

GSA: Text Analytics

Chinese BBS

Hanming (Ethan) Meng

What did we do till now?

- Scraped all ChaseDream.com posts related to:
 - MSBA
 - [REDACTED]
- Extracted key topics discussed within the MSBA field on ChaseDream.com
- Tracked sentiment scores of [REDACTED] on ChaseDream.com
- Analyzed most viewed posts in detail

Web Scraping

- ChaseDream.com
 - MSBA
 - [REDACTED]

How do we scrape ChaseDream.com?

- We use python and beautifulsoup for web scraping
- The results are stored in two csv files for storage

	Dates	Titles	Contents	link
0	2020-4-16	NEU CS Alig	因为SMU和	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1366746&highlight=MSBA
1	2016-3-23	老年人的	随着Duke	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1257089&highlight=MSBA
2	2020-2-10	美英坡MS	我是美国	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1360152&highlight=MSBA
3	2019-4-27	UR MSBA/B	更新 5.22	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1346216&highlight=MSBA
4	2022-1-15	MSBA@UI	帮亲戚小月	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1385670&highlight=MSBA
5	2019-3-27	【已定】N	I 4.29 upc	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1343940&highlight=MSBA
6	2020-5-10	MSBA商业:	更新 : 穷游	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1368210&highlight=MSBA
7	2020-1-27	【可refer】dd	l : 2.8 sr	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1359813&highlight=MSBA
8	2021-11-2	NTU MSBA	已录	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1384327&highlight=MSBA

	Dates	Titles	Contents	link
0	2020-2-10	美英坡MS	我是美国	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1360152&highlight=MSBA
1	2018-4-19	【选校贴】	(谢谢各位)	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1322421&highlight=UCD
2	2022-1-18	更新申请	最近一个	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1385728&highlight=UCD
3	2019-4-9	【总结+选	拒了UCD	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1344870&highlight=UCD
4	2022-1-17	【已录】U	之	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1385699&highlight=UCD
5	2022-1-17	2fall ucd r	bg : 陆本3	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1385686&highlight=UCD
6	2019-2-15	【决定去U	和ucd的小	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1339468&highlight=UCD
7	2021-12-8	【选校】UTor	大家好,	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1384775&highlight=UCD
8	2020-4-21	UCD MSBA	Mark一下	https://forum.chasedream.com/forum.php?mod=viewthread&tid=1367088&highlight=UCD

Fig: Screenshot of the example datas

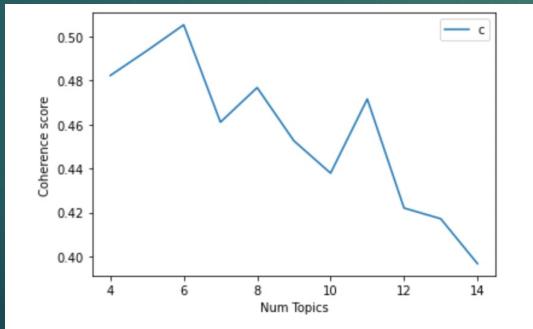
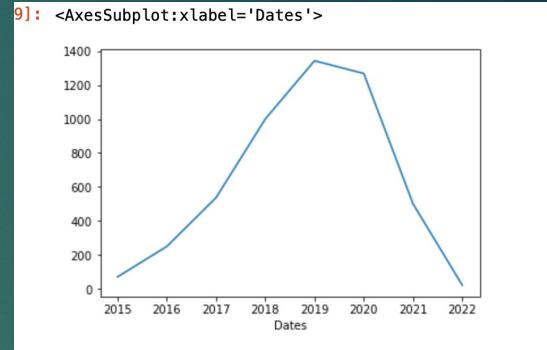
Topic Modelling

- For all MSBA posts on ChaseDream.com

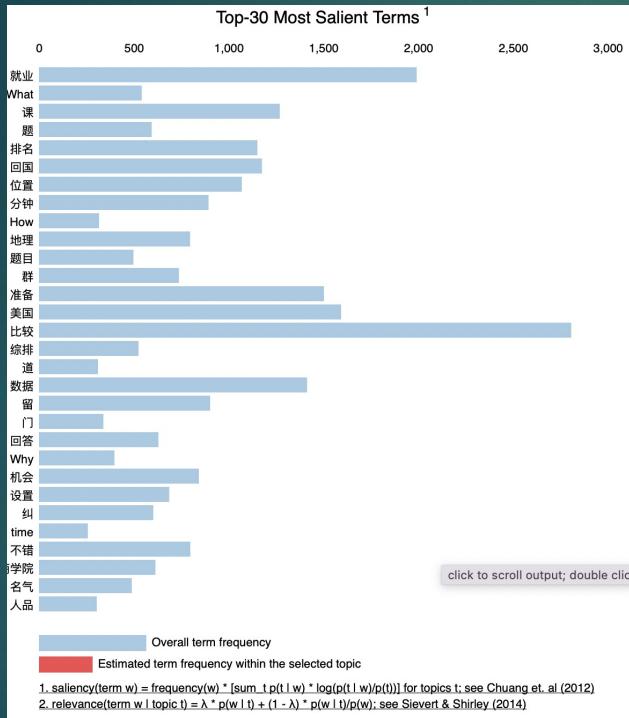
For MSBA posts: Topic Modelling

- Size: 4992 posts on BBS
- Time Frame: 2015-2022
- Parameter Tuning:
 - Coherence Scores gives 6 Clusters

Dates	
2015	71
2016	250
2017	537
2018	1000
2019	1343
2020	1268
2021	501
2022	22



Findings 1: Top words among all posts



Most discussed keyword

#1: Placement

- Ranking
- Back to China
- Stay in US
- Location
- Business school
- Group Chat

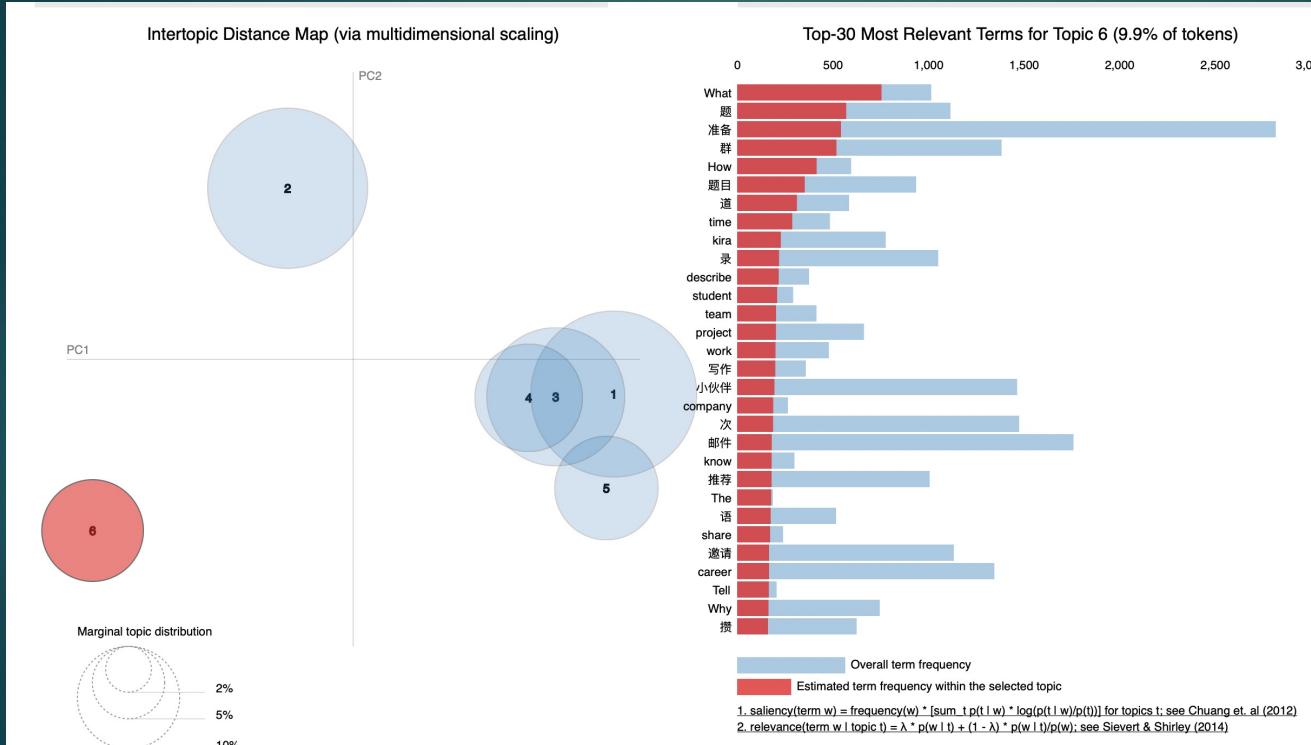
Findings 2: Topic Similarity



Inter-topic Distance

- 4 of them are close together
- 2 are more distant

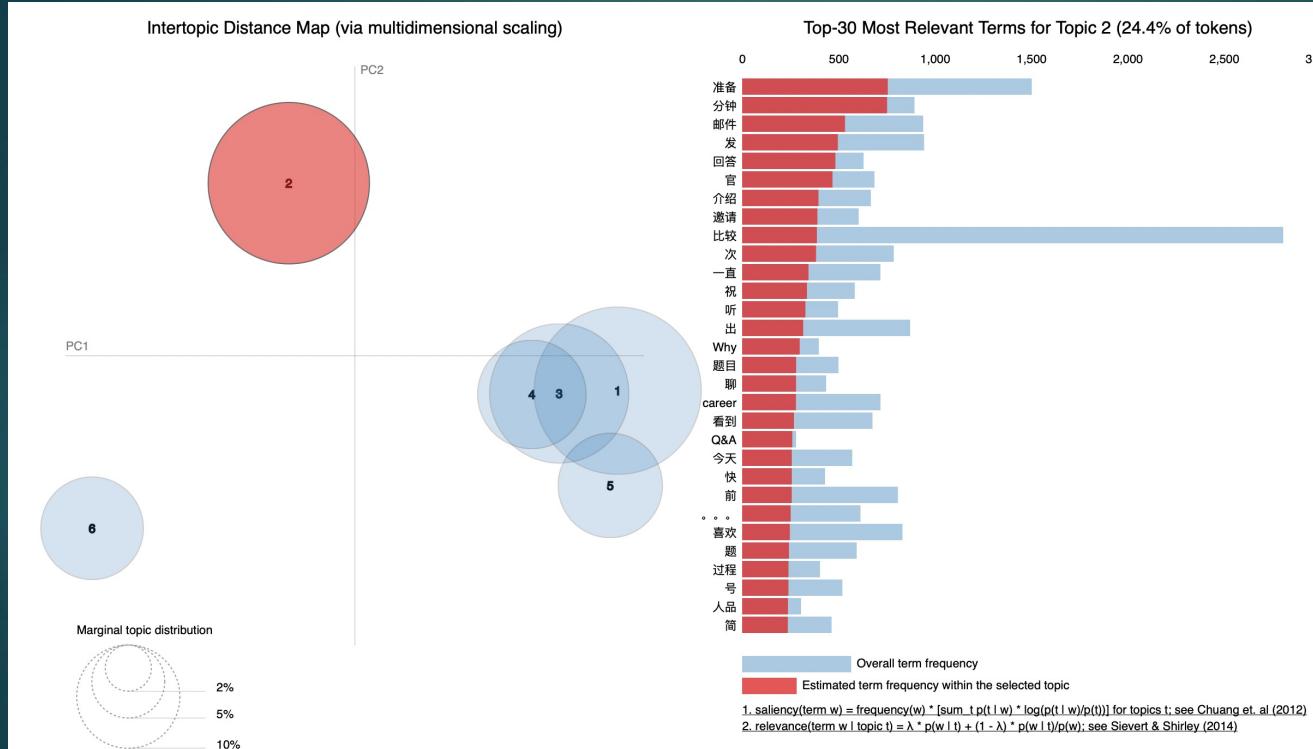
Reasons behind this 😊



This cluster is all about one-way interviews

- Kira
- Record
- Share
- Team
- Project
- Company

Reasons behind this 😊



This cluster is all about in-person interviews

- Interviewer
- Prepare
- Email
- Process
- Q & A
- Personality

Finding 3 : Key words within each cluster

Data Viz in next pages

Topic_Num	Topic_Perc_Contrib	Keywords
0	1.0	0.9992 数据, 比较, 就业, 美国, 排名, 小伙伴, 专业, 位置, 群, 回国
1	2.0	0.9992 比较, 就业, 专业, 回国, 美国, 背景, 本科, 数据, 建议, 金融
2	3.0	0.9990 就业, 位置, 排名, 比较, 回国, 美国, 地理, 机会, 留, 不错
3	4.0	0.9985 What, 题, 准备, 群, How, 题目, 道, time, kira, 录
4	5.0	0.9997 课, 比较, 就业, 美国, 数据, 门, 学习, 机会, 需要, data
5	6.0	0.9977 准备, 分钟, 邮件, 发, 回答, 官, 介绍, 邀请, 比较, 次

Viz: Word Clouds - Cluster #1



In this cluster:

- WFU > UCD \approx WUSTL >
UCSD/WM/UR/USC/GWU/UCSD/UMN
etc.
 - Placement > Location \approx Ranking

Viz: Word Clouds - Cluster #2



In this cluster:

- Location/Placement > Ranking
 - Degree in Data > Degree in Finance
 - USC/WUSTL/UCSD

Viz: Word Clouds - Cluster #3



In this cluster:

- US > China
 - Location and Placement
 - USC/WUSTL/SMU/GWU/GU

Viz: Word Clouds - Cluster #4



In this cluster:

- Questions
 - Prepare
 - Kira
 - Group and Peers

Viz: Word Clouds - Cluster #5



In this cluster:

- Curriculum > Placement
 - Technical courses
 - Pros and Cons
 - Project

Viz: Word Clouds - Cluster #6



In this cluster:

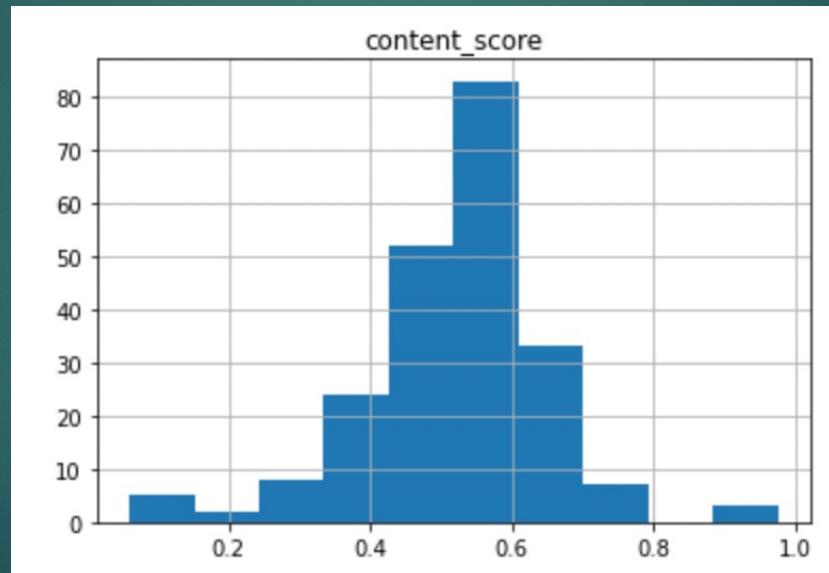
- Interviewer
 - Questions asked and Answers Provided
 - Self-intro, Walk through resume, career goal

Sentiment Analysis

- For all posts on ChaseDream.com relating to [REDACTED]

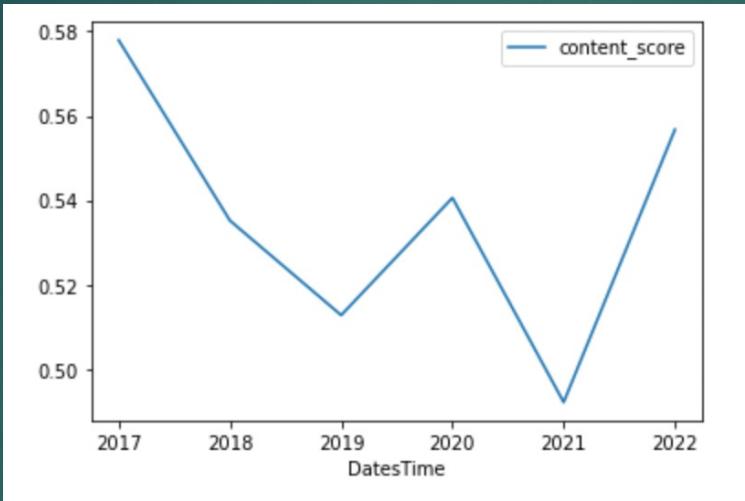
Histogram of all Sentiment Scores

- 540 posts about UCD on ChaseDream.com
- Most are neutral to high



Sentiment Scores by Year

Dates	DateTime
2011	2
2015	2
2016	4
2017	9
2018	43
2019	174
2020	216
2021	84
2022	6

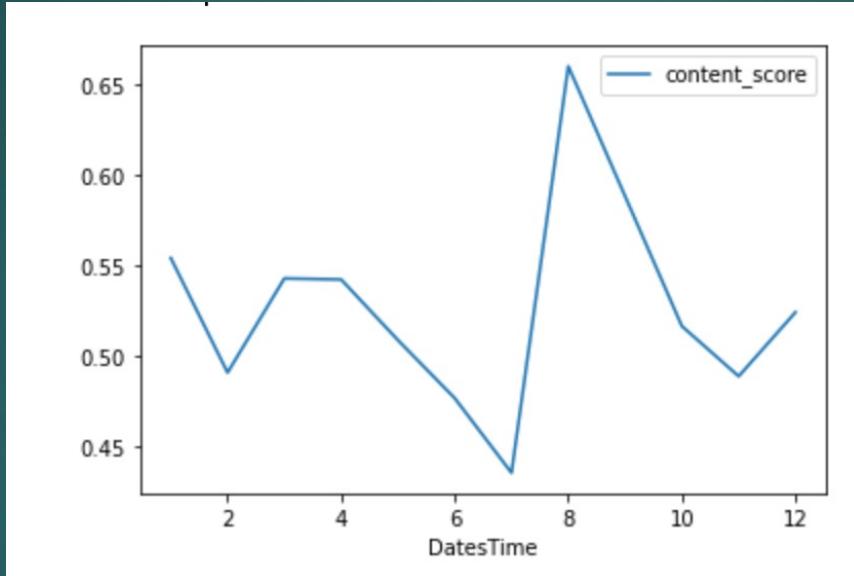


Removed outliers

- Initial marketing effect
- COVID-19 Effect

Sentiment Scores by Month

Dates	1
DatesTime	
1	88
2	126
3	133
4	63
5	22
6	5
7	2
8	2
10	12
11	32
12	55



- Number of posts in line with application cycle
- Summer > Winter
- A mid application cycle depression period: Oct-Nov