CSE 5345/7345

Due: Sunday Sep 23

Quest 5. Pandas 2

Part A. States and area codes

Use Pandas to answer the question "Which states have only one area code?"

Download the file 'states.areacodes.csv" which looks like:

	Area code	State	State code
0	201	New Jersey	NJ
1	202	Washington,DC	DC
2	203	Connecticut	CT
3	205	Alabama	AL
4	206	Washington	WA

The file includes data for Washington,DC which is not a state. Use your Panda skills to remove it from the data set.

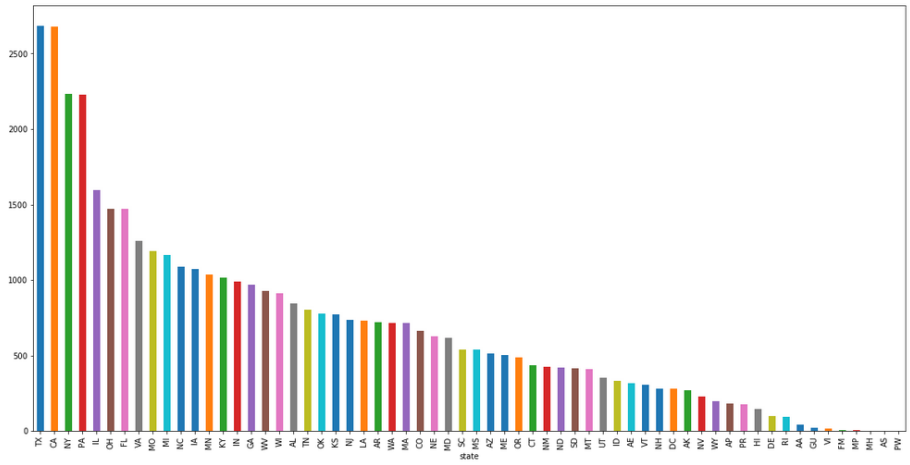
Display only the states that have ONE area code. You may display with a Series or a Dataframe. For example:

```
State
Alaska          1
Delaware        1
Hawaii          1
Idaho           1
Maine           1
Montana         1
New Hampshire   1
New Mexico      1
North Dakota    1
Rhode Island    1
South Dakota    1
Vermont         1
West Virginia   1
Wyoming         1
```

Part B. Zipcodes and States

Load the CSV file: zipcodes.states.gps.csv

Display a bar chart that shows a sorted display of number zipcodes by states and territories (no need to exclude any data)



Part C. Films

- load films2.csv
- There are two similar Subject categories: Western and Westerns. Using Pandas, merge the data from 'Westerns' into "Western'
- Use your knowledge of Panda DataFrames and Series to determine the Actor with the greatest number of movies in each Subject category. For some categories the max value is 1. Remove those entries.
- Display the category, Actor name and number of films (as an integer) where the number of films > 1 as in:

Actor	Actor	nFilms
Subject		
Action	Connery, Sean	15
Comedy	Sellers, Peter	22
Drama	Brando, Marlon	17
Horror	Ford, Wallace	3
Mystery	Connery, Sean	3
Science Fiction	Hamill, Mark	3
War	Wayne, John	10
Western	Wayne, John	46

Note: Your output format may differ but you should display three columns as above. What names you use for columns and rows is up to you.

Part D. Stock Trading Algorithm with Moving Averages (MA)

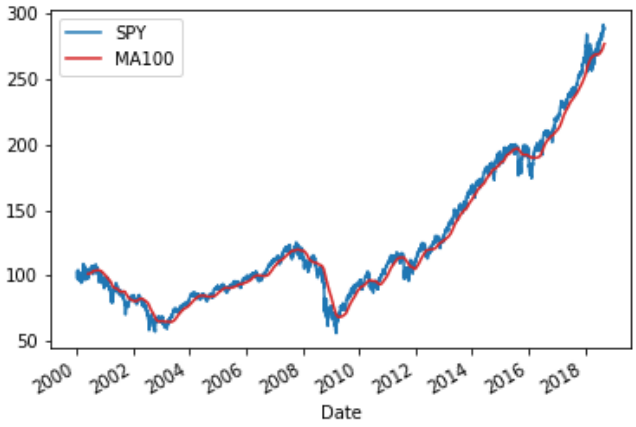
Your objective is to determine whether the profitability of using the moving averages (MA) of stock prices to buy and sell stock is more effective than buy and hold. MA trading is a technique used by traders. The idea is to compute the moving average of a stock over some number of trading days (e.g. 30, 100, 200, 300) and to buy the stock when the price moves above the moving average and then sell the stock when the price drops below the moving average.

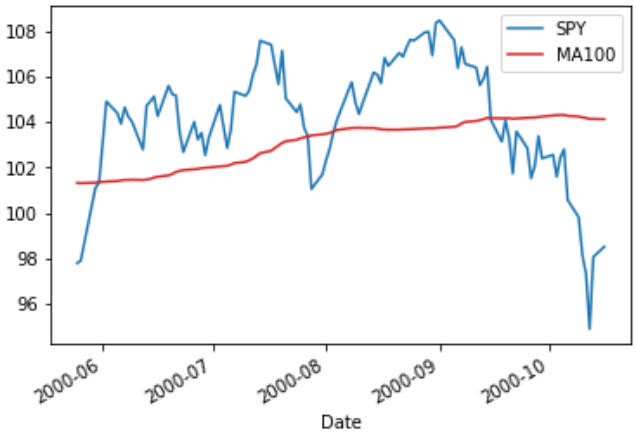
Given a series of prices in chronological order, one can compute a 100-day moving average for a column labeled 'SPY' with:

df['MA100'] = df['SPY'].rolling(100).mean()

This will add a new column to your DataFrame. Your task is to use this information to simulate buying and selling SPY. - buy when the price exceeds the MA and sell when the price drops below the MA.

Below are stock prices plotted against the 100-day MA for 18 years and a segment of the data where you can see the plots diverge.





DO:

- Load a Dataframe the file **SPY.histdata.csv** which contains historical price data for SPY, the ETF that combines the S&P 500 stock prices. SPY is traded on the NY Stock Exchange.
- Convert the dates from the CSV to Panda dates using:
  ```
  df = pd.read_csv('SPY.histdata.csv', parse_dates=['Date'])
  ```
  This will allow your Panda data format to be sorted.
- Sort the DataFrame so it is in ascending order beginning with 2000-01-03. Your df.head() should look like:

	Date	SPY
4701	2000-01-03	102.9308
4700	2000-01-04	98.9057
4699	2000-01-05	99.0826
4698	2000-01-06	97.4902
4697	2000-01-07	103.1521

- Add a new column that is the 100 day moving average for SPY.
- When you do this, the first 99 values will be NaN for the new column. Use your Panda knowledge to remove these rows
- Write code or work Panda magic that tracks buying and selling. Start with $1000. Buy shares (fractional values are expected) when the price first crosses the 100 day MA and sell when the price moves below the moving average. Do this for all the historical data. Keep track of the value of your initial investment.
- At the end, if you are holding the stock, sell at the last price. If you have already sold the stock then you know the value. Compute the percentage gain or loss based on your initial investment of $1000.
- Compare the performance of your algorithm against the Buy and Hold strategy. To determine the gain/loss with Buy-and-Hold, use the first buy and last sale from the Moving Average strategy.
- Compare the results with a bar graph. Include the number of trades you made (each buy or sell) Note: the comparison does not factor in the cost of trading, although that is a factor but for large trades it is small.

Challenge: optional

Can you devise a more profitable trading strategy? Some ideas:

- try different moving averages
- try triggering a buy/sell only when the price moves some percentage above or below the MA

A certificate and a prize for the winning strategy will be awarded.

SUBMIT:

Submit a PDF showing code and results and a zip file of actual code or notebooks