Описание бизнес кейса

Вам будут даны несколько кейсов для разных бизнесов с описанием бизнес задачи, которую необходимо решить:

1. Подведите результаты эксперимента в экселе по следующим данным: ab_stats.csv - Google Диск: \

Стат значимо ли отличается ARPPU в двух группах ? Какие рекомендации дадите менеджеру?

```
In [2]: import pandas as pd
import numpy as np
import scipy.stats as stats
data_path = "ab_stats.csv"
data = pd.read_csv(data_path)
print("Размеры датасета", data.shape)
data.head()
```

Размеры датасета (23652, 5)

```
Out[2]:
             revenue num_purchases purchase ab_group av_site visit
         0
                  0.0
                                    0
                                               0
                                                         Α
                                                                9.040174
         1
                  0.0
                                                                4.811628
         2
                 0.0
                                    0
                                               0
                                                          Α
                                                               7.342623
         3
                 0.0
                                               0
                                                               7.744581
         4
                 0.0
                                    0
                                               0
                                                         Α
                                                              10.511814
```

```
In [3]: # Создаем столбец, в котором вычисляем ARPPU
data['ARPPU'] = np.where(data['num_purchases'] > 0, data['revenue'] / data['num_purchases']
```

Out[3]:		revenue	num_purchases	purchase	ab_group	av_site visit	ARPPU
	0	0.0	0	0	А	9.040174	0.0
	1	0.0	0	0	А	4.811628	0.0
	2	0.0	0	0	А	7.342623	0.0
	3	0.0	0	0	А	7.744581	0.0
	4	0.0	0	0	А	10.511814	0.0

Так как ARPPU в двух группах статистически значимо не отличаются, данная гипотеза не работает.

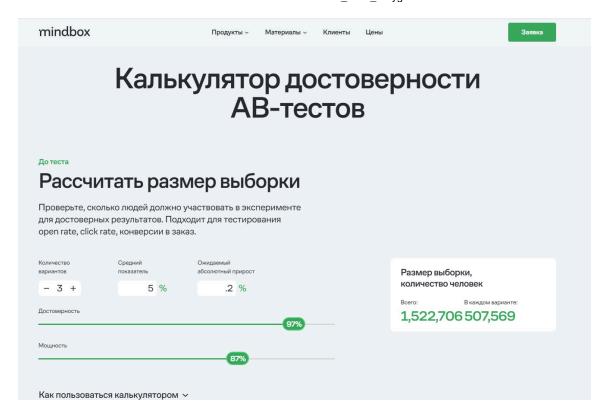
Описание бизнес кейса

2. Мы хотим провести А/Б-тест для трех источников трафика. Нынешняя конверсия равна 5%, мы ожидаем прирост в 0,2%. \

Уровень доверия 97% и уровень мощности 87%. Всего на наш продукт заходит 40 000 пользователей в ме

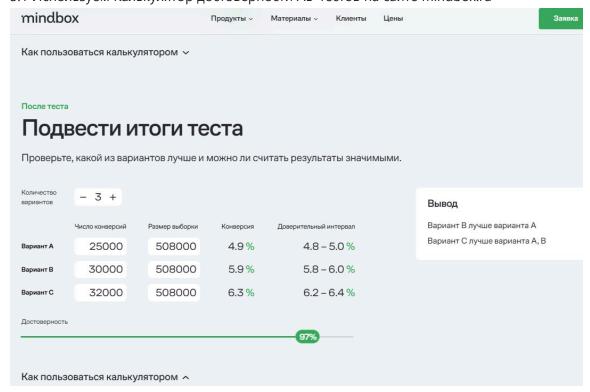
Всего на наш продукт заходит 40 000 пользователей в месяц. 3) За сколько дней мы сможем протестировать гипотезу? И что вы можете посоветовать по результатам подсчета?

- 3.1) Допустим в задаче нет проблемы с количеством посетителей на сайт, тогда подведите результаты тестирования, если у нас следующие результаты по количеству конверсии:
 - 1. 25 000
 - 2.30000
 - 3.32000
 - 3. Используем Калькулятор достоверности AB-тестов на сайте mindbox.ru \



Как мы видим из результатов выборка должна быть - 1,522,706 человек. Если в месяц заходит 40,000 пользователей, то нам понадобится примерно 1142 дня (это более 3-ех лет). Следовательно, от этого теста лучше отказаться, либо же менять какие-то параметры.

3.1 Используем Калькулятор достоверности AB-тестов на сайте mindbox.ru



Как мы видим из результатов вариант В лучше варианта А. И вариант С лучше вариантов А и В. Следовательно, разница в конверсиях

статистически значимы и вариант надо отдать предпочтение варианту С (с конверсией 32000).

Описание бизнес кейса

4. Вы решили сравнивать метрику СРА в двух группах. Размер выборки - 2350 элементов в каждой группе. \

Для проверки нормальности распределения на выборке в 2350 наблюдений применили, критерий Шапиро-Уилка и получили p-value, равный 0.00002, alpha = 5%. Какой бы вывод мы могли сделать в данном случае?

В этом случае какой статистический критерий для проверки первоначальной гипотезы тут лучше всего подойдёт и почему? \

Нулевая гипотеза в тесте Шапиро-Уилка заключается в том, что данные имеют нормальное распределение. Поскольку значение p-value (0.00002) значительно меньше уровня значимости (alpha = 0.05), мы отвергаем нулевую гипотезу о нормальности распределения и не можем использовать t-test или z-test. Следовательно, мы должны использовать Критерий Манна—Уитни (2 группы, независимые выборки, не нормальное распределение)

Описание бизнес кейса

5. Мы провели АБ-тест на увеличение average timespent per user. По итогам тестирования мы получили следующие данные. Является ли результат статистически значимым с уровнем доверия 80%? Какую версию мы выкатим на продакшн? \

```
А) Средняя - 360, отклонение - 40, количество - 9802В) Средняя - 352, отклонение - 58, количество - 9789
```

```
In [13]: import math
   import scipy.stats as stats
```

Чтобы определить, является ли результат статистически значимым с уровнем доверия 80%, нам нужно провести тест на значимость разницы между двумя средними значениями. В данном случае мы будем использовать t-тест для независимых выборок. Вычислим t-статистику и сравним ее с t-критическим.

```
In [15]: m1 = 360
    sd1 = 40
    n1 = 9802
    m2 = 352
    sd2 = 58
    n2 = 9789
    alfa = 1 - 0.8
    df = n1 + n2 -2
In [16]: t = (m1 - m2) / math.sqrt(sd1**2 / n1 + sd2**2 / n2)
```

```
In [16]: t = (m1 - m2) / math.sqrt(sd1**2 / n1 + sd2**2 / n2)
t_kr = stats.t.ppf(1 - alfa / 2, df)
print(f't-статистика = {t:.2f}')
print(f'Критерий Стюдента для alfha = {alfa:.1f} и {df} степеней свободы = {t_kr:.2
if t > t_kr:
    print (f'Результат статистически значим с уровнем доверия 80%')
else:
    print(f'Результат не является статистически значимым с уровнем доверия 80%')
```

t-статистика = 11.24 Критерий Стюдента для alfha = 0.2 и 19589 степеней свободы = 1.28 Результат статистически значим с уровнем доверия 80%

Поскольку средняя для группы А (360) больше средней для группы В (352) и результат статистически значим, мы можем сделать вывод, что версия А (с средней 360) будет выкатана на продакшн.