

```
In [1]: import pandas as pd
import numpy as np
import plotly.express as px
import datetime
```

Summer Sessions Engagement Analysis

```
In [2]: # read in raw data from Sprout Social
df = pd.read_csv('SSALL_IG/sprout_data.csv')
```

```
In [3]: # convert to lowercase for simpler coding
lower = [name.lower() for name in df.columns.to_list()]
df.columns = lower

# separate df into summer sessions and study abroad accounts
ss_df = df[df['instagram profile'] == 'ucberkeleysummer']
# ss_df
sa_df = df[df['instagram profile'] == 'ucbstudyabroad']
sa_df
```

Out[3]:

	date	instagram profile	followers	net follower growth	followers gained	followers lost	following	net following growth	published posts & stories	impression
1	06-30-2021	ucbstudyabroad	2,244	2.0	2.0	0.0	223.0	0.0	0	12
3	07-01-2021	ucbstudyabroad	2,245	1.0	2.0	1.0	223.0	0.0	1	43
5	07-02-2021	ucbstudyabroad	2,241	-4.0	1.0	5.0	223.0	0.0	0	8
7	07-03-2021	ucbstudyabroad	2,241	0.0	2.0	2.0	223.0	0.0	0	6
9	07-04-2021	ucbstudyabroad	2,241	0.0	0.0	0.0	223.0	0.0	0	.
...
723	06-26-2022	ucbstudyabroad	2,707	0.0	1.0	1.0	226.0	0.0	0	19
725	06-27-2022	ucbstudyabroad	2,709	2.0	2.0	0.0	226.0	0.0	1	47
727	06-28-2022	ucbstudyabroad	2,709	0.0	0.0	0.0	226.0	0.0	0	13
729	06-29-2022	ucbstudyabroad	2,709	0.0	0.0	0.0	226.0	0.0	0	6
731	06-30-2022	ucbstudyabroad	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN

Summer Sessions EDA

```
In [4]: # group ss data by month and year
ss_df.index = pd.to_datetime(ss_df['date'])
ss_group = ss_df.groupby(by=[ss_df.index.month, ss_df.index.year]).sum()
ss_group.index = pd.to_datetime(ss_group.index.get_level_values(1).astype(str) + '-' +
                                ss_group.index.get_level_values(0).astype(str),
                                format='%Y-%m')
ss_group.sort_index(inplace=True)
ss_group.index = ss_group.index.strftime('%b %Y')
ss_group
```

```
Out[4]:
```

	net follower growth	followers gained	followers lost	following	net following growth	published posts & stories	engagements	likes	comments	saves	re
date											
Jun 2021	1.0	1.0	0.0	113.0	0.0	0	0	0	0	0	
Jul 2021	2.0	35.0	33.0	3496.0	-1.0	8	73	65	2	6	
Aug 2021	13.0	41.0	28.0	3468.0	0.0	1	60	54	0	6	
Sep 2021	-9.0	19.0	28.0	3248.0	0.0	0	0	0	0	0	
Oct 2021	16.0	51.0	35.0	3457.0	-1.0	1	8	8	0	0	
Nov 2021	32.0	53.0	21.0	3329.0	0.0	0	0	0	0	0	
Dec 2021	32.0	70.0	38.0	3441.0	0.0	5	142	114	1	27	
Jan 2022	47.0	96.0	49.0	3439.0	-1.0	19	212	147	0	65	
Feb 2022	59.0	103.0	44.0	3112.0	2.0	19	358	228	0	130	
Mar 2022	35.0	71.0	36.0	3472.0	0.0	16	185	148	0	37	
Apr 2022	39.0	78.0	39.0	3360.0	0.0	17	110	83	3	24	
May 2022	20.0	65.0	45.0	3468.0	-1.0	22	132	77	1	54	
Jun 2022	27.0	51.0	24.0	3330.0	0.0	4	57	49	0	8	

During which months do users engage the most with Summer Sessions content?

Hypotheses, intuitions, investigations:

- Leading up to summer?
- During enrollment?
- During summer?

- Low during the fall?

Goal:

- To determine the most effective time of the year to spend marketing efforts creating and publishing content.

```
In [5]: # keep only engagement data: engagements, saves, story replies, likes, comments
engagement_ss = ss_group.drop(labels=['net follower growth', 'followers gained',
                                     'followers lost', 'following', 'net following',
                                     'profile actions'], axis=1)

# engagement_ss
```

```
In [8]: # number of posts published per month over the last year in descending order
ss_post_count = engagement_ss.loc[:, ['published posts & stories']].sort_values('publish
ss_post_count
```

Out[8]: **published posts & stories**

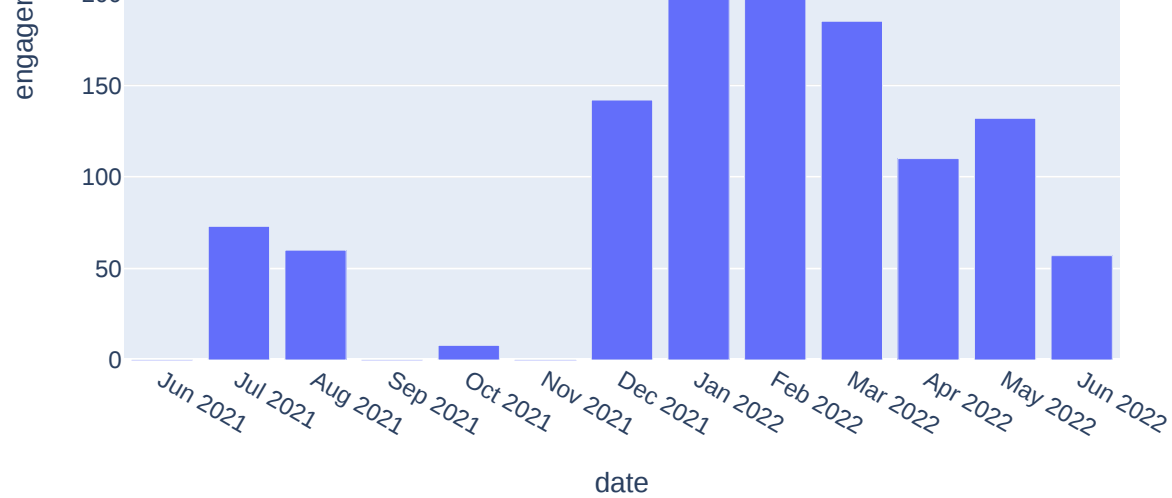
date	
May 2022	22
Jan 2022	19
Feb 2022	19
Apr 2022	17
Mar 2022	16
Jul 2021	8
Dec 2021	5
Jun 2022	4
Aug 2021	1
Oct 2021	1
Jun 2021	0
Sep 2021	0
Nov 2021	0

```
In [9]: # plot engagement vs month to see when users were most engaged in the past year
fig1 = px.bar(data_frame = engagement_ss, x = engagement_ss.index, y = 'engagements', ti
fig1.show()
```



Summer Sessions: Total Engagements Per Month





Commentary/Analysis

May saw the most published content but low engagement. February 2022 was the second month with the most published content. February also saw the most engagement at 358 interactions. Engagement falters during months where no content is published (September and November). While the amount of content published is important to engagement, the month is most influential to the data. We see valleys and peaks that give insight into engagement per month.

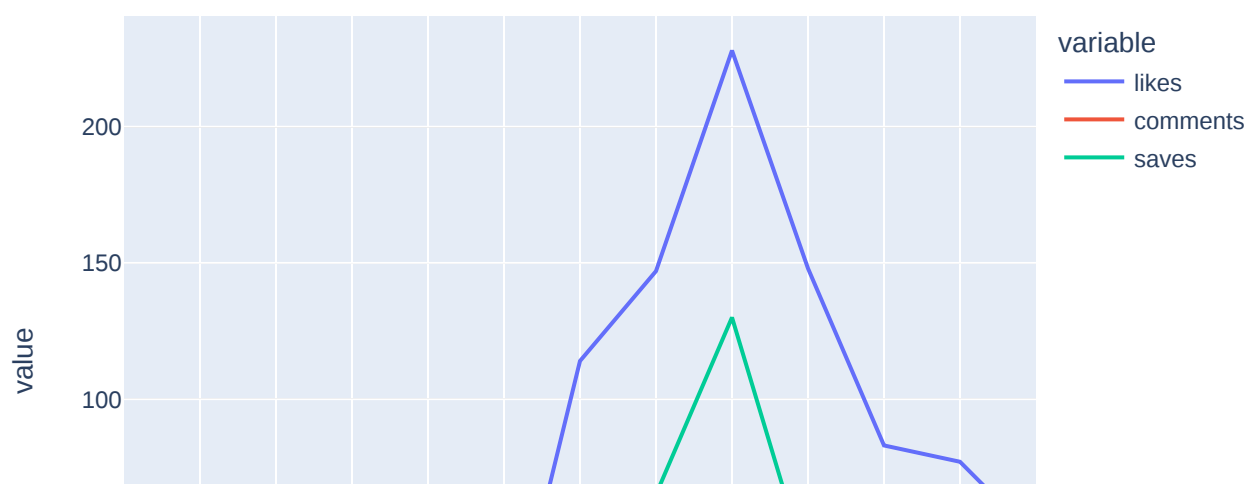
Posting content drives engagement, but engagement is optimized when posts are published during relevant months. For example, the months September, October, and November were only met with 8 engagements overall. This is consistent with the hypothesis that engagement is low during the fall, the period post-Summer Sessions and pre-enrollment. During summer in July and August, we also see minimal engagement. Following the decline in the fall, engagement picks up in December and peaks in February. Enrollment begins February 1 explaining the spike in engagement.

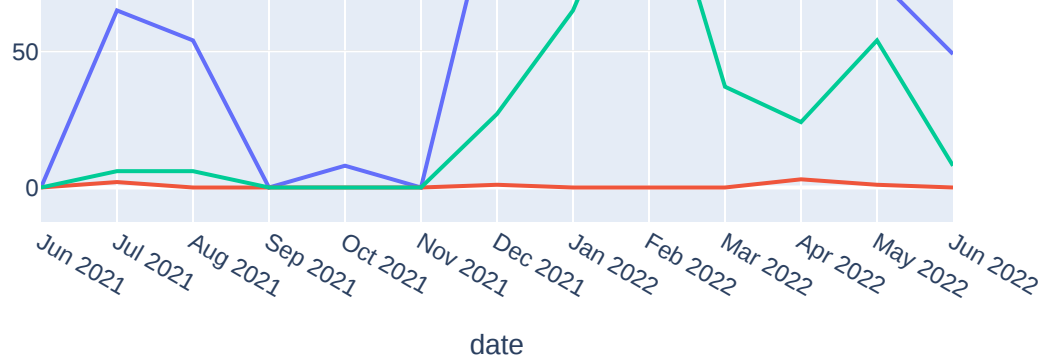
Post enrollment, engagement declines with a local maximum during May 2022, right before summer sessions begin.

```
In [10]: # break down engagement into parts
fig_2 = px.line(data_frame=engagement_ss, y=['likes', 'comments', 'saves'], x = engagement_ss['date'])
fig_2.show()
```



Summer Sessions: Likes, Comments, & Saves per Month





Commentary/Analysis

Engagement is an umbrella term for likes, comments, and saves. Likes are not considered a quality metric for our purposes since users often carelessly like posts on Instagram while scrolling. Nevertheless, they are a fine indication of reach.

Saves, here, are the most valuable metric as it indicates a user's desire to return to the information.

We see that the trends follow the same analysis above. Peaking in February, hitting rock bottom post-Summer Sessions, and gaining traction during early December.

How does Summer Sessions' following grow throughout the year?

Hypotheses, intuitions, investigations:

- Grow leading up to enrollment?
- Grow during enrollment?
- Grow during the beginning of summer?
- Decrease during fall months?

Goal:

- We want to discover when students follow Summer Sessions and indicate an interest in receiving additional relevant information.

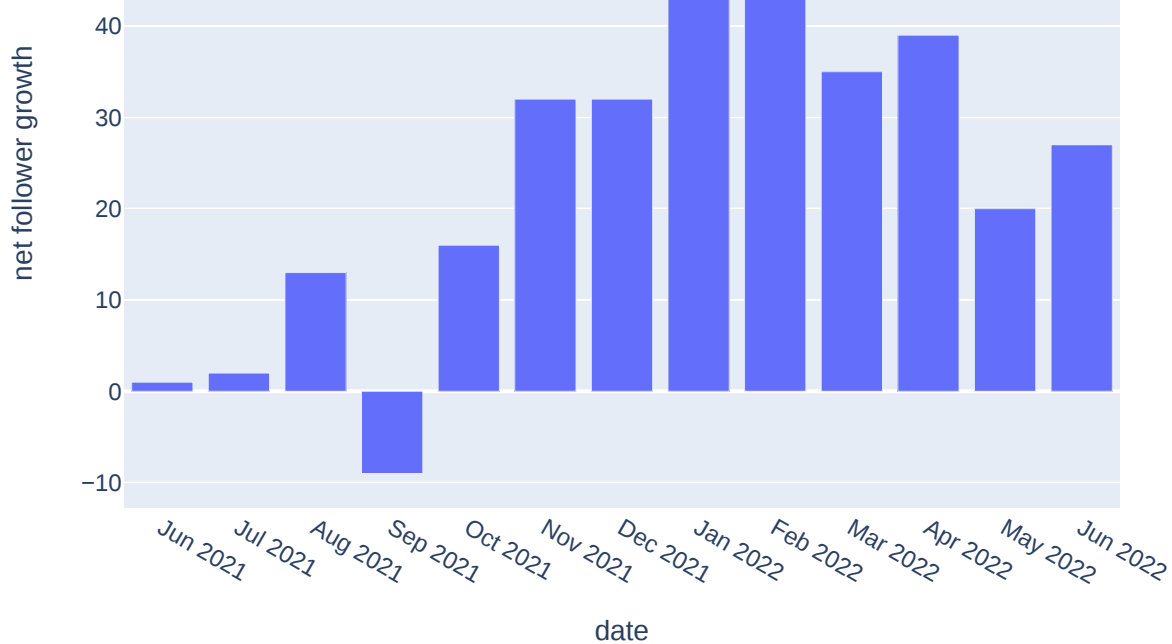
Follower Visualization

```
In [11]: # Net followers gained/lost per month
fig3 = px.bar(data_frame=ss_group, x=ss_group.index, y='net follower growth', title='Sum
fig3.show()
```



Summer Sessions: Follower Growth Per Month





Commentary/Analysis

The Summer Sessions account experiences steady follower growth during November through June. There is lower net growth during July through October. September experienced a negative net growth, indicating that students are not seeking Summer Sessions information during this time.

Summer sessions gained the most net followers during February. This, again, aligns with our hypothesis that students seek information the most during the summer enrollment phase.

Reccomendation/Conclusion

The optimal time to publish content on the Summer Sessions account to engage the most students is during the beginning of enrollment in February. During this month, students find the information the most relevant.

When is Summer Sessions' reach optimized?

Hypotheses, intuitions, investigations:

- During summer sessions when students are looking for course info.
- During enrollment when students are looking for course offerings.

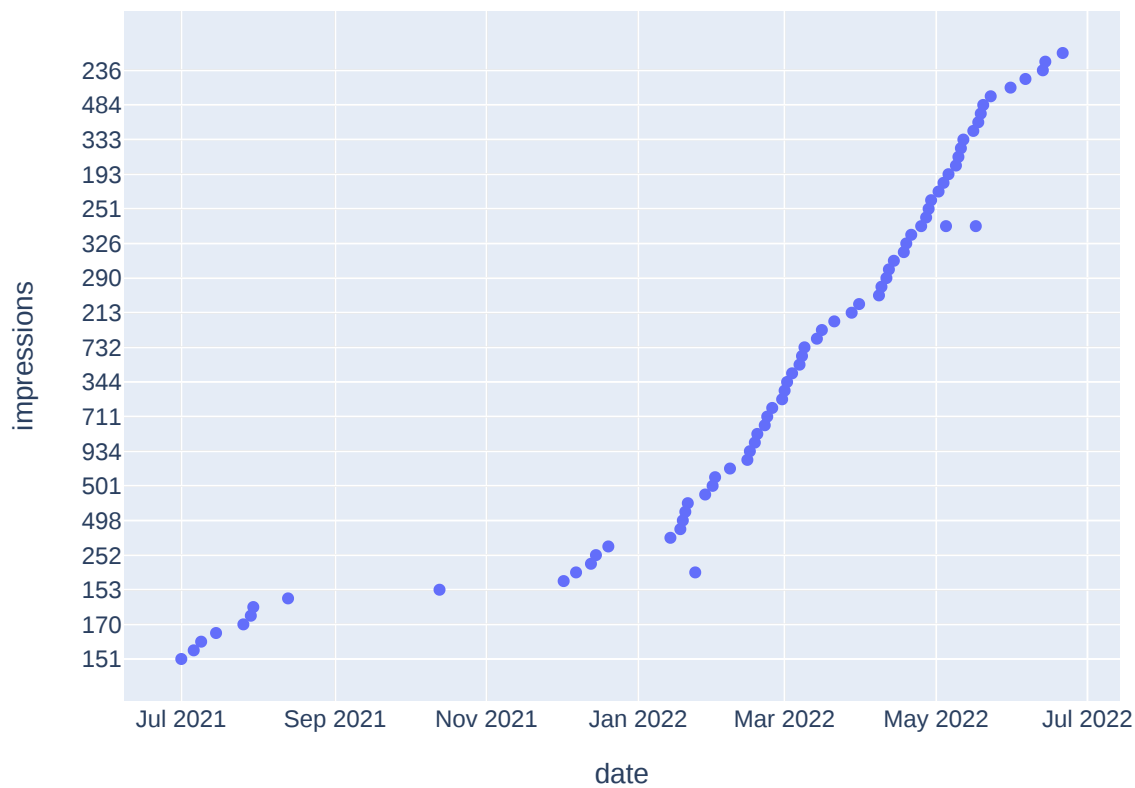
Goal:

Explore trends relating to impressions and reach to get a better sense of summer sessions outreach via IG.

```
In [68]: imp = ss_df.loc[:, ['impressions', 'reach', 'published posts & stories']]
imp = imp[imp['published posts & stories'] != 0]
imp = imp.reset_index()
imp
px.scatter(data_frame=imp, x='date', y='impressions', title="Summer Sessions Post Impres
```



Summer Sessions Post Impressions per Day



Findings:

Summer Sessions' impressions per month followed an upward trend from July 2021 to July 2022. What happens in the content throughout the year to explain this trend?

Study Abroad EDA

```
In [59]: # group sa df by month and year
sa_df.index = pd.to_datetime(sa_df['date'])
sa_group = sa_df.groupby(by=[sa_df.index.month, sa_df.index.year]).sum()
sa_group.index = pd.to_datetime(sa_group.index.get_level_values(1).astype(str) + '-' +
                                sa_group.index.get_level_values(0).astype(str),
                                format='%Y-%m')
sa_group.sort_index(inplace=True)
sa_group.index = sa_group.index.strftime('%b %Y')
# sa_group
```

When do users engage the most with Study Abroad content?

Hypotheses, intuitions, investigations:

- During enrollment?
- During their time abroad?

Goal:

- To determine the most effective time of the year to spend marketing efforts creating and publishing content.

Engagements include actions such as likes, comments, saves, and story replies.

```
In [60]: # keep only engagement data: engagements, saves, story replies, likes, comments
engagement_sa = sa_group.drop(labels=['net follower growth', 'followers gained',
                                     'followers lost', 'following', 'net following',
                                     'profile actions'], axis=1)
```

```
In [61]: # number of posts published per month over the last year in descending order

sa_post_count = engagement_sa.loc[:, ['published posts & stories']].sort_values('published posts & stories')
sa_post_count
```

Out[61]:

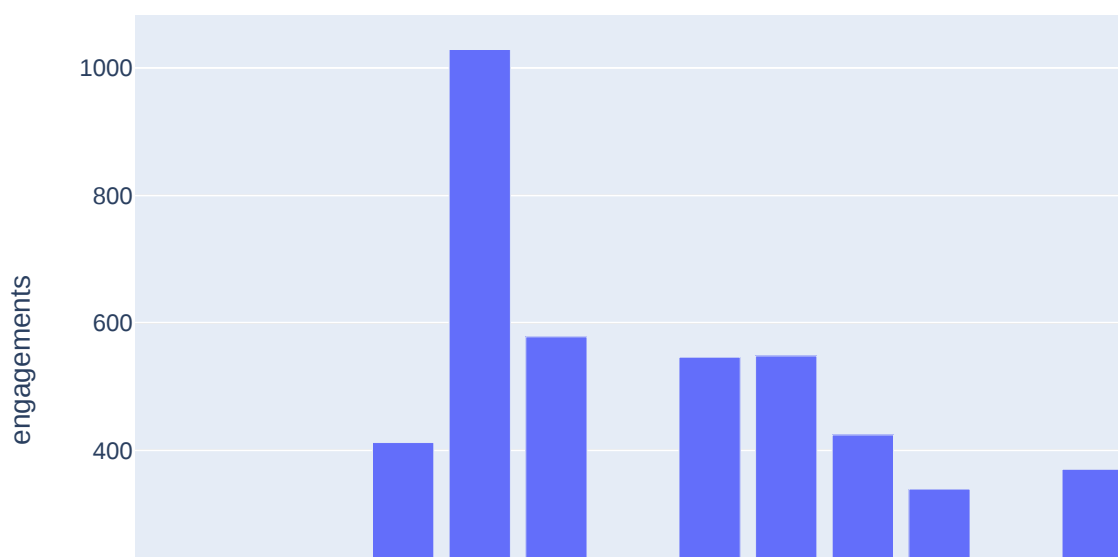
published posts & stories

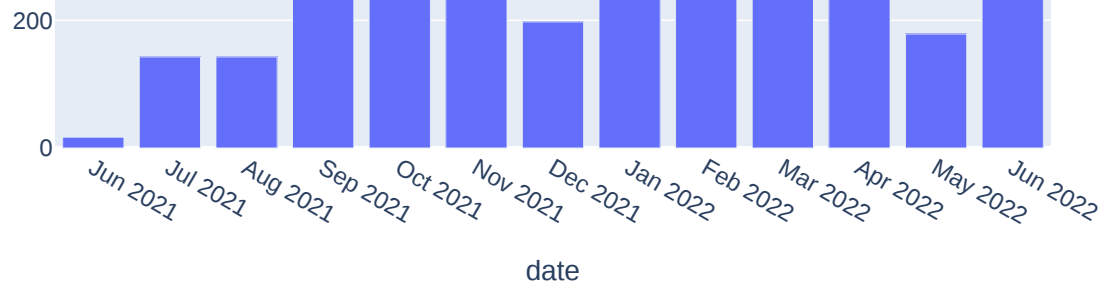
date	
Nov 2021	28
Feb 2022	28
Oct 2021	27
Dec 2021	19
Sep 2021	15
Mar 2022	15
Jan 2022	13
May 2022	10
Jun 2022	10
Apr 2022	9
Jul 2021	7
Aug 2021	6
Jun 2021	0

```
In [62]: # plot engagement vs month to see when users were most engaged in the past year
fig3 = px.bar(data_frame = engagement_sa, x = engagement_sa.index, y = 'engagements', title='Study Abroad: Total Engagements Per Month')
fig3.show()
```



Study Abroad: Total Engagements Per Month





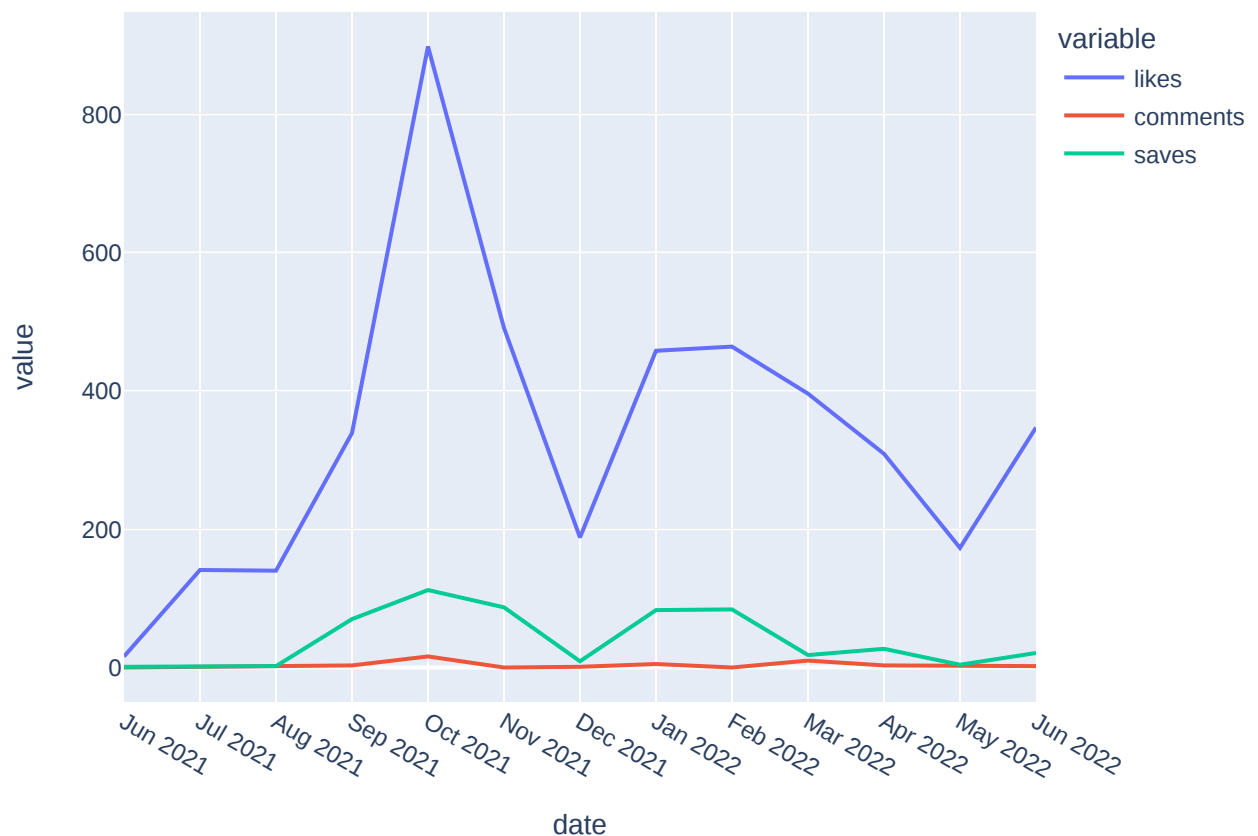
Commentary/Analysis

Engagement peaks in October - the Fall months when study abroad week takes place and we push large marketing campaigns. Engagement falls during winter where students are not engaging likely due to finals and winter break. Engagement remains consistent during the early spring term and dips in May at the end of the term.

```
In [63]: # break down engagement into parts
fig_4 = px.line(data_frame=engagement_sa, y=['likes', 'comments', 'saves'], x = engagement_sa['date'])
fig_4.show()
```



Study Abroad: Likes, Comments, & Saves per Month



Commentary/Analysis

During what time of the year does Study Abroad's following increase?

Hypotheses, intuitions, investigations:

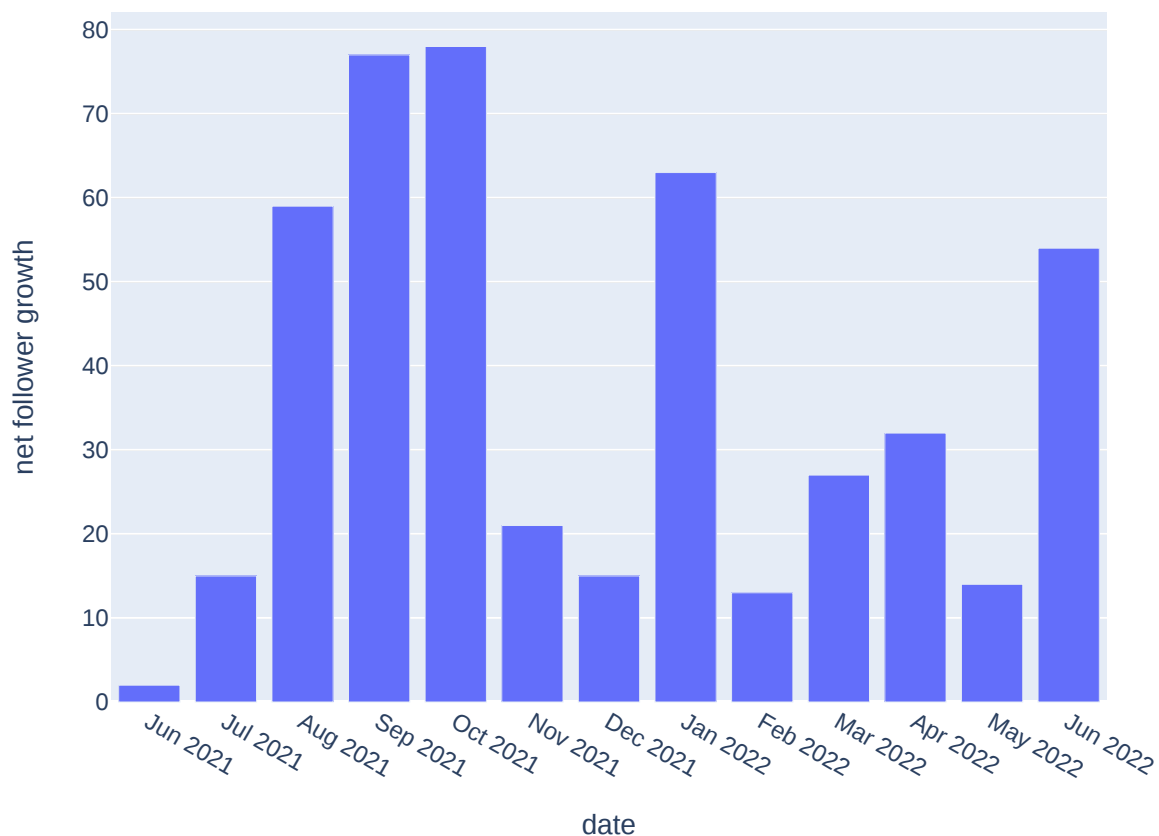
- Leading up to enrollment?
- During enrollment?

Goal: Using the analysis of follower trends to understand when users are looking for our information, subscribing, or gaining interest for the first time.

```
In [64]: # Net followers gained/lost per month
fig5 = px.bar(data_frame=sa_group, x=sa_group.index, y='net follower growth', title='Stu
fig5.show()
```



Study Abroad: Follower Growth Per Month



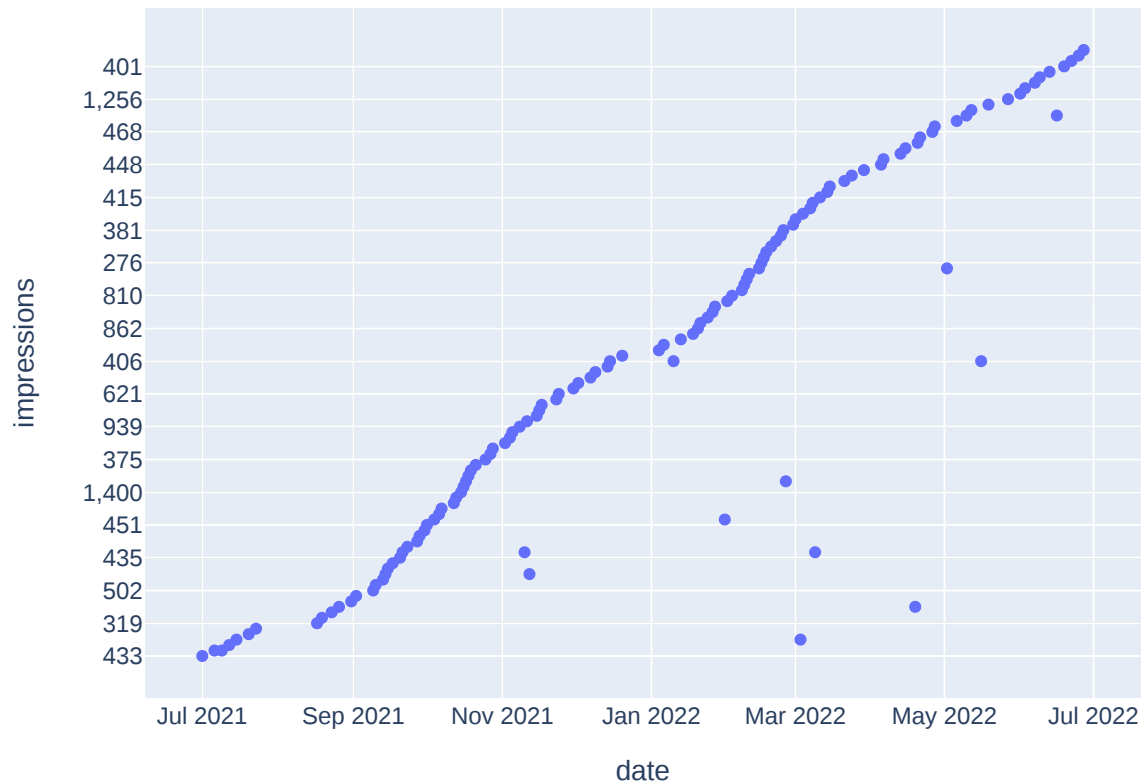
During what time of the year is Study Abroad's reach optimized?

Goal: Explore trends relating to impressions and reach to get a better sense of Study Abroad's outreach via IG.

```
In [69]: saimp = sa_df.loc[:, ['impressions', 'reach', 'published posts & stories']]
saimp = saimp[saimp['published posts & stories'] != 0]
saimp = saimp.reset_index()
saimp
px.scatter(data_frame=saimp, x='date', y='impressions', title="Study Abroad Post Impress
```



Study Abroad Post Impressions per Day



Findings:

Study Abroad's post impressions per day, similar to Summer Sessions, follows an upward trend. There are some outliers signifying low impressions. What happens in the content throughout the year to explain this trend?

Further Research

To further explore the nuances of the Summer Sessions and Study Abroad Instagram Account engagement trends, an analysis of content format and type is the next logical step. What content was relevant to months of high engagement? What kind of posts (story, post, reel?) and what type of content (travel, student life, informational?) led to the peaks and valleys we see in the graphs above?