Задание

Вы работаете аналитиком данных в крупном ретейлере *SkyLenta*, который присутствует во многих российских городах.

Вы работаете в отделе продуктовой аналитики, и ваша текущая задача — проанализировать A/B-тест, проведенный во всех городах.

Описание данных

**Вкладка «Данные»**

* *id\_order —* уникальный идентификатор покупки.
* *id\_client —* уникальный идентификатор клиента.
* *amt\_payment* — размер платежа.
* *dtime\_pay* — дата и время оплаты.

**Вкладка Clients**

* *id\_client —* уникальный идентификатор клиента.
* *dtime\_ad —* дата и время показа рекламного объявления (или баннером в приложении, или пуш-уведомлением, в зависимости от группы).
* *nflag\_test* — группа эксперимента (0 — контроль, 1 — тест).
* *id\_trading\_point* — идентификатор торговой точки, к которой прикреплен данный пользователь.

**Вкладка Region\_dict**

* *id\_trading\_point* — уникальный идентификатор торговой точки, к которой прикреплен данный пользователь.
* *city —* название города.

Цель эксперимента

Исследование альтернативного метода воздействия на клиентские покупки с помощью пуш-уведомлений.

Воздействие **«контроль»** — уведомление о новых товарах и скидках с помощью баннера в приложении.

Воздействие **«тест»** — уведомление с помощью пуша (сообщение о товарах и скидках появится в уведомлениях приложения).

Дизайн эксперимента

Длительность эксперимента: 3 месяца.

География: в эксперименте задействованы все города присутствия в России.

Сплит-система

Клиенты разбиты на две группы одинакового размера случайным образом.

Таргет-метрики

* Конверсия из рекламы в покупку.
* Средний чек.

Вам необходимо проанализировать и визуализировать результаты, провести сегментацию, а также сделать выводы и сформулировать рекомендации для дальнейших запусков A/B-теста.

Также нужно построить таблицу, которая будет в удобной форме хранить результаты A/B-теста.

Таблица, которая должна быть сформирована в результате исследования, должна показывать для каждой торговой точки результат исследования. Обратите внимание, что перед составлением этой таблицы должны быть очищены данные, то есть удалены торговые точки, имеющие нулевые платежи или нулевое количество клиентов в контрольной или тестовой группе.

**Поля таблицы должны быть следующими:**

* *city —* город.
* *id\_trading\_point —* уникальный идентификатор торговой точки, к которой прикреплен данный пользователь.
* *count\_test* — кол-во наблюдений в тестовой группе.
* *count\_control* — кол-во наблюдений в контрольной группе.
* *count\_all —* суммарное кол-во наблюдений.
* *percent\_count —* процент кол-ва наблюдений данной ТТ от всех наблюдений.
* *avg\_payment\_test —* средний платеж в тестовой группе.
* *avg\_payment\_control —* средний платеж в контрольной группе.
* *diff —* разница между средними платежами.
* *sigma\_test —* стандартное отклонение платежей в тестовой группе.
* *sigma\_control —* стандартное отклонение платежей в контрольной группе.
* *ttest —* значение статистического критерия (критерий Стьюдента для сравнения средних платежей).
* *pvalue\_ttest —* pvalue статистического критерия (критерий Стьюдента для сравнения средних платежей).

В таблице также должен присутствовать *флаг (лейбл),* по которому можно будет разделить торговые точки по результатам эксперимента.

* Если результат (поле diff) *положительный* и разница на самом деле *есть* (на основании *p\_value*), то это **положительный исход.**
* Если результат (поле *diff*) *отрицательный* и разница на самом деле *есть* (на основании *p\_value*), то это **отрицательный исход.**
* Если разницы на самом деле *нет* (на основании *p\_value*), то это **нейтральный исход.**

Полученную таблицу необходимо выгрузить в Excel.

Три различных исхода должны быть выгружены на отдельные листы.

**ТЗ для калькулятора SkyLenta**

1. **Положительные исходы.**

Внешним параметром должно быть количество потенциальных клиентов-покупателей за период *N*.

Опираясь на поля *diff*, а также на долю покупателей в этой торговой точке среди всех, определите, какая выгода может быть получена от замены механики *A* на *B* при условии, что ей воспользуются *N* клиентов.

1. **Отрицательные исходы.**

Конфигурация калькулятора такая же, как на положительных исходах. Необходимо рассчитать, какая сумма может быть потеряна, если мы заменим механику *А* на механику *В*.

1. **Нейтральные исходы.**

Мы предполагаем, что результаты на этих торговых точках не видны, так как мы собрали слишком мало наблюдений. Мы хотим понять, какое количество наблюдений нам потребуется в каждой из этих торговых точек, чтобы разглядеть определенную разницу в средних платежах.

Внешним параметром должно быть *MDE —* сумма в рублях (разница между средними платежами в группах), которая считается минимально значимой с точки зрения бизнеса.

Для каждой торговой точки должно быть рассчитано кол-во наблюдений, необходимое для обнаружения разницы масштаба *MDE* (рассчитывать на основании стандартного отклонения платежей).