* 1. На самом деле здесь все гораздо проще. В таблице есть столбец под названием instrument\_type и тут нужно просто посчитать количество уникальных значений.

Пример:

Instrument\_type count

digital-option 235699

crypto 22300

forex 1934

И тем самым можно узнать самые используемые инструменты. Это можно реализовать без сложных функций. Вроде в pandas как раз есть функция, которая так считает.

И отдельно можно сделать тоже самое с группировкой по пользователям.

То есть в конечно итоге мы получим топ-инструментов в общем и топ по юзерам.

1.2 Проделать тоже самое, что и для 1.1, только для region\_name

1.3 PNL это просто deal\_investment – deal\_result. Таким образом мы получим PNL для одной строчки. Теперь можем суммировать этот PNL по различным столбцам.

Пример:

User\_id … asset deal\_investment deal\_result ---------------🡪 PNL

435456 … EURUSD 5 2 3

323344 … GBPJPY 12 6 6

434534 … EURUSD 8 0 8

233445 … USDCHF 4 1 3

322323 … EURUSD 23 2 21

131231 … GBPJPY 11 0 11

Если посчитаем сумму по asset, получим:

Asset PNL

EURUSD (3+8+21) 32

GBPJPY 17

USDCHF 3

32, 17, 3, тоесть получается по EURUSD больше всего денег. PNL.

Это только для стоблца asset. Теперь тоже самое нужно проделать для instrument\_type, platform, region\_name.

1.4 Нужно по deal\_close отсечь все даты кроме мартовских, а потом посчитать сумму PNL по юзерам.

Пример:

User\_id … PNL

234235 4

676578 -3

234353 10

5683748 1

34434 -12

234235 -1

234353 2

5683748 -23

34434 6

…. ….

Получим:

User\_id PnlSum

234235 3

676578 -3

234353 12

5683748 -22

34434 -6

Получается, что самый успешный трейдер 5683748, потому что он выигрывает, значит мы проигрываем и значит знак минус.

1.5 Сначала нужно для каждого пользователя посчитать его Trading Volume. То есть посчитать для каждого user\_id сумму всех его deal\_investment. И присоединить этот новый столбец с trading volume к нашей таблице. Потом составить новый столбец, где у нас будут (deal\_investment – deal\_result)/trading volume. И по этому столбцу вместе с deal\_close\_time построить график.

И по графику можно будет посмотреть изменения.

2. Эта задачка посложнее. Тут нужно прям поанализировать. Но что-то в принципе можно сделать. Загрузить всю эту таблицу в пандовский dataframe. Сделать так, чтобы места, где пропуски заполнялись нулями.

То есть там, где

,,, 🡪 0,0,0

$12,, 🡪 12,0,0

$11,,$2 🡪 11,0,2

И соответсвенно знаки долларов убрать. Чтобы формат был численный и с ним можно было работать.

Посчитать количество строк disabled, min\_10, min\_20.

Посчитать количество уникальных user\_id для каждой из этих 3 групп

После этого посчитать сумму стоблца deposits по стоблцу state (там всего 3 значения disabled, min\_10, min\_20).

После этого посчитать сумму стоблца binary\_PNL по стоблцу state (там всего 3 значения disabled, min\_10, min\_20).

После этого посчитать сумму стоблца instrument\_PNL по стоблцу state (там всего 3 значения disabled, min\_10, min\_20).