aspiration.ai

Machine Learning internship

Introduction

Instructions

Select an Industry

Module 1

Module 2 🔒

Module 3 🔒

Module 4 🔒

Module 5 🔒

Module 6 🔒

My Jupyter Notebook

Module 1

Introduction to the problem

In Module 1, you are going to get familiar with pandas, the python module which is used to process and analyse data. Processing could include removing unknown values from the data or replacing unknown values with values which make sense, maybe 0. Analysing the data could include finding out the trend of a stock price, e.g. how the stock price changes with respect to the Nifty 50 basket of stocks.

Please go through the reference material suggested for module 1 before attempting the tasks in module 1.

You should target to finish module 1, including the prerequisites, in 1 week

Problem Statements

1.1 Import the csv file of the stock you have been allotted using 'pd.read_csv()' function into a dataframe.
Shares of a company can be offered in more than one category. The category of a stock is indicated in the 'Series' column.
If the csv file has data on more than one category, the 'Date' column will have repeating values. To avoid repetitions in the date, remove all the rows where 'Series' column is NOT 'EQ'.
Analyze and understand each column properly.
You'd find the head(), tail() and describe() functions to be immensely useful for exploration. You're free to carry out any other exploration of your own.

1.2 Calculate the maximum, minimum and mean price for the last 90 days. (price=Closing Price unless stated otherwise)

1.3 Analyse the data types for each column of the dataframe. Pandas knows how to deal with dates in an intelligent manner. But to make use of Pandas functionality for dates, you need to ensure that the column is of type 'datetime64(ns)'. Change the date column from 'object' type to 'datetime64(ns)' for future convenience. See what happens if you subtract the minimum value of the date column from the maximum value.

1.4 In a separate array , calculate the monthwise VWAP (Volume Weighted Average Price) of the stock.
(VWAP = sum(price*volume)/sum(volume))
To know more about VWAP , visit - VWAP definition (https://www.investopedia.com/terms/v/vwap.asp)
{Hint : Create a new dataframe column 'Month'. The values for this column can be derived from the 'Date" column by using appropriate pandas functions. Similarly, create a column 'Year' and initialize it. Then use the 'groupby()' function by month and year. Finally, calculate the vwap value for each month (i.e. for each group created).

1.5Write a function to calculate the average price over the last N days of the stock price data where N is a user defined parameter. Write a second function to calculate the profit/loss percentage over the last N days.
Calculate the average price AND the profit/loss percentages over the course of last -
1 week, 2 weeks, 1 month, 3 months, 6 months and 1 year.
{Note : Profit/Loss percentage between N days is the percentage change between the closing prices of the 2 days }

1.6 Add a column 'Day_Perc_Change' where the values are the daily change in percentages i.e. the percentage change between 2 consecutive day's closing prices. Instead of using the basic mathematical formula for computing the same, use 'pct_change()' function provided by Pandas for dataframes. You will note that the first entry of the column will have a 'Nan' value. Why does this happen? Either remove the first row, or set the entry to 0 before proceeding.

1.7 Add another column 'Trend' whose values are:

- ✔ 'Slight or No change' for 'Day_Perc_Change' in between -0.5 and 0.5

- ✔ 'Slight positive' for 'Day_Perc_Change' in between 0.5 and 1

- ✔ 'Slight negative' for 'Day_Perc_Change' in between -0.5 and -1

- ✔ 'Positive' for 'Day_Perc_Change' in between 1 and 3

- ✔ 'Negative' for 'Day_Perc_Change' in between -1 and -3

- ✔ 'Among top gainers' for 'Day_Perc_Change' in between 3 and 7

- ✔ 'Among top losers' for 'Day_Perc_Change' in between -3 and -7

- ✔ 'Bull run' for 'Day_Perc_Change' >7

- ✔ 'Bear drop' for 'Day_Perc_Change' <-7

1.8 Find the average and median values of the column 'Total Traded Quantity' for each of the types of 'Trend'.
{Hint : use 'groupby()' on the 'Trend' column and then calculate the average and median values of the column 'Total Traded Quantity'}

1.9 SAVE the dataframe with the additional columns computed as a csv file week2.csv. In Module 2, you are going to get familiar with matplotlib, the python module which is used to visualize data.

Submit

Getting stuck and need a little guidance. Get access to the guidance program (https://careerlauncher.com/cl-online/ProductDesc.jsp?prodeid=0K4XsvNcqbM%3D&prodCat=DATA%20ANALYTICS&prodGroup=Online%20Coaching) from our machine learning faculty!

.ai (/)

About Us (/about.jsp)   Contact Us (/contact-us.jsp)   Help (/support.jsp)

f   🐦   ▶   ⊙

(//www.facebook.com/CareerLauncher