

# Connections

Predicting Venmo expansion

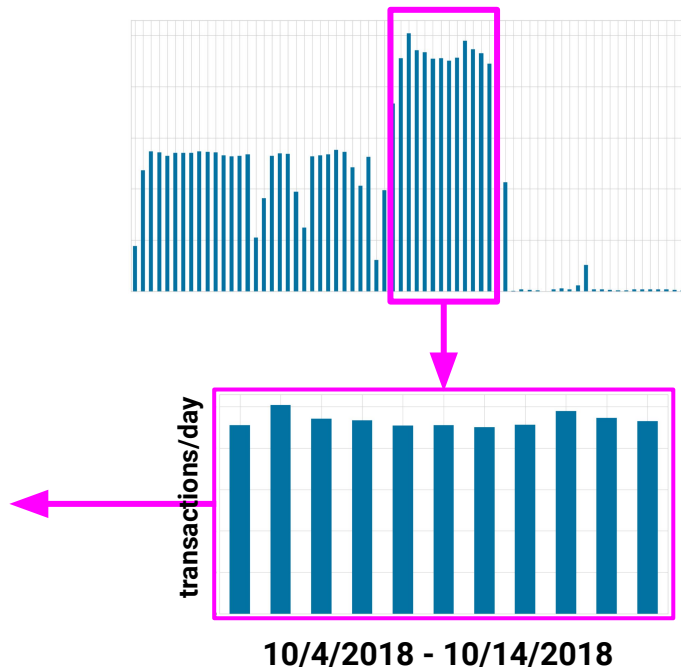








What expands  
Venmo's  
network?

# Data Description

- **1.8M public Venmo transactions in 11 consecutive days (October 4 - 14th, 2018)**
- **2.6M Venmo users in slice**
- **Public API features are few:**
  - **who paid** (*payee\_name*)
  - **who received** (*payer\_name*)
  - **when they joined** (*payee\_join\_date*, *payer\_join\_date*)
  - **when they paid** (*date*)
  - **what they said** (*note*)
  - **what app they used** (*app\_name*)



# NLP: Topic Modeling Venmo Notes

note	count
Food	20359
	14346
	8269
:uber:	7469
	7430
Uber	6184
	5925
	5834
	5399
	4920

GSDMM  
deals well  
with single-  
document  
corpus

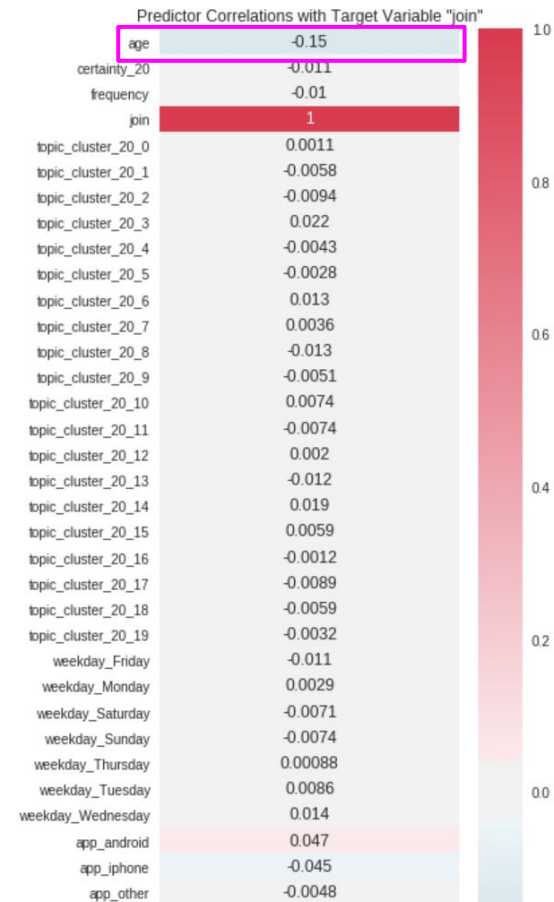
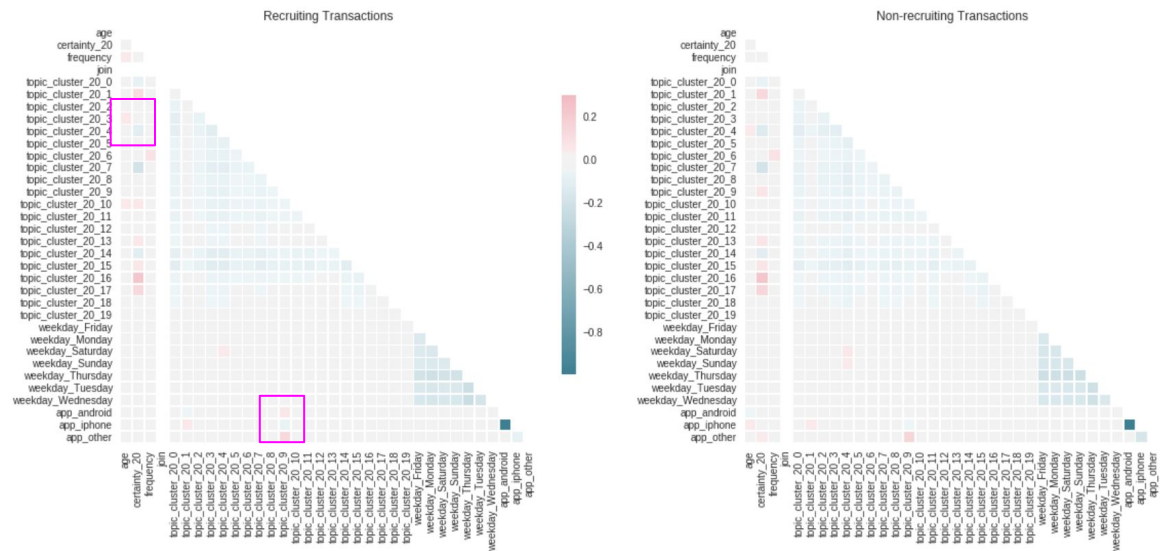
**GSDMM**  
(spaCy)

**Hard to sort:**  
Top 10 notes  
= 0.05%  
unique notes

TOPIC	ESTIMATE	20 MOST FREQUENT TOKENS IN CLUSTER
0	<i>eating_out</i>	['hamburg', 'poultry_leg', 'french_fri', ' ', 'cooki', 'sushi', 'pizza', 'bread', 'steaming_bowl', 'green_salad', 'cheese_wedg', 'cook', 'soft_ice_cream', 'bacon', 'cup_with_straw', 'red_appl', 'sandwich', 'beer_mug', 'spaghetti', 'chicken']
1	<i>friends</i>	['thi', 'n't', 's', '!', 'm', 'thank', 'get', 'money', 'sorri', 'back', 'know', 'pay', 'love', 'one', 'wa', 'go', 'time', 'face_with_tears_of_joy', 'let', 'lol']
2	<i>reoccurring</i>	['due', 'octob', 'thank', 'rent', 'fee', 'deposit', 's', 'payment', 'clean', 'oct', 'car', 'week', 'shirt', 'lesson', 'lawn', 'servic', '10', 'class', 'hous', 'tutor']
3	<i>transport</i>	['food', 'uber', 'fuel_pump', 'automobil', 'fork_and_knif', 'groceri', 'taxi', 'venmo_dollar', 'ga', 'oncoming_automobil', 'lyft', 'stuff', 'car', 'person_getting_haircut', 'lunch', 'sport_utility_vehicl', 'pizza', 'thank', 'Hi', 'Za']
4	<i>sports_event</i>	['american_footbal', 'ticket', 'admission_ticket', 'movie_camera', 'basketbal', 'basebal', 'fantasi', 'week', ' ', 'popcorn', 'soccer_bal', 'flag_in_hol', 'football', 'ice_hockey', 'game', 'pick', 'movi', 'clapper_board', 'Go', 'tix']
...	...	...

See **Appendix B** for full topic cluster list

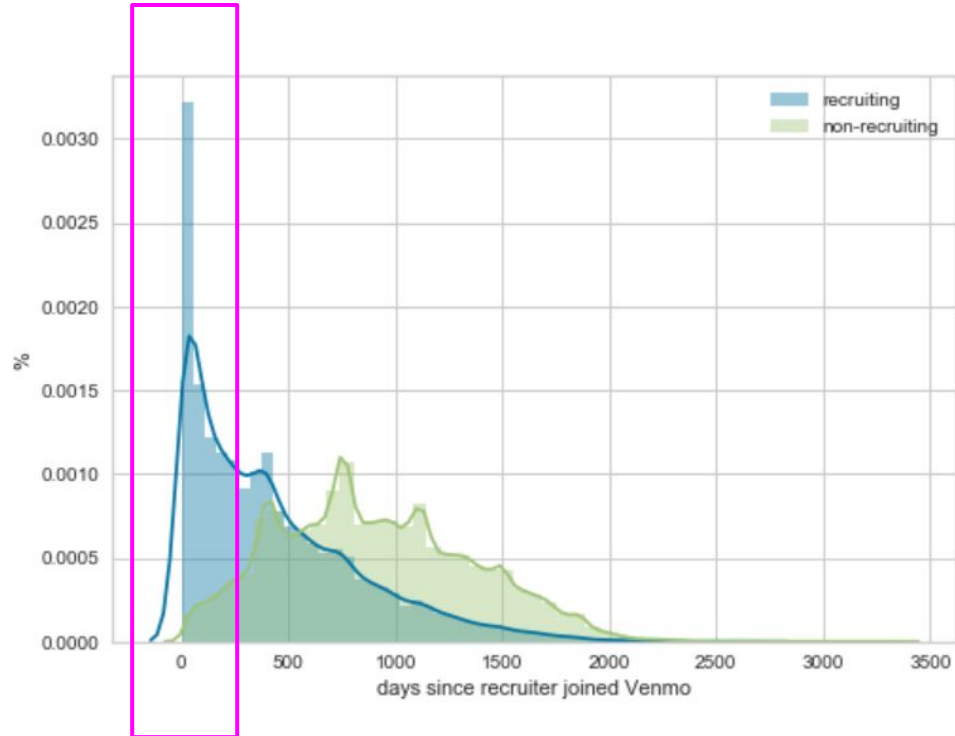
# Class Comparisons



## Age: a closer look

**New users seem more likely to bring in other new users...**

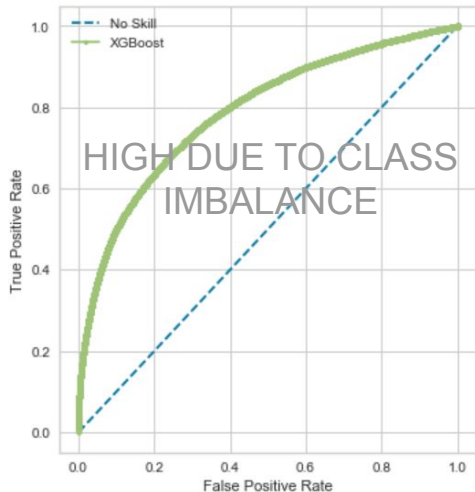
**...but did the reason for payment matter?**



# Recruitment Model (XGBoost)

79.0%  
HIGH DUE TO CLASS  
IMBALANCE

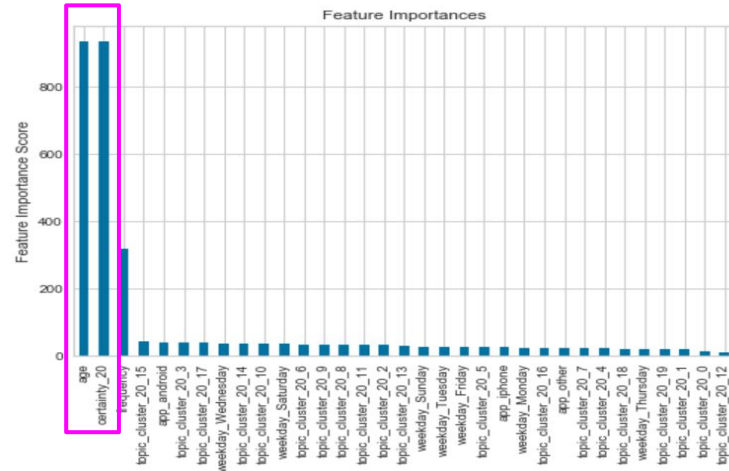
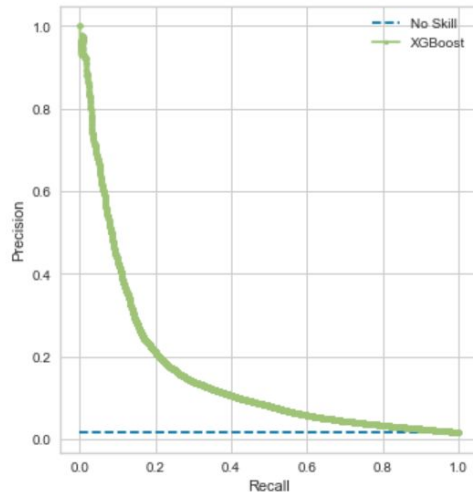
AUROC



15.8%  
AUPRC

AUPRC

2% baseline



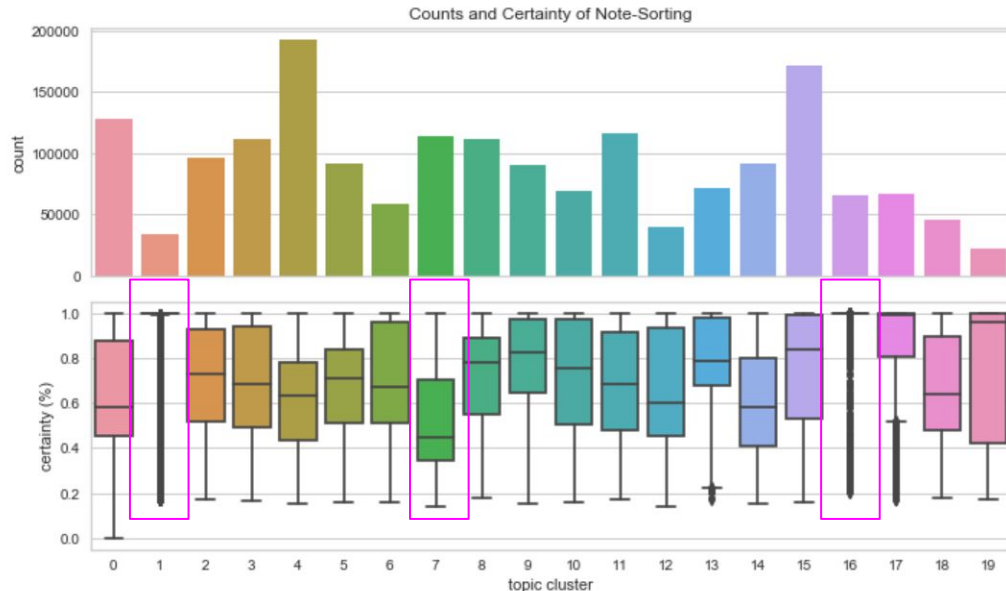
<b>age</b> ✓	30.3%
<b>topic</b> ?	
<b>certainty</b>	30.3%
<b>frequency</b>	10.12%



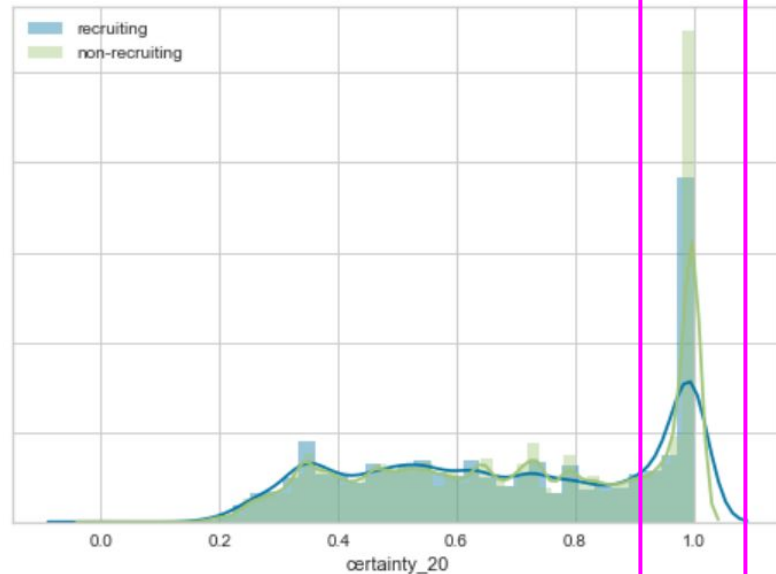
# Certainty\_20: a closer look

Predictor Correlations with Target Variable "join"	
age	-0.15
certainty_20	-0.011
frequency	-0.01

**Topic\_7 (relationships) seems least certain**  
**\_1(friends) and \_16 (fast\_food) the most**



**Messages in recruiting transactions harder to sort**





# Recommendations & Limitations

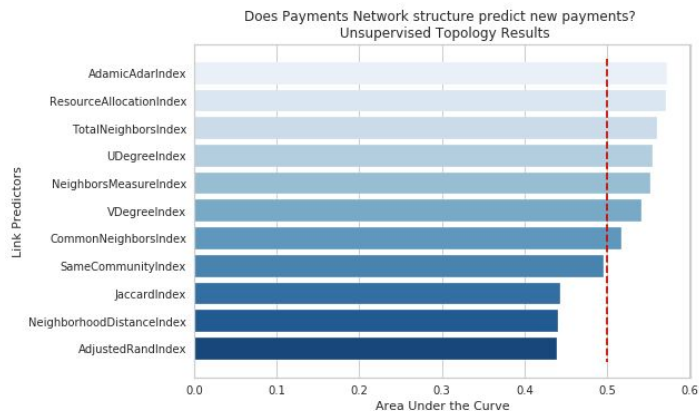
## Recommendations

- ❖ Offer group recruitment incentives
- ❖ Advertise Venmo for group events
- ❖ Further analysis on uncertain notes

## Limitations

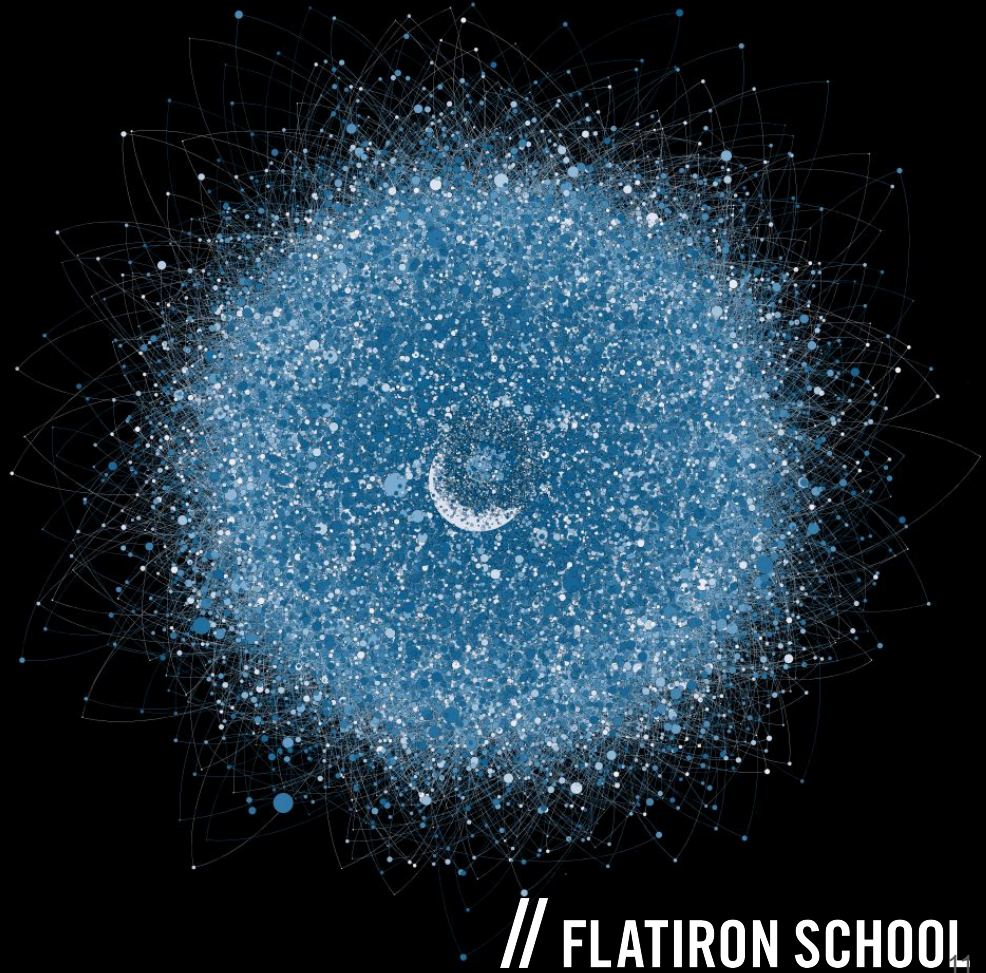
- Public API
- Uncertain source
- GSDMM evaluation

# Further Research and Next Steps







1. **Validate further on August data**
2. **Isolate and analyze uncertain messages**
3. **Construct supervised topological prediction (ongoing)**

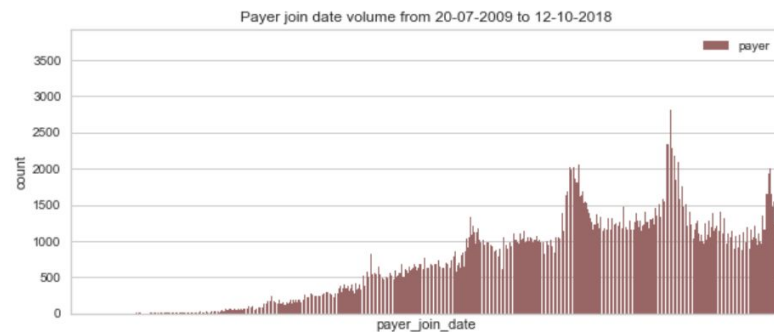
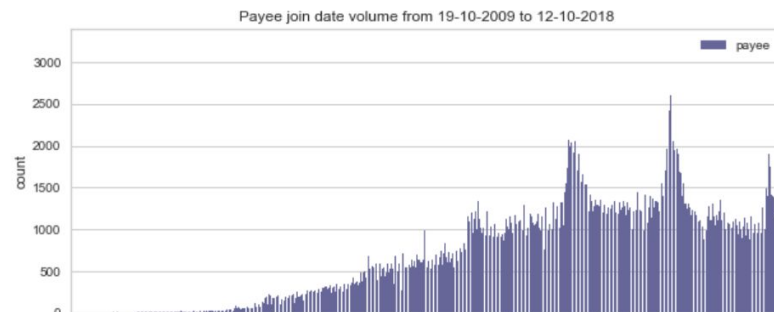
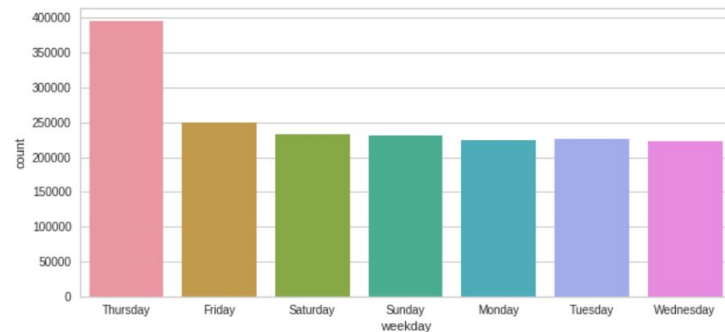
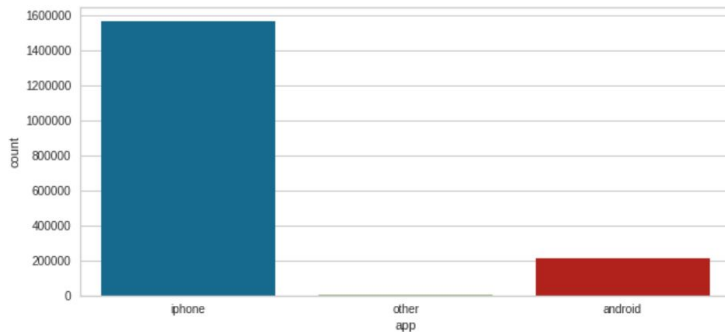
***Thank You!***



**// FLATIRON SCHOOL**

# Appendix A: Initial EDA

note	count
Food	20359
	14346
	8269
:uber:	7469
	7430
Uber	6184
	5925
	5834
	5399
	4920



TOPIC	ESTIMATE	MOST FREQUENT TOKENS
0	eating_out	['hamburg', 'poultry_leg', 'french_fri', ' ', ' ', 'cooki', 'sushi', 'pizza', 'bread', 'steaming_bowl', 'green_salad', 'cheese_wedg', 'cook', 'soft_ice_cream', 'bacon', 'cup_with_straw', 'red_appl', 'sandwich', 'beer_mug', 'spaghetti', 'chicken']
1	friends	['thi', 'n't', 's', ' ', ' ', 'm', 'thank', 'get', 'money', 'sori', 'back', 'know', 'pay', 'love', 'one', 'wa', 'go', 'time', 'face_with_tears_of_joy', 'let', 'lol']
2	reoccurring	['due', 'octob', 'thank', 'rent', 'fee', 'deposit', 's', ' ', 'payment', 'clean', 'oct', 'car', 'week', 'shirt', 'lesson', 'lawn', 'servic', '10', 'class', 'hous', 'tutor']
3	transport	['food', 'uber', 'fuel_pump', 'automobil', 'fork_and_knif', 'groceri', 'taxi', 'venmo_dollar', 'ga', 'oncoming_automobil', 'lyft', 'stuff', 'car', 'person_getting_haircut', 'lunch', 'sport_utility_vehicl', 'pizza', 'thank', 'Hi', 'Za']
4	sports_event	['american_footbal', 'ticket', 'admission_ticket', 'movie_camera', 'basketbal', 'basebal', 'fantasi', 'week', ' ', ' ', 'popcorn', 'soccer_bal', 'flag_in_hol', 'footbal', 'ice_hockey', 'game', 'pick', 'movi', 'clapper_board', 'Go', 'tix']
5	exercise	['light_skin_ton', 'dark_skin_ton', 'baby_medium', 'woman_dancing_medium', 'thumbs_up_medium', 'raising_hands_medium', 'folded_hands_medium', 'flexed_biceps_medium', ' ', ' ', 'woman_tipping_hand_medium', 'ok_hand_medium', 'call_me_hand_medium', 'girl_medium', 'woman_getting_haircut_medium', 'boy_medium', 'woman_shrugging_medium', 'baby_light_skin_ton', 'red_heart', 'bride_with_veil_medium', 'thank']
6	groceries	['chicken', 'stuff', 'clean', 'paper', 'food', 'chees', 'water', 'hot', 'juic', 'bowl', 'folded_hand', 'thank', 'chip', 'fri', 'bag', 'towel', 'wing', 'hous', 'pod', 'beer']
7	relationships	['red_heart', 'eye', 'thank', 'dog_fac', 'face_blowing_a_kiss', 'two_heart', 'smiling_face_with_heart', 'star', 'struck', 'dog', 'sparkli', 'din', 'purple_heart', 'yellow_heart', 'growing_heart', ' ', ' ', 'blue_heart', 'lipstick', 'kiss_mark', 'smiling_face_with_smiling_ey']
8	foreign_communication	['-', '1', 'middle_finger_dark_skin_ton', 'wub', 'ree', 'posti', 'skurt', 'goggl', 'Wt', 'starb', 'idfk', 'mhmm', 'amsterdamm', 'thumbs_up_dark_skin_ton', 'doof', 'poverti', 'yia', 'dadada', '机票', 'bizza']
9	celebration	['s', 'money', 'musical_not', 'mouth_fac', 'babi', 'gift', 'wrapped_gift', 'ticket', '"s', 'bday', 'microphon', 'shower', 'day', 'thank', 'kale', 'mcdonald', 'birthday', 'festival_stag', 'festival_b', 'salad']
10	thanks_coffee	['thank', 'good', 'love', 'red_heart', 'best', 'face_blowing_a_kiss', 'coffe', 'life', 'hot_beverag', 'get', 'ice', 'food', 'stuff', 're', 'time', 'two_heart', 'My', 'cream', 'thx', 'happi']
11	travel	['night', 'uber', 'food', 'dinner', 'ticket', 'fight', 'hotel', 'last', 'fun', 's', ' ', 'stuff', 'trip', 'upsid', 'down_fac', 'drink', 'parti', 'beer', 'weekend', 'groceri', 'airbnb']
12	sports_international	['zany_fac', 'boxing_glov', 'know', 'fight', 'cowboy_hat_fac', 'mcgregor', 'ireland', 'khabib', 'b_button_(blood_typ', 'ufe', 'running_sho', 'smirking_fac', 'yee', 'bet', 'Ya', 'russia', 'côte_d'ivoir', 'air', 's', 'hedgehog']
13	friends_	['fire', 'sweat_droplet', 'pile_of_poo', 'eggplant', 'peach', ' ', ' ', 'smiling_face_with_sunglass', 'evil_monkey', 'tongu', 'electric_plug', 'hundred_point', 'clown_fac', 'face_vomit', 'ok_hand', 'dashing_away', 'shushing_fac', 'smiling_face_with_horn', 'see', 'alien', 'book']
14	house_services	[' ', 'airplan', 'money_bag', 'house_with_garden', 'hous', 'automobil', 'maple_leaf', 'fuel_pump', 'fallen_leaf', 'leaf_fluttering_in_wind', 'palm_tre', 'snowflak', 'sun', 'bouquet', 'sport_utility_vehicl', 'clinking_beer_mug', 'hotel', 'evergreen_tre', 'water_wav', 'blossom']
15	party	['party_popp', 'wine_glass', 'beer_mug', 'clinking_beer_mug', 'bottle_with_popping_cork', 'birthday_cak', 'wrapped_gift', 'clinking_glass', 'cocktail_glass', 'happi', 'birthday', ' ', ' ', 'tropical_drink', 'tumbler_glass', 'balloon', 'ring', 'pizza', 'bu', 'woman_dancing_light_skin_ton', 'red_heart']
16	fast_food	['taco', 'hot_beverag', 'burrito', 'tropical_drink', 'doughnut', 'cook', 'bell', ' ', ' ', 'flexed_bicep', 'stuffed_flatbread', 'breakfast', 'person_run', 'brunch', 'pancak', 'beer_mug', 'red_heart', 'frog_fac', 'tuesday', 'donut', 'bacon']
17	unknown	['pizza', 'shirt', 'woman's_cloth', 'dress', 'jean', 'chick', 'yum', 'fil', 'pie', 'thank', 'phi', 'cancer', 'red_heart', 'medium', 'larg', 's', 'chi', 'de', 'ribbon', 'la']
18	utilities	['money_with_w', 'hous', 'bill', 'electric_plug', 'high_voltag', 'light_bulb', 'dollar_banknot', 'mobile_phon', 'sweat_droplet', 'laptop_comput', 'octob', 'electr', 'septemb', 'rent', 'wifi', 'ga', 'televis', 'splitwise.com', 'fire', 'internet']
19	fall_events	['jack', 'lantern', 'eggplant', 'ghost', ' ', ' ', 'dumpl', 'duck', 'exploding_head', 'tombston', 'american_footbal', 'skull', 'steaming_bowl', 'halloween', 'skull_pop', 'n\ln', 'hous', 'baby_bottl', 'hot_dog', 'ram', 'crown']