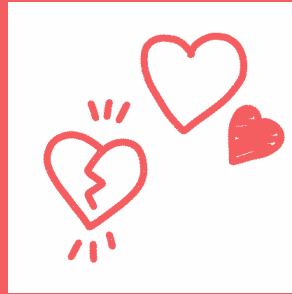


The Cycle of Love

The HeartBreakers



Meet The HeartBreakers

Ahmed Shah

Kaz Kazarinoff

Abeer Fatima

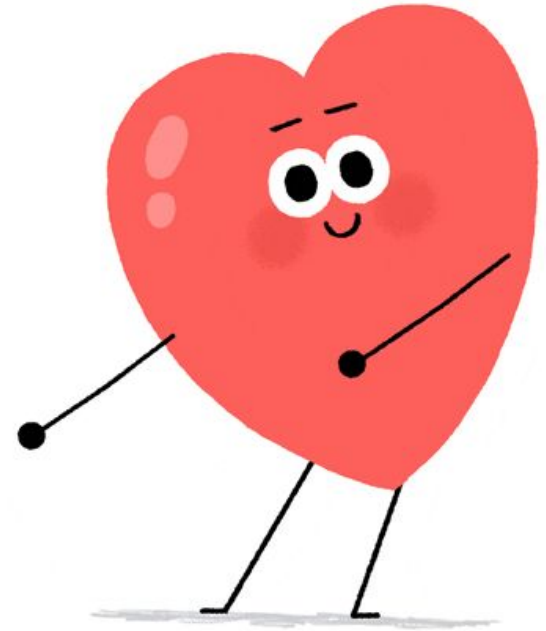
John Ray Alcano

Inspiration and Goals

Our project stemmed from observing dating patterns within our friend circle, and we could anecdotally sort these trends into two distinct sets.

- A mass ending of relationships at the beginning of semesters, dubbed “Heartbreak Row”
- A mass beginning of relationships during the months of October and November, dubbed “Cuffing Season”

The main aim of this project was to determine if this anecdotal evidence had any real substantial claim to reality.



YOU MAKE MY HEART FLOSS

Collecting Data

We conducted a unique study by collecting data firsthand through a couple of means:

- The flyer to the right being placed at high foot traffic locales around campus
- The link and QR code being spread around on various social media platforms.



Gathering data via Google Surveys

Cycle of Love

We have all heard the talk of cuffing season and heartbreak row—monikers for times of supposed mass entanglement events and the equivalent of the French revolution for hearts instead of heads, respectively—but do these names have any actual real life credibility? We have observed these “pseudo-cycles” within the confines of our own friend groups and would like to know if it is anything more than random coincidence.

Our main hypothesis is that there will be at least two distinct timeframes where both of these events happen at higher rates, which will most likely coincide with the beginning and end of the school year. Therefore, we are gathering and processing data to see if our hypothesis is actually true.

This is going to be an anonymous submission so please feel comfortable answering these questions. Thank you.

ahmed2000alishah@gmail.com [Switch account](#)

Not shared

* Indicates required question

What are you sharing info about? *

What are you sharing info about? *

☐ Breakup

☐ New Relationship/situationship

☐ Both

How old are you? *

☐ Below 18

☐ 18-19

☐ 19-20

☐ 21-22

☐ 23-24

☐ 25-28

☐ 29-33

Gathering data via Google Surveys

What month was your last breakup?

- ☐ N/A
- ☐ January
- ☐ February
- ☐ March
- ☐ April
- ☐ May

What month did you start your new relationship/situationship?

- ☐ Currently Single
- ☐ January
- ☐ February
- ☐ March
- ☐ April
- ☐ May

How long was your last relationship/situationship before it ended?

- ☐ N/A
- ☐ Less than 3 months
- ☐ 3 - 6 months
- ☐ 6 - 9 months
- ☐ 9 months to a 1.5 years
- ☐ 1.5 to 2.5 years
- ☐ 2.5 to 4 years
- ☐ 4+ years

How long have you been in the new relationship/situationship?

- ☐ N/A
- ☐ Less than 3 months
- ☐ 3 - 6 months

Preprocessing Data for Love

```
love_pruned.rename(columns = {'What month did you start your new relationship/situation?': 'relationship_start', 'How long was your last re
```

```
new_love_data = new_love_data.fillna(0)
```

```
new_love_data['relationship_start'] = new_love_data['relationship_start'].map(monthTranslator)  
new_love_data = new_love_data[new_love_data['relationship_start'] != 0]
```



Preprocessing Data for Love

```
def monthTranslator(toTranslate):  
    return monthTranslationTable[toTranslate]  
def timeTranslator(toTranslate):  
    return timeTranslationTable[toTranslate]  
def satisfactionTranslator(toTranslate):  
    return satisfactionTranslationTable[toTranslate]
```

```
monthTranslationTable = {'January': 1, 'February':2, 'March':3, 'April': 4, 'May':5, 'June':6, 'July':7, 'August':8, 'September':9, 'October':10, 'No
```



Preprocessing for Hate

```
new_hate_data = hate_pruned.rename(columns = {'What month was your last breakup?': 'breakup_start', 'How long was your last relationship/situationshi
```

```
new_hate_data = new_hate_data.fillna(0)
```

```
new_hate_data['breakup_start'] = new_hate_data['breakup_start'].map(monthTranslator)  
new_hate_data = new_hate_data[new_hate_data['breakup_start'] != 0]
```



Preprocessing for Hate

```
def monthTranslator(toTranslate):  
    return monthTranslationTable[toTranslate]  
def timeTranslator(toTranslate):  
    return timeTranslationTable[toTranslate]  
def satisfactionTranslator(toTranslate):  
    return satisfactionTranslationTable[toTranslate]
```

```
monthTranslationTable = {'January': 1, 'February': 2, 'March': 3, 'April': 4, 'May': 5, 'June': 6, 'July': 7, 'August': 8, 'September': 9, 'October': 10, 'No
```



Exploratory Data Analysis

```
hate_count = new_hate_data.shape[0]
love_count = new_love_data.shape[0]
total_count = survey_responses.shape[0]
unusableDFHate = new_hate_data.index.values.tolist()
unusableDFLove = new_love_data.index.values.tolist()
```

```
usable_list = list(set(unusableDFLove + unusableDFHate))
hate_percent = float(hate_count / total_count)
love_percent = float(love_count / total_count)
unusable_percent = float((total_count - len(usable_list)) / total_count)
```

Percent of respondees that provided valid data on their breakups is: 0.8452380952380952

Percent of respondees that provided valid data on their new relationships is: 0.5595238095238095

Percent of respondees that provided data that is impossible to parse without resurveying is: 0.15476190476190477



Statistical Distribution

```
#MODES
```

```
hate_start_mode = monthReverser(new_hate_data['breakup_start'].mode()[0])
```

```
def monthReverser(toReverse):  
    return reverseMonthTable[toReverse]  
def lengthReverser(toReverse):  
    return reverseTimeTable[toReverse]  
def satisfactionReverser(toReverse):  
    return reverseSatisfactionTable[toReverse]
```

```
#Reverse translator dicts
```

```
reverseMonthTable = {1:'January', 2:'February', 3:'March', 4:'April', 5:'May', 6:'June', 7:'July', 8:'August',9: 'September', 10:'October', 11:'  
reverseTimeTable = {1:'Less than 3 months', 2:'3 - 6 months',3:'6 - 9 months',4:'9 months to a 1.5 years',5:'1.5 to 2.5 years',6:'2.5 to 4 years  
reverseSatisfactionTable = {1:'Strongly Unsatisfied', 2:'Somewhat Unsatisfied', 3:'Neutral',4:'Somewhat Satisfied',5:'Strongly Satisfied'}
```

Statistical Distribution

```
#MEANS
```

```
hate_start_mean = new_hate_data['breakup_start'].mean()
```

Breakup Data:

Breakup Start-- Mode: August Mean: 7.070422535211268

Breakup Length-- Mode: 3 - 6 months Mean: 3.2535211267605635

Breakup Satisfaction-- Mode: Strongly Unsatisfied Mean: 2.6338028169014085

Relationship Data:

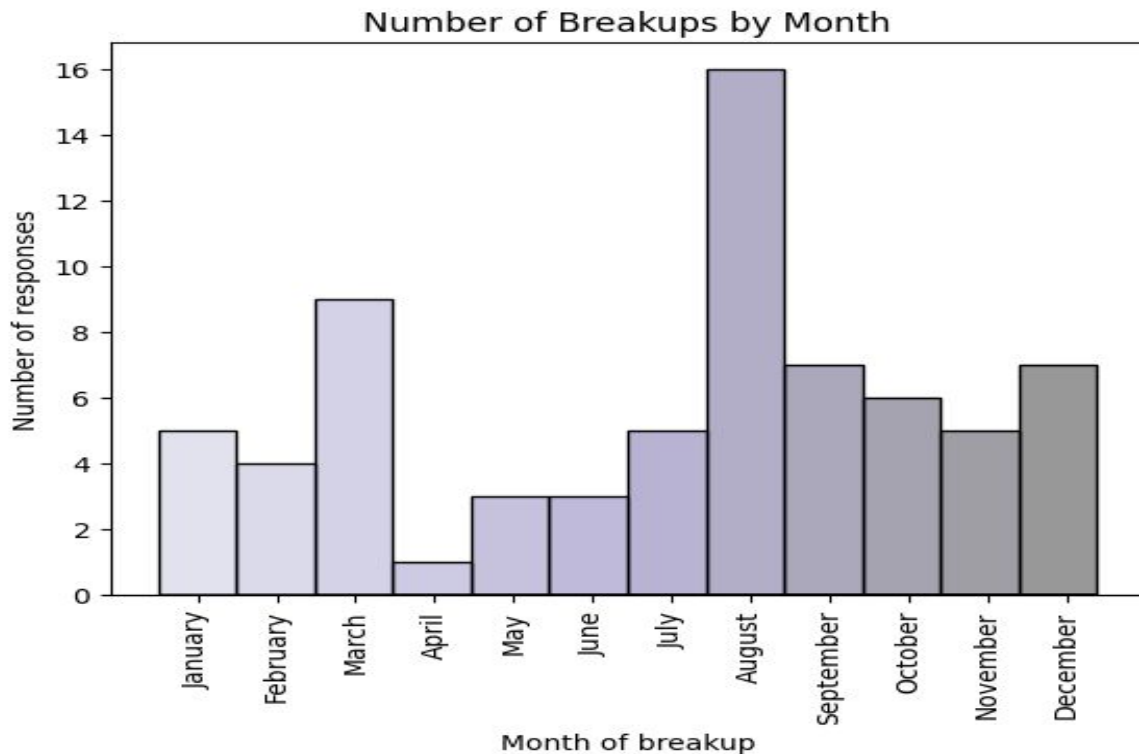
Relationship Start-- Mode: October Mean: 7.127659574468085

Relationship Length-- Mode: Less than 3 months Mean: 3.4680851063829787

Relationship Satisfaction-- Mode: Strongly Satisfied Mean: 4.361702127659575

```
axes = sns.histplot(data=new_hate_data, x='breakup_start', binwidth=0.99, binrange=(1, 12), hue='breakup_start', palette='Purples_d')
axes.legend([],[], frameon = False)
monthList = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December']
```

Visualization: Breakups per month



Possible Ramifications

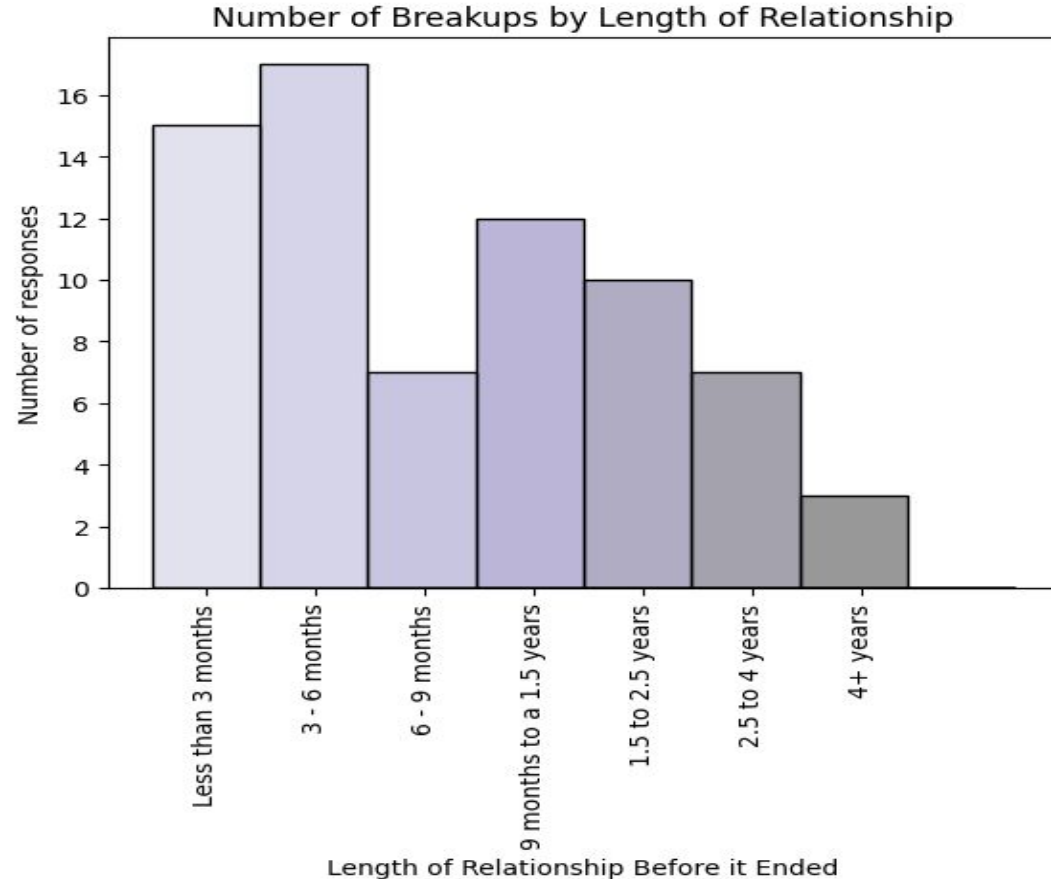
By far the single most common month to find yourself newly single is August.

This could be linked to the massive increase of new possible love interests as people meet new people during the start of Fall Semesters, which falls in line with the age distribution of our dataset

Long distance is not for the weak.


```
axes = sns.histplot(data = new_hate_data, x = 'breakup_length', binwidth = .99, binrange = (1,8), hue = 'breakup_length', palette = 'Purples_d')
axes.legend([],[], frameon = False)
axes.set_xticks([1.5,2.5,3.5,4.5,5.5,6.5,7.5])
axes.set_xticklabels(monthList)
axes.tick_params(axis='x', rotation=90)
```

Visualization: Breakups by Length of Relationship



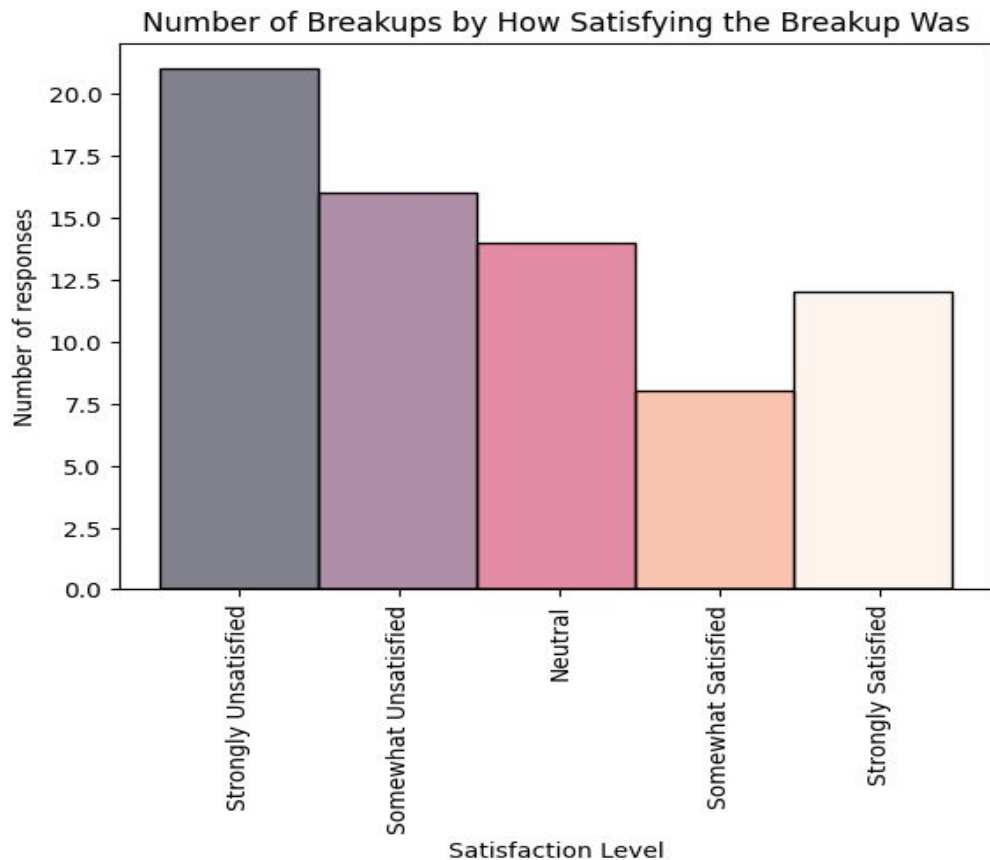
Possible Ramifications

Most relationships end within six months, which actually aligns with a known psychological trend called the feeling gap.

The six month line is known to separate infatuation from more long term affection, so if you can make it past this mark your relationship is much more likely to last.

```
axes = sns.histplot(data = new_hate_data, x = 'breakup_length', binwidth = .99, binrange = (1,8), hue = 'breakup_length', palette = 'Purples_d')
axes.legend([],[], frameon = False)
axes.set_xticks([1.5,2.5,3.5,4.5,5.5,6.5,7.5])
```

Visualization: Breakup by Satisfaction



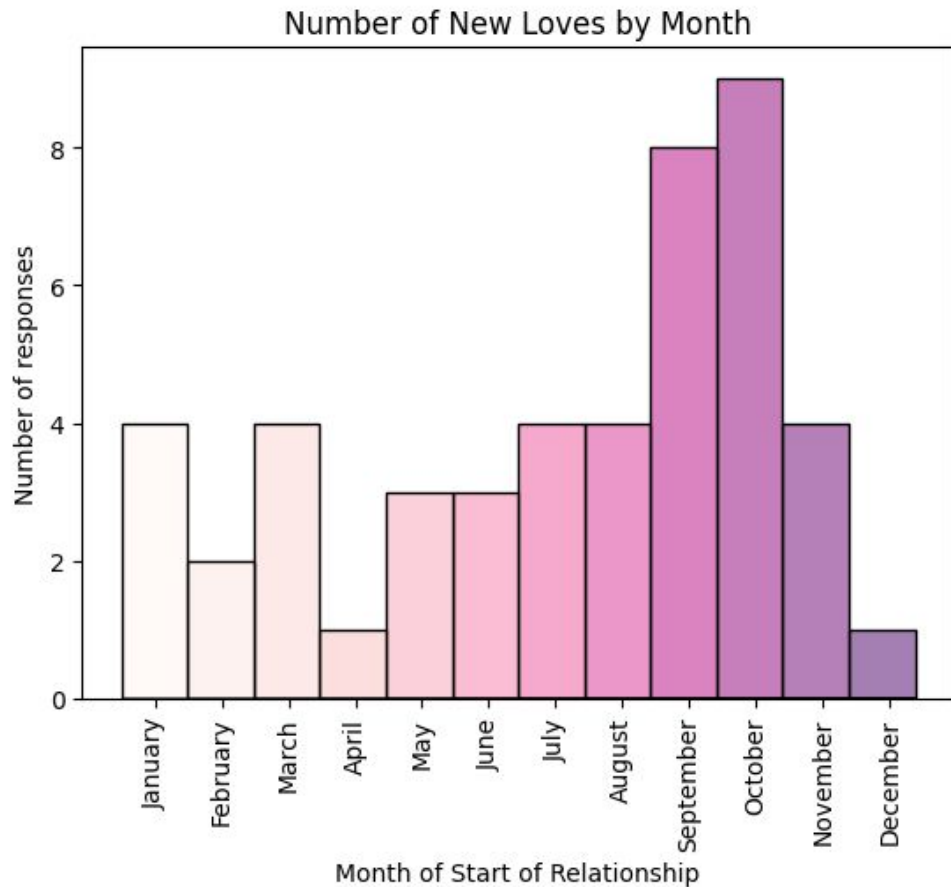
Possible Ramifications

People are generally unsatisfied with breakups.

However, this is a much, much more uniform distribution than what we will see for satisfaction in new relationships.

```
kes = sns.histplot(data = new_love_data, x = 'relationship_start', binwidth = .99, binrange = (1,12), hue = 'relationship_start', palette = 'RdPu')
kes.legend([],[], frameon = False)
kes.set_xticks([1.5,2.5,3.5,4.5,5.5,6.5,7.5,8.5,9.5,10.5,11.5,12.5])
kes.set_xticklabels(monthList)
kes.tick_params(axis='x', rotation=90)
```

Visualization: New Loves by Month



Possible Ramifications

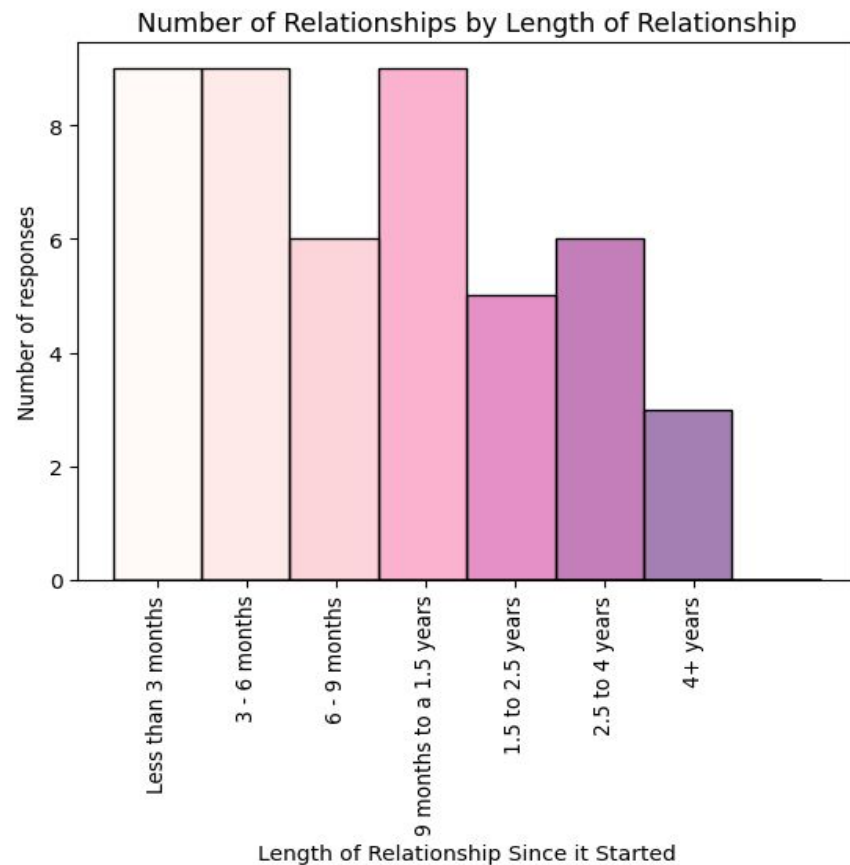
If you recall the most popular month to ditch your current boo, August, the most popular months for getting together being September and October would make a lot of sense.

The dating pool is massively inflated every August and naturally returns to a more balanced state just a bit later in the year.

Bundle up with a new boo?

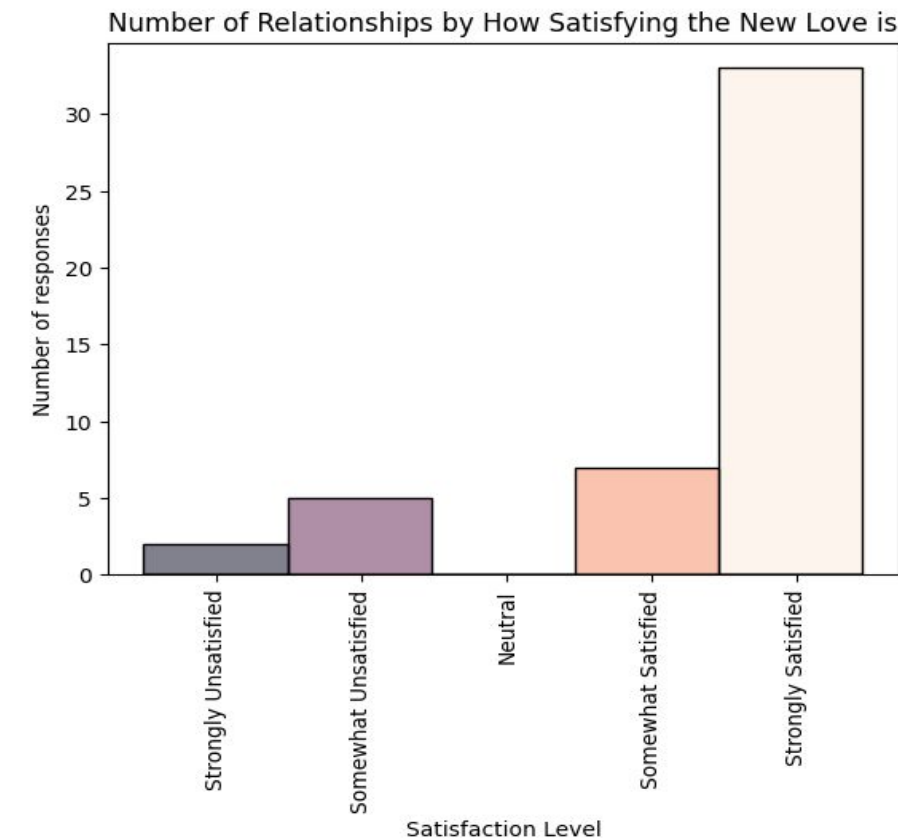
```
axes = sns.histplot(data = new_love_data, x = 'relationship_length', binwidth = .99, binrange = (1,8), hue = 'relationship_length', palette = 'RdPu')
axes.legend([],[], frameon = False)
monthList = ['Less than 3 months', '3 - 6 months','6 - 9 months','9 months to a 1.5 years','1.5 to 2.5 years','2.5 to 4 years','4+ years']
axes.set_xticks([1.5,2.5,3.5,4.5,5.5,6.5,7.5])
axes.set_xticklabels(monthList)
axes.tick_params(axis='x', rotation=90)
```

Visualization: New Loves by Relationship Length



```
axes = sns.histplot(data = new_love_data, x = 'relationship_satisfaction', binwidth = .99, binrange = (1,5), hue = 'relationship_satisfaction', palette = 'magma', axes.legend([],[], frameon = False))
axes.set_xticks([1.5,2.5,3.5,4.5,5.5])
axes.set_xticklabels(satisfaction_list)
axes.tick_params(axis='x', rotation=90)
```

Visualization: Satisfaction Within New Love



Outlier or Personal Bias?

Turns out people are overwhelmingly more likely to report that their new relationship is going just swell, even though almost 50% of relationships are doomed to die in the first 6 months from our data.

This might be just because we tend to be a bit more loyal to our significant others until problems have proven to be too much to hand

Additionally, absolutely no one we surveyed was neutral about their new relationship status, which is bewildering from a statistical point of view but maybe less so from a relationship perspective.

ML Preprocessing

The issue at first is that our two datasets that we would like to use as labels currently have mismatching columns, so we need to manage that first.

- 1) Create uniform labels across both datasets and add the label column to both
- 2) Combine the datasets to train_test_split across.

```
total_data = love_data_uniform + hate_data_uniform  
total_train, total_test = sklearn.model_selection.train_test_split(total_data, test_size = .8, train_size = .2)
```

Using the Prediction model

```
learn_classifier(total_train_pruned, total_train_labels, poly)
```

```
accuracy = accuracy_score(total_test_binary, returned_predictions)
```

```
Accuracy score is .94238369564136258
```

Future of Heartbreakers

We can just keep rerunning the same jupyter file as more data comes in, so collecting additional data is a way to endlessly improve our rationale.

We also hope to increase the amount of different metrics we are able to predict based upon the age of the person surveyed through our Machine Learning Model.

More visualizations–The most fun part of data science!



Any Questions?

