

НИС

Танирбергенова А. , Пентюхов В.И.

11 ноября 2023 г.

Содержание

1 Цели проекта	1
2 Что мы использовали	1
3 Инструменты	2
4 Задачи	2
4.1 Предобработка	2
4.2 Поиск выбросов, ошибок в данных	2
4.3 Расчет показателей для каждого инструмента: для работы с первым датасетом: YTM, Base For Accruals, Annualizer Profitability. Для второго: Base For Accruals, Annualizer Profitability, Purchase price.	2
4.4 Сортировка наших инструментов по мере их убывания по показателю PP.	2
4.5 Поиск решения проблемы с данными.	2
5 Результаты	3

1 Цели проекта

Целью нашего проекта является рассмотрение цены каждого из активов, проанализировать их, отсортировать по убыванию цены, а также понять, есть ли в данной нам информации ошибки, и если есть, то попробовать их исправить.

2 Что мы использовали

Yield to maturity - это доходность к погашению облигации, которая учитывает ее текущую цену, номинальную стоимость, срок погашения и выплаты купонов. Это показатель, который позволяет инвесторам оценить потенциальную доходность от инвестиции в облигацию, если они будут держать ее до даты погашения. Purchase price - это цена покупки, по которой инвестор приобретает облигацию. Данные для расчета этих значений предоставлены в наших датасетах. Мы можем просто воспользоваться формулами:

$$YTM = \frac{\text{annualizedprofitability}}{\text{"base" foraccruals}}$$

$$AP = \frac{C \cdot N + (N - PP)}{T}$$

$$BFA = \frac{N + PP}{2}$$

$$PP = \frac{2 \cdot C \cdot N + 2 \cdot N - YTM \cdot T \cdot N}{YTM \cdot T + 2}$$

AP - annualized profitability

BFA - base for accruals

C - coupon rate

N - face value or nominal

PP - purchase price

T - years to maturity

3 Инструменты

Мы используем библиотеки языка Python, это позволяет работать со всеми инструментами сразу и помогает визуализировать наши результаты графически: pandas(библиотека, необходимая нам для работы с таблицами вида xlsx (Excel-таблицы) и красивой визуализации таблиц), matplotlib.pyplot(библиотека для визуализации данных, построения графиков).

4 Задачи

У нас 2 датасета, но принцип работы с ними схож.

1. **Предобработка данных**
2. **Поиск выбросов, ошибок в данных.**
3. **Расчет показателей для каждого инструмента: для работы с первым датасетом: YTM, Base For Accruals, Annualizer Profitability. Для второго: Base For Accruals, Annualizer Profitability, Purchase price.**
4. **Сортировка наших инструментов по мере их убывания по показателю PP и YTM.**
5. **Поиск решения проблемы с данными.**

4.1 Предобработка

Загружаем необходимые нам библиотеки и выгружаем наши данные в формате таблицы.

4.2 Поиск выбросов, ошибок в данных

Здесь мы с помощью графиков показываем, что у нас имеются ”выбросы” или ошибки в данных, так как они принимают необычные значения, отличные от всех остальных точек, значения которых находятся в каком-то определенном диапазоне.

4.3 Расчет показателей для каждого инструмента: для работы с первым датасетом: YTM, Base For Accruals, Annualizer Profitability. Для второго: Base For Accruals, Annualizer Profitability, Purchase price.

Реализуем данные функции "def AnnualizerProfitability", "def BaseForAccruals", "def YTM", "def PurchasePriceFromYTM", найдем все необходимые нам значения. Вновь реализуем таблицу с нашими полученными результатами.

4.4 Сортировка наших инструментов по мере их убывания по показателю PP.

Производим сортировку элементов по нужному признаку в порядке убывания для того чтобы выяснить какой имеет лучшие значения. Воспользуемся функцией sorted и отсортируем массив в порядке убывания по цене покупки, далее просчитываем значения, которые мы должны были посчитать с помощью функции для новой таблицы уже с отсортированными значениями.

4.5 Поиск решения проблемы с данными.

Решением проблемы будет либо удалить инструмент который выбивается, либо заменить его неверное значение медианным, которое мы тоже можем увидеть на графике и которое как раз входит в один диапазон значений со всеми другими инструментами.

5 Результаты

- Dataset №1: Мы нашли выброс в датасете и решили просто исключить инструмент с такими значениями, т.е. Инструмент №6 удаляется, его значение face value аномально мало. Также, находим самый выгодный инструмент с точки зрения годовой прибыли. После всех наших махинаций этим элементом является Инструмент №16, но это не говорит о том, что остальные Инструменты менее прибыльные в общем.
- Dataset №2: В данном датасете встречается не только выброс, а ошибочное значение, так как оно не просто отличается от всех остальных, а имеет неверный тип значения. УТМ Инструмента №13 принимает текстовый тип значения, а не числовой, также Инструмент №18 имеет очень малое значение coupon rate, оба этих элемента в последствии удаляются. Производим сортировку по РР и относительно годовой прибыли выбираем самый прибыльный - это Инструмент №16.