### Построение аналитики для корпоративных клиентов IT Resume

Корпоративные клиенты постоянно приходят к нам и просят выгрузить им то одну, то другую аналитику по своим студентам. Какие задачи решают, сколько попыток, сколько времени, какой прогресс и так далее. С разбивкой по месяцам. А теперь за последний месяц. А теперь за весь период сразу. Ну и так далее - каждый раз что-то новое.

Кстати, частично вы уже анализировали это в одной из глав - но это только кусочек.

Ваша задача - поставить себя на место преподавателя корпоративного клиента, который очень переживает за процесс обучения своих студентов. Ему очень важно знать, какие сложности испытывают студенты, что у них получается, в каком месте нужно допилить обучающую программу и так далее. А еще поставьте себя на место финансового директора - возможно, ему будет интересно, а за что именно они нам платят деньги. А может быть еще что-то будет интересно посмотреть генеральному директору? А маркетологу? А методистам?

Короче, вам нужно подготовить разностороннее исследование, чтобы наши корпоративные клиенты были довольны и все вопросы у них сразу отпали. В качестве подопытного можете взять клиента с id=1.

### Краткое саммари

**Цель:** Провести масштабный анализ успеваемости и прогресса студентов корпоративного клиента (id = 1) для предоставления исчерпывающей аналитики, которая поможет улучшить процесс обучения, оптимизировать программу и повысить удовлетворенность клиента.

**Метрики:**

1. **Прогресс студентов**:
   * Количество завершенных задач из начатых.
2. **Сложности решения задач у студентов**:
   * Задачи с наибольшим количеством попыток.
   * Среднее количество попыток на задачу.

**Гипотезы**:

1. Студенты испытывают сложности в определенных темах.
2. Некоторые задачи требуют доработки из-за высокой сложности или недостаточной ясности.

### Часть 1. Анализ прогресса студентов

Из таблицы codesubmit определим сколько задач решил каждый студент, а из таблиц coderun+codesubmit выясним суммарное количество попыток на выполнение задач.

Определим максимальное количество задач, решённое пользователями компании 1.

**SELECT** **COUNT**(**DISTINCT** *c*.problem\_id) **AS** *count* **FROM** codesubmit *c*

**JOIN** users *u* **ON** *u*.id = *c*.user\_id

**WHERE** *u*.company\_id = 1



Посмотрим, сколько задач решил каждый пользователь. Отсортируем по убыванию количества задач, чтобы было понятно сколько максимально решили задачи данные пользователи.

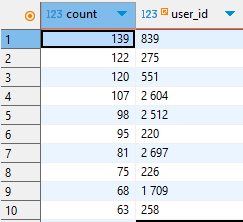
**SELECT** **COUNT**(**DISTINCT** *c*.problem\_id) **AS** *count*, *c*.user\_id **FROM** codesubmit *c*

**JOIN** users *u* **ON** *u*.id = *c*.user\_id

**WHERE** *u*.company\_id = 1

**GROUP** **BY** *c*.user\_id

**ORDER** **BY** *count* **DESC**



Изменим запрос так, чтобы посчитать среднее количество задач на пользователя:

**WITH** *cnt* **AS** (

**SELECT** **COUNT**(**DISTINCT** *c*.problem\_id) **AS** *count*, *c*.user\_id **FROM** codesubmit *c*

**JOIN** users *u* **ON** *u*.id = *c*.user\_id

**WHERE** *u*.company\_id = 1

**GROUP** **BY** *c*.user\_id

**ORDER** **BY** *count* **DESC**

)

**SELECT** **AVG**(*count*) **FROM** *cnt*



Получается, что в среднем пользователь компании 2 решает 31 задачу из всех 205 (то есть примерно 15%) и 31 задачу из 139 максимальных (то есть примерно 22%).

### Часть 2. Анализ сложности решения задач у студентов

Определим суммарное количество попыток на решение задач у данных пользователей:

**WITH** *codes* **AS** (

**SELECT** *c*.created\_at, *c*.problem\_id, *c*.user\_id, 'cr' **AS** *TYPE*

**FROM** coderun *c*

**UNION** **ALL**

**SELECT** *cs*.created\_at, *cs*.problem\_id, *cs*.user\_id, 'cs' **AS** *TYPE*

**FROM** codesubmit *cs*

), *total\_table* **AS** (

**SELECT** created\_at, problem\_id, user\_id, type

**FROM** *codes* *c*

**RIGHT** **JOIN** users *u* **ON** *u*.id = *c*.user\_id

**WHERE** *u*.company\_id = 1 **AND** created\_at **IS** **NOT** **NULL**

**ORDER** **BY** user\_id, problem\_id, created\_at

), *solutions* **AS** (

**SELECT** **COUNT**(created\_at) **AS** *count*, problem\_id, user\_id

**FROM** *total\_table*

**GROUP** **BY** user\_id, problem\_id

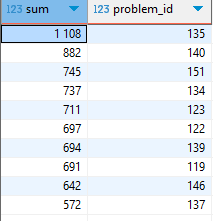
)

**SELECT** **SUM**(*count*) **AS** *sum*, problem\_id

**FROM** *solutions*

**GROUP** **BY** problem\_id

**ORDER** **BY** **SUM**(*count*) **DESC**



**Вывод:** Задачи 135, 140, 151, 134, 123, 122, 139, 119, 146, 137 решали большее количество раз. Можно сделать вывод, что данные задачи сложные.

Найдём среднее количество попыток на решение задач (включая обе таблицы coderun и codesubmit):

**WITH** *codes* **AS** (

**SELECT** *c*.created\_at, *c*.problem\_id , *c*.user\_id, 'cr' **AS** *type*

**FROM** coderun *c*

**UNION** **ALL**

**SELECT** *cs*.created\_at, *cs*.problem\_id, *cs*.user\_id, 'cs' **AS** *type*

**FROM** codesubmit *cs*

),

*total\_table* **AS** (

**SELECT** created\_at, problem\_id, user\_id, *type*

**FROM** *codes* *c*

**RIGHT** **JOIN** users *u* **ON** *u*.id = *c*.user\_id

**WHERE** *u*.company\_id = 1 **AND** created\_at **IS** **NOT** **NULL**

**ORDER** **BY** user\_id, problem\_id, created\_at

),

*solutions* **AS** (

**SELECT** **COUNT**(created\_at) **AS** *count*, problem\_id, user\_id

--, type

**FROM** *total\_table*

**GROUP** **BY** user\_id, problem\_id

--, type

)

**SELECT** **AVG**(*count*)::**NUMERIC** **AS** *avg*, problem\_id

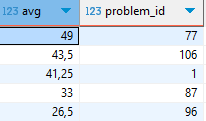
**FROM** *solutions*

**-- WHERE *"type"* = 'cr'**

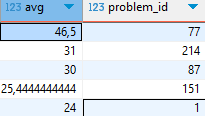
**GROUP** **BY** problem\_id

**ORDER** **BY** **AVG**(*count*)::**NUMERIC** **DESC**

**LIMIT** 5



Если добавить условие type = ‘cr’, то получим среднее количество попыток:



**Вывод:** Следует обратить внимание на задачи с id 77, 106, 1, 87, 96, 151 и 214, возможно добавить подсказок или обратить внимание на формулировки.