|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ «Информатика, искусственный интеллект и системы управления» (ИУ)

КАФЕДРА «Системы обработки информации и управления» (ИУ5)

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

***НА ТЕМУ:***

**«Анализ и прогнозирование цен на акции»**

Студент группы ИУ5-21М **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** Ся Бэйбэй

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** Ю. Е. Гапанюк

Оценка

*2022 г.*

**Министерство образования и науки Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технический университет имени Н.Э. Баумана**

**(национальный исследовательский университет)»**

**(МГТУ им. Н.Э. Баумана)**

УТВЕРЖДАЮ

Заведующий кафедрой ИУ5

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **В.И. Терехов**

« \_\_\_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 22 \_\_\_\_ г.

**ЗАДАНИЕ**

**на выполнение научно-исследовательской работы**

по теме «Анализ и прогнозирование цен на акции»

Студент группы ИУ5-31М

Ся Бэйбэй

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

исследовательская

Источник тематики (кафедра, предприятие, НИР) \_\_\_\_\_\_\_\_\_\_НИР\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

График выполнения НИР: 25% к \_\_\_ нед., 50% к \_\_\_ нед., 75% к \_\_ нед., 100% к \_\_\_ нед.

***Техническое задание*** Выполнять анализ, прогнозирование и анализ нескольких портфелей акций на основе данных об акциях. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка на 30 листе формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата выдачи задания « \_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 22\_\_ г.

**Руководитель НИР**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ю. Е. Гапанюк

**Студент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** Ся Бэйбэй

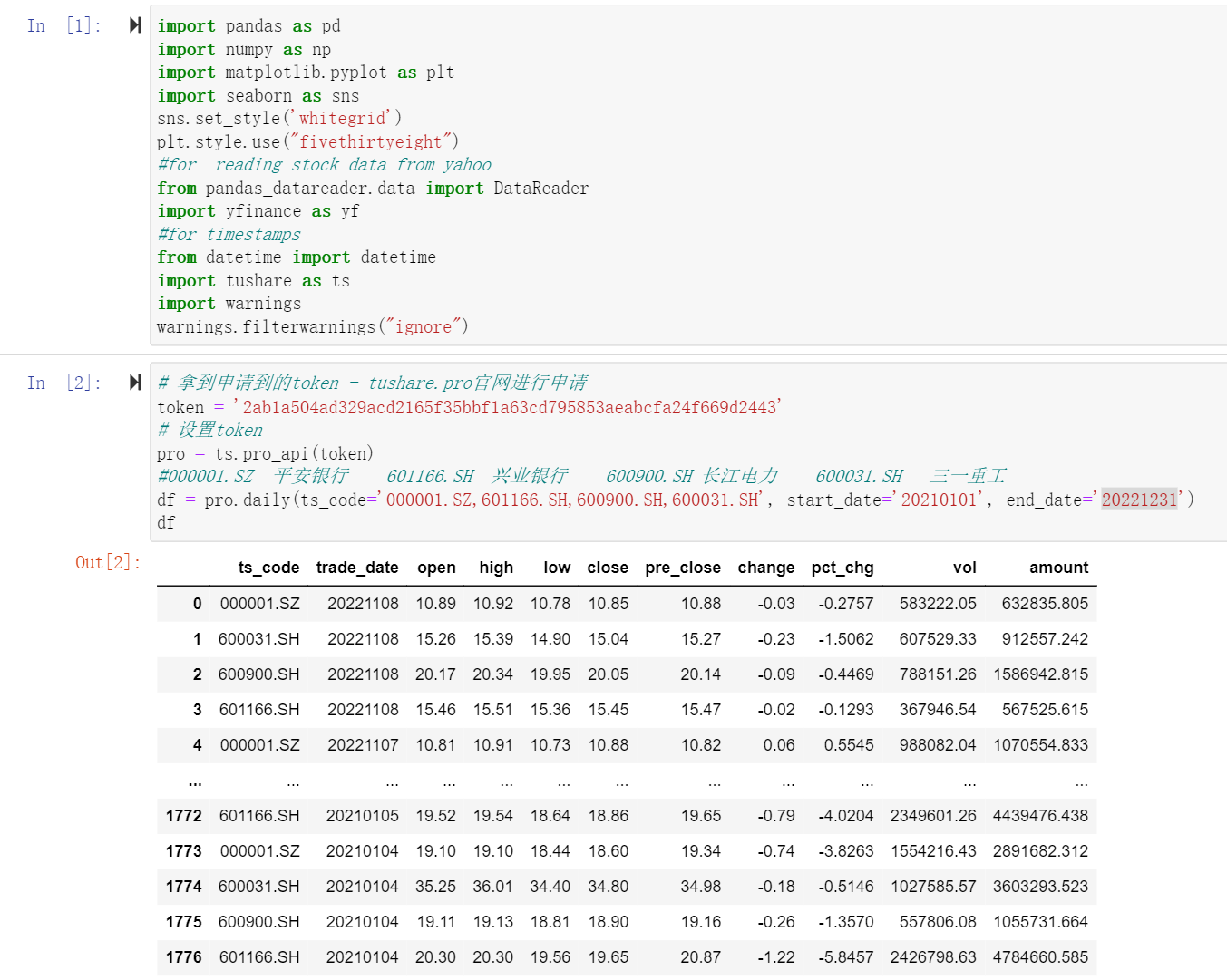
Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

# Введение

Передовые математические модели используются для замены субъективных суждений человека, а компьютерные технологии используются для выбора различных событий с «высокой вероятностью», которые могут принести избыточную отдачу от огромных исторических данных для формулирования стратегий, что значительно снижает влияние колебаний настроений инвесторов. иррациональные инвестиционные решения, когда рынок крайне маниакальный или пессимистичный. Количественная торговля, как правило, проверяется с помощью тестов с моделированием массивных данных и смоделированных операций, а позиции и средства распределяются в соответствии с определенными алгоритмами управления рисками для минимизации рисков и максимизации доходов, но часто существуют определенные потенциальные риски. При разработке алгоритма компьютерная программа используется для выдачи торговых инструкций. В торговле диапазон, который может определить программа, включает в себя выбор времени торговли, цену сделки и даже количество активов, которыми необходимо торговать в конце. Используйте компьютер как инструмент для анализа, оценки и принятия решений с помощью набора фиксированной логики. Количественные стратегии могут выполняться как автоматически, так и вручную. Полная стратегия должна включать ввод, логику обработки стратегии и вывод; логика обработки стратегии должна учитывать такие факторы, как выбор акций, выбор времени, управление позицией, тейк-профит и стоп-лосс. Используйте количественные методы для выбора определенных инвестиционных портфелей, ожидая, что такие инвестиционные портфели принесут доход от инвестиций, превышающий рыночный.

# Анализ данных о запасах

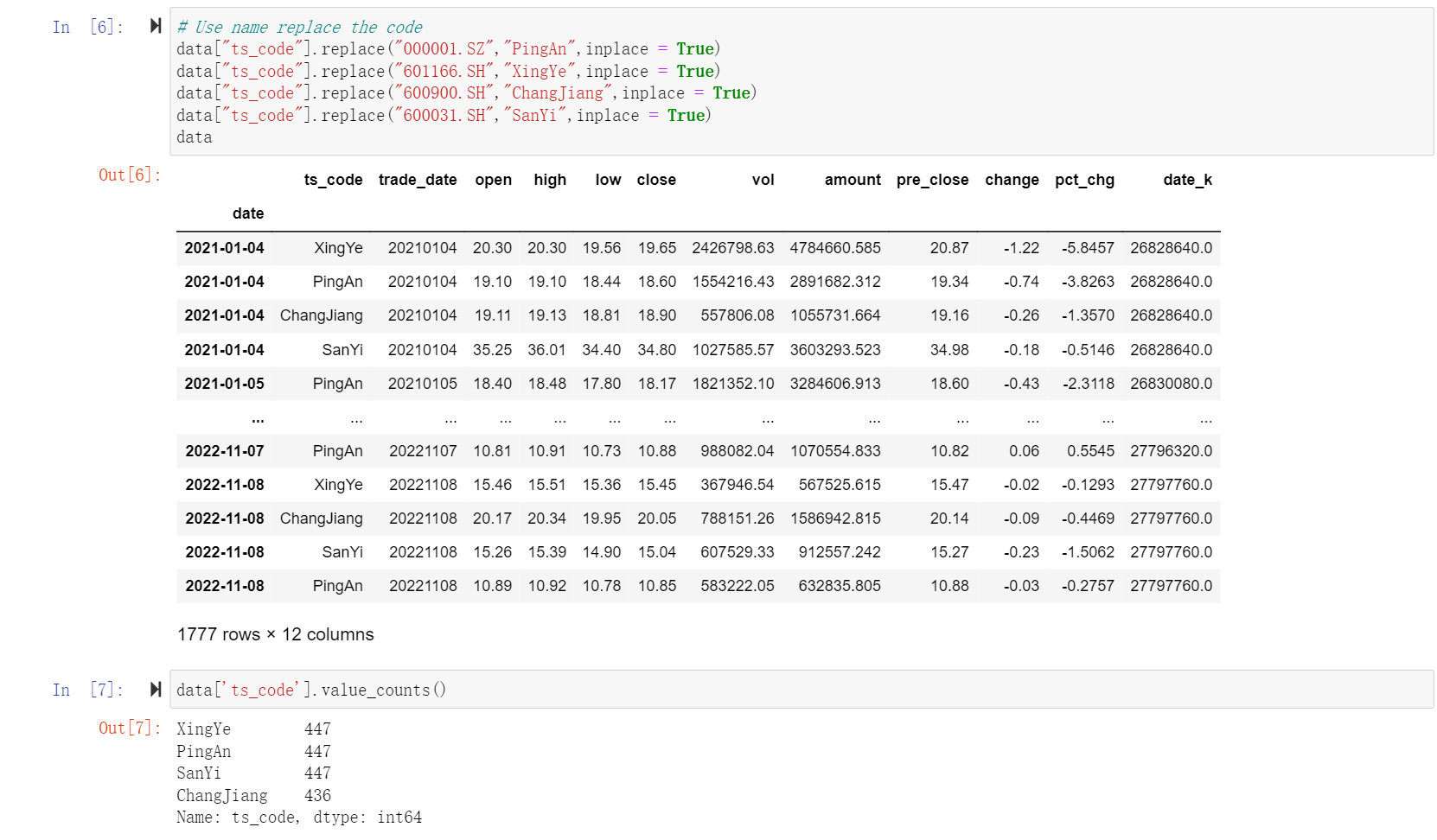
Открытое сообщество больших данных Tushare предоставляет все виды данных бесплатно, чтобы помочь промышленности и количественным исследованиям. Содержание данных будет расширено за счет включения акций, фондов, фьючерсов, облигаций, иностранной валюты, отраслевых больших данных и платформы финансовых больших данных с полным набором данных, которая включает данные блокчейна, такие как условия рынка цифровой валюты, предоставляя различные финансовые инвестиции и исследования. Применимые данные и инструменты. Он поддерживает получение исторических данных и фундаментальных данных об акциях и является лучшим выбором для бесплатного доступа к биржевым данным Шэньчжэня и Шанхая. Используйте Tushare, чтобы получить биржевые данные Ping An Bank, Industrial Bank, Yangtze Power, Sany Heavy Industry с 01.01.2021 по 31.12.2022. Всего данных 11 столбцов, и перечислены имя и формат данных каждого столбца, и нет пропущенных значений. ts\_code: код акции trade\_date: дата сделки open: цена открытия high: самая высокая цена low: самая низкая цена close: цена закрытия pre\_close：вчерашняя цена закрытия (до восстановления) change：взлеты и падения pct\_chg：взлеты и падения  vol：объем amount：оборот (тыс. юаней).



Во-первых, измените формат времени транзакции в данных со строки на дату и установите столбца date в качестве индекса.

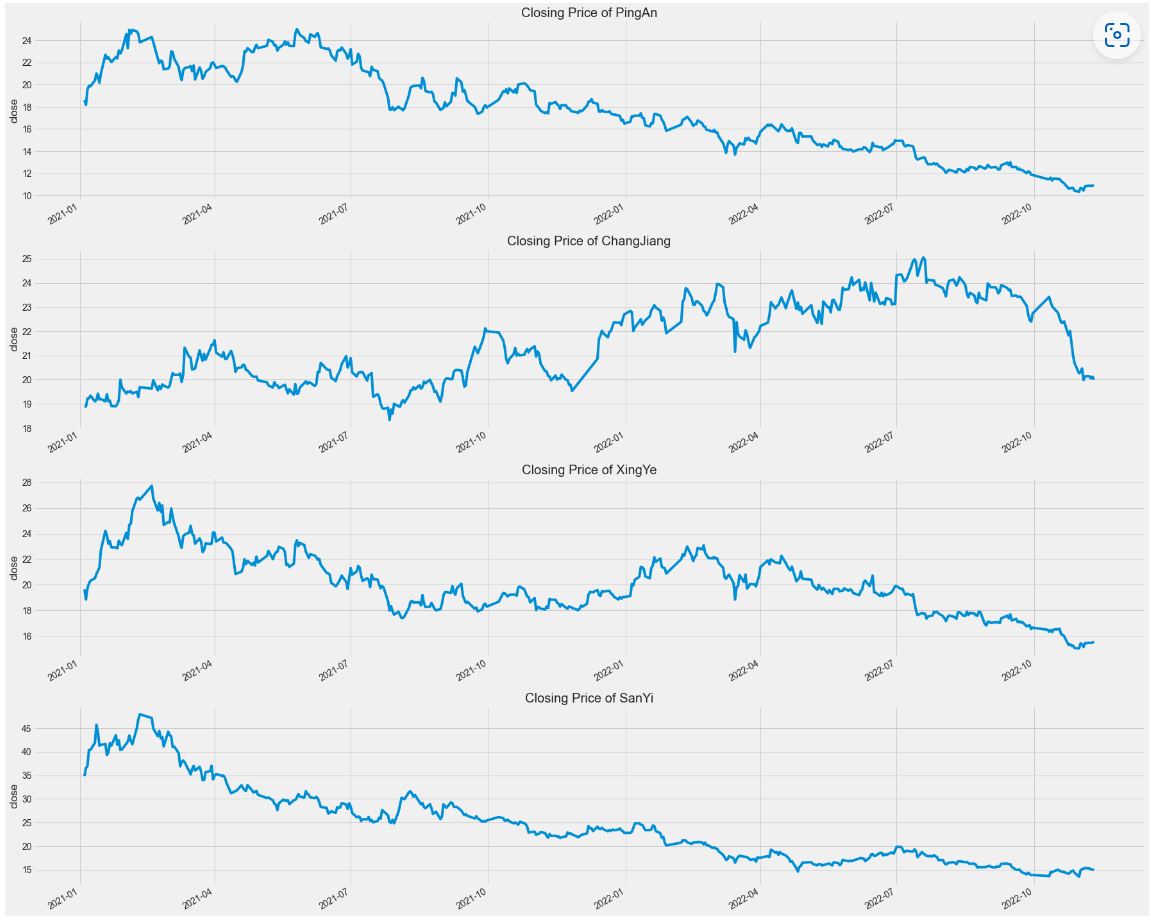


Замените символы ts\_code строки более читаемыми именами.



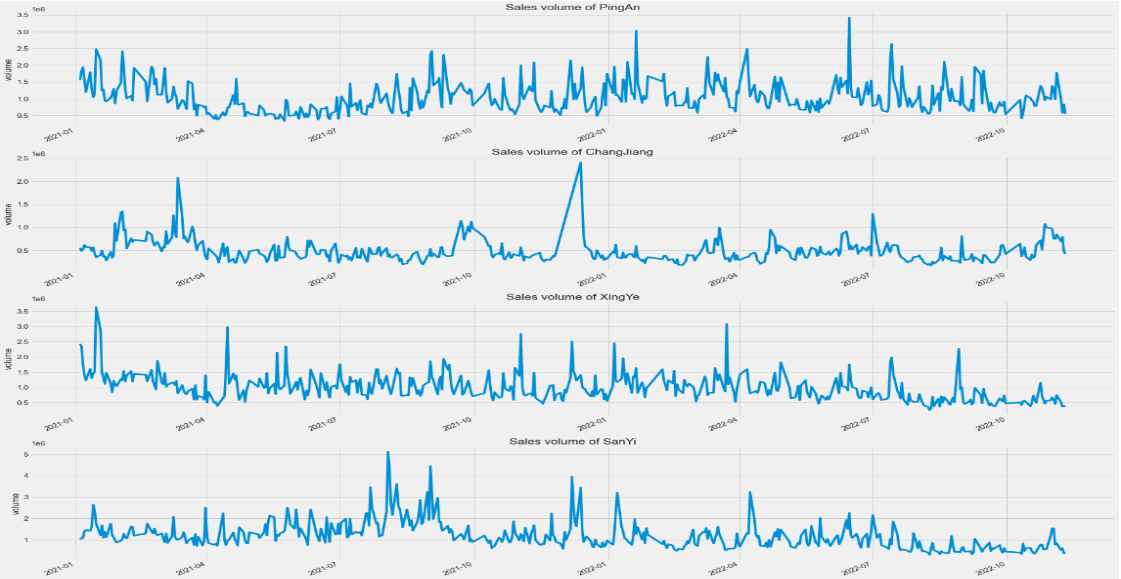
## Постройте график временных рядов цен закрытия акций

Приняв время за абсциссу, а дневную цену закрытия за ординату, построив линейный график, вы сможете наблюдать процесс изменения дневной цены закрытия акций с течением времени.



Из линейного графика видно, что, за исключением ChangJiang, цены закрытия акций остальных трех компаний в прошлом году показали тенденцию к снижению.

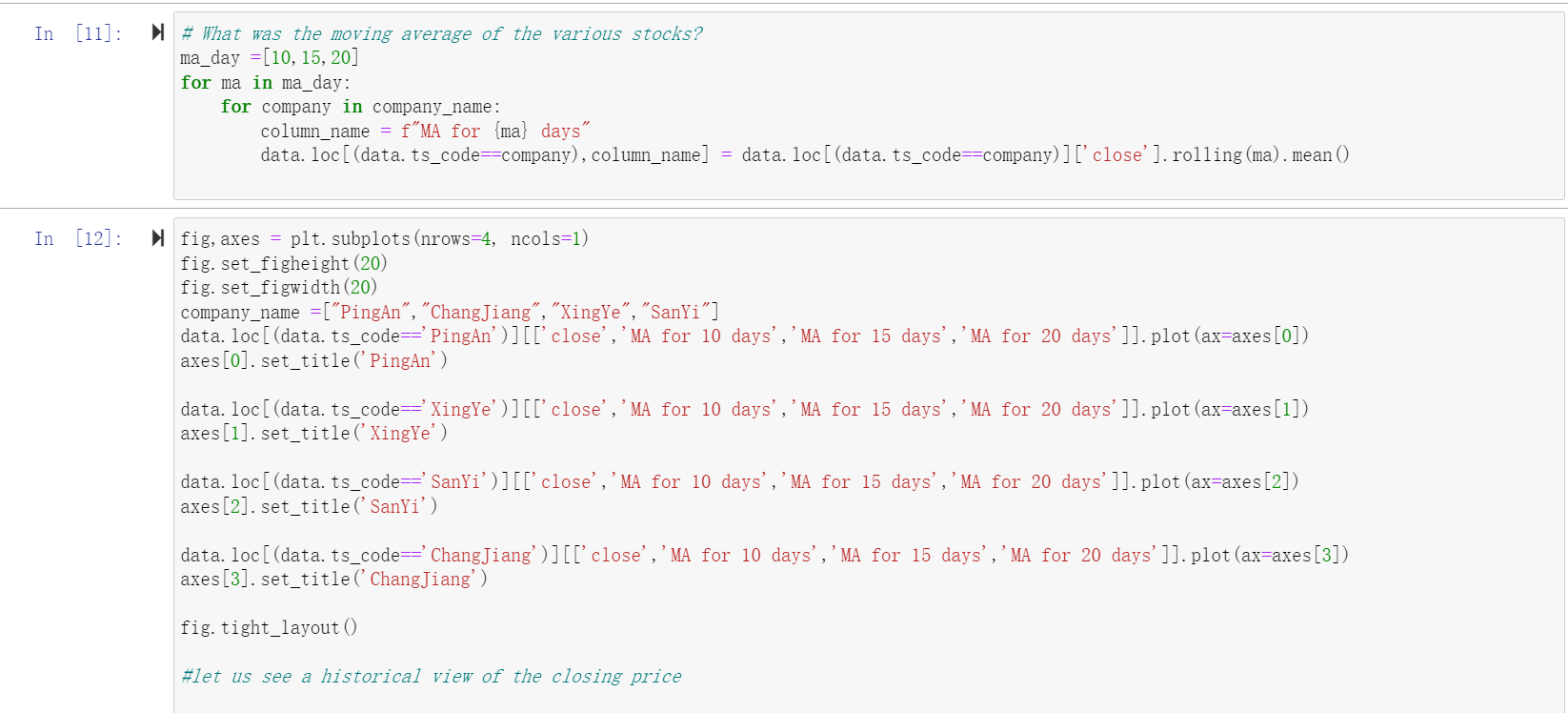
Приняв время за абсциссу, а дневной объем торгов за ординату, построив линейный график, вы сможете наблюдать процесс изменения дневного объема торгов акциями с течением времени.

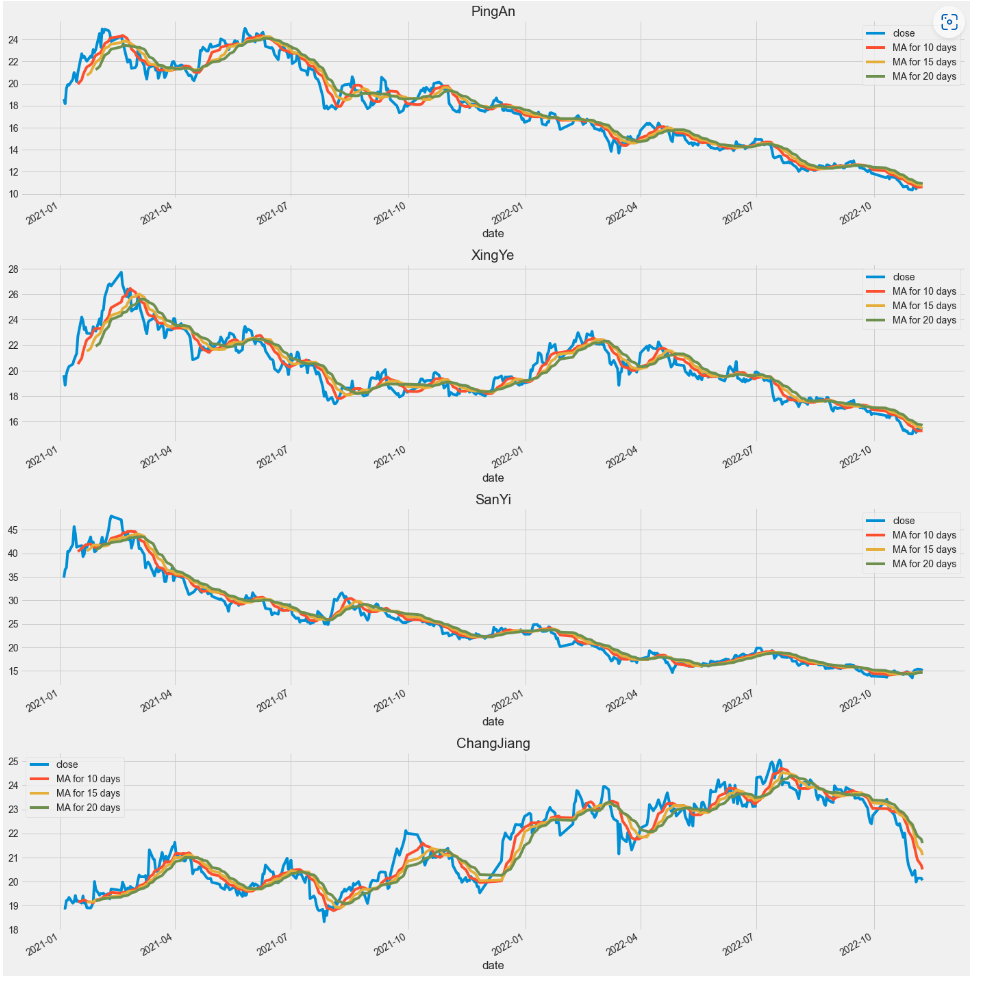


Как видно из линейного графика, ежедневный объем торгов этими четырьмя акциями в прошлом году был относительно стабильным, без особо значительных изменений.

**Скользящая средняя**

Скользящее среднее (MA) — это метод статистического анализа средних цен ценных бумаг (индексов) за определенный период времени и соединения средних значений разного времени для формирования MA для наблюдения за изменениями цен ценных бумаг. Тенденция. Используйте дневную цену закрытия в биржевых данных, чтобы рассчитать средние цены за 10, 15 и 20 дней, и нарисуйте линейный график (также называемый скользящей средней) средней цены.



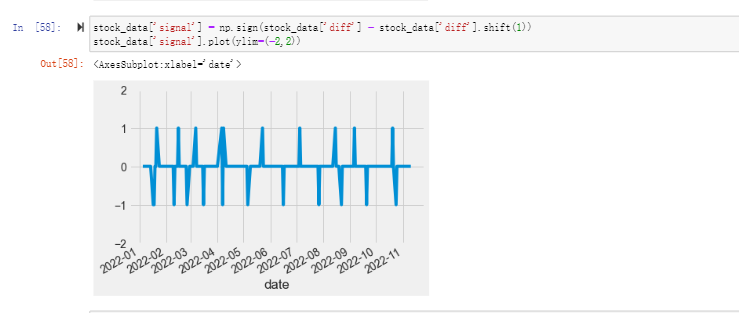


Как видно из линейного графика, цены закрытия акций этих четырех компаний были относительно стабильными.

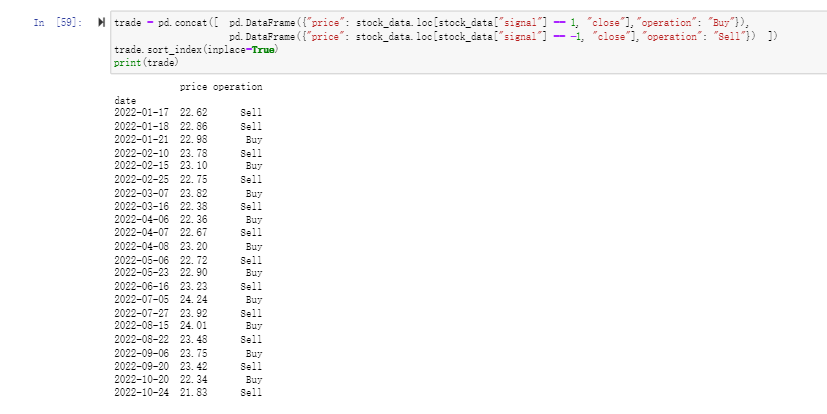
Скользящие средние сглаживают краткосрочные колебания и могут лучше отражать долгосрочные тренды. Посмотрите на график выше и сравните 10-дневную и 15-дневную скользящие средние, и особенно обратите внимание на их пересечения, это время для торговли. Самый простой способ стратегии скользящей средней: когда 10-дневная скользящая средняя превышает 15-дневную скользящую среднюю снизу, покупайте акции, а когда 10-дневная скользящая средняя пересекает 15-дневную скользящую среднюю сверху вниз, продавайте. акции. Для того, чтобы узнать время совершения сделки, мы вычисляем разницу между 10-дневной средней ценой и 15-дневной средней ценой, и берем ее знак и делаем на рисунке ниже. Когда горизонтальная линия на графике прыгает, пора торговать. Выберите Анализ запаса ChangJiang.



Для облегчения наблюдения рассчитанная выше средняя разница цен используется для получения сигнального индикатора путем взятия разницы между соседними датами. Когда сигнал равен 1, это означает покупку акции, когда сигнал равен -1, это означает продажу акции, когда сигнал равен 0, никаких операций не выполняется.

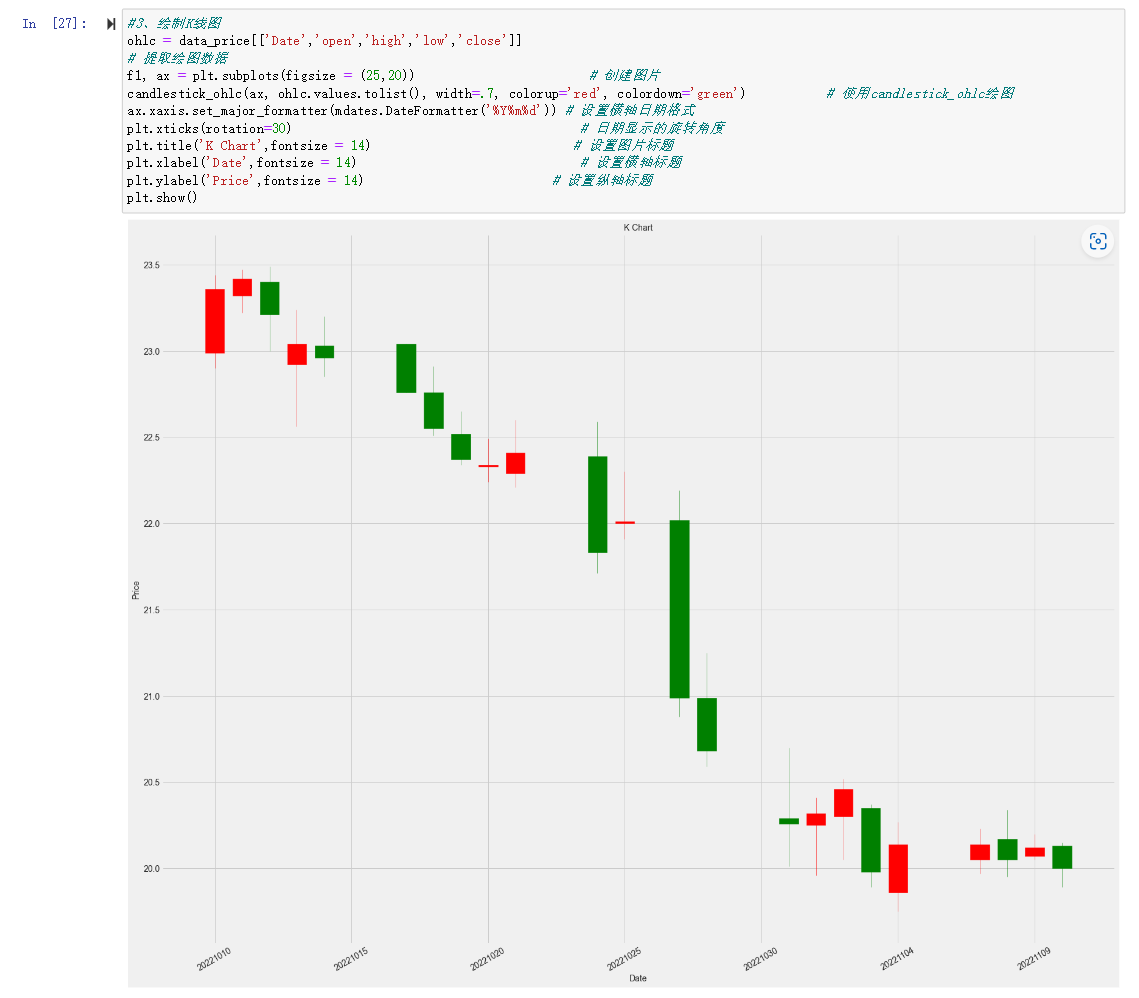


Как видно из диаграммы выше, всего существует 10 раундов возможностей покупки и продажи. Взгляните на детали этих 10 раундов транзакций:



В приведенной выше таблице указаны дата транзакции, операция и цена дня. Однако было обнаружено, что цена продажи двух раундов торгов была меньше цены покупки.По сути, вышеописанный метод был убыточным! Анализ здесь просто для того, чтобы продемонстрировать идею стратегии скользящей средней, а не реальную инвестиционную рекомендацию. Если вы рассматриваете более длительный промежуток времени, например 2 года, 5 лет, и рассматриваете более длинное скользящее среднее, например, сравнивая 20-дневную скользящую среднюю с 50-дневной скользящей средней, хотя в процессе есть потери, вероятность выигрыша больше.

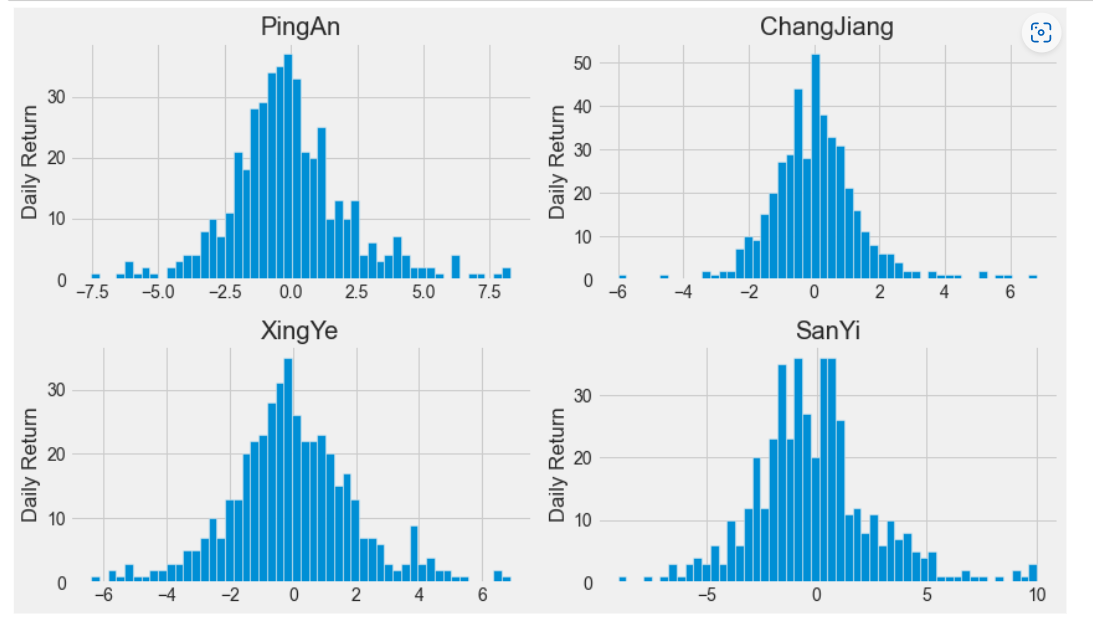
Метод построения графика К-линии: сначала находим максимальную и минимальную цены дня или определенного периода и соединяем их по вертикали в прямую линию, затем находим цены открытия и закрытия дня или определенного периода , и поместите эти две цены вместе, чтобы сформировать длинный и узкий прямоугольный столбец. Если цена закрытия дня или определенного периода выше цены открытия (т. е. минимум открытия и максимум закрытия), мы покажем это красным или оставим незаполненным столбец, который называется «Ян Сянь». Если цена закрытия дня или определенного цикла ниже цены открытия (т. е. максимум открытия и минимум закрытия), мы покажем это зеленым цветом или закрасим черным цветом бар, который представляет собой «линию инь». Значимость: Возможность полностью и подробно наблюдать за реальными изменениями на рынке. На графике К-линии мы можем не только увидеть тренд цены акции (или рынка), но и понять колебания ежедневных рыночных условий.



## Анализ выгод и рисков

Ежедневную норму доходности можно понимать как разницу в цене между двумя днями, деленную на цену предыдущего дня. В pandas, pct\_change() используется для расчета дневной нормы прибыли, хотя загруженные данные содержат ставку возвращения. Давайте сначала посмотрим на изменение и распределение доходности акций с течением времени.

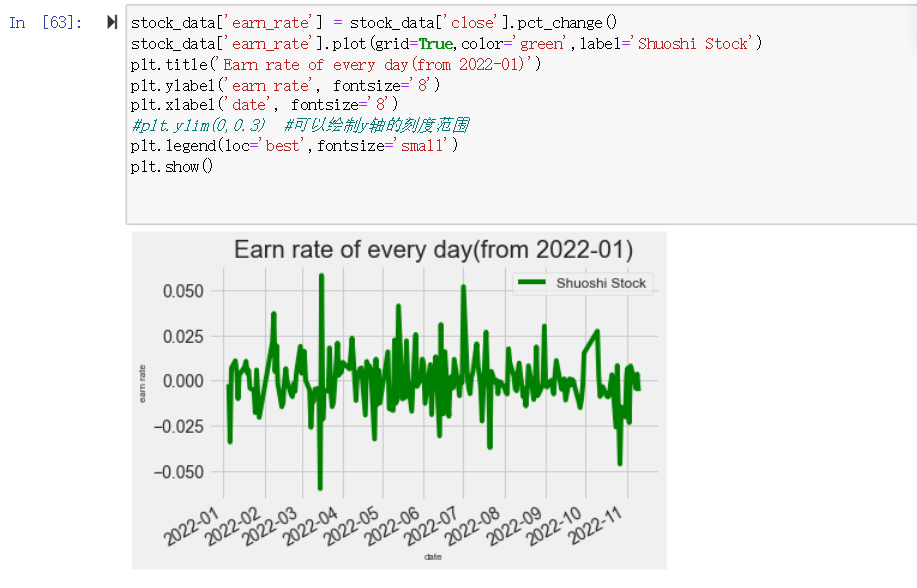




Как вы можете видеть на графике выше, эти четыре акции находятся в анализе Gaussian .

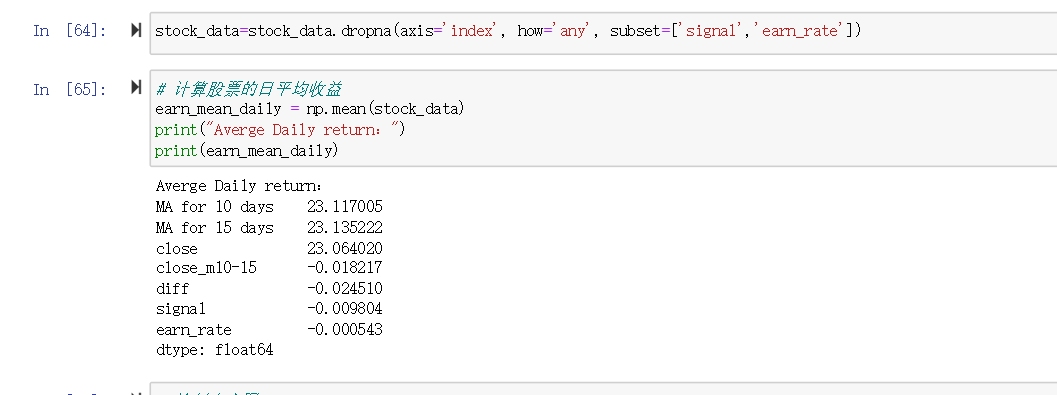
Далее мы в основном анализируем одну ChangJiang из акций.

Во фрейме данных есть дополнительный столбец Earn\_rate, который представляет собой дневную норму доходности акций. Обратите внимание, что доходность за первый день представляет собой отсутствующее значение NaN, поскольку нет данных за предыдущий день, которые можно было бы использовать для расчета. Для удобства последующих расчетов выбираем столбец Earn\_rate, отбрасываем недостающие значения и сохраняем их в новой переменной Earn\_rate\_data. Используйте метод .dropna(), чтобы удалить отсутствующие значения.



**Средняя дневная доходность**

Среднее значение является наиболее часто используемой статистикой.Оно усредняет строку данных и объединяет ее в одно значение, но при этом теряется информация о волатильности данных. Среднее значение исторической доходности акций можно рассчитать с помощью функции mean() в пакете numpy. Можно видеть, что средняя доходность этой акции очень низкая, со значением -0,001890.

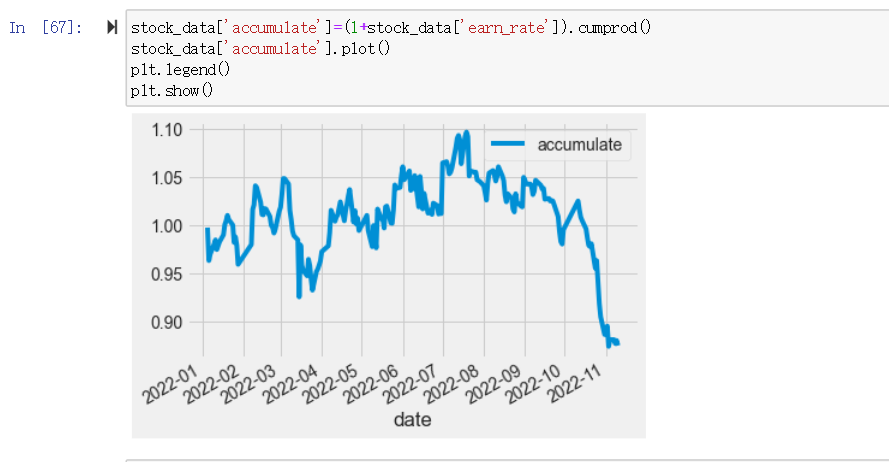


Постройте гистограмму ежедневных доходностей, чтобы понять их распределение, а также наблюдать выбросы в ежедневных доходностях. Как правило, по обеим сторонам распределения дохода есть два длинных хвоста.При инвестировании мы обычно стараемся избегать выбросов на левом хвосте, потому что они представляют собой более крупные потери, а выбросы, распределенные на правом хвосте, обычно вещь, это представляет большую прибыль. Используйте функцию hist() из пакета построения графиков matplotlib, чтобы нарисовать гистограмму.



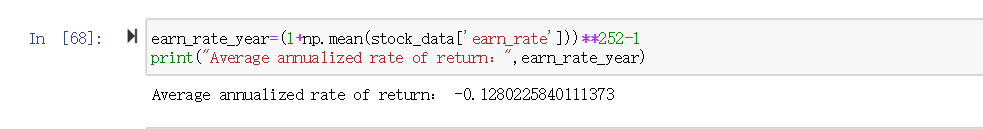
**Рассчитайте совокупную дневную доходность.**

Совокупная дневная норма прибыли помогает определить стоимость инвестиций на регулярной основе. Совокупную дневную норму доходности можно рассчитать, используя числовое значение дневного процентного изменения, просто добавьте к нему 1 и рассчитайте совокупный продукт. Совокупная дневная доходность рассчитывается и строится следующим образом:



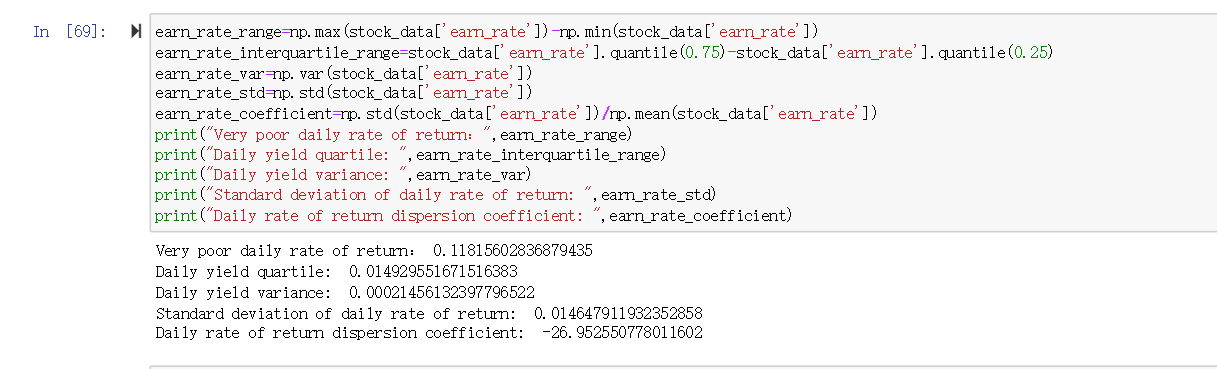
**Средняя годовая доходность акций**

Дневная ставка доходности преобразуется в годовую ставку доходности (обычно предполагается 252 торговых дня в году), где μ — средняя дневная ставка доходности. Формула расчета среднегодовой нормы прибыли выглядит следующим образом:

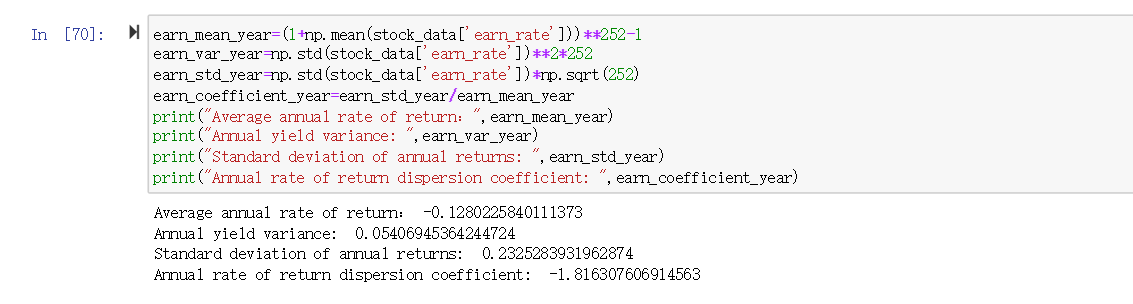


**Оценка фондового риска**

Риск на финансовых рынках – это мера неопределенности, выражающаяся в изменчивости доходности. Как правило, для представления могут использоваться следующие статистические данные: размах, межквартильный размах, разность средних, дисперсия, стандартное отклонение и коэффициент дисперсии, асимметрия, эксцесс и т. д. 1. Вычисление размаха, межквартильного размаха, разности средних, дисперсии, стандартного отклонения и коэффициента дисперсии. Диапазон: Диапазон — это разница между максимальным и минимальным значением в выборке данных. Это самый простой из всех методов. Он отражает числовой диапазон выборки данных и является самым основным способом измерения степени дисперсии данных. Экстремальные значения оказывают большее влияние. Межквартильная разница: то есть разница между верхним и нижним квартилями выборки данных, которая отражает степень дисперсии в средних 50% данных.Чем меньше значение, тем более концентрированы данные, и чем больше значение, тем более концентрированы данные. Чем более дискретны данные, и поскольку медиана находится между квартилями, межквартильный диапазон также отражает степень представления медианы выборке данных. Чем меньше представление, тем выше представление, и чем больше представительство, тем ниже. Средняя разница: сумма абсолютных значений разницы между значениями каждой переменной и среднего, деленная на общее число п. Средняя разница центрируется на среднем, что может полностью и точно отражать дисперсию набор данных. Чем больше средняя разница, тем более дискретны данные. Чем больше значение, тем меньше степень дисперсии. Дисперсия: Дисперсия представляет собой сумму квадратов различий между каждой переменной и средним значением, деленную на общее число n-1, меру степени дисперсии данных. Стандартное отклонение: стандартное отклонение, также известное как среднеквадратическая ошибка, представляет собой арифметический квадратный корень из дисперсии. Более высокое стандартное отклонение доходности инвестиций означает более высокий риск, поскольку данные распределяются дальше от среднего значения, а доходность более изменчива. И дисперсия, и стандартное отклонение могут очень хорошо отражать степень дисперсии данных. Коэффициент дисперсии: коэффициент вариации, представляющий собой отношение стандартного отклонения набора данных к среднему значению. Для стандартного отклонения и дисперсии разных выборок данных результаты нельзя сравнивать напрямую из-за разных единиц измерения данных.Чтобы получить один и тот же индекс измерения, рассчитывается коэффициент дисперсии. Диапазон дневной доходности, межквартильный диапазон, дисперсия, стандартное отклонение и коэффициент дисперсии рассчитываются следующим образом:

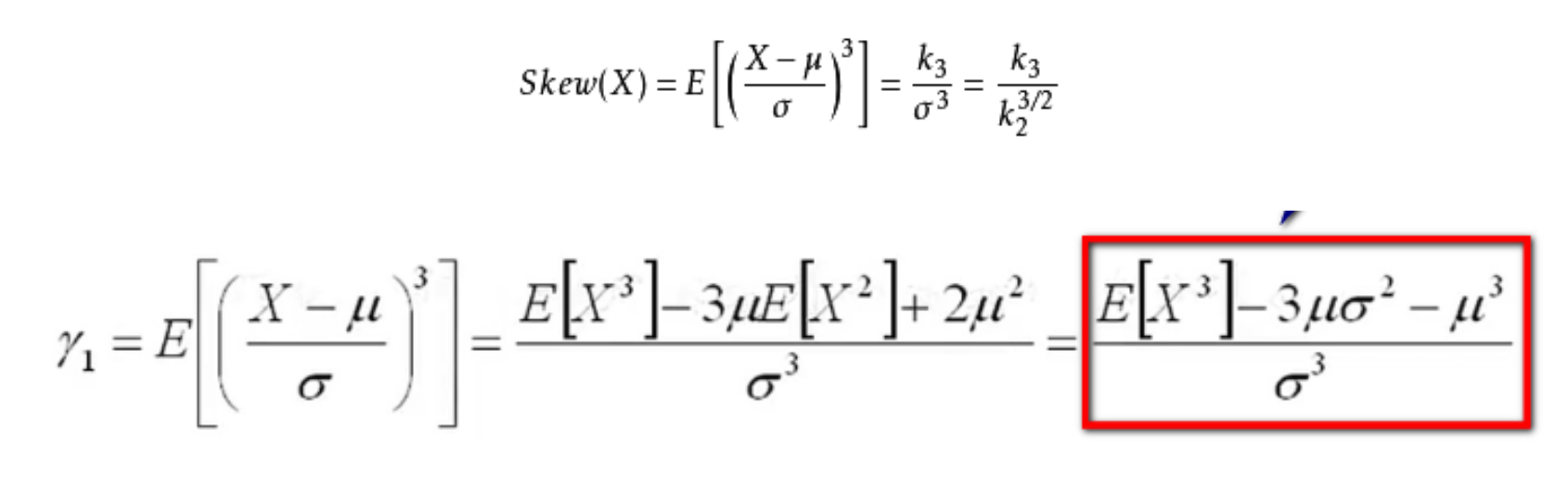


Умножьте стандартное отклонение на квадратный корень из числа торговых дней, чтобы получить годовое стандартное отклонение. Годовая дисперсия получается путем возведения в квадрат годового стандартного отклонения. То есть: средняя годовая норма прибыли, дисперсия годовой нормы прибыли, стандартное отклонение годовой нормы прибыли и коэффициент дисперсии годовой нормы прибыли рассчитываются следующим образом:

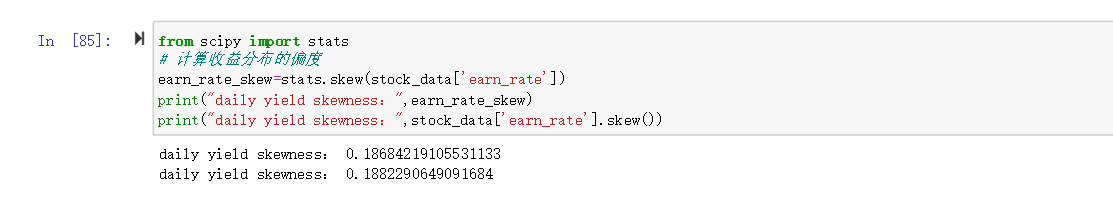


**Асимметрия**

Асимметрия — это мера направления и степени асимметрии в распределении статистических данных, а также числовая характеристика степени асимметрии в распределении статистических данных. Асимметрия по определению представляет собой нормированный момент третьего порядка выборки. Асимметрия измеряет асимметрию распределения вероятностей случайной величины и является мерой степени симметрии относительно среднего значения. Асимметрия нуля указывает на то, что значения относительно равномерно распределены по обе стороны от среднего, но это не обязательно означает, что распределение должно быть симметричным.



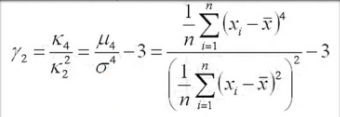
Определение пристрастия включает в себя нормальное распределение (смещение = 0), распределение правого смещения (также называемое распределение положительного смещения, его смещение> 0), распределение левого смещения (также известное как распределение отрицательного смещения, его смещение <0). В финансовой сфере люди более склонны к положительному предвзятости, потому что это означает, что вероятность высокой прибыли больше. Вы можете использовать функцию skew (), предоставленную Scipy.stats для расчета распределения смещения распределения урожайности, или вы можете напрямую использовать функцию Pandas.



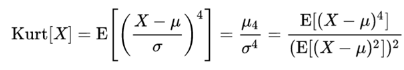
После расчета асимметрии видно, что она имеет небольшую положительную асимметрию.

**Эксцесс**

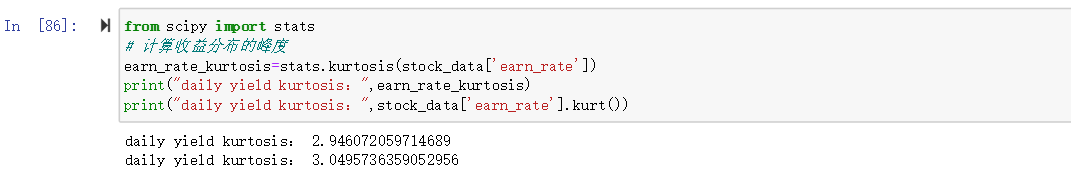
Расчет эксцесса Эксцесс также известен как коэффициент эксцесса. Эксцесс характеризует характерное число высоты пика кривой распределения плотности вероятности при среднем значении, которое отражает резкость пиковой части. . Метод расчета эксцесса случайной величины: отношение центрального момента четвертого порядка случайной величины к квадрату дисперсии минус 3. (Вычтите 3, чтобы сделать эксцесс нормального распределения равным 0). Эту статистику необходимо сравнить с нормальным распределением. Эксцесс, равный 0, указывает на то, что общее распределение данных имеет такую ​​же крутизну, как и нормальное распределение; эксцесс больше 0 указывает на то, что общее распределение данных круче, чем нормальное распределение, т. е. пик; эксцесс менее 0 указывает на то, что общее распределение данных является относительно плоским по сравнению с нормальным распределением и представляет собой плоский пик. Чем больше абсолютное значение эксцесса, тем больше разница между крутизной его распределения и нормальным распределением.



Следующая формула также известна как избыточный эксцесс:

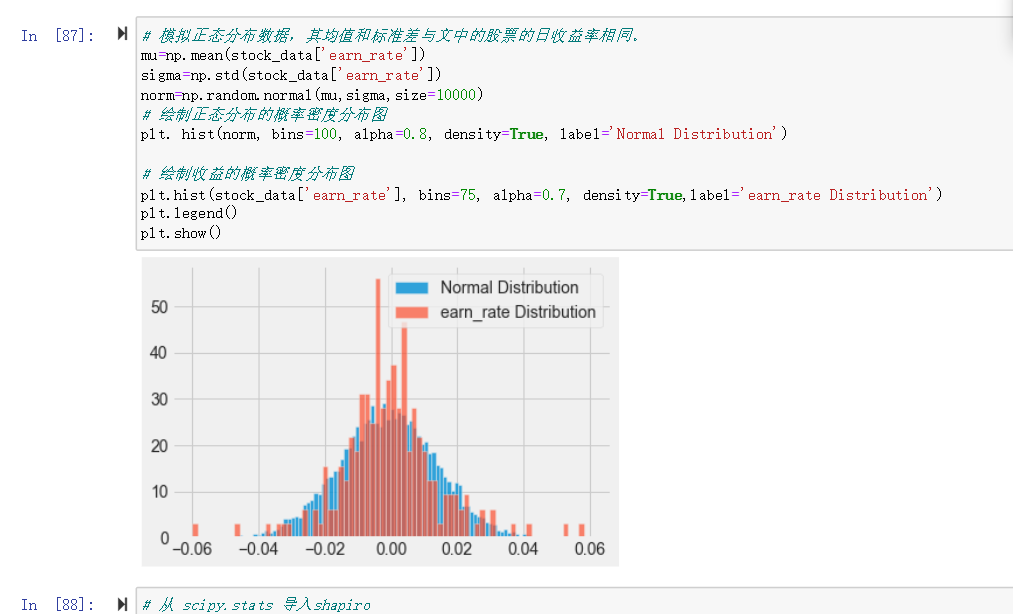


В практических приложениях значение эксцесса обычно вычитается на 3, так что эксцесс нормального распределения равен 0. Поэтому при использовании статистического программного обеспечения для расчета следует обратить внимание на формулу расчета значения эксцесса программного обеспечения по умолчанию. Например, эксцесс нормального распределения по умолчанию для Eviews равен 3.

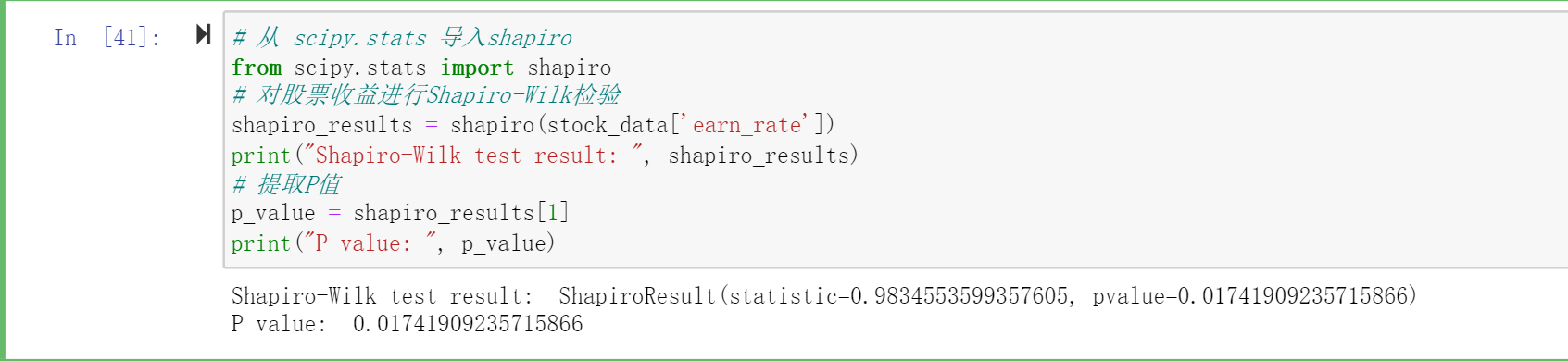


Поскольку эксцесс нормального распределения равен 3, часть, превышающая 3, называется избыточным эксцессом. Большинство финансовых доходов имеют положительный избыточный эксцесс. Избыточный эксцесс распределения можно рассчитать с помощью функции эксцесса (), предоставляемой scipy.stats, или с помощью функции pandas. Результаты расчета приведенного выше эксцесса показывают, что пик дневной доходности акции ниже нормального распределения.

Из сравнения распределения плотности вероятности на рисунке ниже видно, что оранжевый цвет на рисунке представляет собой распределение дневной доходности, а синий цвет представляет собой нормальное распределение.

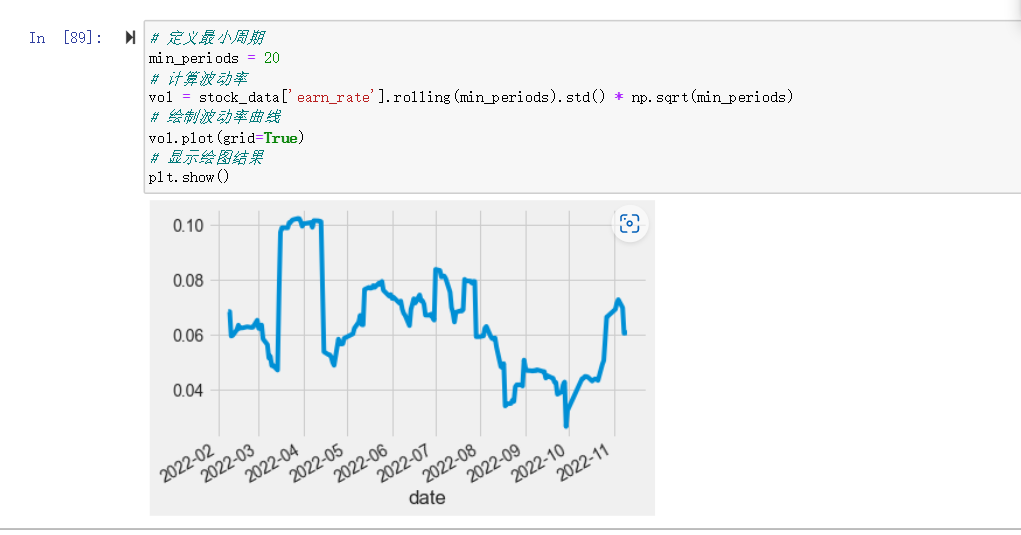


Чтобы судить о нормальности распределения доходности акций, нам нужно использовать настоящие статистические тесты, а не просто проверять эксцесс или асимметрию. При этом используется функция shapiro(), предоставляемая scipy.stats, для выполнения теста Шапиро-Уилка по распределению доходности акций. Функция имеет два возвращаемых значения: одно — это t-статистика теста, а другое — p-значение. Чем он ближе к 1, тем лучше данные соответствуют нормальному распределению.



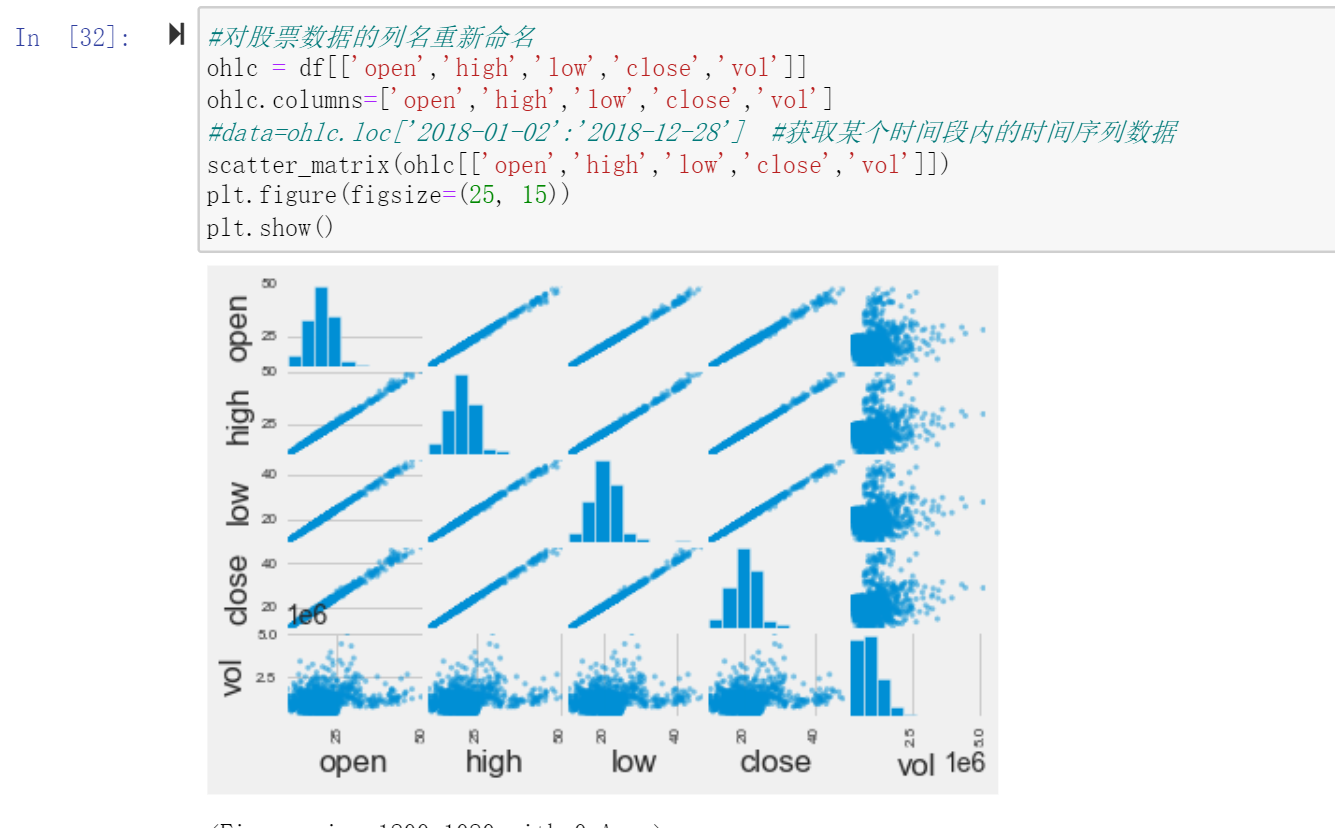
**Расчет волатильности для акций**

Волатильность акции измеряет изменение доходности акции за определенный период времени. Волатильность одной акции часто сравнивают с другой, чтобы найти акции с меньшим риском, или с рыночным индексом, чтобы изучить, как акция движется по рынку. В целом, чем выше волатильность, тем рискованнее вкладывать средства в эти акции, что побуждает людей инвестировать в другие акции. Волатильность получается путем расчета стандартного отклонения скользящего окна процентного изменения акций. Его можно рассчитать и нанести на график, используя следующий код:



## Корреляционный анализ биржевых показателей.

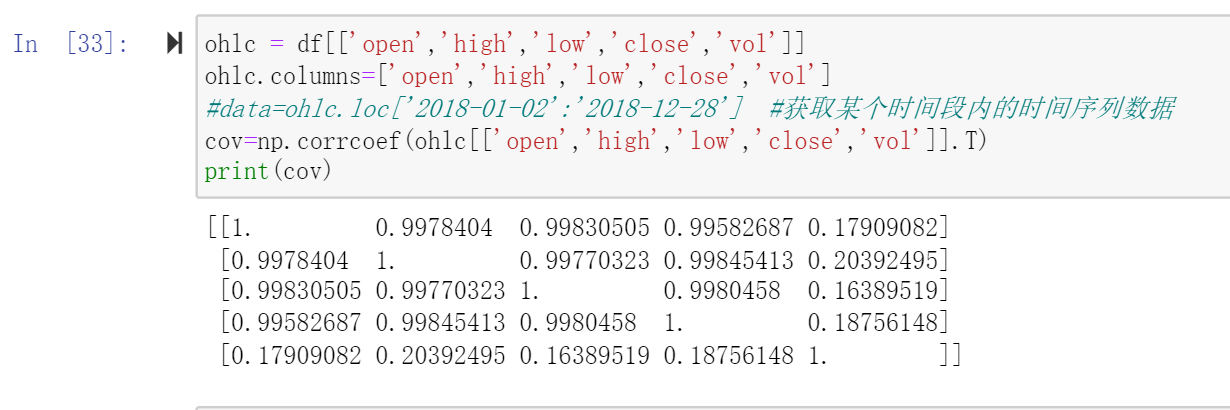
Некоторые репрезентативные индикаторы выбраны ниже, а функция pandas.scatter\_matrix() используется для связывания данных каждого индикатора для построения графика рассеивания, а диагональная линия представляет собой гистограмму данных каждого индикатора. Индикаторы включают в себя: открытие, максимум, минимум, закрытие, объем.



Из рисунка ясно видно, что между открытием, максимумом, минимумом и закрытием существует очень очевидная линейная зависимость; vol не имеет линейной связи с другими параметрами.

**Анализ коэффициента корреляции**

Коэффициент корреляции — это статистический показатель, используемый для отражения тесноты корреляции между переменными. Мы можем использовать numpy.corrcof() для прямого вычисления коэффициента корреляции между данными каждого индикатора.



# Анализ портфеля нескольких акций

В настоящее время финансовый рынок всегда непредсказуем, полон неопределенностей и представляет собой рынок со многими инвестиционными рисками. Это связано с ее собственными рыночными законами и непредвиденными обстоятельствами.Финансовые кризисы, национальная политика и стихийные бедствия - все это повлияет на финансовый рынок и повлияет на возврат инвестиций. Поэтому инвесторы всегда надеются найти способ снизить инвестиционные риски и увеличить доходность. Поскольку простые люди остро нуждаются в теории рационального распределения богатства, обучение оптимизации инвестиций и финансового управления, а также достижение рационального инвестирования является наиболее важным вопросом для инвесторов в настоящее время. Основная проблема оптимизации инвестиций заключается в том, как разумно распределить существующее богатство между инвестируемыми рисковыми активами для достижения инвестиционных целей, таких как максимизация доходов при заданных рисках или максимизация совокупной прибыли. Одной из основных тем современных финансов, а именно проблемой оптимизации инвестиционного портфеля, является изучение того, как рационально покупать и рационально распределять финансовые продукты в условиях неопределенности, чтобы достичь баланса между доходностью и риском. Теорию портфеля также часто называют теорией диверсификации.Ее основная идея состоит в том, чтобы не класть все яйца в одну корзину, то есть изучать, как инвесторы должны распределять имеющиеся средства, когда средства ограничены, а ожидаемая доходность неопределенна.Таким образом избегая рисков в финансовый рынок и максимизация прибыли. Суть инвестирования заключается в том, чтобы брать на себя определенные риски в обмен на компенсацию риска (доходность), и вообще, чем больше доход, тем больший риск приходится нести. Поэтому при принятии инвестиционных решений инвесторы должны идти на компромисс между рисками и доходностью в соответствии со своими обстоятельствами. Для достижения оптимальных инвестиций необходимо количественно оценить инвестиции. Количественные инвестиции — это процесс использования компьютерных технологий и определенных моделей для отработки инвестиционных концепций и реализации инвестиционных стратегий. По сути, количественное инвестирование заключается в попытке найти взаимосвязь между различными факторами и будущей доходностью акций, соблюдая законы рынка, и найти более успешную, то есть закон большей вероятности успеха.

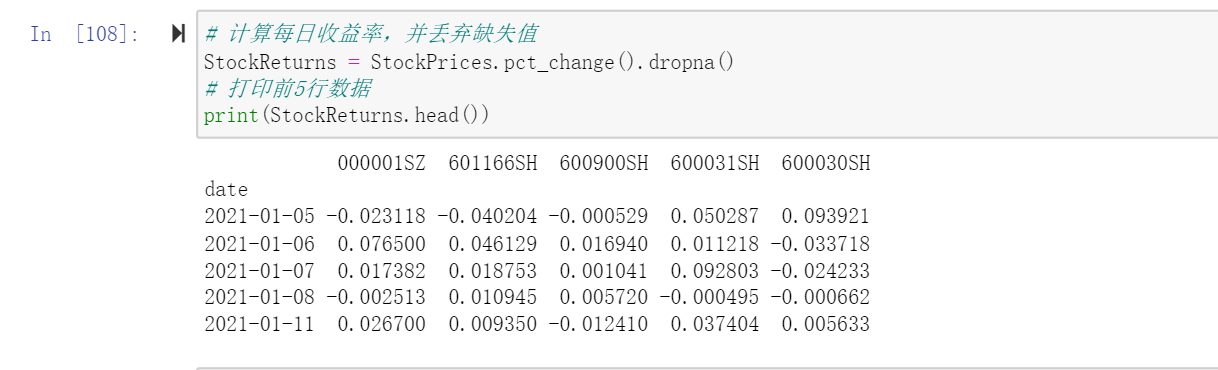
**Получите цену закрытия каждой акции**

Сохраните дневную цену закрытия акций в переменной StockPrices фрейма данных.



**Рассчитать дневную норму доходности акции**

Рассчитайте дневную доходность акции, сохранив данные в переменной StockReturns фрейма данных.



На данный момент мы подготовили для анализа данные StockReturns, в которых зафиксирована ежедневная доходность 5 акций в 2021-2022 гг.

## Расчет доходности портфеля.

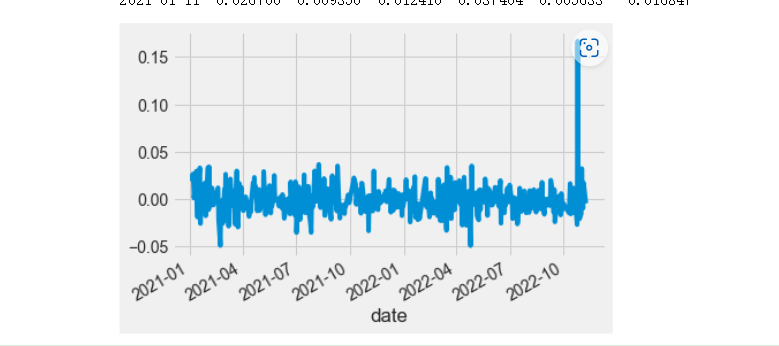
Мы выбрали 5 акций, но как будут распределены средства? Какой купить больше, какой меньше? Для этого необходимо задать им соответствующие веса.Ниже мы используем три схемы распределения весов для расчета доходности инвестиций при различных комбинациях.

**Портфель с заданным весом**

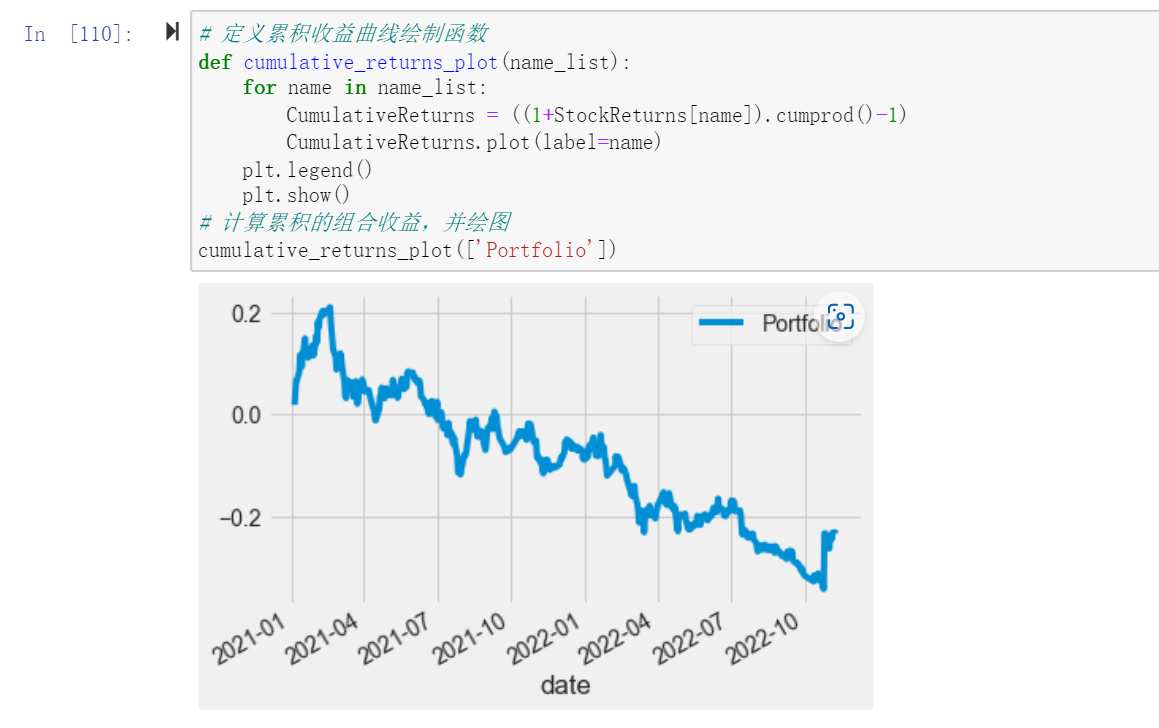
Первый вариант — предварительно установить набор весов (сумма весов всех запасов равна 1). Соответствующие веса этих 5 акций следующие: [0.32, 0.15, 0.10, 0.18, 0.25].



Мы умножаем доходность каждой акции на соответствующий вес, чтобы получить взвешенную доходность акции, затем суммируем взвешенную доходность всех акций, чтобы получить доходность портфельных инвестиций. График доходности портфеля с течением времени выглядит следующим образом:

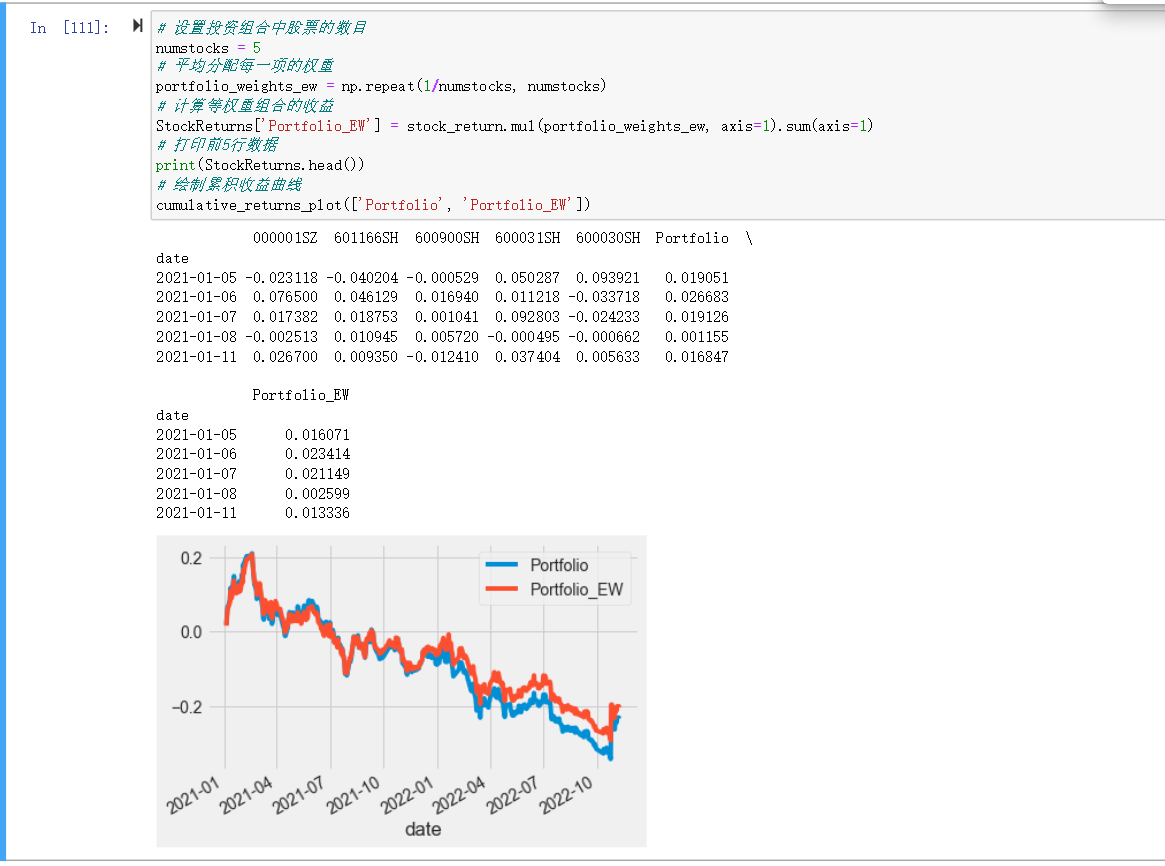


Определите функцию построения кривой кумулятивной доходности cumulative\_returns\_plot() и нарисуйте кривую кумулятивной доходности портфеля с заданным весом.



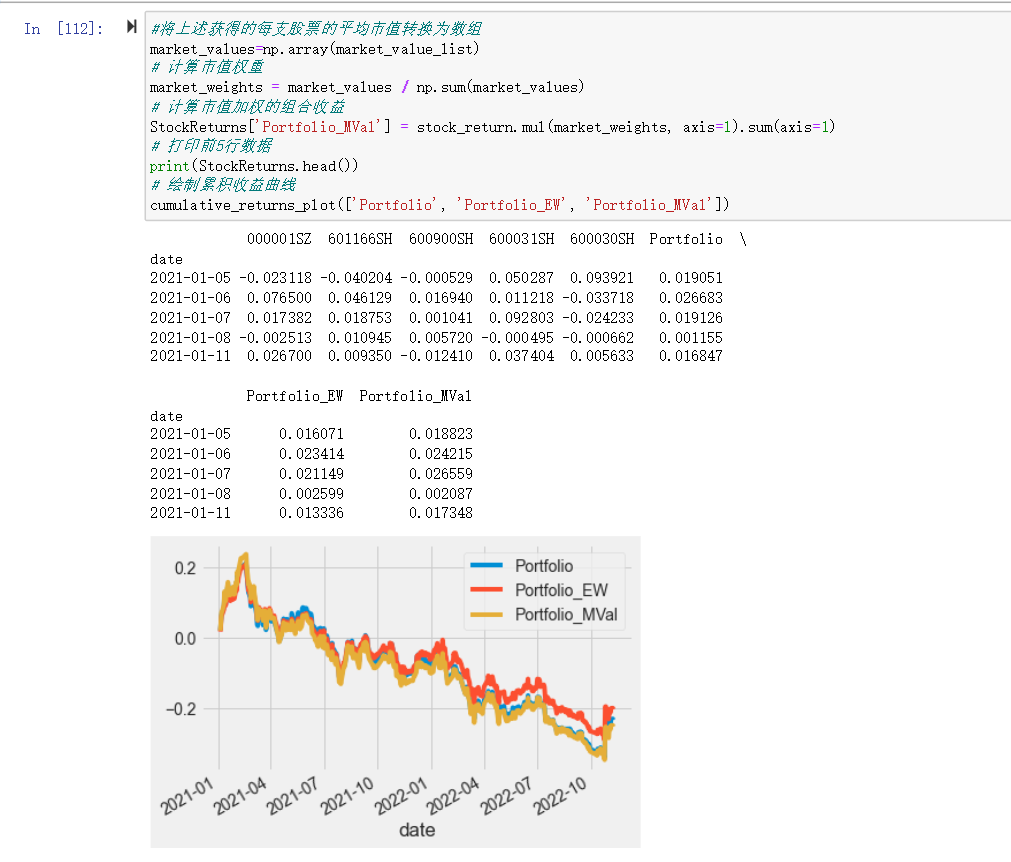
**Равновзвешенный портфель**

Второй вариант — равномерно взвесить каждую акцию, чтобы все они были равны. Это самый простой способ инвестирования, который служит эталоном для других портфелей. Метод расчета такой же, как и выше, просто измените массив, в котором хранятся веса.



**Взвешенный портфель рыночной капитализации**

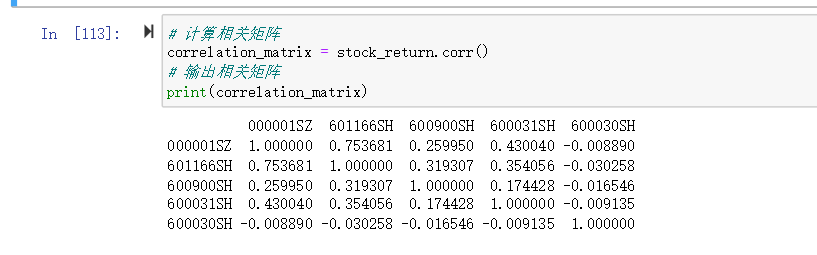
Третий законопроект учитывает рыночную стоимость акций и присваивает веса в соответствии с пропорцией рыночной стоимости. Следовательно, акции с высокой рыночной капитализацией имеют более высокий вес, и когда эти акции с высокой рыночной капитализацией показывают хорошие результаты, портфель работает лучше.



## Корреляционный анализ портфеля.

**Корреляционная матрица для портфеля.**

Матрица корреляции используется для оценки линейной зависимости между доходностью нескольких акций и может быть рассчитана с помощью метода .corr(), встроенного в фреймы данных pandas.

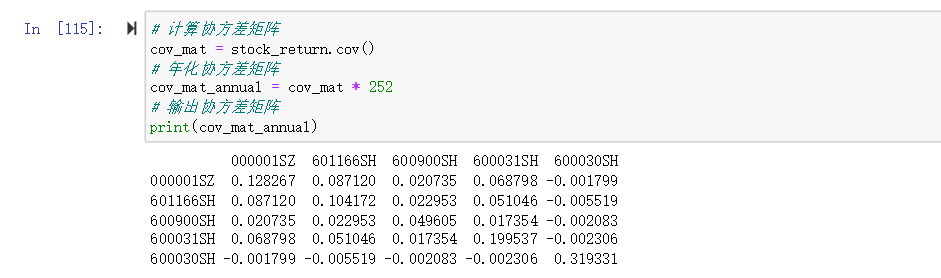


Каждый элемент в матрице представляет собой коэффициент корреляции соответствующей акции в диапазоне от -1 до 1, где положительные числа представляют положительные корреляции, а отрицательные числа представляют отрицательные корреляции. Заметим, что диагональ матрицы всегда равна 1, потому что self и self идеально коррелированы. Кроме того, корреляционная матрица также симметрична, то есть верхний треугольник и нижний треугольник зеркально-симметричны. Для облегчения наблюдения численную корреляционную матрицу можно отобразить в виде тепловой карты. Далее используется пакет seaborn для рисования тепловых карт.



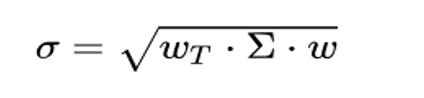
**Ковариационная матрица портфеля**

Коэффициент корреляции лишь отражает линейную зависимость между акциями, но не говорит нам о волатильности акций, а матрица ковариации содержит эту информацию. Матрицу ковариации можно рассчитать с помощью встроенного метода .cov() для фреймов данных pandas.



**Стандартное отклонение портфеля**

Риск портфеля можно измерить стандартным отклонением, которое можно рассчитать по следующей формуле, если известны веса портфеля и ковариационная матрица.



σ：Стандартное отклонение портфеля

⅀：Ковариационная матрица доходности

·： Операция скалярного произведения

W：вес власти портфеля(T Транспонирования)

В NumPy массивы транспонируются с использованием свойства .T, а функция np.dot() используется для вычисления скалярного произведения двух массивов.



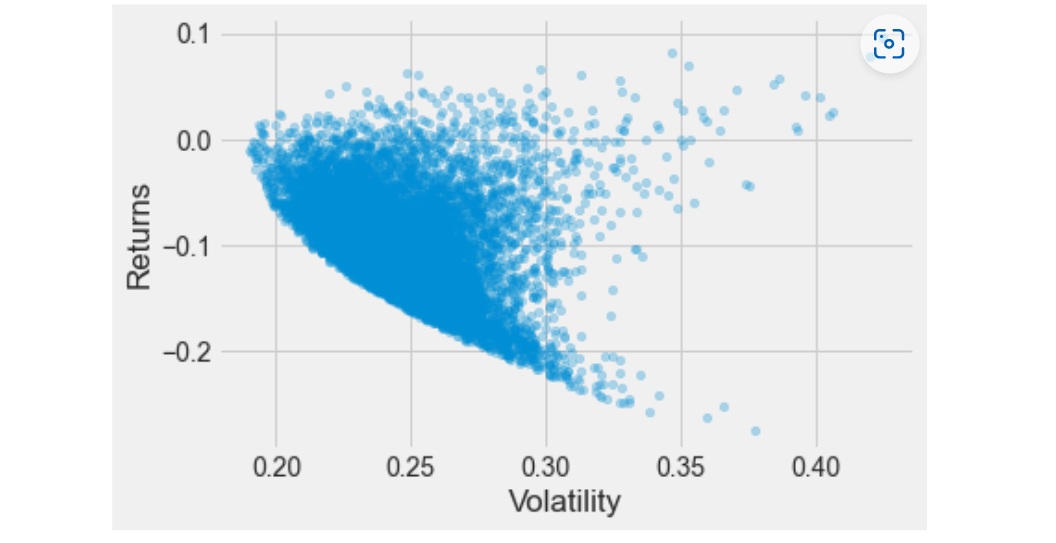
## Исследуйте оптимальный портфель акций

Какой вес портфеля лучше выбрать, чтобы максимизировать доход или минимизировать риск. Нам нужно сбалансировать два фактора, риск и доходность.

Портфельная теория, предложенная лауреатом Нобелевской премии по экономике Марковицем, широко используется при выборе портфеля и распределении активов. Метод анализа средней дисперсии и модель эффективной границы в этой теории могут быть использованы для нахождения оптимального портфеля.

**Используйте метод Монте-Карло для имитации модели Markowitz.** Используйте моделирование Монте-Карло для анализа, то есть случайным образом сгенерируйте набор весов, рассчитайте доходность и стандартное отклонение для комбинации, повторите этот процесс много раз (например, 10 000 раз) и рассчитайте доходность и стандартное отклонение каждой комбинации. График в виде диаграммы рассеяния.

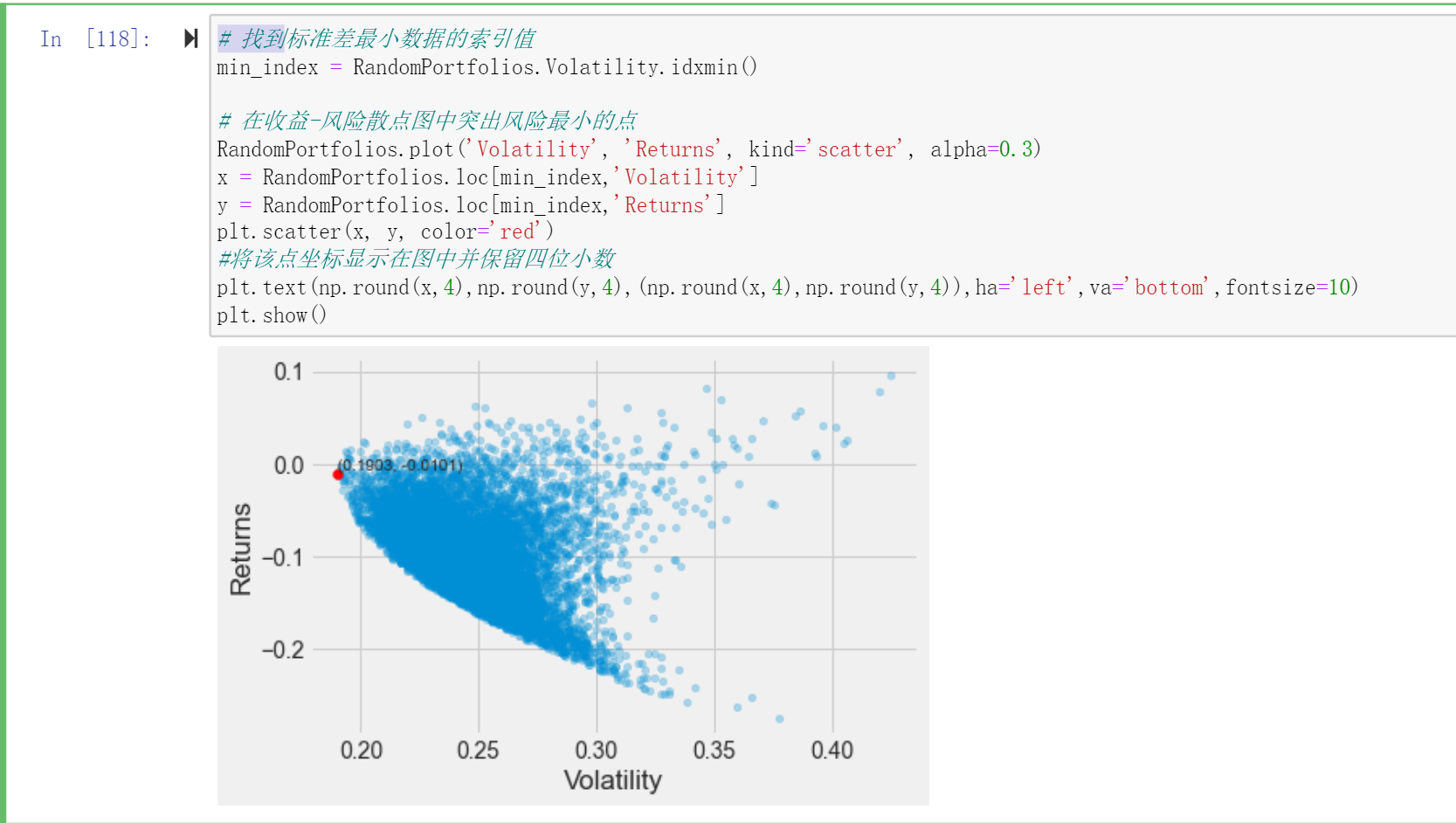




Суть инвестиций заключается в выборе между риском и доходностью, и на картинке выше изображены эти два элемента. Каждая из этих точек представляет ситуацию с портфелем, абсцисса — это стандартное отклонение, представляющее риск, а ордината — норма доходности. Согласно портфельной теории Markowitz , рациональные инвесторы всегда максимизируют ожидаемый доход при заданном уровне риска или минимизируют ожидаемый риск при заданном уровне дохода. На рисунке, который представляет собой границу эффективности, показанную красной кривой, только точки на границе эффективности являются наиболее эффективными портфелями.

**Портфель с минимальным инвестиционным риском**

Одна стратегия состоит в том, чтобы выбрать портфель с наименьшим риском и самой высокой доходностью при этом уровне риска, который называется портфелем с минимальным риском (портфель GMV). Давайте найдем портфель с наименьшим риском и нанесем его на точечный график, представляющий доходность в зависимости от риска.



Веса для получения наименее рискованного портфеля следующие:



**Инвестируйте в лучший портфель**

Коэффициент Шарпа

Коэффициент Шарпа был разработан лауреатом Нобелевской премии Уильямом Шарпом, чтобы помочь инвесторам сравнить доходность и риск инвестиций. Рациональные инвесторы обычно фиксируют риск, который они могут вынести, и добиваются максимальной прибыли, или фиксируют ожидаемую доходность и стремятся к минимальному риску. Таким образом, коэффициент Шарпа рассчитывает избыточную доходность на единицу общего принятого риска. Рассчитывается следующим образом:

Коэффициент Шарпа =

: Ожидаемая норма прибыли.

: Безрисковая ставка.

: Стандартное отклонение избыточной доходности.

Числитель вычисляет разницу, которая представляет собой избыточный доход, сравнивающий инвестиции с эталоном, представляющим весь класс инвестиций. Стандартное отклонение знаменателя представляет собой волатильность доходности, которая соответствует риску, поскольку большая волатильность указывает на больший риск.

Просто разделите среднее значение избыточной доходности на его стандартное отклонение, чтобы получить коэффициент Шарпа, который измеряет доходность и риск. Кроме того, умножьте на sqrt(252) (252 торговых дня в году), чтобы получить годовой коэффициент Шарпа.

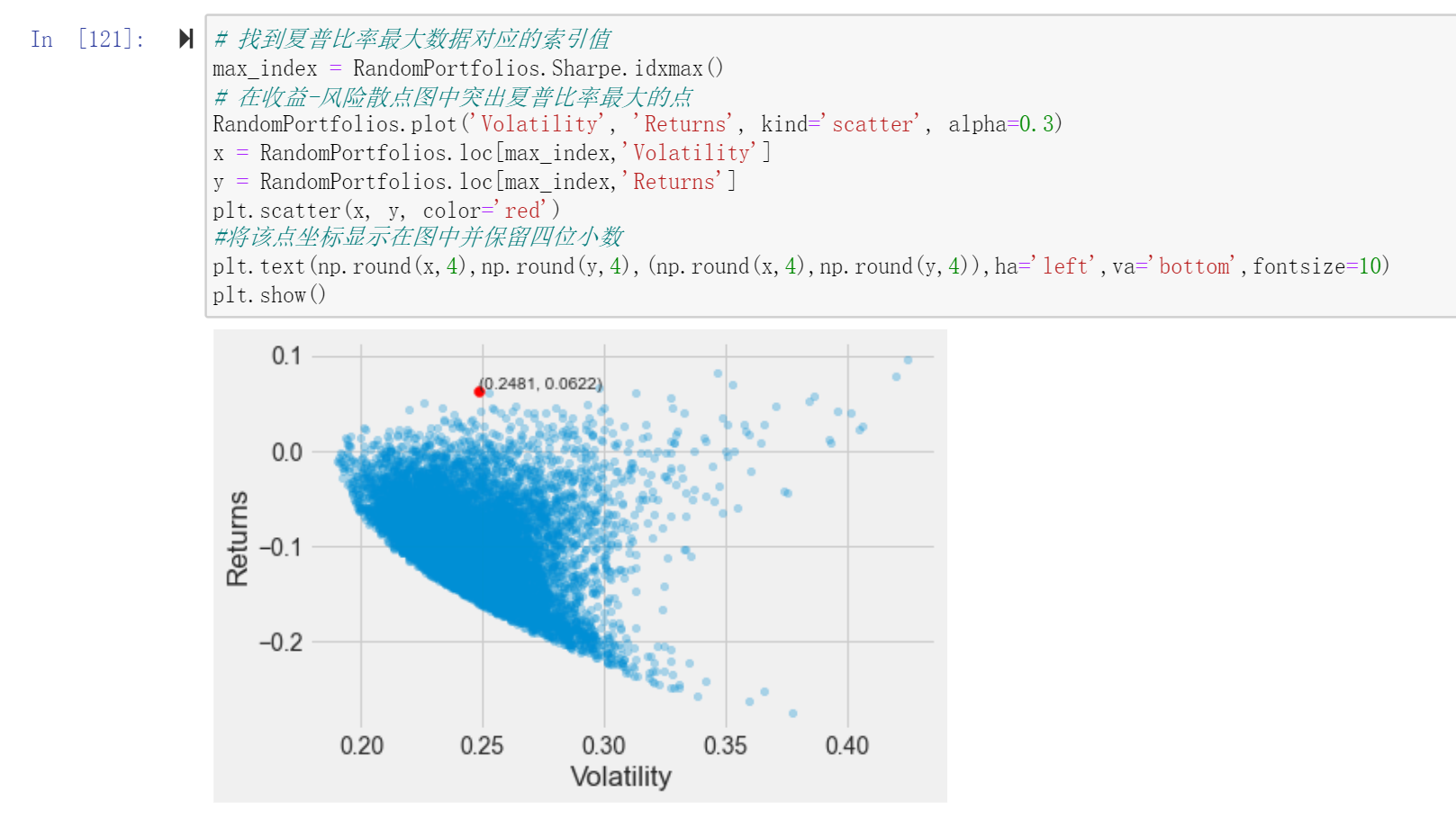
**Выбор оптимальной комбинации Sharp**

На самом деле, мы хотим найти баланс между доходностью и риском.Коэффициент Sharp—это переменная, которая помогает мне принимать более правильные решения.Он вычисляет избыточную доходность на единицу принятого риска.

Сначала мы вычисляем коэффициент Sharp для комбинации вышеприведенных симуляций Монте-Карло и наносим его как третью переменную на диаграмме рассеяния риска доходности, используя цвет в качестве визуальной подсказки для характеристики коэффициента Sharp.



Было обнаружено, что комбинации на верхнем краю графика рассеяния имеют более высокие коэффициенты Sharp. Затем найдите портфель с наибольшим коэффициентом Sharp и нанесите его на диаграмму рассеяния риска доходности.



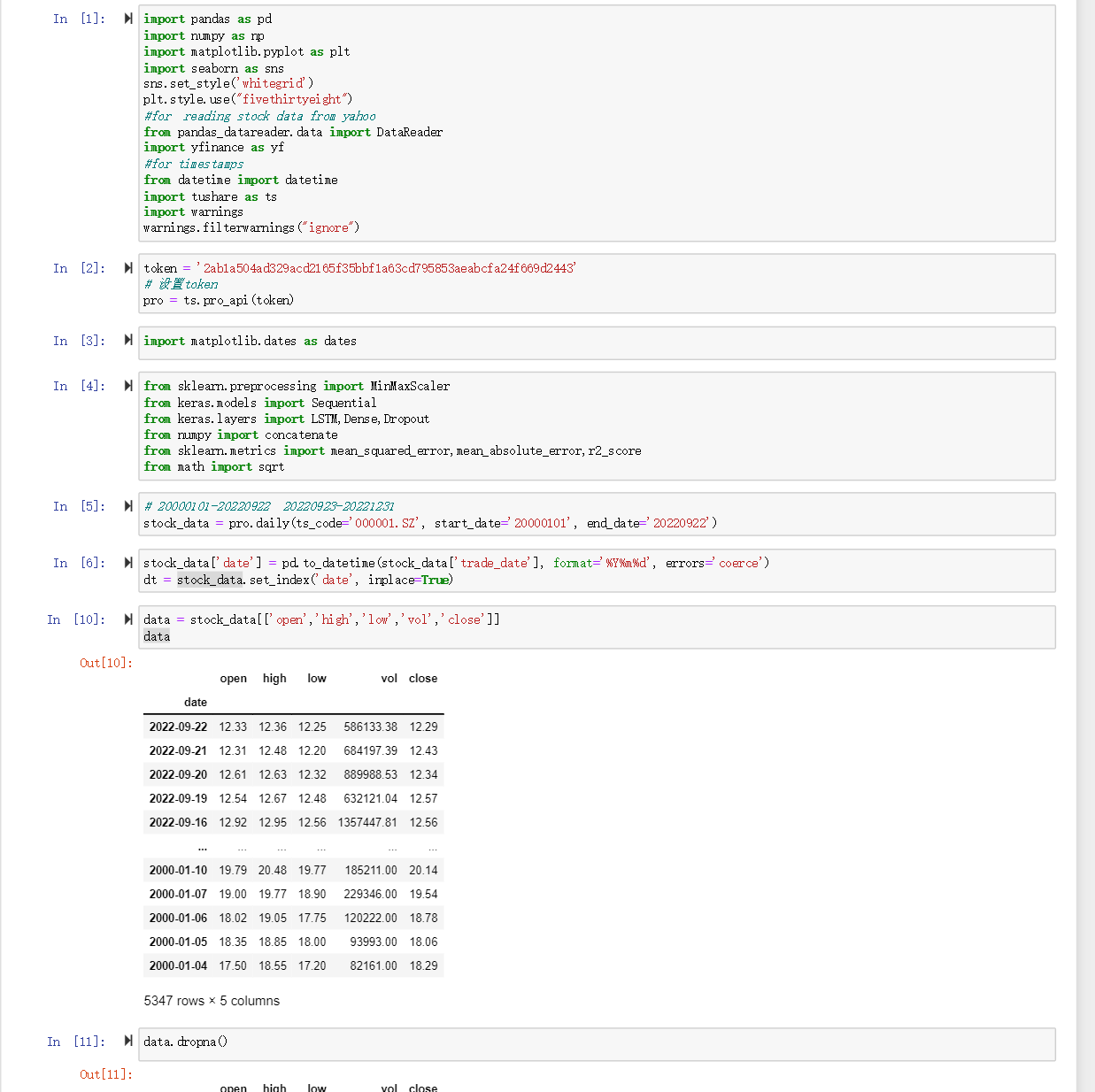
Веса для получения комбинации с наибольшим коэффициентом Шарпа следующие:



# Прогноз данных о запасах

LSTM — это рекуррентная нейронная сеть, способная обрабатывать долгосрочные зависимости. Допустим, вы смотрите фильм, и когда в фильме что-то происходит, вы уже знаете, что было раньше, и понимаете, что происходит что-то новое из-за того, что было в прошлом. RNN работают таким же образом, они запоминают прошлую информацию и используют ее для обработки текущего ввода. Проблема с RNN заключается в том, что они не могут запоминать долгосрочные зависимости из-за исчезающих градиентов. Поэтому LSTM разработан, чтобы избежать проблемы долгосрочной зависимости.

Импорт данных и простая обработка：



Разделите данные на тестовые и обучающие наборы с коэффициентом распределения 0,2.

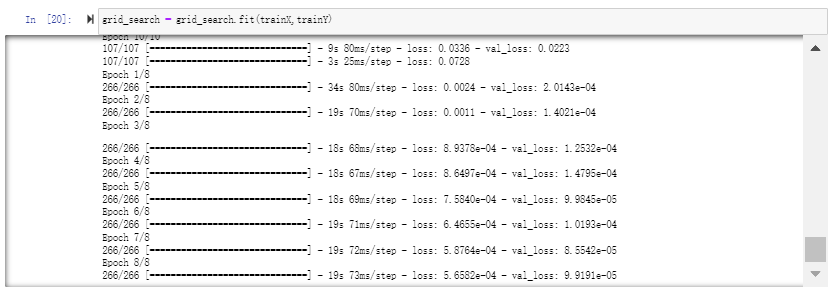
Можно заметить, что диапазоны данных очень большие и они не масштабируются в одном диапазоне, поэтому, чтобы избежать ошибок прогнозирования, давайте сначала масштабируем данные с помощью MinMaxScaler. (также можно использовать StandardScaler).

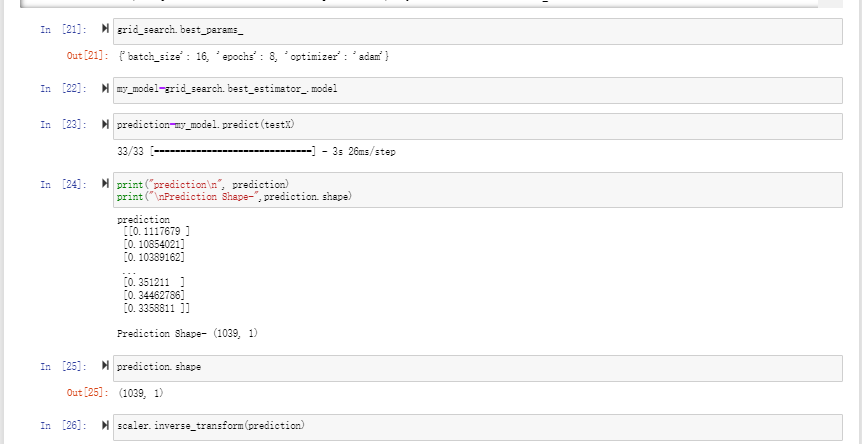


Создайте модель нейронной сети с помощью LSTM:

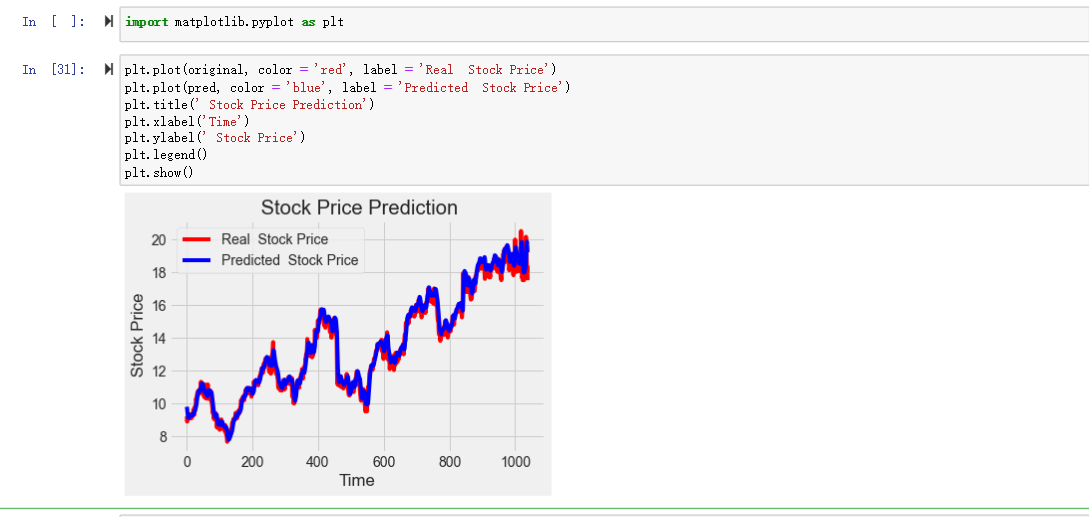


Обучение модели, параметр batch\_size: 16 эпох: 8 оптимизатор: адам можно увидеть на правом рисунке, это настройка параметра с лучшими результатами.

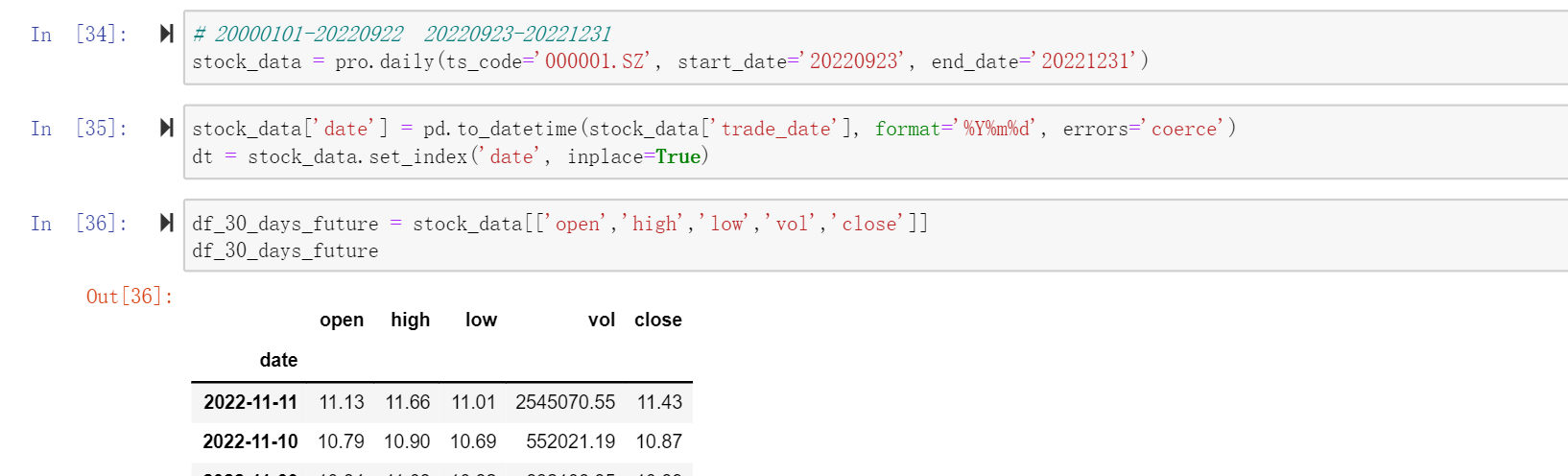




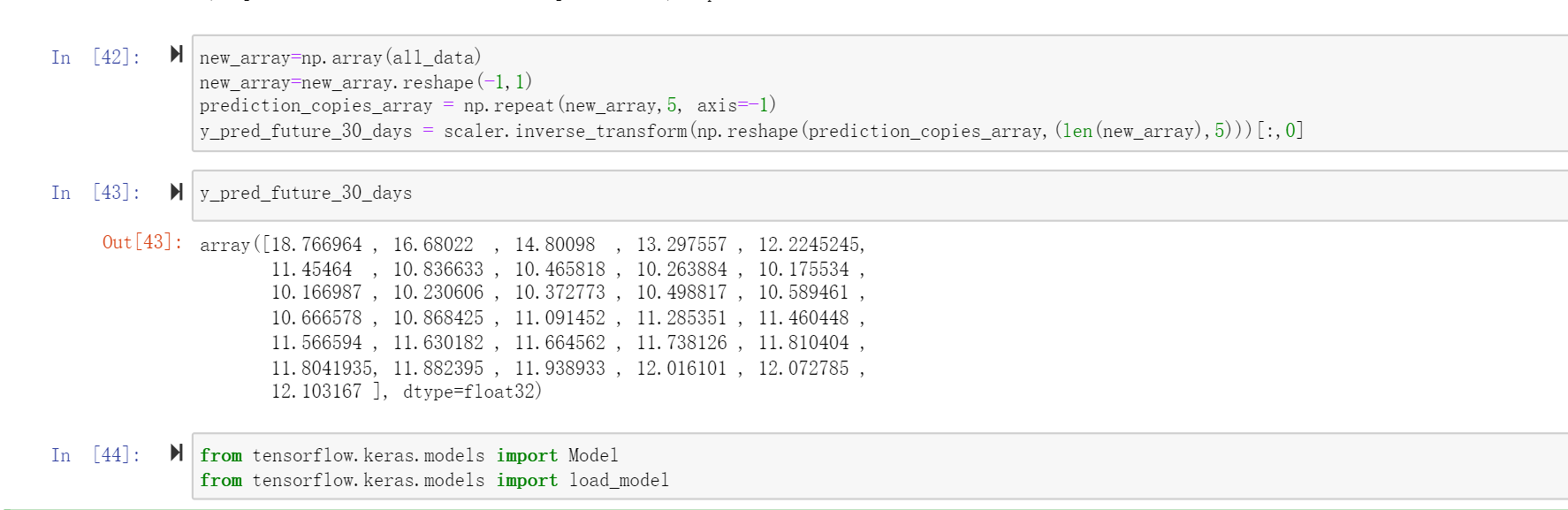
Просмотр реальных и тестовых значений:



Предсказать значение на 30 дней и сохраните модель







# Заключение

Риск инвестирования в акции имеет очевидную двойственность, то есть его существование объективно и абсолютно, но также субъективно и относительно, он одновременно неустраним и управляем. Стратегия портфеля средней дисперсии ненадежна при оценке выборки, и даже ее эффективность не так хороша, как у классической стратегии равных весов, которая создает много проблем для оптимизации портфеля. Полный процесс создания портфеля состоит из двух частей: выбора акций и установки инвестиционных весов. Необходимо объединить данные оценки акций из нескольких источников, то есть исторические данные транзакций, технические индикаторы, данные социальных сетей и новости, для анализа инвестиционной стоимости акций из двух частей: выбор акций и установка веса портфеля, а также стремится создать более эффективный инвестиционный портфель. Ориентируясь на крайне нелинейный, высокий уровень шума и динамический характер цен на акции, модель сети с долговременной кратковременной памятью (LSTM) используется для прогнозирования цен на акции.Экспериментальные результаты показывают, что использование нейронных сетей LSTM для прогнозирования цен на акции имеет хороший эффект и может быть использован для инвестиций.дайте ссылку.

# Список использованных источников

1.https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm/notebook.

2.https://blog.csdn.net/qq\_35240555/article/details/124652270

3.https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm.

4. https://tushare.pro/document/2