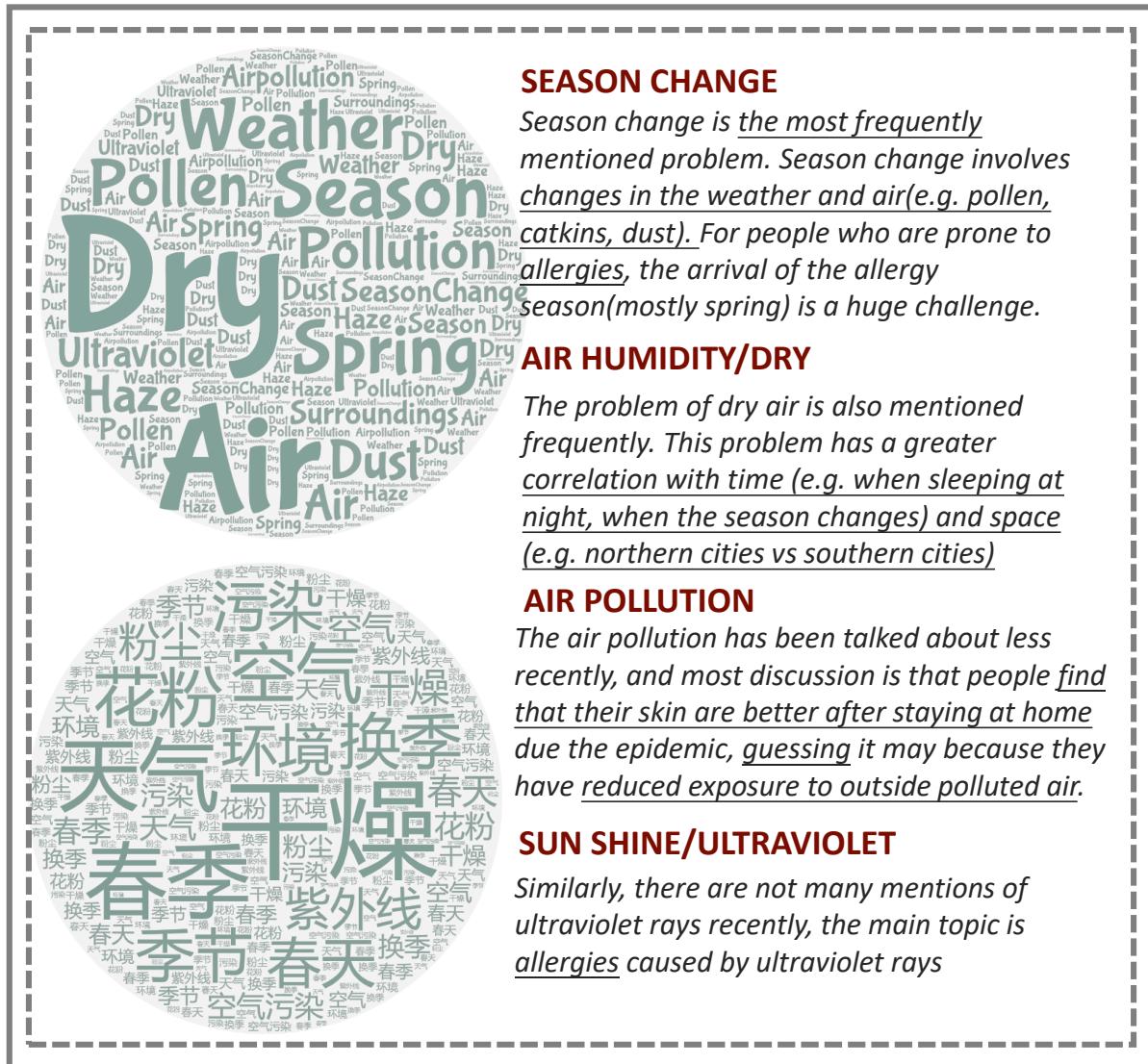
Number of Weibo Posts over Time



Data Source: Weibo Search

- *Gender*
More female than male care about the environment issue which may because man have less awareness of the environment change
- *Target Age & City*
The people being analyzed are mainly from 20 to 40 years in Tier 1 or Tier 2 cities
- *Posting Time*
Unlike the work stress group, environment related contents posts are more stable during the day, indicating the awareness of this problem may not have time-related triggers.

Stressor Deep Dive – Environment: specific stress source and skin problems



Stressor Deep Dive – Environment: common treatment/products for work related skin problem

COMMON NEEDS OF SKINCARE

HYDRATION/MOISTURIZING

Hydration is a commonly mentioned requirement. Whether it is because of allergies or because of the imbalance of water and oil, moisturizing is everyone's concern.

@Kasi-quihong *

I dare not provoke my skin during season change. Treat in carefully and use more mask! Otherwise, it is super dry. Without hydrating the skin, it is easy to be allergic and red, etc.

FACIAL CLEANSING

Skin cleansing is a basic requirement, especially for those care about the air pollution. Moreover, for allergic groups, it is urgent to clean and protect fragile skin in the meantime.

@Renxiaguida

The air is so dirty, the dust is so much, I can't stand it if my skin is not cleaned.

@AimeeMeow

I have to wash my face with Evian during the allergy season, which I won't drink usually.

SKIN REPAIR

For people with eczema, the need for skin repair is particularly prominent. Safe and repairable skin care products are attractive to them

@Xunyouyunjian

I don't know how it happened, I just got eczema on my face. The doctor said that no makeup for at least 3 months. It is best to apply only Yuze 's repair cream for skin care products.

MENTIONED PRODUCT SUB-CATEGORYS

FACIAL MASK/ MASQUE

Due to the moisturizing function, the mask is mentioned by many people. It is a skin care product that is not used daily, but is indispensable for a long time, especially for people with dry skin and allergies

@Xiaoyushi

For a while, I didn't buy anything except daily necessities. Later, when spring came, the air became extremely dry, so I bought body lotion. And recently I also bought a leave-in mask for moisturizing too.

@Juzixiaojiejie

The weather is getting better and better but the air is still a bit dry. Weekly mask is essential

OTHER PRODUCTS

The other mentioned categories are more scattered, and there is no concentrated mention of a certain type of product, and it is basically advertorial post and marketing ID post. The mentioned post are :

- facial cleanser/face wash(foaming, milky, cream, gel)
- sun screen/sun block
- Day/ night creams
- foundation primer



Typical Scenario Deep Dive: Moment 1

Typical Scenario #1: Season Change

Season change is a huge source of stress, especially the transition from winter to spring, which means the arrival of many people's allergy season. A large group of people are troubled by skin problems during the season change.

- **Stress Influence:** Season change will bring changes in temperature and air humidity. The composition of the air will also change, for example, when spring comes, allergens will increase, such as pollen, catkins, etc.
- **Skin Problem:** Allergies, eczema, fragile skin, redness, dryness, pain and acne
- **Mentioned Solutions:** Reduce the variety of skin care, use safe and low-burden skin care products, insist on using moisturizing products such as masks, and go out with a mask; seek doctor's or other professional help.
- **Interested Skincare Products:** Products that can maintain skin stability during the season change, reduce the occurrence of skin redness and itching problems; in the case where the season change skin problems have already occurred, they need to have a strong oil moisturizing effect without increasing the burden on the skin; products with skin repair function; safe product with less side effect
- **Occasion of awareness:** After washing, find that the face was dry; acne and itching continued to be unsolved; the same skin problems were found every year at a similar time
- **Potential Barriers to Purchase Skincare Products:**
 - Fear of unsafe ingredients, exacerbating existing skin problems
 - Not sure about skincare effects
 - Hard to choose products that claim to be "allergic skin" / "sensitive skin" friendly on the market



“

@Aiyounigedashacha

With my rhinitis and face acne guarantee, the allergy season is really here

@-anpingshisanmei

Hate spring, spring is allergy season

@Nichihongshaogourouma

It's my skin sensitive season again, I can't apply toner and lotion, it will cause a hot pain, and acne will rot half the face

@5i33ydg

Spring eczema, itching! Scratching ... Itching ...

@Woshiaxina

It's been a long time since I spent the spring in Laiyuan, I can't stand my face in another season, dry and painful

”

Typical Scenario Deep Dive: Moment 2

Typical Scenario #2: air pollution and dryness

Due to air pollution and dryness, there is a lot of dust in the air, leading to more dirt in the pores, which causes a series of skin problems. This problem is more strongly realized by everyone, when people find that their skin has improved after isolation at home.

- **Stress Influence:** Dirty air leads to a lot of burden on people's skin when they go out
- **Skin Problem:** Acne, dryness, comedones, rough skin
- **Mentioned Solutions:** Facial cleansing and moisturizing
- **Interested Skincare Products:** Products that can help resist the damage of turbid air to the skin; products that can deeply clean the skin against the dust in the air
- **Occasion of awareness:** Isolated at home for a while due to the covid-19 epidemic and found that the skin condition has greatly improved. Under the premise of the same skin care steps as before, and even lower frequency of skin care, the skin condition is obviously refreshing. Therefore, people think that it is because of the outdoor air turbidity that caused their original skin problems
- **Potential Barriers to Purchase Skincare Products:**
 - Worried that the deep cleansing products are too irritating to the skin
 - Because masks are worn when going out, the damage caused by air pollution to the skin is a little relieved, and people lack the motivation to buy related products.



“

@Lingxiaocangtongxue

I found that my comedones were really because of the dirty air outside, not because of me. In the past, if I ignored it for a week, the blackheads could be full of nose. I just remembered that I hadn't been concerned about blackheads for more than a month, but when I looked in the mirror, my face and nose were completely clean.

@Zailushang_Joanna

After staying at home for a long time, my skin has improved, probably without the stimulation of external air pollution. Even if I don't use a lot of masks, and only use a little skin care products after washing my face every day. My skin still looks very moist!

@Llubai0197

If you don't wash your hair nor tie it up, you will end up with acne. The air humidity here is not high. My skin has transformed from absolute oily acne type to mixed oil and there is more oil secretion from the nose (and I am afraid of oil secretion dissonance so that I dare not wash my face). Ha ha, my face is the Loess Plateau, pitted and dusty.

”

2. ANALYTICAL REPORT

2.4 Results Summary



Results Recap & Summary

Analysis in former sections is summarized in this part, and the *research questions* mentioned in Page 7 are answered.

- Identify Major Source of Stress and Related Skin Problems of Interest in China (through social intelligence)

- Major source of stress: *work, environment, study, family*
- Related Skin Problems:

Stressor	Related Skin Problems
Study & Work	acnes, dark circles, eye bags, puffy eyes
Environment	allergy, eczema, rash, skin cleansing problem, dry skin
Family	skin relaxation, wrinkle

Results for this research question are obtained from observing Weibo posts, and justified by the clustering analysis and Baidu Index.

- Find Consumer Demands, Preference, Motivations and Barriers to Related Skincare Products (through Text mining & social intelligence)

- Qualitative analysis on two robust categories – *work & environment* – is conducted
- Demands & Purchasing Barriers

Work

- Demand: Products providing overall skin protection efficiently
- Purchasing Barriers: Allergies, lack of time and product information
- Purchasing Motivation: Friends'/online recommendation

Environment

- Demand: Multiple demands such as hydration, cleansing and repair
- Purchasing Barriers: Allergies, worry about effectiveness
- Purchasing Motivation: Friends'/experts'/online recommendation

Results for this research question are obtained from word cloud based on Weibo posts and qualitative analysis

BUSINESS PROPOSAL:

New Product Compass



New Product Compass

Based on the results in last section, the design strategy of new products is explained

Work Stressor			
Demand	Potential Product	Purchasing Barriers	Marketing Strategy
<ul style="list-style-type: none"><i>Overall protect the skin against staying up late, computer radiation and psychological stress in limited time</i><i>No side effect</i>	<ul style="list-style-type: none"><i>Essence & Eye masks with safe Ingredients</i>	<ul style="list-style-type: none"><i>Lack of time selecting products</i><i>Allergies</i>	<ul style="list-style-type: none"><i>Identify target group</i><i>Targeted advertising with short contents on generally used social media (e.g. Weibo)</i><i>Put more ads in late night, especially on transportation tools, such as the subway and night shift taxi</i><i>Expert endorsement to guarantee the anti-allergy effect</i>
Environment Stressor			
Demand	Potential Product	Purchasing Barriers	Marketing Strategy
<ul style="list-style-type: none"><i>Safe first</i><i>Hydration & moisturize</i><i>Skin repair</i><i>Facial cleansing</i>	<ul style="list-style-type: none"><i>Anti-allergy facial masks</i><i>Other anti-allergy products</i>	<ul style="list-style-type: none"><i>Not sure if the product is suitable for the skin</i><i>Worry about the effectiveness of the product</i><i>Allergies</i>	<ul style="list-style-type: none"><i>Provide samples for target group and let consumers prove the effectiveness</i><i>Expert endorsement to guarantee the anti-allergy effect</i><i>Dermatologist recommendation.</i>

APPENDIX



APPENDIX A: Data Collection and Mining Techniques

Weibo Data Collection (the codes can be found in attachment):

使用种子关键词进行微博搜索，根据结果扩充关键词库，总结压力源和皮肤问题种类，再根据（压力源+皮肤问题）的关键词组合进行正式爬虫。爬虫使用的工具为Python的Selenium工具包。

Step1: 种子关键词扩充词库

首先打开微博搜索主页（s.weibo.com），根据微博的html源码结构进行模拟登陆操作，根据表1所示的种子关键词进行微博搜索，获得微博数据如表2所示，包括用户名、发文日期、内容、转发数、点赞数、评论数等数据。

类别	关键词
压力	崩溃，不安，烦，焦虑，情绪，心累，压力，抑郁，忧虑
皮肤	皮肤，护肤

表1：种子关键词表

Index	压力词	皮肤词	用户名	日期	内容	转发数	评论数	点赞数	数据源
1	压力	皮肤	淮子历险记	2020-04-07 01:28:00	青春青春活力乐趣乐趣在于	0	0	0	https://weibo.com/6001419332/IChMJ68YM?refer_flag=1001030103
2	压力	皮肤	黄冈晓琼	2020-04-07 22:23:00	情绪管理舒压。人体所承受	0	12	6	https://weibo.com/2281870457/ICqzK825o?refer_flag=1001030103
4	压力	皮肤	御尚坊颜连锁美业	2020-04-07 13:28:00	柔滑的皮肤，每个人都想要	0	0	0	https://weibo.com/2704094385/ICd4otlf3?refer_flag=1001030103
5	压力	皮肤	爱吃番茄啦	2020-04-07 12:17:00	别人在疫情期间：烘焙做饼	0	0	3	https://weibo.com/3967409891/ICcBO5vEX?refer_flag=1001030103
6	压力	皮肤	蔓越阑珊没有外彙	2020-04-07 09:47:00	长发它不香吗？为什么要剪	0	9	13	https://weibo.com/5669801520/IC8tPeFpZ?refer_flag=1001030103

表2：爬虫数据结构

将表2内容列去重、切词、去除停用词后，得到所有词的词频，将其整理后选取高词频词汇，可将压力源分为学习、工作、环境、家庭四类，皮肤问题约20种，详见表3（下页）。切词工作均利用Python的Jieba工具包完成，停用词库来源于公开渠道整理。

APPENDIX A: Data Collection and Mining Techniques

Weibo Data Collection (the codes can be found in attachment):

使用种子关键词进行微博搜索，根据结果扩充关键词库，总结压力源和皮肤问题种类，再根据（压力源+皮肤问题）的关键词组合进行正式爬虫。爬虫使用的工具为Python的Selenium工具包。

Step2：正式爬虫

根据Step 1中得到的关键词组合，进行正式的爬虫操作，时间段为2020年3月10日00: 00到2020年4月10日00:00，内容为原创，每次搜索的链接格式如图1: 得到的数据格式与表2相同。

类别	关键词
压力源	工作 学习 环境 家庭
	上班,工作,加班,赚钱,公司 学习,学校,考试,培训,上课,论文,作业 空气,雾霾,季,污染 宝宝,孩子,孕期,父母,爸妈,催婚
皮肤问题	皮肤,痘,过敏,干燥,皱纹,闭口,黑眼圈,衰老,毛孔,湿疹,松弛,粉刺,皮炎,黑头,细纹

表3：正式爬虫关键词表

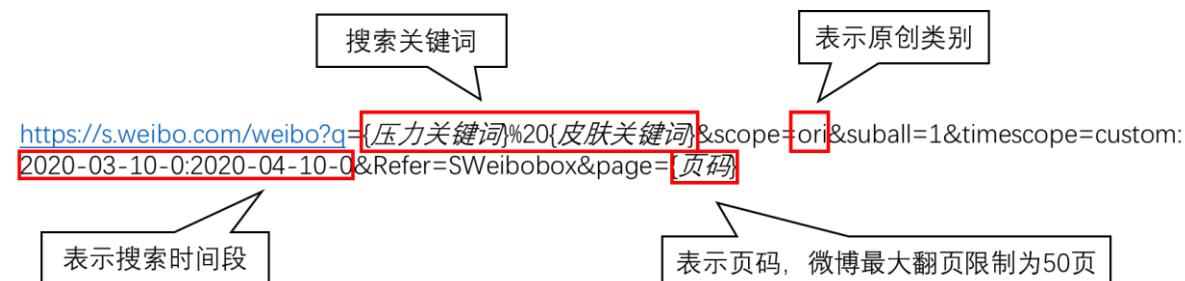


图1：爬虫链接格式

APPENDIX A: Data Collection and Mining Techniques

Weibo Data Collection (the codes can be found in attachment):

使用种子关键词进行微博搜索，根据结果扩充关键词库，总结压力源和皮肤问题种类，再根据（压力源+皮肤问题）的关键词组合进行正式爬虫。爬虫使用的工具为Python的Selenium工具包。

Step2：正式爬虫

根据Step 1中得到的关键词组合，进行正式的爬虫操作，时间段为2020年3月10日00: 00到2020年4月10日00:00，内容为原创，每次搜索的链接格式如图1: 得到的数据格式与表2相同。

类别	关键词
压力源	工作 学习 环境 家庭
	上班,工作,加班,赚钱,公司 学习,学校,考试,培训,上课,论文,作业 空气,雾霾,季,污染 宝宝,孩子,孕期,父母,爸妈,催婚
皮肤问题	皮肤,痘,过敏,干燥,皱纹,闭口,黑眼圈,衰老,毛孔,湿疹,松弛,粉刺,皮炎,黑头,细纹

表3：正式爬虫关键词表

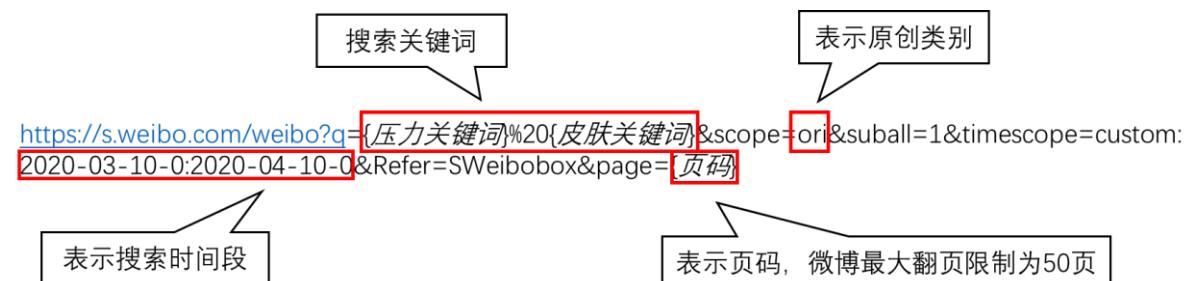


图1：爬虫链接格式

APPENDIX A: Data Collection and Mining Techniques

Data Preprocessing (the codes can be found in attachment):

针对数据各列进行的预处理操作见表4:

列名	预处理操作
用户名	删除一些非UGC用户的发文，如用户名包含“新闻”、“医院”等字眼
日期	同一日期格式为“YYYY-MM-DD h:m:s”
内容	去重，并根据特殊字符、无关文本去噪
转发、点赞、评论数	将文本储存的数字转换为数字格式

表4：数据预处理

Text Mining:

Step1：切词并制作词云

使用Python的jieba工具对预处理后的数据切词后得到的结果样例如表5，与此同时去掉了与压力源、皮肤问题及解决方法、护肤品等内容无关的词汇，并删除了词频较小的数据。基于该数据制作了报告中所展示的词云

关键词	工作	环境	家庭	学习	总词频
上班	2774	40	52	72	6116
孩子	118	109	2833	118	6421
宝宝	116	277	2820	30	5455
学习	164	15	61	1972	4294
空气	78	1920	65	19	2082

表5：切词结果样例

APPENDIX A: Data Collection and Mining Techniques

Text Mining (the codes can be found in attachment):

Step2: 关键词聚类

1) 关键词标注

为了找到压力源、皮肤问题、护肤品三者的对应关系，对词频高的关键词进行了标注。

2) 计算TF-IDF指数

根据每个关键词在四种压力源中的词频直接聚类，会由于词数的不均衡导致误差较大，TF-IDF指数则是对这种不均衡的一种标准化处理。

TF (Term Frequency, 词频)，计算公式见图2，以“上班”在工作类别的TF为例，等于“上班”在工作类别中的词频除以工作类别的总词频数，该指标是为了防止聚类结果偏向总词频高的类别，是对词数的归一化。

IDF (Inverse Document Frequency)，计算公式见图2，是衡量每个词普遍性的指标，等于总类别数目除以包含该词的类别数目。如“工作”一词，在4个类别中都有出现，且总类别数目为4，则其IDF= $\log(4/4)=0$ 。该指标的意义在于过滤掉普遍存在的词语，提取出对分类有价值的词语。TF-IDF指标为上述两个指标的综合，计算公式见图2。

根据该指数计算出的各词四维行向量如表6。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$idf_i = \lg \frac{|D|}{|\{j : t_i \in d_j\}|}$$

$$tfidf_{i,j} = tf_{i,j} \times idf_i$$

图2: TF-IDF指数计算公式

关键词	工作	环境	家庭	学习
皮肤	0.59	0.47	0.47	0.46
上班	1	0.02	0.02	0.03
黑眼圈	0.7	0.04	0.36	0.61
痘	0.66	0.14	0.39	0.63
口罩	0.93	0.3	0.15	0.14

图2: TF-IDF计算结果样例

APPENDIX A: Data Collection and Mining Techniques

Text Mining (the codes can be found in attachment):

Step2: 关键词聚类

3) 聚类

根据上一步中得到的向量对筛选后的126个词语进行K-means聚类， $k=4$ ，得到四个聚类中心，将各词与类簇中心点进行相关关系的Pearson显著性检验，只保留有显著相关关系的词79个。经检查聚类结果与预期的4大类压力源家庭、工作、环境、学习大体吻合，详见本报告13页。

4) 降维可视化

使用TSNE降维将Step3中4维向量结果映射到二维平面上作为可视化的横纵坐标。得到基于关键词的“压力-皮肤问题-护肤品” Social Landscape，结果见本报告13页。

APPENDIX B: Chinese Verbatim of Quotes(1)

P20:

@怎么不让人发现我

上班前→下班后

戴口罩上班闷出痘痘，

有没有姐妹安利对痘痘 毛孔效果好的精华

@张咪瓶

昨天老公照了半天镜子，说自己工作辛苦黑眼圈大，然后羞答答的说该贴眼膜了，半夜0: 27开始，陪他保养到了一点...

@高雅轩

上班太难了 我最近九点多就睡觉不过早睡皮肤状态倒是好了很多

@MISS无敌兔

因为疫情 两个月没怎么正经上班了 也不知道为什么 最近的皮肤 好的发亮 白的反光 可能是因为我最近坚持运动的原因

P21:

@钮祜禄氏如花

因为长期投注工作，长时间熬夜加班，导致皮肤软组织下垂，花了2W大洋安排了自体脂肪填充后就很甜啦，年龄感都小了。看起来木有敲大变化，但是美就丸了。

@怎么不让人发现我

上班前→下班后

戴口罩上班闷出痘痘，

有没有姐妹安利对痘痘 毛孔效果好的精华

P22:

@不想加班的柚子你个柠檬

最近一直中国时间上班德国时间下班，烦躁焦虑到疯狂爆痘，项目同事各种令人窒息的操作我原地就是一个爆炸，带不动是真的带不动.....

终于在今天难得的按时下班之后一通胡乱吃喝玩乐到舍不得睡觉.....

熬夜吃喝令人精神愉快，也令人头秃

@浪里个浪儿里个浪

今天有被问到会喝酒么，我的回答是酒精过敏，喝酒会起疹子，我的想法或许很单纯很简单，工作是工作，生活是生活，我想分开，虽然我知道很多时候很难分得开，其实很矛盾，我不想让太熟悉的人看到自己狼狈的一面，所以喝酒这个事，既然酒精过敏开了头，那就不会再在同事面前去沾碰

APPENDIX B: Chinese Verbatim of Quotes(2)

P25:

@卡丝-秋红
@仁小桂大
@AimeeMeow
@巡游暮云間
@小于市
@Ju子小姐姐

转季的皮肤惹不起，要好生伺候着 多敷面膜！不然真的是无敌干燥 不多补水肌肤很容易过敏泛红脱皮卡粉等等
空气那么脏，灰尘那么重，皮肤不清洁我受不了啊
一到过敏季就得天天用依云洗脸平时喝都舍不得 能有啥办法呢
不知道咋回事脸上起了湿疹，医生说至少3个月不能化妆，护肤品最好只涂玉泽的修护面霜，刚拆封的娇韵诗双萃和obaji都给老妈用了，实惨
有一段时间什么都没买，只是采购食物和日用品。后来春天来了，空气特别干燥，采购了一波润肤露。最近又买了免洗的面膜，也是为了保湿。
天气越来越好不过空气还是有些干燥。每周面膜不可少

P26:

@哎呦你个大傻叉
@我是阿辛啊
@-安平十三妹
@你吃红烧狗肉吗
@ 5i33ydq

用我的鼻炎和满脸痘担保，过敏季真是来了
太久没在涞源过春天了，换个季脸都受不了了，干燥 疼痛
讨厌春天 春季即过敏季
又到了我的皮肤敏感季水乳不能涂，涂了就会火辣辣的疼，并且还痘痘烂了半边脸
春季湿疹，爆痒！挠挠……更痒……

P26:

@留白0197

不洗头还披着头的下场就是长痤疮，家这边空气湿度没那么高，从绝对油痘肌变成了混油，鼻子油分泌变多了，居然还起皮（而且我怕油脂分泌失调，还不敢洗脸）。呵呵，我的脸就是黄土高原，坑坑洼洼，满面灰尘。

@La_Collectionneuse

很真实 在外面奔波 皮肤会沾染很多污染 在家待两周皮肤真实变好

@凌小苍同学

我发现长黑头真的是外面空气脏不是我的原因。以前一周不管，黑头能密密麻麻长得满鼻子都是，刚才突然想起一个多月没管黑头了，一照镜子居然干干净净

@在路上_Joanna

在家待久了，皮肤也变好了，大概是没有外界空气污染的刺激，即便不怎么做面膜，每天洗完脸只擦一点护肤品，皮肤依然看起来很水润！

Thank you!

