**ЦЕЛЬ РАБОТЫ**

Нам необходимо проанализировать и визуализировать результаты, провести сегментацию, а также сделать выводы и сформулировать рекомендации для дальнейших запусков АБ Теста. Также надо построить таблицу, которая будет в удобной форме хранить результаты АБ Теста.

**Алгоритм решения**

* 1. ***Импорт и анализ таблиц***

1.1. Импортируйте данные в окружение *Jupyter Notebook*.

1.2. Изучите все три таблицы и проверьте их на наличие пустых значений.

1.3. Исключите из таблиц все строки, в которых есть нулловые значения. Можно также подумать об альтернативных действиях - замене нуллов ну нули или на другие значения. Обоснуйте свой выбор.

1.4. Исследуйте количество торговых точек в каждом городе - постройте группировку по количеству в каждом городе и визуализируйте с помощью гистограммы.

* 1. ***Объединение таблиц***

2.1. У одного клиента может быть несколько платежей. Постройте агрегацию таблицы с платежами, где вычислите сумму платежей на каждого клиента.

2.2. Соедините (по клиенту) сгруппированную таблицу с платежами с клиентской таблицей. Убедитесь, что в таблице остались все клиенты из клиентской таблицы.

2.3. Заполните нулями суммы покупок тех клиентов, которые их не совершали

2.4. С помощью словаря регионов подтяните к каждой торговой точке город, в котором она находится.

2.5. Создайте поле флаг платежа, который принимает значения 0 или 1 в зависимости от того, заплатил клиент или нет.

* 1. ***Автоматизация статистических вычислений***

*Если вы уже создавали эти функции в рамках домашнего задания или курсовой работы, то можете использовать их.*

3.1. Создайте функцию ***test\_calc***, которая будет вычислять значение t-критерия (критерия Стьюдента) и *p\_value* для сравнения средних и с помощью функции *print* выводить сообщение о том, существует ли разница между средними (на основании *p\_value*).

*Аргументы функции:* *r1* (первая выборка), *r2* (вторая выборка), *alpha* (уровень значимости - имеет значение по умолчанию 5%).

3.2. Создайте функцию ***mann\_whitney\_func***, которая будет рассчитывать значение критерия Манна Уитни и p\_value для сравнения распределений и с помощью функции *print* выводить сообщение о том, существует ли разница между средними (на основании *p\_value*).

*Аргументы функции:* *r1* (первая выборка), *r2* (вторая выборка), *alpha* (уровень значимости - имеет значение по умолчанию 5%).

* 1. ***Чистка неверно заполненных точек***

4.1. Создайте пустой список. Запустите цикл по всем торговым точкам и добавьте к этому списку все торговые точки, в которых не было заплачено ни одного рубля ни одним клиентом.

4.2. Создайте еще один пустой список. Запустите цикл по всем торговым точкам и добавьте к этому списку все торговые точки, в которых пустует или контрольная, или тестовая группа.

* 1. ***Расчет общих результатов АБ Теста***

5.1. Отбросьте все торговые точки, которые были вами обнаружены в **пункте 4**.

5.2. Изобразите гистограмму платежей, на которой различными цветами изображены группы “тест” и “контроль”.

5.3. Примените функцию ***test\_calc*** и сделайте вывод (сравниваем средние платежи)

5.4. Примените функцию ***test\_calc*** и сделайте вывод (сравниваем конверсию в платеж, то есть *r1* и *r2* - это ряды нулей и единиц, в зависимости от того, оплатил ли что-то клиент или нет).

5.5. Примените функцию ***mann\_whitney\_func*** и сделайте вывод.

* 1. ***Сегментация результатов АБ Теста***

6.1. ***Москва***

6.1.1. Повторите аналитику из пункта 4 в отношении Москвы. Сделайте выводы.

6.2. ***Санкт-Петербург***

6.2.1. Повторите аналитику из пункта 4 в отношении Санкт-Петербурга. Сделайте выводы.

6.3. ***Другие города***

6.3.1. Запустите цикл по оставшимся городам. Проведите такой же анализ.

* 1. ***Отчет по АБ Тесту***

Создайте пустой датафрейм. Запустите цикл по всем городам и торговым точкам.

Заполните для каждой торговой точки все поля, указанные в требовании к таблице.

Не забудьте задать один из трех лейблов, в зависимости от результата теста в данной торговой точке.

Выгрузите полученные результаты в Excel. Разнесите результаты по трем разным листам в зависимости от лейбла.

* 1. ***Калькулятор***

ТЗ на построение калькулятора описано в основной части задания.

* + - **Положительные исходы**: умножение *diff* на кол-во клиентов (параметр) и умножение на долю данной торговой точки среди всех торговых точек.
    - **Отрицательные исходы**: такая же механика, только с отрицательными и статистически подтвержденными *diff.*
    - **Нейтральные исходы**: воспользуйтесь формулой минимального кол-ва наблюдений для АБ Теста. В качестве сигмы можно взять сигму контроля или минимум из сигм контроля и теста.