Twitter Sentiment Analysis to Predict Stock Movements

Group Members:
Andrew Smith-Mui
Olivia Greenberg
Paul Kren
Walid Nofal

KEY QUESTIONS

(1) Can tweets impact share prices?
(2) Can sentiment anticipate a shift in a stock's price trajectory?

MOTIVATION

- Common interest in markets and trading
- Intent to use new knowledge of Python to take advantage of open source data
- Shared curiosity if any causal link could be identified

HYPOTHESIS

By scraping relevant tweets, sentiment analysis will be an indicator of a stock's short term market movements.

Financial Markets, FinTech, & Findings

TRENDS:

Markets are changing at an ever increasing pace.

Old methods of trading using “gut instinct” have given way to a new breed of traders.

The ability to use open source data may allow for investors to identify unique ways to outperform the market.

IN REVIEW:

-The process was far more difficult and less straightforward than anticipated.

-Collecting, cleaning, and manipulating the data proved to be the most challenging aspect.

-We were able to successfully gather and analyze twitter data as well as intraday stock price movements.

-The extremely limited data set we had to work from did not adequately help to explain or predict stock price movements.

Our Process

Selecting the Stocks

We focused on airline stocks because we thought airlines would be a key source for emotional/passionate twitter postings.

Selecting the Tweets

We pulled two days worth of real time tweets - focusing on keywords that represent each specific airline, we were able to pull an average of 3000 -3500 tweets per airline.

Jupyter Notebook - Python Libraries

```
import pandas as pd
import os
import dotenv
import tweepy
from tweepy import OAuthHandler
from textblob import TextBlob
from tweepy import Stream
from tweepy.streaming import StreamListener
import time
```

```
import pandas as pd
import numpy as np
import datetime as dt
from pathlib import Path
import seaborn as sns
%matplotlib inline
from textblob import TextBlob
```

Twitter API

- Requires setting up a developer account
- Must set up an app to obtain access tokens
- Allows for querying new tweets based on keywords
- API Key, API Secret Key, Access Token, and Access Token Secret needed for requests
- Standard requests are limited

New Python Libraries & Twitter API

Tweepy:

- Python library for accessing twitter's API

Textblob:

- Python library for simplified natural language processing (NLP)
- Used for determining sentiment values from tweets (Polarity & Subjectivity)
- Polarity is assigned a value equal to or between -1 and 1
  - -1 = most negative
  - 0 = neutral
  - 1 = most positive
- Subjectivity is assigned a value between 0 and 1
  - The closer to 0, the more objective the text is (based more on factual information)
  - The closer to 1, the more subjective the text is (based more on emotions/feelings/opinions i.e. sentiment)

Jupyter Notebook - Data Cleanup

```
count = 18000


def iterate_scrape():

    print(time.ctime())

    text_query = "United Airlines"
    scrape_tweets(login_to_twitter(), text_query, count)

    text_query = "UAL"
    scrape_tweets(login_to_twitter(), text_query, count)
```

```


 while True:

     iterate_scrape()

     time.sleep(7200)

```

```
def login_to_twitter():

    dotenv.load_dotenv()

    consumer_key = os.getenv("twitter_api_key")
    consumer_secret = os.getenv("twitter_api_secret_key")
    access_token = os.getenv("twitter_access_token")
    access_token_secret = os.getenv("twitter_access_token_secret")

    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_token, access_token_secret)

    api = tweepy.API(auth,wait_on_rate_limit=True)

    return api
```

Jupyter Notebook - Sentiment Analysis

```
#Conversion of tweets to srings

airline_tweets['Text'] = airline_tweets['Text'].astype(str)

airline_tweets[['polarity', 'subjectivity']] = airline_tweets['Text'].apply(lambda Text: pd.Series(TextBlob(Text).sentiment))
```

```
airline_tweets.head()
```

Date	Number	Location	Location Coordinates	TweetID	Text	Airline	polarity	subjectivity
2020-08-11 19:18:26	99	Whitney, Texas USA	NaN	1.293266e+18	RT @RossanaWyatt: Good idea, and good to know ...	Alaska Airlines	0.175000	0.533333
2020-08-11 19:23:38	98	Los Angeles, CA	NaN	1.293267e+18	Alaska Airlines flight #ASA278 spotted at 2,02...	Alaska Airlines	0.000000	0.000000
2020-08-11 19:24:10	97	Los Angeles, CA	NaN	1.293267e+18	Alaska Airlines flight #ASA658 spotted at 2,72...	Alaska Airlines	0.000000	0.000000
2020-08-11 19:28:08	96	NaN	NaN	1.293268e+18	Alaska airlines has BOGO free plane tickets ri...	Alaska Airlines	0.495238	0.745238
2020-08-11 19:28:47	95	Santa Barbara, CA	NaN	1.293268e+18	RT @glaciermt: Beginning in March 2021, Alaska...	Alaska Airlines	0.000000	0.000000

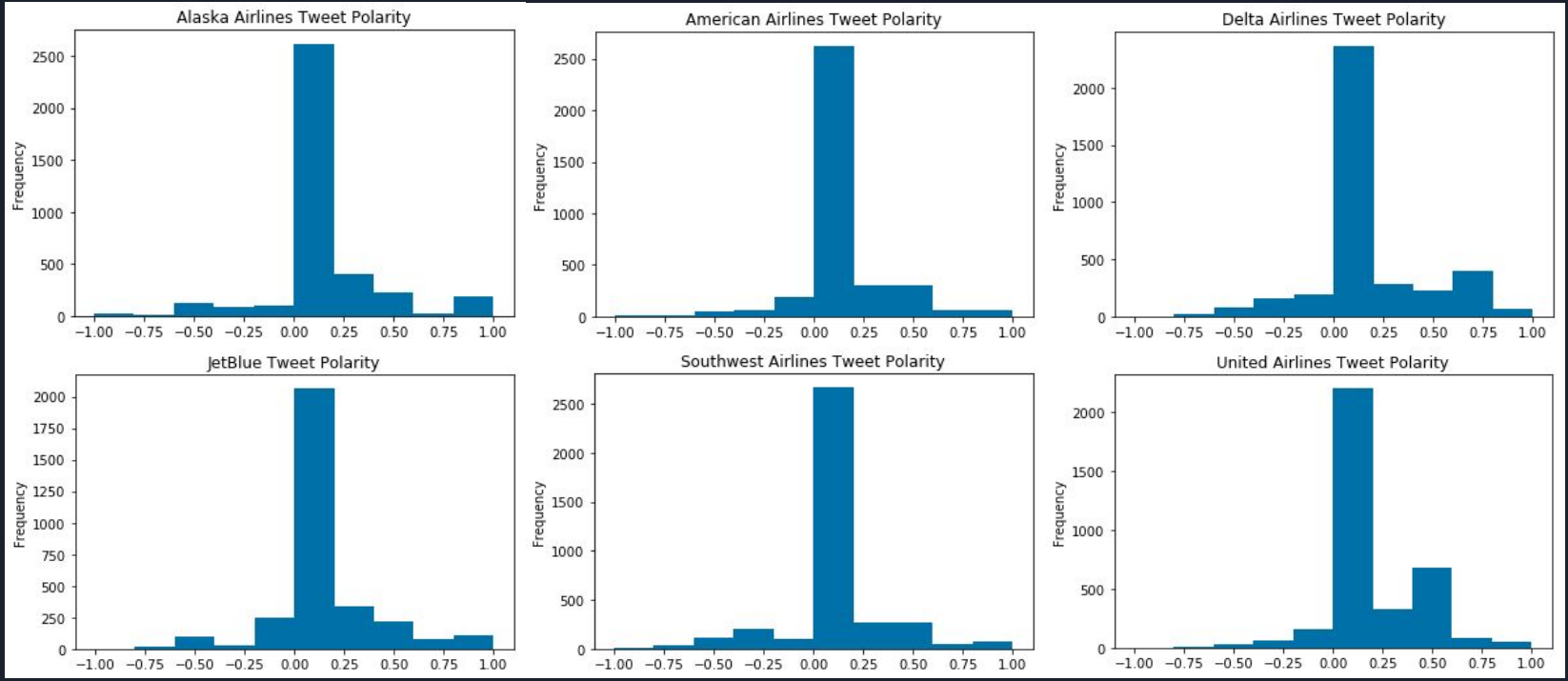
Jupyter Notebook - Plots

```
american_polarity = airline_tweets[airline_tweets['Airline']=='American Airlines']
american_polarity_hist = american_polarity['polarity'].plot(kind='hist',
                                                            bins=10,
                                                            title='American Airlines Tweet Polarity')
```

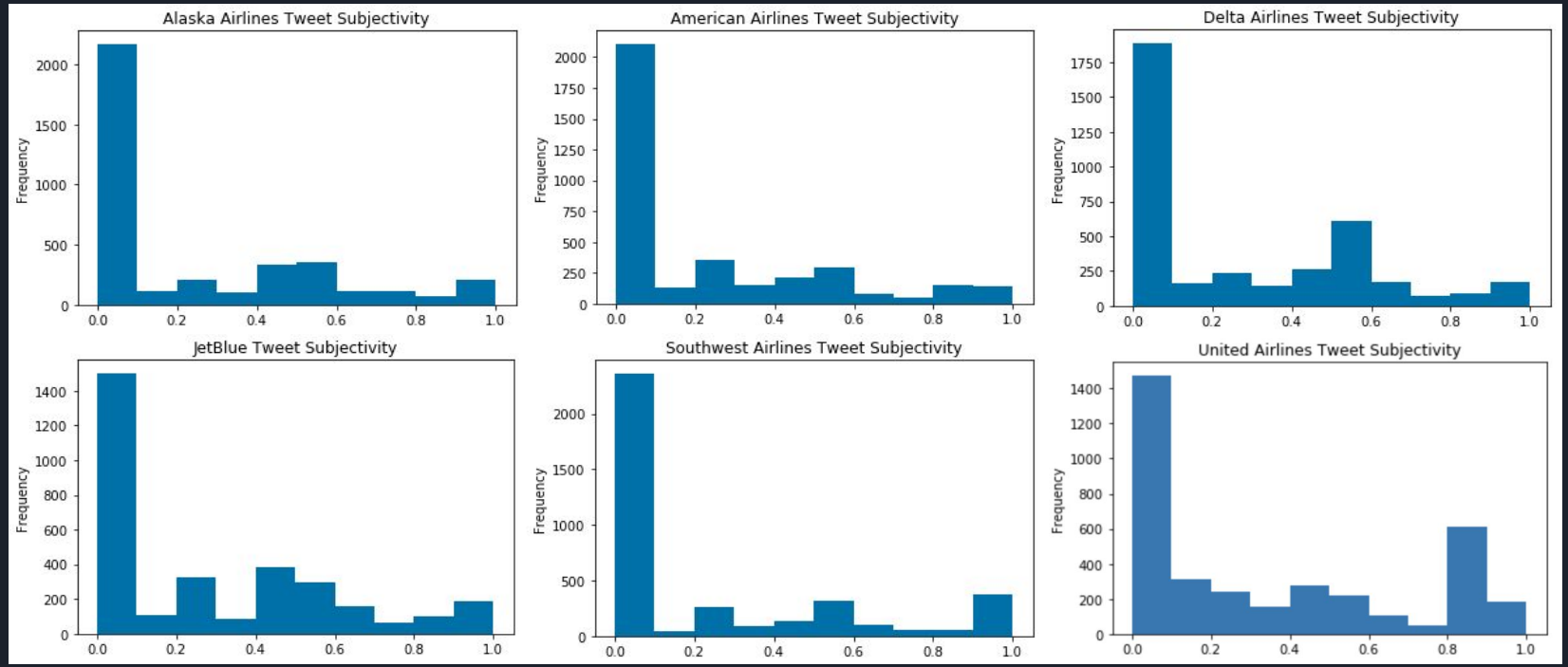
```
american_subjectivity = airline_tweets[airline_tweets['Airline']=='American Airlines']
american_subjectivity_hist = american_subjectivity['subjectivity'].plot(kind='hist',
                                                                        bins=10,
                                                                        title='American Airlines Tweet
Subjectivity')
```

```
american_polarity.plot(kind="scatter",
                       x='polarity',
                       y='subjectivity',
                       title='American Airlines')
```

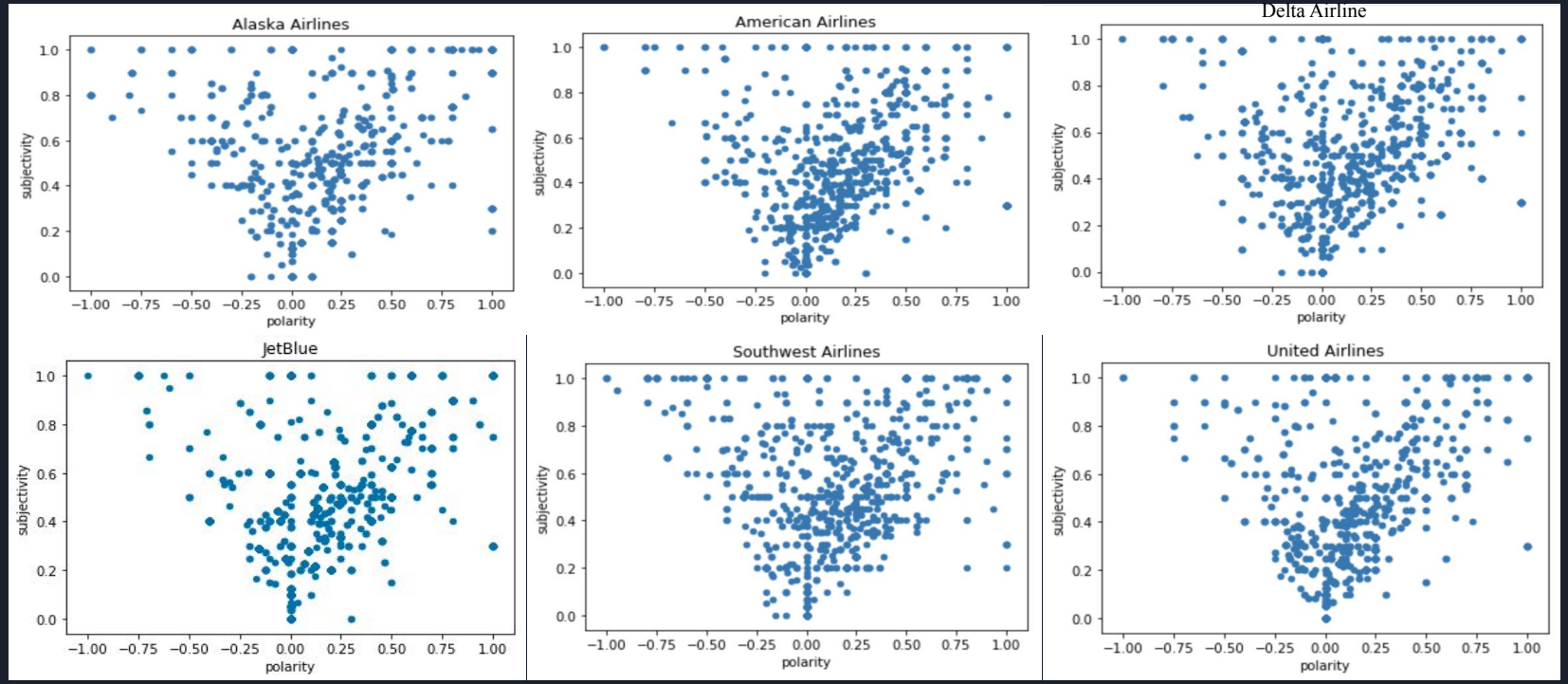
Histograms - Tweet Polarity



Histograms - Tweet Subjectivity



Scatter Plots



Stock Price Plots

[18]
```
fig = px.line(stock_prices_plot, x='Period', y='value', color='variable', title='Airline Stock Prices(USD)
from August 11 - 14 (2020)',
              labels={"value": "Price (USD)", "variable": "Stock"},
              width=800, height=500)

fig.show()
```

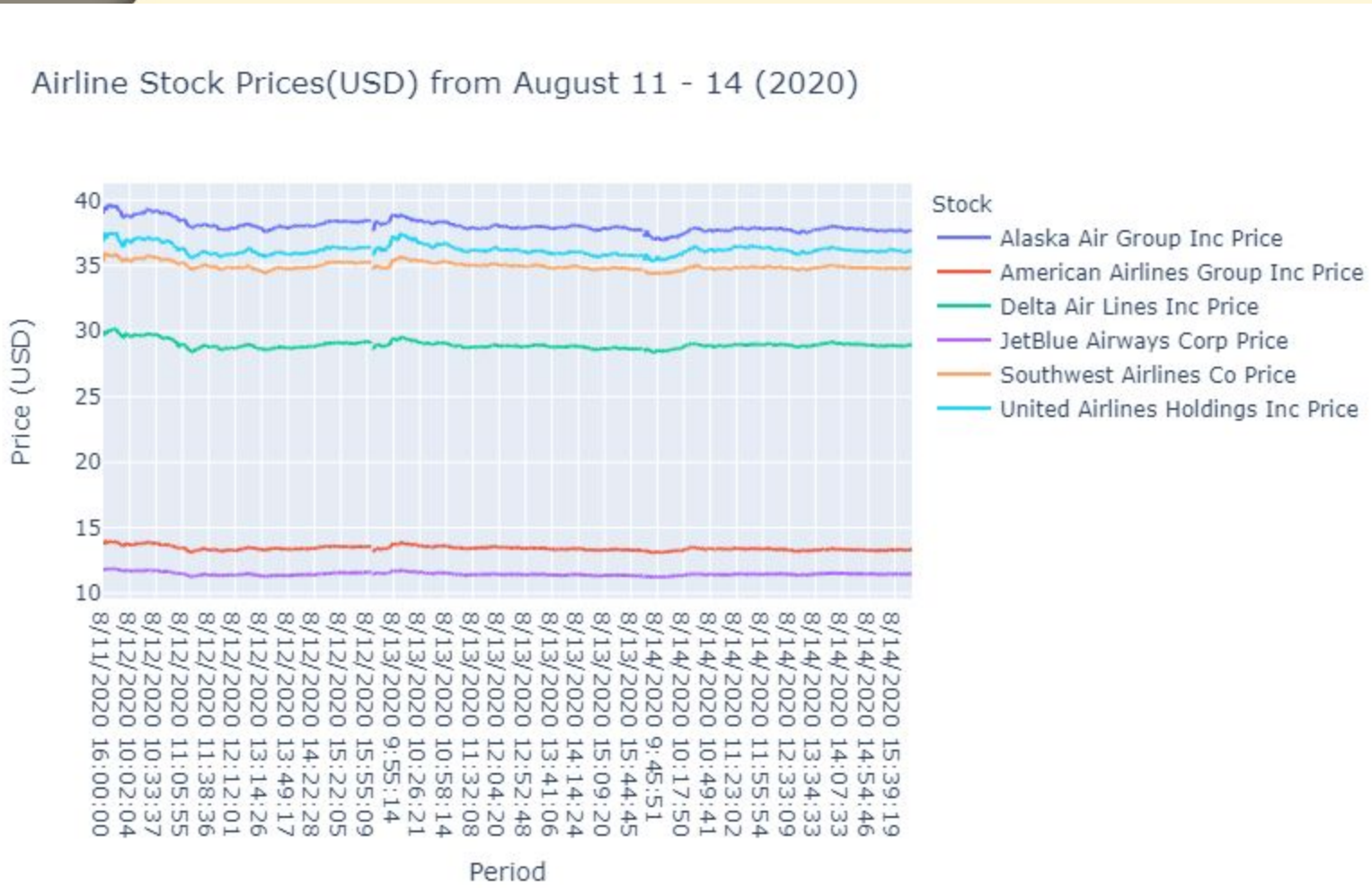
[43]
```
stock_prices2 = stock_prices.set_index(['Period'])
pct_change = stock_prices2.pct_change()
pct_change.dropna(inplace=True)
```

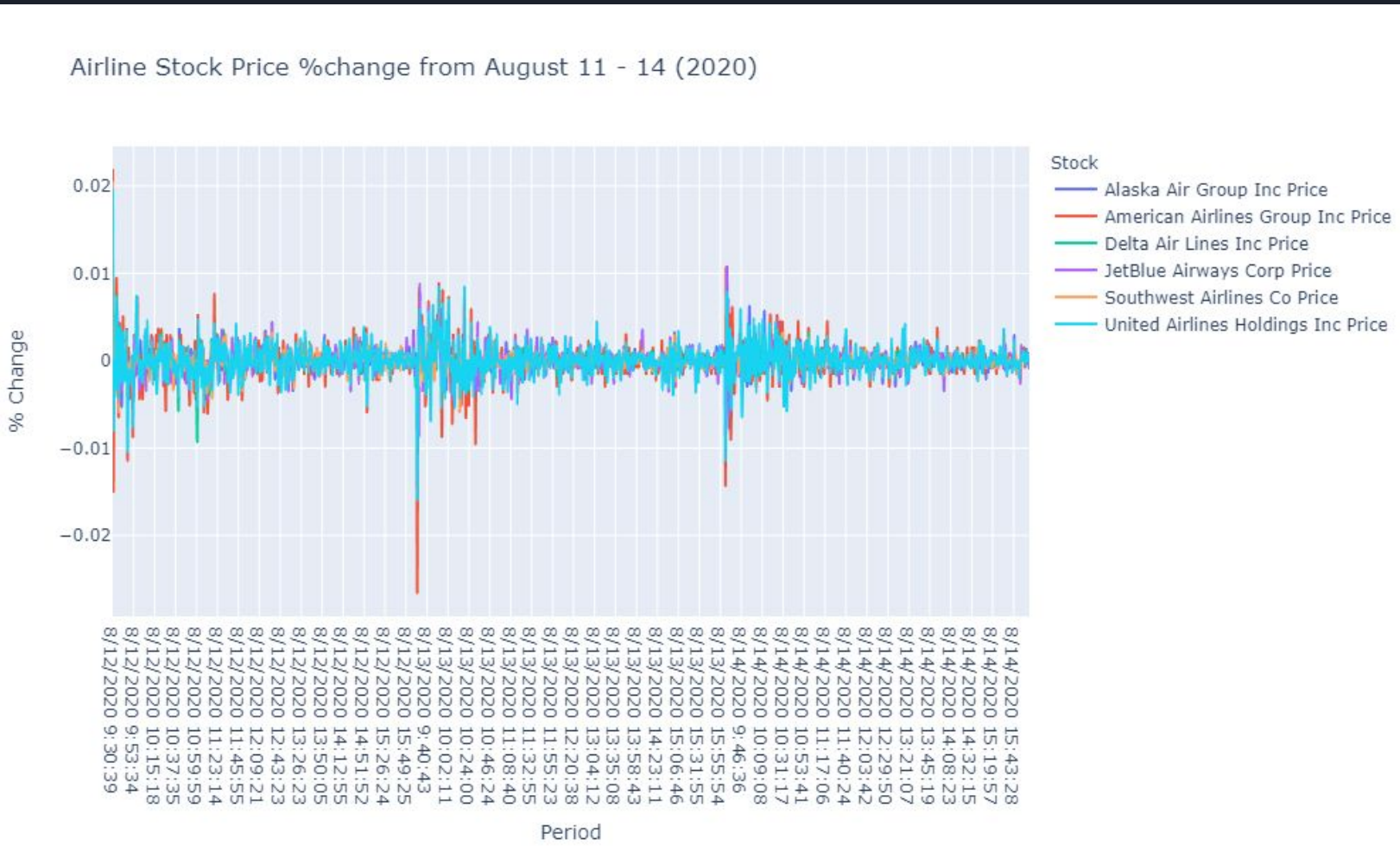
[46]
```
pct_change_fixed = pct_change.reset_index()
```

[47]
```
pct_change_plot = pd.melt(pct_change_fixed,
                          id_vars=['Period'],
                          value_vars=['Alaska Air Group Inc Price', 'American Airlines Group Inc Price',
'Delta Air Lines Inc Price', 'JetBlue Airways Corp Price', 'Southwest Airlines Co Price', 'United Airlines
Holdings Inc Price'])
```

[51]
```
fig = px.line(pct_change_plot, x='Period', y='value', color='variable', title='Airline Stock Price %change
from August 11 - 14 (2020)',
              labels={"value": "% Change", "variable": "Stock"},
              width=1000, height=600)

fig.show()
```


Airline Stock Prices(USD) from August 11 - 14 (2020)
Price (USD)
40
35
30
25
20
15
10
8/11/2020 16:00:00
8/12/2020 10:02:04
8/12/2020 10:33:37
8/12/2020 11:05:55
8/12/2020 11:38:36
8/12/2020 12:12:01
8/12/2020 13:14:26
8/12/2020 13:49:17
8/12/2020 14:22:28
8/12/2020 15:22:05
8/12/2020 15:55:09
8/13/2020 9:55:14
8/13/2020 10:26:21
8/13/2020 10:58:14
8/13/2020 11:32:08
8/13/2020 12:04:20
8/13/2020 12:52:48
8/13/2020 13:41:06
8/13/2020 14:14:24
8/13/2020 15:09:20
8/13/2020 15:44:45
8/14/2020 9:45:51
8/14/2020 10:17:50
8/14/2020 10:49:41
8/14/2020 11:23:02
8/14/2020 11:55:54
8/14/2020 12:33:09
8/14/2020 13:34:33
8/14/2020 14:07:33
8/14/2020 14:54:46
8/14/2020 15:39:19
Period
Stock
Alaska Air Group Inc Price
American Airlines Group Inc Price
Delta Air Lines Inc Price
JetBlue Airways Corp Price
Southwest Airlines Co Price
United Airlines Holdings Inc Price


Airline Stock Price %change from August 11 - 14 (2020)
% Change
0.02
0.01
0
−0.01
−0.02
Period
8/12/2020 9:30:39
8/12/2020 9:53:34
8/12/2020 10:15:18
8/12/2020 10:37:35
8/12/2020 10:59:59
8/12/2020 11:23:14
8/12/2020 11:45:55
8/12/2020 12:09:21
8/12/2020 12:43:23
8/12/2020 13:26:23
8/12/2020 13:50:05
8/12/2020 14:12:55
8/12/2020 14:51:52
8/12/2020 15:26:24
8/12/2020 15:49:25
8/13/2020 9:40:43
8/13/2020 10:02:11
8/13/2020 10:24:00
8/13/2020 10:46:24
8/13/2020 11:08:40
8/13/2020 11:32:55
8/13/2020 11:55:23
8/13/2020 12:20:38
8/13/2020 13:04:12
8/13/2020 13:35:08
8/13/2020 13:58:43
8/13/2020 14:23:11
8/13/2020 15:06:46
8/13/2020 15:31:55
8/13/2020 15:55:54
8/14/2020 9:46:36
8/14/2020 10:09:08
8/14/2020 10:31:17
8/14/2020 10:53:41
8/14/2020 11:17:06
8/14/2020 11:40:24
8/14/2020 12:03:42
8/14/2020 12:29:50
8/14/2020 13:21:07
8/14/2020 13:45:19
8/14/2020 14:08:23
8/14/2020 14:32:15
8/14/2020 15:19:57
8/14/2020 15:43:28
Stock
Alaska Air Group Inc Price
American Airlines Group Inc Price
Delta Air Lines Inc Price
JetBlue Airways Corp Price
Southwest Airlines Co Price
United Airlines Holdings Inc Price

Discussion

Questions