Stock recommendation system building by using Prophet Model

Name: Yukang Xu

Ⅰ. Introduction

Stock price forecasting is a popular and important topic in financial and academic studies. Time series analysis is the most common and fundamental method used to perform this task Investing in the stock. Investing in the stock market is the most reliable way to create wealth over long time periods. It might surprise you to learn that a $10,000 investment in the S&P 500 index 50 years ago would be worth nearly $1.2 million today. My goal is to create a recommendation system for a portfolio of 5 stocks (Facebook, Amazon, Apple, Netflix and Google).

You can invest in individual stocks if -- and only if -- you have the time and desire to thoroughly research and evaluate stocks on an ongoing basis. If this is the case, we 100% encourage you to do so -- it is entirely possible for a smart and patient investor to beat the market over time. On the other hand, if things like quarterly earnings reports and some moderate mathematical calculations don't sound appealing, there's absolutely nothing wrong with taking a more passive approach.

The stock market is no place for money that you might need within the next five years, at a minimum. While the stock market will almost certainly rise over the long run, there's simply too much uncertainty in stock prices in the short term -- in fact, statistically speaking, a drop of 20% in any given year wouldn't even be considered an unusual occurrence.

Considering available money and background, different people may allocate different money in stock investment. In this project, I will differentiate our target clients then give them recommendations separately.

Ⅱ. Data Analysis

1. Identify our data

Since we are pulling data from Yahoo Finance directly and considering a stock that is updated regularly, we can eliminate a couple of steps for data preprocessing or cleaning, converting Excel file data into indexes, or accounting for empty values.

After loading the data, the first dimension consists of the various fields which Yahoo Finance returns for a given instrument, namely, the Open, High, Low, Close and Adj Close prices for each date. The second dimension contain the dates. The third one contains the instrument identifiers.

Let us assume that we are interested in working with the Close prices which have been already been adjusted by Google finance to account for stock splits. We want to make sure that all weekdays are included in our dataset, which is very often desirable for quantitative trading strategies.

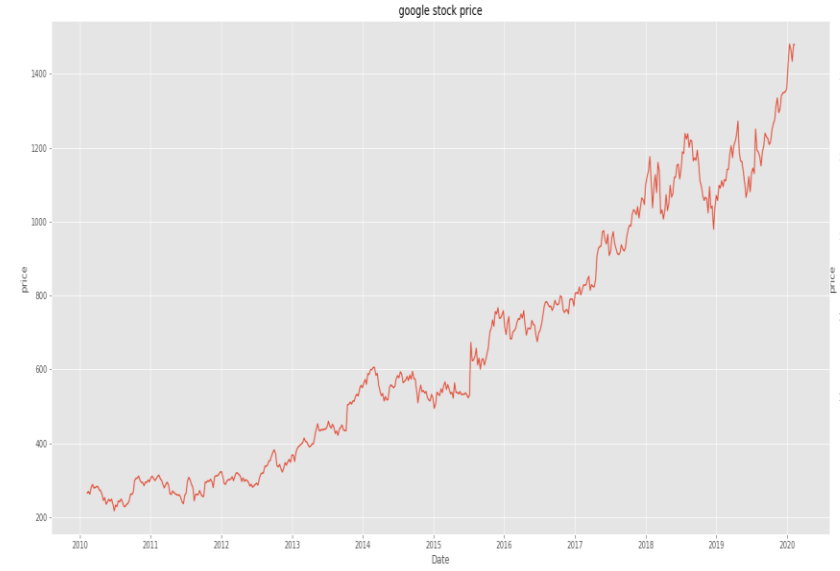
	Date	Open	High	Low	Close	Adj Close	Volume
0	2010-02-08	119.379997	121.000000	116.000000	119.660004	119.660004	41763400
1	2010-02-15	120.059998	120.500000	114.820000	117.519997	117.519997	34793200
2	2010-02-22	117.370003	119.800003	115.849998	118.400002	118.400002	36520400
3	2010-03-01	118.699997	129.449997	117.529999	128.910004	128.910004	46113400

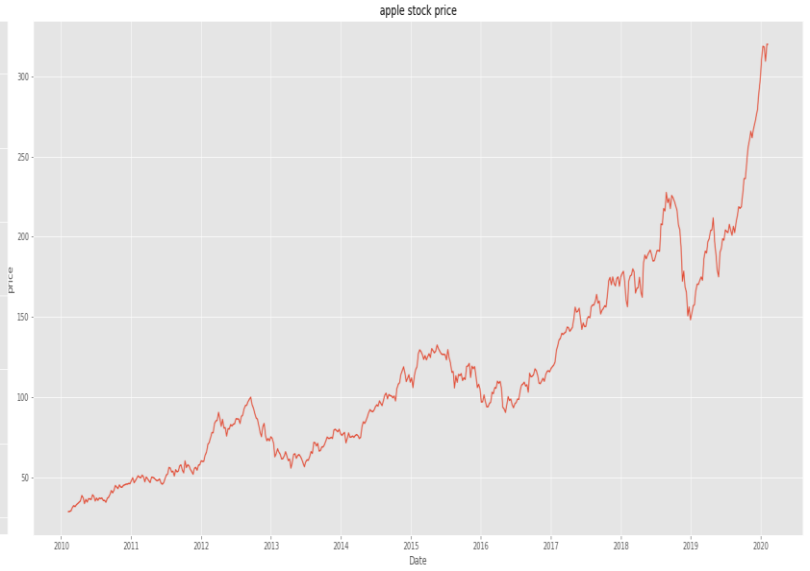
2.Explore the dataset

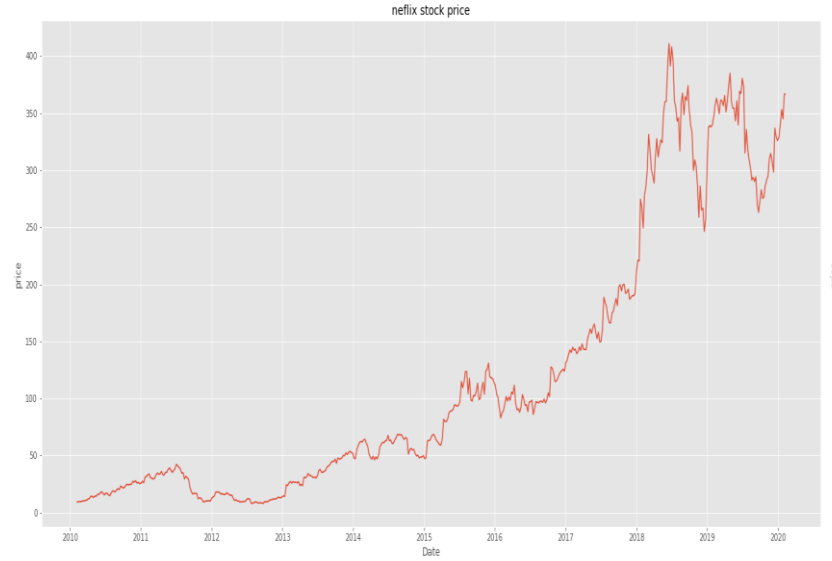
Plotting data is arguably the most critical step in the exploratory analysis phase. (We chose to emphasize the time-series object that has intervals from 2010 to 2019, a choice we will explain later!). Visualizing our time-series data enables us to make inferences about important components, such as trend, seasonality, heteroskedasticity, and stationarity. You can easily see that our time series has instances of both positive and negative trend. Overall, it is going up.

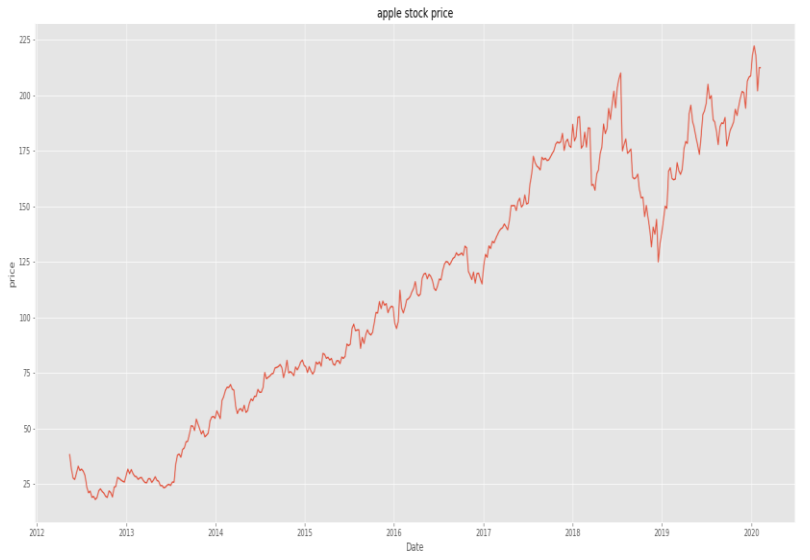
Upon briefly looking at the data, we can see that the stock value increased over time since 2015, but spiked during 2018. There is no apparent pattern that can be used to scale the value of the stock price because the trends are not linear, and there is no mathematical formula to describe the change in curve or the fluctuations between increasing or decreasing prices.

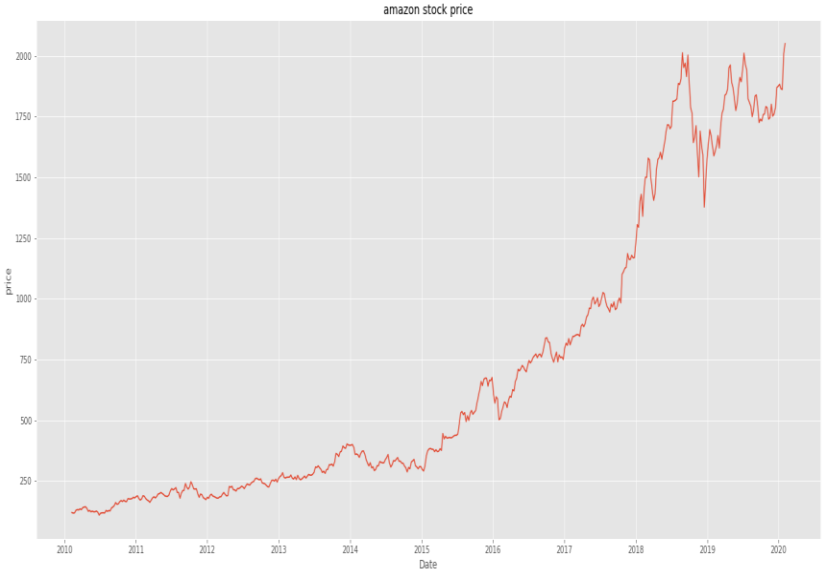
As the stock price spikes in 2018, the variance between datapoints also appears to increase, with stock prices being especially volatile at the most recent dates.


google stock price
price
Date


apple stock price
price
Date


neflix stock price
price
Date


apple stock price
price
Date


amazon stock price
price
Date

3. Stationary Test

I list all important things as below in stationary test including stationary character, how to detect and solve.

Stationarity rule	Description	How to detect	How to solve
Constant mean	Follow one specific pattern	Pattern detection: duller test and KPSS test	Differencing
Autocovariance does not depend on time			
Constant variance	Not stable	Variance and mean: rolling statistic	Transformation

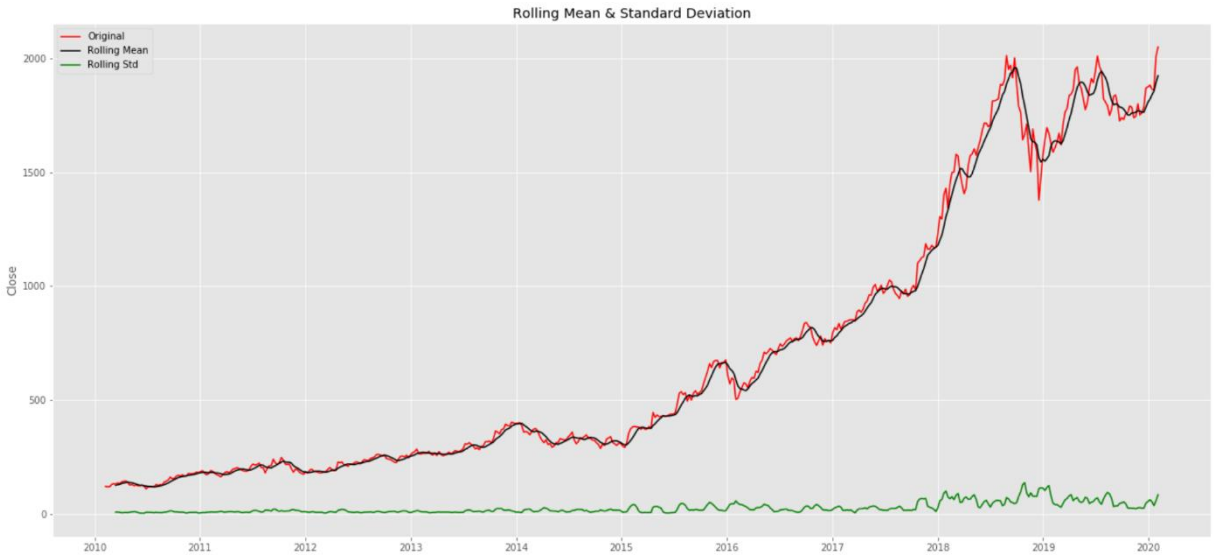
One more thing, Rolling statistic is to smooth out short-term fluctuations and highlight longer-term trends. As you can see from plot above, our time series has trend variation. Overall the price of stock is going up. Now let us check stationary of time series. We can check stationarity using the following methods: duller test, kpss test and rolling statistic graph.

Plotting Rolling Statistics: We have a window lets say window size is 6 and then we find rolling mean and variance to check stationary.

Dickey-Fuller Test: The test results comprise of a Test Statistic and some Critical Values for difference confidence levels. If the test statistic is less than the critical value, we can say that time series is stationary.

KPSS Test: this is another test for checking the stationarity of a time series (slightly less popular than the Dickey Fuller test). The null and alternate hypothesis for the KPSS test is opposite that of the Dickey-Fuller Test, which often creates confusion.

```
####################Rolling statistics#######################
```



```
####################Dickey-Fuller test#######################
Test statistic:  1.1215724599233596
p-value:  0.9953965939935294
Critical Values: {'1%': -3.4429638415440995, '5%': -2.867103482127635, '10%': -2.569733131582922}
####################KPSS test#######################
Test Statistic            2.342418
p-value                   0.010000
Lags Used                19.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

From the rolling statistic graph and statistical test, we find that the graph without constant mean did not pass the Dickey-Fuller test. So we have to deal with trend issue by differencing. Since variance appears to be stable, so we do not have to consider transformation except for differencing.

Time series datasets may contain trends and seasonality, which may need to be removed prior to modeling. Trends can result in a varying mean over time, whereas seasonality can result in a changing variance over time, both which define a time series as being non-stationary. Stationary datasets are those that have a stable mean and variance, and are in turn much easier to model. Differencing is a popular and widely used data transform for making time series data stationary.

Differencing can help stabilize the mean of the time series by removing changes in the level of a time series, and so eliminating (or reducing) trend and seasonality. Differencing is performed by subtracting the previous observation from the current observation. The result after differencing is as below. Since rolling statistic graph appear to be stable and we pass both of statistical test, we can move on to the next part.

```
####################Rolling statistics#######################
```

```
####################Dickey-Fuller test#######################
Test statistic:  -16.773579349600244
p-value:  1.2810548032392496e-29
Critical Values: {'1%': -3.4430126933746767, '5%': -2.8671249839002764, '10%': -2.569744590233924}
####################KPSS test#######################
Test Statistic            0.387763
p-value                   0.082430
Lags Used                19.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

As stock price of these five companies apparently have positive trend, I apply differencing method to deal with this issue. Since the procedure is similar, I only show the entire handling process for Amazon stock. As what I attach above, differencing helps our data to be stationary and pass ADF test, KPSS test.

4. Model Development using Prophet

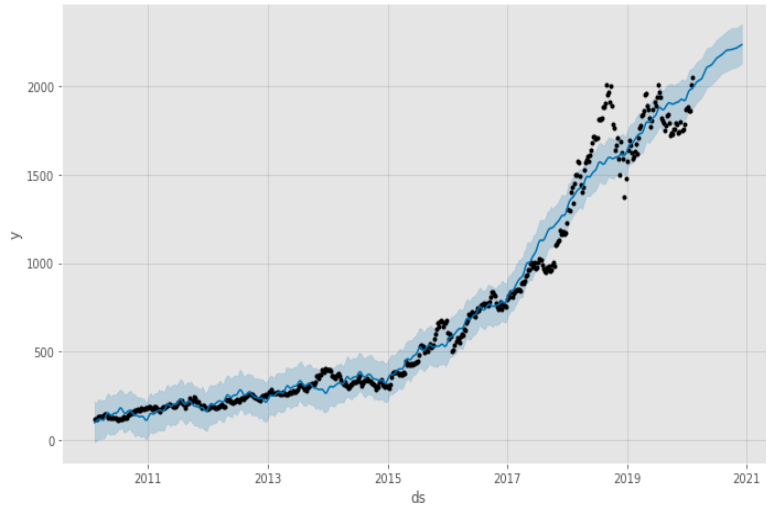
Prophet is an open source library published by Facebook that is based on decomposable (trend+ seasonality+ holidays) models. It provides us with the ability to make time series predictions with good accuracy using simple intuitive parameters and has support for including impact of custom seasonality and holidays! Since we already prove that our stock data has specific trend which can be removed. Prophet is a good choice to make a time series prediction.

Now that we are well versed with nuts and bolts of this amazing tool. Let's dive into a real dataset to see its potential. The input to Prophet is always a data frame with two columns: ds and y. The ds (date stamp) column should be of a format expected by Pandas, ideally YYYY-MM-DD for a date or YYYY-MM-DD HH:MM:SS for a timestamp. The y column must be numeric, and represents the measurement we wish to forecast.

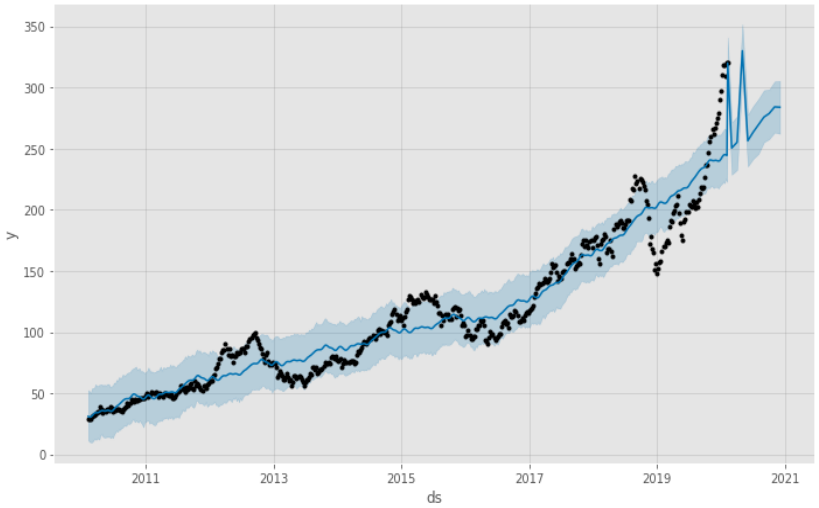
The predict method will assign each row in future a predicted value which it names "yhat". If you pass in historical dates, it will provide an in-sample fit. The forecast object here is a new data frame that includes a column "yhat" with the forecast, as well as columns for components and uncertainty intervals. After loading our stock data, we get five graphs as follows which are our prediction and historical data.

Our model definitely can predict long-term time series dataset. In this project, we apply Prophet to predict one-year future stock price. Then we are all set to build our recommendation system.

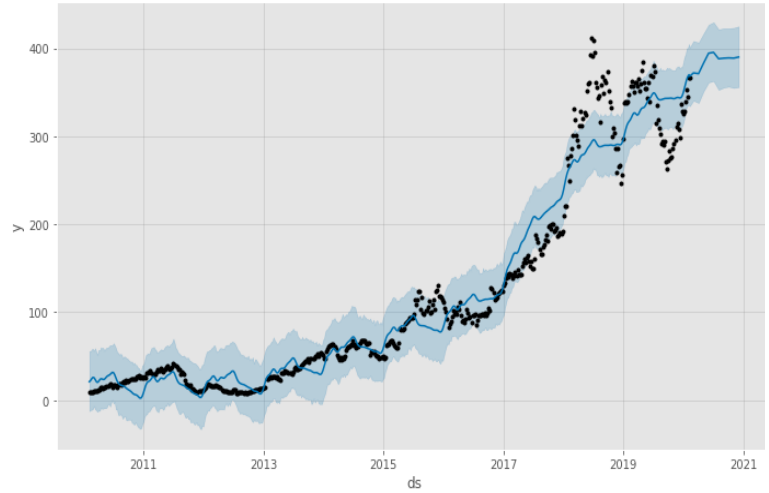
Amazon stock price prediction



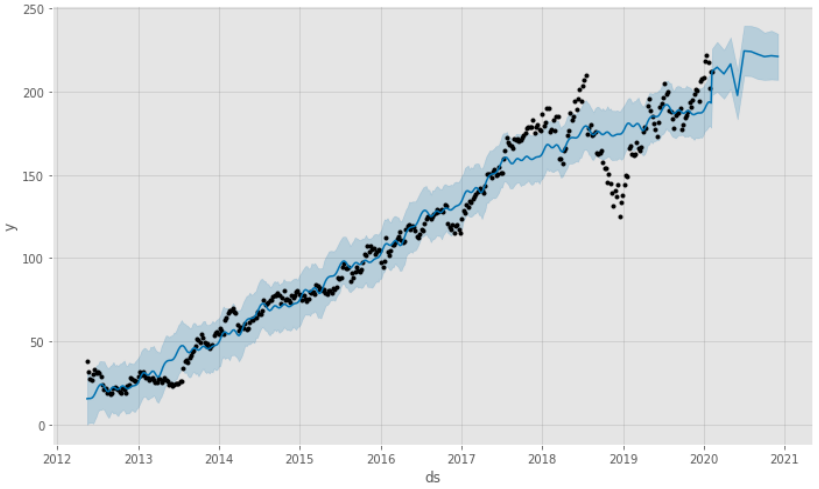
Apple stock price prediction



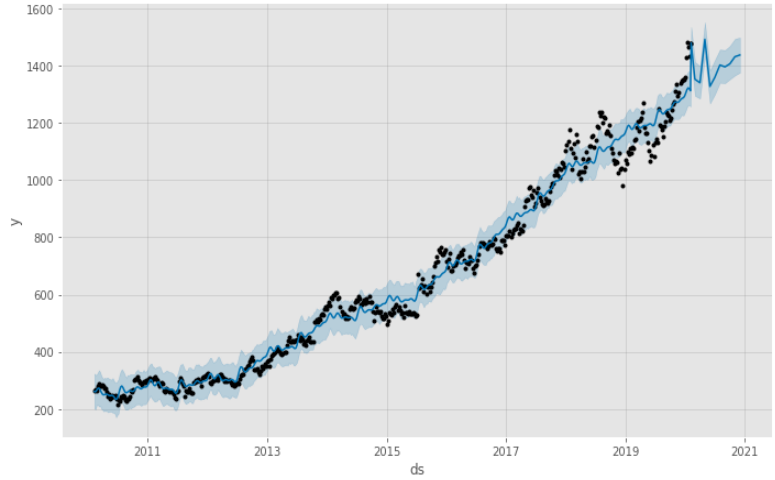
Netflix stock price prediction



Facebook stock price prediction



Google stock price prediction



Ⅲ. Conclusion

Now, we have stock price for year 2020. How do we use these predictions? If we classify our potential customers as three groups like Active Investors, Passive Investors and Speculators, they apparently need different strategies. Assuming they are living in Atlanta, they are all in middle class. From DATAUS, median salary in Atlanta is 65345 dollars per year while cost of living is about 931 dollars per month without rental. Based on these information, let's say our clients having 10,000 dollars for investment First of all, let me introduce our clients, Tom, Jeff and Mike. Let us make a summary of our clients.

Name	Type	Level	Strategy	Recommendation	Profits in 1 year
Tom	Active Investors	Medium	Following trend	Netflix + Apple	$1226
Jeff	Passive Investors	Entry	Long-term investment	Amazon	$1275
Mike	Speculators	Advanced	Short-term investment	Mix	$1393

Active Investors--Tom

Tom just got his bachelor's degree last year and landed a job in a small bank. He has basic financial knowledge. Since he is still single, he could assign his extra money in stock investment.

Tom stays abreast of their stocks' performance, does a lot of research and keep up with the daily financial news. Active investors like Tom, they don't necessarily buy one day and sell the next, but they do pay attention to changes in trends and buy or sell based on those trends. This person is an avid investor who takes a great deal of care with each investment decision and does not necessarily hold an investment long term.

After researching up-to-date news, Tom wants to pick one or two companies in Amazon, Netflix and Apple to invest considering not too low profits. Apple announced that they expect to make a major change in features in 2020. Netflix will be super stable since they have a creative producers' team. Meanwhile, Amazon's price was around $2000 which is so high. We expect that Netflix and Apple have more space for growing. Our recommendations for Tom will be 50% Apple's stock and 50% Netflix's stock. Tom expects to earn profits about $1226 at the beginning of 2021.

Passive Investors --Jeff

Jeff is an engineer in a global construction company. He did not have so much financial knowledge. So he prefer to do some safe investment.

This kind of investor doesn't try to go for the biggest possible gains at all times. Instead, the passive investor accepts reasonable gains in exchange for a lower stress level and more free time. This person may invest in mutual funds so the funds' money

managers can make buy and sell decisions. She may buy individual stock in established companies and hold that investment for a year or more. Passive investors tend to remove stress from investment decisions by setting parameters for adding more stock to their portfolios. For example, when their stocks rise 20 percent, they may sell some to take profits.

As a “lazy” investor, Jeff only pays attention to profits. Based on prediction model, we recommend Jeff to go with 100% Amazon’s stock. Jeff expects to earn profits about 1275 at the beginning of 2021.

Speculators—Mike

Mike is a senior manager in an insurance company. Stock investment is his personal interest. To get money back as soon as possible, he will allocate a lot of time on exchanging stocks.

These investors look for a chance to make money fast. They search the market for stocks that are poised to go up because of an impending deal. They scour the news for announcements about mergers that could affect a company positively, and then they pounce on the stocks of those companies. They tend to sell after a stock makes them a little money, reasoning that they can repeat the process of buying and selling frequently and therefore outperform the market.

Mike is willing to spend a lot of time on selling and buying for exchanges. To make profits as much as possible, we will recommend a complicated strategy as below. Mike expects to earn around $1393.

Operation Date	Selling	Buying
First Month	NA	Amazon
Second Month	Amazon	Apple
Third Month	Apple	Amazon
Fourth Month	NA	NA
Fifth Month	Amazon	Netflix
Sixth Month	Netflix	Apple
Seventh Month	Apple	Netflix
Eighth Month	Netflix	Amazon
Ninth Month	Amazon	Netflix
Tenth Month	Netflix	Apple
Eleventh Month	Apple	NA