

Ad Hoc в такси агрегаторе

Ссылка на презентацию в notion

Ad Hoc в такси агрегаторе

▼ Условие

Кейс 1. Ad hoc в агрегаторе такси

Вы работаете продуктовым аналитиком в агрегаторе такси SkyTaxi.

У вас есть бэклог из нескольких задач Ad hoc.

Для их решения рассмотрите основную конверсию из заказа в поездку *Order2Ride*, а также промежуточные три звена конверсии, из которых состоит *Order2Ride* (*Order2Assign*, *Assign2Arrive*, *Arrive2Ride*).

Скачайте отсюда (<https://drive.google.com>) следующие csv-файлы:

- *Data_Skytaxi.csv*
- *city_dict.csv*
- *tariff_dict.csv*

Оформите решение в Jupyter-тетради в формате .ipynb. Подготовьте единую презентацию (в формате .pptx) с результатами всех задач кейса. Добавьте свою презентацию и ссылку на Miro (ссылка находится в материалах урока). Проголосуйте за наиболее понравившиеся вам работы.

Задание 1

Изучите утренние и вечерние часы пик.

1. Постройте график с количеством заказов по суточным часам (на оси X — часы от 0 до 23).
2. На сколько за один час в среднем делается больше заказов в часы пик, чем в обычное время? (По всем городам вместе.)

Подсказка

1. Рассмотрите города по отдельности: для каждого города выведите разницу в количестве заказов (среднечасовом) между часами пик и обычными часами.
2. В каком городе наблюдается наибольшее отклонение конверсии *Order2Ride* в часы пик по сравнению с обычными часами?

Задание 2

Изучите заказы в Хабаровске и Тюмени. Эти города интересуют менеджеров, так как в них наша компания начала работать относительно недавно.

1. На сколько процентных пунктов *Order2Ride* в среднем в них ниже, чем в других городах?
2. За счет какого звена воронки достигается эта разница? Сделайте выводы по Хабаровску и по Тюмени по отдельности.
3. Какие рекомендации вы можете дать локальным менеджерам?

Задание 3

К вам обратились коллеги из отдела антифрода (отдела, предотвращающего мошенничество). У них есть подозрение, что некоторые водители мимикрируют координаты, то есть на самом деле не приезжают в точку А своего заказа, но в приложение посылают сигнал, что они в эту точку А приехали. Таким образом они вынуждают клиента отменить заказ после «прибытия ими в точку А».

Постарайтесь локализовать данную проблему и выделить города, в которых такое происходит чаще всего.

Подсказка

Задание 4

Тариф «Доставка» был запущен недавно по всей России. Изучите конверсии в рамках данного тарифа по городам, локализируйте просадку

конверсии *Order2Ride* в рамках данного тарифа и дайте рекомендации отделу операционистов, которые занимаются этим тарифом.

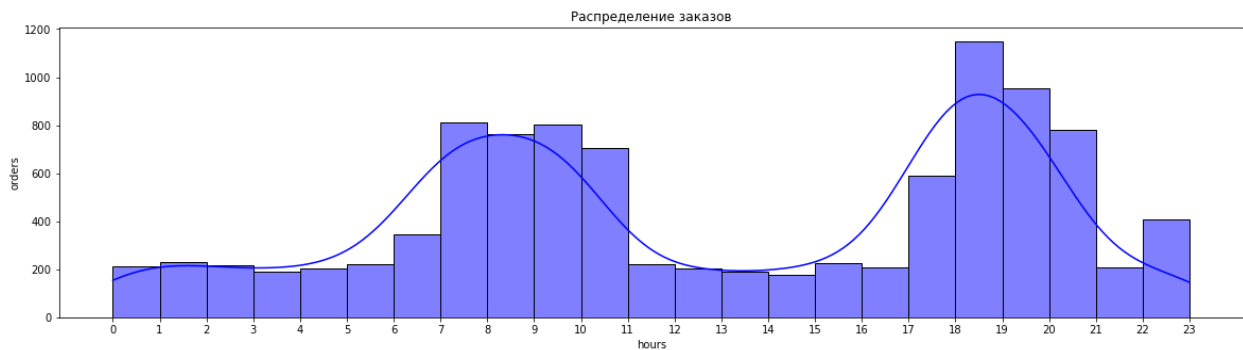
Подсказка

Дополнительное задание*

Вы обратились к дата-инженерам, и они согласились создать таблицы в «Хранилище», соответствующие .csv-файлам из этого кейса. Подготовьте SQL-скрипты, которые будут производить расчеты данных заданий Ad hoc на уровне «Хранилища».

В первую очередь изучим наши данные - утренние и вечерние часы пик.

Распределение заказов в течении дня следующее



В день наблюдается 2 час пика: с 7 до 10 утра и с 17 до 20 часов вечера включительно.

Приблизительно на 600 заказов за один час в среднем больше в часы пик, чем в обычное время по всем городам вместе.

Проанализировав каждый город точно видно, что самая большая дельта наблюдается в МСК и СПб. В Красноярске дельта меньше всего

city	delta_peak_hour
Казань	56.9375
Москва	182.8125
Тюмень	36.6250
Краснодар	46.9375
Санкт-Петербург	119.7500

city	delta_peak_hour
Екатеринбург	35.4375
Хабаровск	38.8125
Красноярск	30.6250
Новосибирск	54.5000



Order2Ride

В каком городе наблюдается наибольшее отклонение конверсии Order2Ride в часы пик по сравнению с обычными часами?

В Москве наблюдается наибольшее отклонение O2R - на 11,6 п.п. O2R в часы пик меньше, чем в обычное время.

city	o2r	o2r_peak_hour	o2r_not_peak_hour	delta_o2r_peak_hour
Казань	0.56	0.55	0.56	-0.01
Москва	0.48	0.44	0.56	-0.12
Тюмень	0.27	0.27	0.29	-0.02
Краснодар	0.58	0.57	0.60	-0.03
Санкт-Петербург	0.50	0.46	0.55	-0.10
Екатеринбург	0.57	0.58	0.56	0.02
Хабаровск	0.26	0.28	0.22	0.06
Красноярск	0.39	0.41	0.36	0.05
Новосибирск	0.59	0.59	0.61	-0.02

Заказы в Хабаровске и Тюмени

На сколько процентных пунктов Order2Ride в среднем в них ниже, чем в других городах?

Хабаровск - на 0.22 п.п. O2R в среднем ниже, чем в других городах.

Тюмень - на 0.215 п.п. O2R в среднем ниже, чем в других городах

За счет какого звена воронки достигается эта разница?

В Хабаровске проседает O2A=0,34.

В Тюмене тоже сниженный O2A=0,56, но в большей степени просел A2R=0.57 (по сравнению с A2R=0.90 в остальных городах)

city	O2A	A2A	A2R	O2R
Хабаровск	0.34	0.81	0.95	0.26
Тюмень	0.56	0.84	0.57	0.27
Общий	0.67	0.79	0.90	0.48

Время назначения водителя (O2A) в Хабаровске и Тюмене незначительно отличается от времени назначения в остальных городах. Аналогичная ситуация и со временем поездки (A2R) в Тюмене. Необходимо смотреть глубже, кто инициатор отказов, из-за которых просели эти этапы воронки

city	time O2A	time A2R
Москва	00:03:53	00:22:53
Санкт-Петербург	00:04:00	00:17:23
Красноярск	00:04:05	00:10:46
Екатеринбург	00:04:08	00:14:07
Казань	00:04:13	00:13:59
Тюмень	00:04:15	00:10:37
Краснодар	00:04:21	00:13:55
Новосибирск	00:04:22	00:10:39
Хабаровск	00:04:31	00:10:27

Антифрод

Короткий временной промежуток между назначением водителя (assign_time) и его прибытием (arrive_time) может быть следствием мимикрии координат. Будем считать, что если дельта между прибытием и назначением меньше или равна 2 минутам,

тогда это причина подозревать водителя в мошенничестве. Данные заказы стоит детально проанализировать (проверить изменения геолокации).

Чаще всего данные кейсы встречаются в Москве (10), СПб (4) и Тюмене (4)

Конверсия в тарифе “Доставка”

City Tariff	O2A	A2A	A2R	O2R
Екатеринбург Доставка	0.79	0.84	0.90	0.60
Казань Доставка	0.76	0.81	0.97	0.60
Новосибирск Доставка	0.66	0.91	0.97	0.58
Санкт-Петербург Доставка	0.73	0.74	0.95	0.51
Москва Доставка	0.70	0.76	0.95	0.51
Общий Доставка	0.66	0.78	0.93	0.48
Красноярск Доставка	0.82	0.81	0.72	0.47
Хабаровск Доставка	0.34	0.75	0.92	0.23
Тюмень Доставка	0.25	0.73	0.88	0.16
Краснодар Доставка	0.00	NaN	NaN	0.00

Тюмень и Хабаровск показывают наименьшие показатели O2R в тарифе Доставка. В расчетах не учитываем Краснодар, т.к. в нет данных для анализа

Отличий тарифа Доставка от остальных тарифов нет, значения конверсии равны.

▼ Сравнение Доставки с Эконом

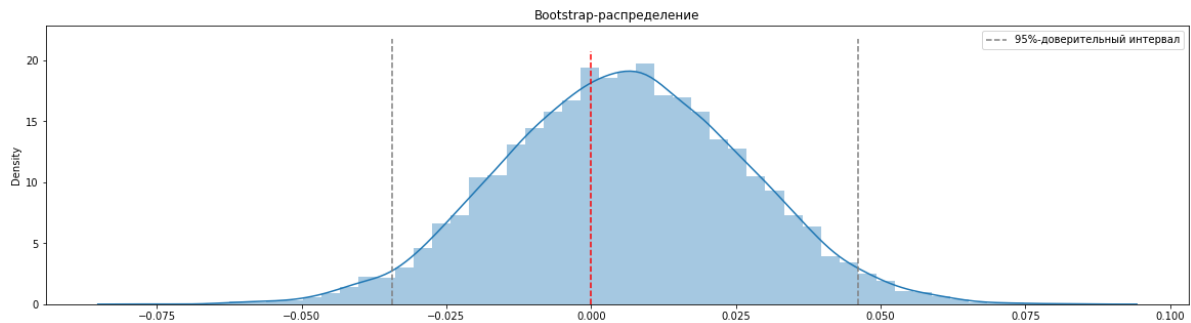
----Сравнение Доставки с Эконом----

Среднее значение 1 ряда 0.48396946564885496

Среднее значение 2 ряда 0.4897626803164262

Разница средних = -0.005793214667571267

Нулевая гипотеза о равенстве средних принимается, т.к. p_value = 0.7823735177378696, что больше чем 0.05. Значение стат критерия -0.27624227703747206



▼ Сравнение Доставки с Комфорт

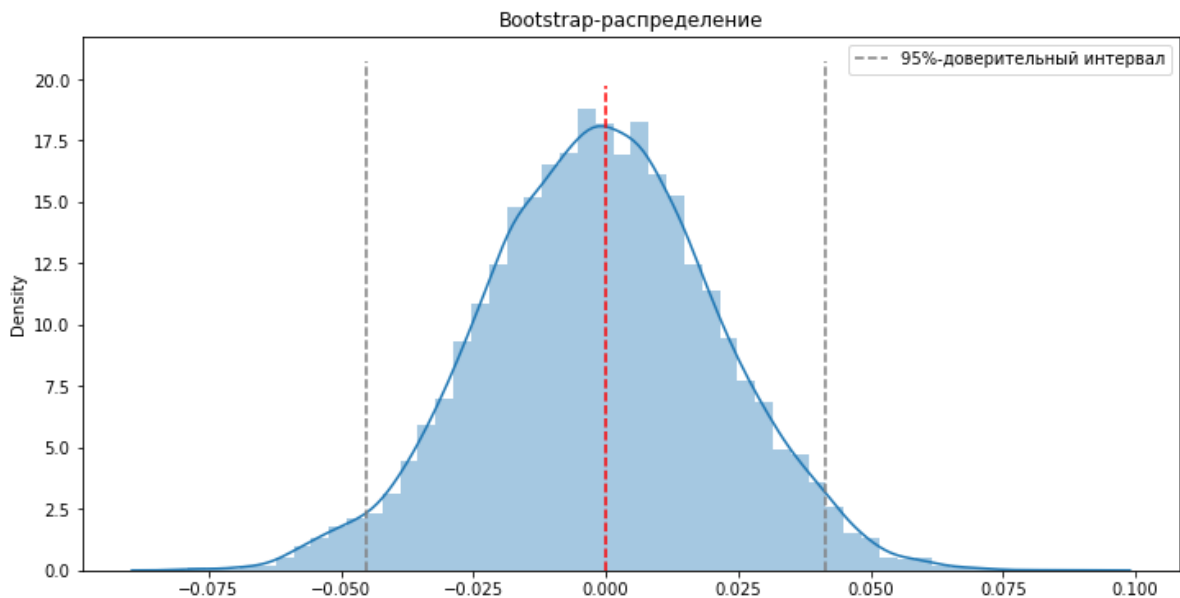
----Сравнение Доставки с Комфорт----

Среднее значение 1 ряда 0.48396946564885496

Среднее значение 2 ряда 0.4826577517760134

Разница средних = 0.001311713872841569

Нулевая гипотеза о равенстве средних принимается, т.к. $p_value = 0.9525526035142586$, что больше чем 0.05. Значение стат критерия 0.05950648551817487



▼ Сравнение Доставки с Бизнес

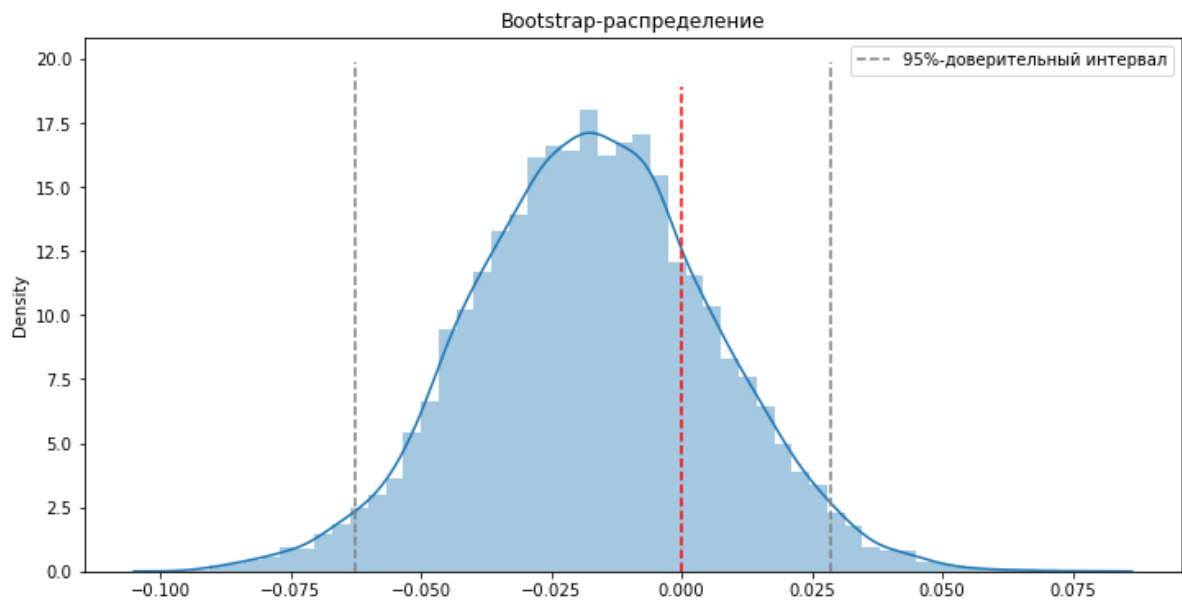
----Сравнение Доставки с Бизнес----

Среднее значение 1 ряда 0.48396946564885496

Среднее значение 2 ряда 0.4668109668109668

Разница средних = 0.01715849883788817

Нулевая гипотеза о равенстве средних принимается, т.к. $p_value = 0.46877918435996335$, что больше чем 0.05. Значение стат критерия 0.7246022994886587



▼ Сравнение Доставки с Комфорт+

----Сравнение Доставки с Комфорт+----

Среднее значение 1 ряда 0.48396946564885496

Среднее значение 2 ряда 0.47360248447204967

Разница средних = 0.01036698117680529

Нулевая гипотеза о равенстве средних принимается, т.к. $p_value = 0.6655767544647266$, что больше чем 0.05. Значение стат критерия 0.4322926422688504

