* Необходимо было решить - стоит ли проводить акцию в дальнейшем,

если игроки участвовавшие в акции принесли больше денег, чем игроки

у которых акции не было?

* К решению подошёл серьёзно, проанализировать каждый из исходных источников
* Делал проект с помощью Pandas, т.к. долго загружались данные в базу данных

Исходники кодов:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import scipy.stats

from scipy.stats import kurtosis

from scipy.stats import skew

from scipy import stats

df\_money = pd.read\_csv('Money.csv')

df\_cash = pd.read\_csv('Cash.csv')

df\_cheaters = pd.read\_csv('Cheaters.csv')

df\_AB\_group = pd.read\_csv('AB\_group.csv')

df\_platforms = pd.read\_csv('Platforms.csv')

* Вычистил читеров предварительно ознакомившись с таблицей по читерам

-df\_cheaters[df\_cheaters['cheaters'] == 1]

Присвоив переменную – true\_cheaters

-true\_cheaters = df\_cheaters[df\_cheaters['cheaters'] == 1]

Далее df\_money[df\_money['user\_id'].isin(true\_cheaters['user\_id'])==False].reset\_index(drop=True)

Касаемо ещё остальных читеров, нашёл их путём аномального кэша, но без платежа

df\_cash\_money[df\_cash\_money['money'] == 0]

df\_without\_payers = df\_cash\_money[df\_cash\_money['money'] == 0]

Затем исходя из гистограммы, убрал читеров у которых был кэш от 50000

df\_without\_payers['cash'].hist()

df\_without\_payers[(df\_without\_payers['cash']>=50000) & (df\_without\_payers['money'] == 0)]

df\_cash\_from\_50000 = df\_without\_payers[(df\_without\_payers['cash']>=50000) & (df\_without\_payers['money'] == 0)]

df\_cash\_money = df\_cash\_money[df\_cash\_money['user\_id'].isin(df\_cash\_from\_50000['user\_id'])==False].reset\_index(drop=True)

* Доверительные интервалы по ARPU и ARPPU нашёл именно по платежам, так как от платежей зависит доход по акции

Использовал функцию

def interval(data,confidence = 0.95):

a = 1.0 \* np.array(data)

n = len(a)

k, se = np.mean(a), scipy.stats.sem(a)

h = se \* scipy.stats.t.ppf((1+confidence)/2.,n-1)

return k,k-h,k+h

Потом в последствии сравнивая три значения интервала по группам

Пример:

interval(df\_pc\_money\_c\_g['money'])

(5.646169929840374, 5.637770160202389, 5.654569699478359)

interval(df\_pc\_money\_t\_g['money'])

(6.269046157268379, 6.26012751674264, 6.277964797794119)

* Файл финальной работы, графики по платежам и внутриигровой валюте в PowerBI, сводная таблица Excel по ARPU прилагаются к сообщению

**Вывод:** анализируя результаты, пришёл к выводу, что интервальные значения по показателям ARPU и ARPPU по контрольным и тестовым группам по платформам PC и XBox не равны, результаты неслучайны и акция была удачна, касаемо платформы PS4 показатели ARPU и ARPPU тестовых групп соответственно больше, чем у контрольных групп, однако надо заметить, что интервальные значения показателя ARPPU тестовой и контрольной группы равны, что означает, что результаты случайны и акция не принесла результатов. Считаю, что акцию проводить можно по PC и XBox