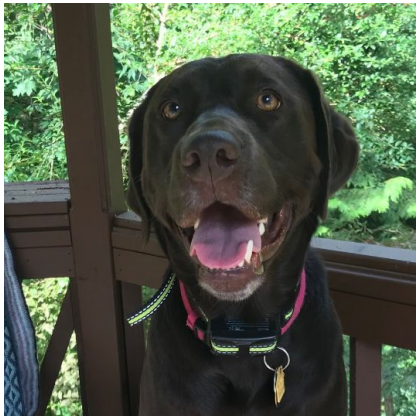
The correlation between COVID-19 cases and the number of pets being adopted.

Group Members:
Jen Johansson
Mabel Alamu
JB Kinlacheeny
Radhika Sagar

“At the beginning of the pandemic, people across the country were trying everything to beat the quarantine blues. Flour and eggs flew off the shelves for baking, workout equipment was out of stock for miles and art supplies orders were backed up for weeks. But who knew the most in demand quarantine find was a new furry friend?” (Jackson,2020).



Why Covid-19 and Pet Adoption Rates?

Our team looked at available COVID-19 data and cross-referenced it with publicly available data related to pet adoptions to determine if there is an increase in pet adoptions as the pandemic continues.

Null Hypothesis:

There is no correlation between COVID positive cases and pet adoption rates.

Alternative Hypothesis:

There is a positive correlation between adoption rates in 2020 (after March) due to more individuals staying at home (as indicated by COVID cases).

Defining Research Questions

1. What is the baseline adoption rate by month/year?

2. Is there an increase in adoption rates after the pandemic hit the US (ie after March 2020)?

3. If there is a correlation to adoptions, does it correlate to documented COVID cases?

4. Extend the analysis to multiple cities. Do correlations to COVID differ between cities?

Norfolk, VA Data Sources:

- Norfolk,VA COVID-19 Data (pulled API data source)

	report_date	fips	locality	vdh_health_district	total_cases	hospitalizations	deaths
0	2020-12-09T00:00:00.000	51710	Norfolk	Norfolk	7106	490	98
1	2020-12-08T00:00:00.000	51710	Norfolk	Norfolk	7038	483	98
2	2020-12-07T00:00:00.000	51710	Norfolk	Norfolk	6894	475	93
3	2020-12-06T00:00:00.000	51710	Norfolk	Norfolk	6783	476	93

Norfolk COVID-19 Data: 271 rows × 7 columns

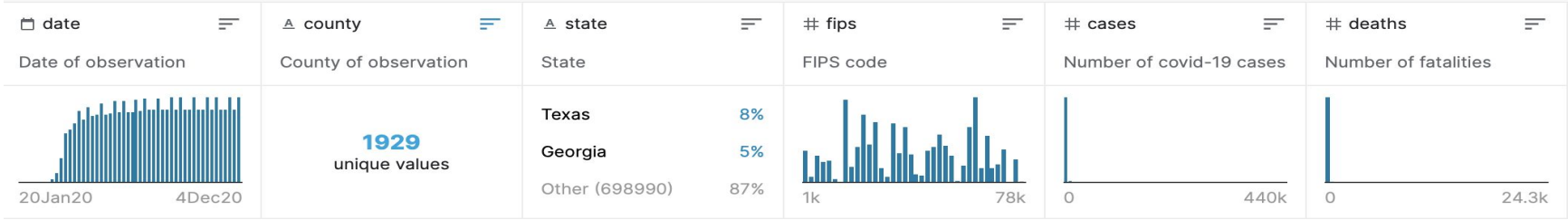
Norfolk Pet Adoption Data: 10842 rows × 13 columns

- Norfolk, VA Pet Adoption Data (pulled .csv data source)

	Animal ID	Animal Type	Sex	Years Old	Months Old	Primary Breed	Primary Color	Intake Type	Intake Date	Outcome Type	Outcome Date	Outcome Month	Outcome Year
0	A000455	Dog	Neutered Male	4.0	0.0	POODLE MIN	WHITE	Return	2017-08-16	Adoption	2017-08-19	08	2017
1	A000455	Dog	Neutered Male	4.0	0.0	POODLE MIN	WHITE	Owner Surrendered	2017-05-03	Adoption	2017-05-13	05	2017
2	A000601	Cat	Spayed Female	0.0	1.8	DOMESTIC SH	CALICO	Transfer	2016-08-20	Adoption	2016-10-02	10	2016

Dallas, Texas Data Sources

- Dallas, Texas COVID-19 Data (pulled .csv data source from Kaggle)



- Dallas, Texas Pet Adoption Data (pulled .csv data source)

	Animal Type	Animal Breed	Outcome Type	Outcome Date
0	DOG	HAVANESE	RETURNED TO OWNER	11/09/2019
1	DOG	TERRIER MIX	DEAD ON ARRIVAL	11/10/2019
2	DOG	CATAHOULA	RETURNED TO OWNER	10/03/2019
3	DOG	GERM SHEPHERD	TRANSFER	10/15/2019

Dallas COVID-19 Data:
271 rows (per day)

Dallas Pet Adoption Data:
10087 rows × 6 columns

Overall Approach

Clean Up Process:

- Compiled the data
- Import and read CSV
- GET requests to retrieve COVID data
- Merge CSV with our API data
- Change any NaN values to Zero
- Converting our dates to pandas date format
- Used plotting and statistics functions

```
In [65]: Dallas_by_month = pd.DataFrame({
             "2019": Dallas_2019_adoptions,
             "2020": Dallas_2020_adoptions
         })
         #Dallas_by_month= Dallas_by_month.fillna(0)
         Dallas_by_month
```

Out[65]:

	2019	2020
01	NaN	1314.0
02	NaN	1140.0
03	NaN	922.0

```
In [66]: Dallas_by_month = pd.DataFrame({
             "2019": Dallas_2019_adoptions,
             "2020": Dallas_2020_adoptions
         })
         Dallas_by_month= Dallas_by_month.fillna(0)
         Dallas_by_month
```

Out[66]:

	2019	2020
01	0.0	1314.0
02	0.0	1140.0
03	0.0	922.0

Data Cleanup & Exploration

For each dataset, we had to pull apart the date field and make new columns for Outcome Year and Outcome Month

```
1 Dallas_COVID['date'] = pd.to_datetime(Dallas_COVID['date'],format='%Y/%m/%d')
2 Dallas_COVID['Month'] = Dallas_COVID['date'].map(lambda x: x.strftime('%m'))
3 Dallas_COVID['Year'] = Dallas_COVID['date'].map(lambda x: x.strftime('%Y'))
4 Dallas_COVID.head()
```

	date	county	state	fips	cases	deaths	Month	Year
1133	2020-03-10	Dallas	Texas	48113.0	2	0.0	03	2020
1329	2020-03-11	Dallas	Texas	48113.0	3	0.0	03	2020

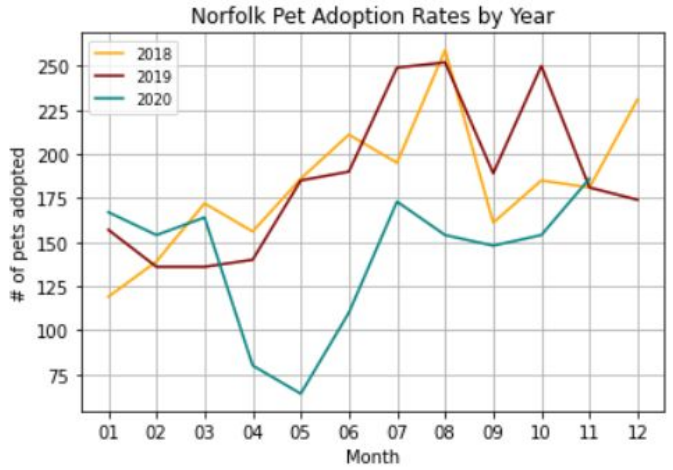
Data Analysis - Pet Adoption Rates - Norfolk, VA

Q: What is the baseline adoption rate by month/year? Is there an increase in adoption rates after the pandemic hit the US (ie after March 2020)?

In [211]:
```
#saving years 2018 through 2020 as a df
adoption_by_month = pd.DataFrame({
    "2018": year1_adoptions,
    "2019": year2_adoptions,
    "2020": year3_adoptions
})
adoption_by_month= adoption_by_month.fillna(0)
adoption_by_month
```

Out[211]:

	2018	2019	2020
01	119	157	167.0
02	139	136	154.0
03	172	136	164.0
04	156	140	80.0



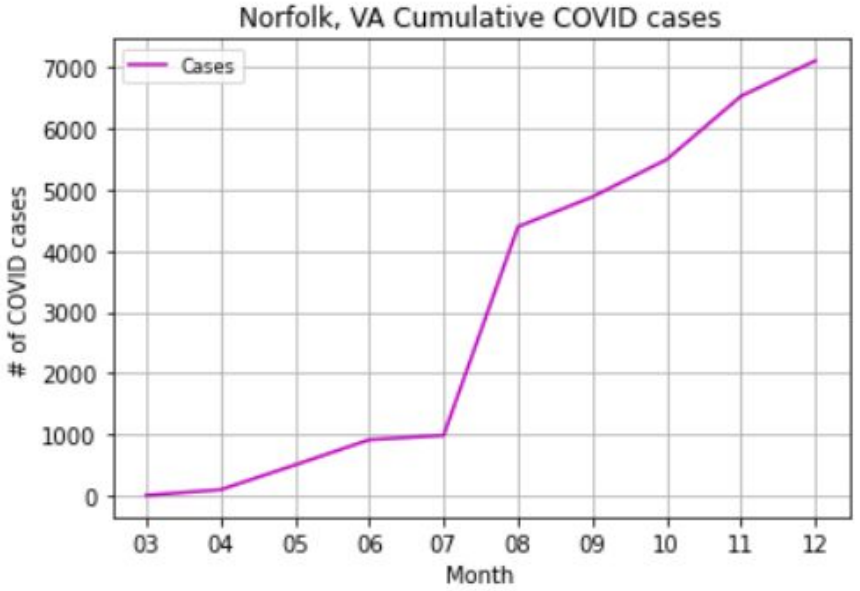
We notice that adoption rates are lower in 2020 compared to previous years.

Data Analysis - COVID cases -Norfolk, VA

Plotted cumulative COVID cases by month

Out[95]:

Outcome Month	# of Cases
03	8
04	99
05	503
06	915
07	987
08	4394
09	4881
10	5493
11	6528
12	7106



Data Analysis - COVID cases - Norfolk, VA

Realized we should plot and compare **deltas** in cases, not cumulative, using the diff() function

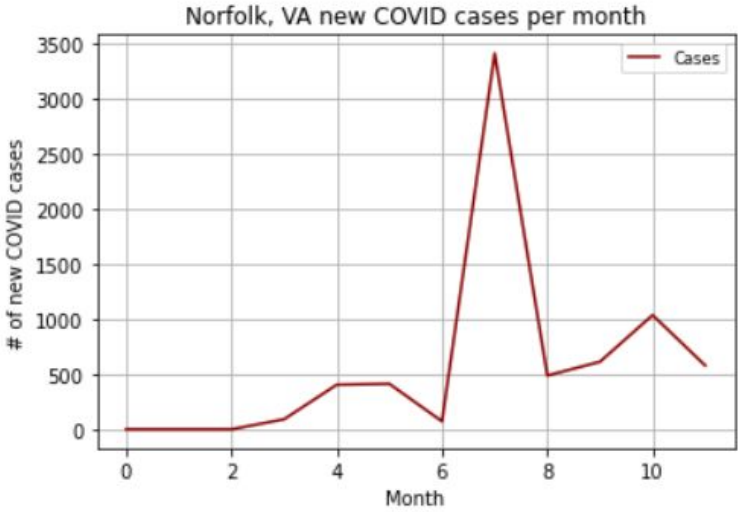
Plotting the delta number of cases month over month

In [97]:

```
1 COVID_by_month['# of Cases'] = COVID_by_month['# of Cases
2
3 COVID_diff = COVID_by_month['# of Cases'].diff().fillna(0
4
5 COVID_diff
6
7 COVID_per_month = pd.DataFrame({"# of Cases": COVID_diff,
8                                 "Month": COVID_diff.index})
9 COVID_per_month
```

Out[97]:

Outcome Month	# of Cases	Month
03	0.0	03
04	91.0	04
05	404.0	05
06	412.0	06



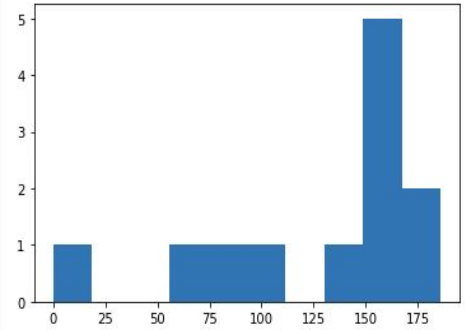
Data Analysis - analyzing for normal distribution - Norfolk

In [112]:
```
population1= adoption_by_month['2020'].mean()
population2= COVID_per_month['# of Cases'].mean()

population1_std= np.std(adoption_by_month['2020'])
population2_std= np.std(COVID_per_month['# of Cases'])

population1_std, population2_std
```

Out[112]: (53.005502858979966, 902.4332385279257)

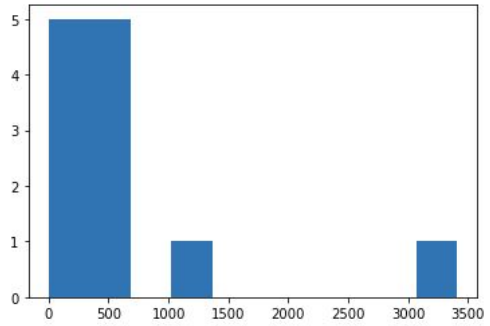
In [115]:
```
plt.hist(adoption_by_month['2020'])
print(st.normaltest(adoption_by_month['2020']))
```

```
NormaltestResult(statistic=6.020364134826666, pvalue=0.04928270516893548)
```



In [116]:
```
plt.hist(COVID_per_month['# of Cases'])
print(st.normaltest(COVID_per_month['# of Cases']))
```

```
NormaltestResult(statistic=25.367620197303985, pvalue=3.1009250162270873e
-06)
```



Data Analysis - T-test - Norfolk

Independent t-test: ○ Also known as two sample t-test , determines whether there is a statistically significant difference between the means in two unrelated groups. Assumptions include:

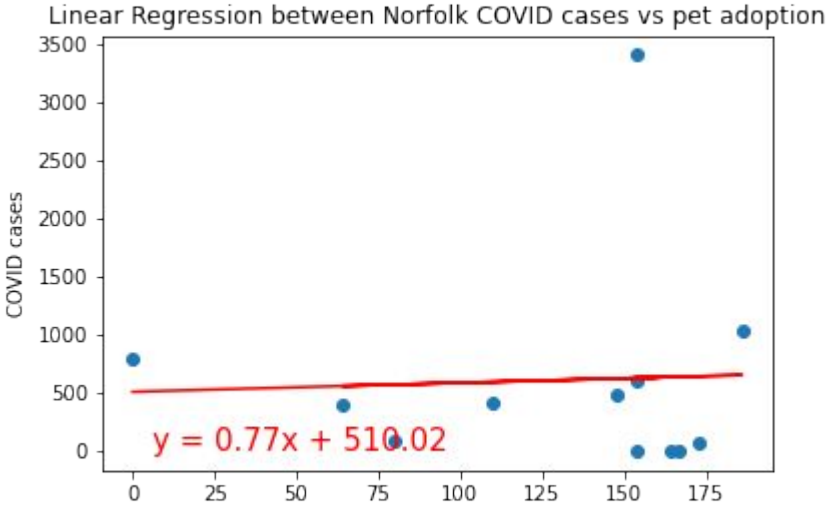
- Data is normally distributed
- Data is independent
- Data is homogenous (the standard deviations are roughly equal)

```
In [106]: 1 st.ttest_ind(adoption_by_month['2020'],COVID_per_month['# of Cases'],equal_var= False)

Out[106]: Ttest_indResult(statistic=-1.6950221554424156, pvalue=0.11796733678932625)
```

With the p-value of 11.7%, we fail to reject the null hypothesis. Hence, there is no correlation between COVID positive cases and pet adoption rates in Norfolk.

Data Analysis - Adoptions vs COVID cases ... any correlation?

Q: If there is a correlation to adoptions, does it correlate to documented COVID cases?



Absolute Value of r	Strength of Correlation
$r < 0.3$	None or very weak
$0.3 \leq r < 0.5$	Weak
$0.5 \leq r < 0.7$	Moderate
$r \geq 0.7$	Strong

```
In [61]:  1 orfolk {round(st.pearsonr(adoption_by_month['2020'],COVID_per_month['# of Cases'])[0],2)}")
```

```
The correlation coefficient between pet adoptions and COVID cases in Norfolk 0.05
```

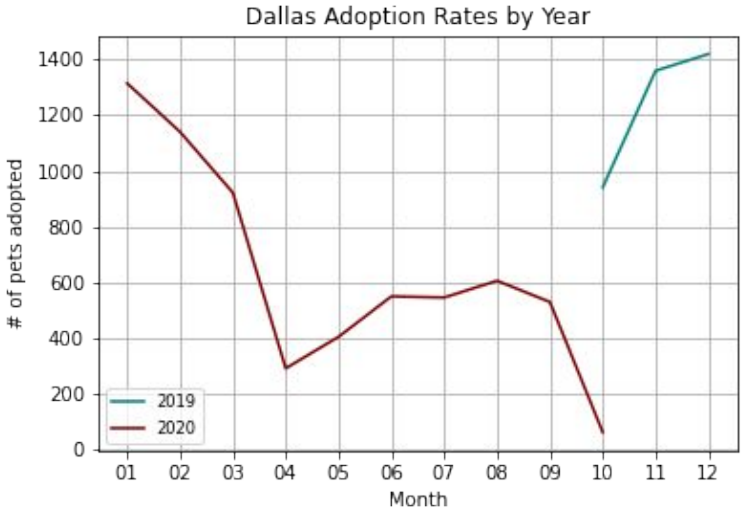
Data Analysis - Pet Adoption Rates - Dallas

Q: Extend the analysis to multiple cities. Do correlations to COVID differ between cities?

In [53]:
```
#Saving adoption data for 2019 and 2020 into a df
Dallas_by_month = pd.DataFrame({
    "2019": Dallas_2019_adoptions,
    "2020": Dallas_2020_adoptions
})

Dallas_by_month= Dallas_by_month.fillna(0)
Dallas_by_month
```

Out[53]:

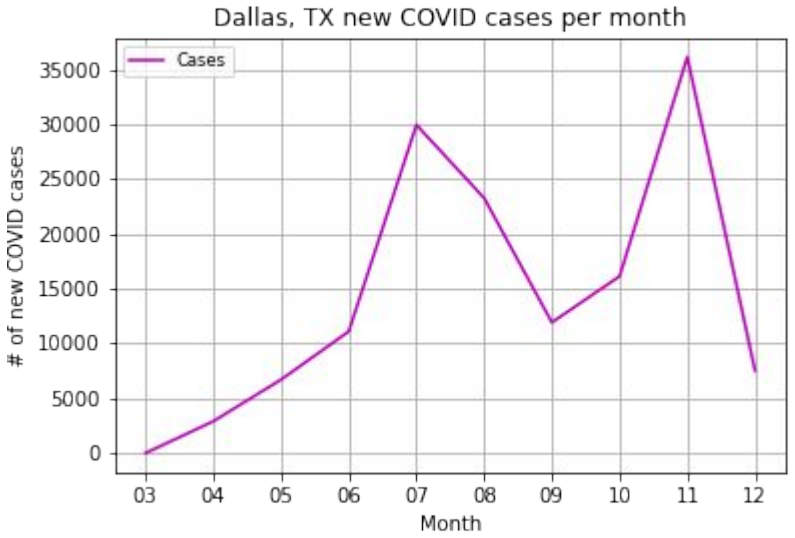
	2019	2020
01	0.0	1314.0
02	0.0	1140.0



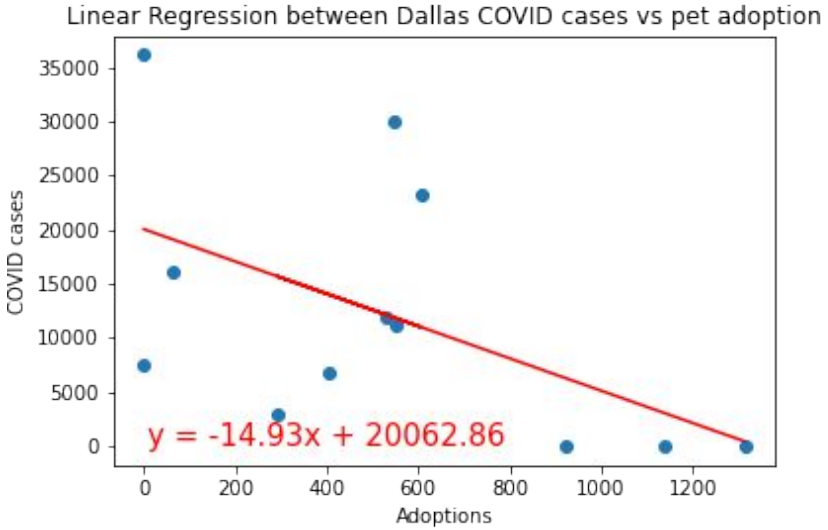
Data Analysis - COVID cases -Dallas, Texas

Out[46]:

Month	cases	Month
03	0.0	03
04	2900.0	04
05	6703.0	05
06	11104.0	06
07	29961.0	07
08	23277.0	08
09	11924.0	09
10	16142.0	10
11	36156.0	11
12	7522.0	12



Data Analysis - COVID cases vs Pet Adoption - Dallas, TX



Absolute Value of r	Strength of Correlation
r < 0.3	None or very weak
0.3 ≤ r < 0.5	Weak
0.5 ≤ r < 0.7	Moderate
r ≥ 0.7	Strong

```
1 print(f"The correlation coefficient between pet adoptions and COVID cases in Dallas {round(
```

```
The correlation coefficient between pet adoptions and COVID cases in Dallas -0.53
```

Data Analysis - T-test - Dallas

```
In [185]: 1 st.ttest_ind(Dallas_by_month['2020'],Dallas_per_month['cases'],equal_var= False)

Out[185]: Ttest_indResult(statistic=-3.3182674870712097, pvalue=0.006830329816449098)
```

With the p-value of 0.68%, we reject the null hypothesis. Hence, there is a correlation between COVID positive cases and pet adoption rates in Dallas.

Summary/Conclusion

- Adoption rates appear to increase over the summer months in Norfolk. There was a big dip in adoptions in April/May/June 2020, right after the pandemic started.
- For Norfolk, we did not find a correlation between COVID cases and pet adoptions.
    a. The Pearson correlation coefficient was .05 (very weak)
- For Dallas, we did find a **moderate** correlation but it was negative.
    a. The Pearson correlation coefficient was -.5
- Inferences/Conclusions:
    a. The adoption dataset for both cities is small, which impacts the analysis and test results.
    b. Other factors could be influencing the results. For example, were the shelters closed during any time period? Were people adopting pets through other sources like rescue organizations?

Post Mortem

- Discuss any difficulties that arose, and how you dealt with them
    a. Locating datasets that were available via API, and were large enough to show conclusive results. The adoption data available is by city not state.
    b. Finding the raw data for COVID. There is a lot of analysis available but we were looking for raw data.
- Discuss any additional questions that came up, but which you didn't have time to answer: What would you research next, if you had two more weeks?
    a. Look for ways to enlarge the adoption dataset by including rescue organizations and other cities.
    b. Other factors influencing pet adoption rates (ie increased unemployment)
    c. Were fewer pets given up to shelters during the COVID pandemic? Did people keep their pets (and why: more time to care, or are pets more of a disturbance while working from home)

Questions

- Open-floor Q&A with the audience