

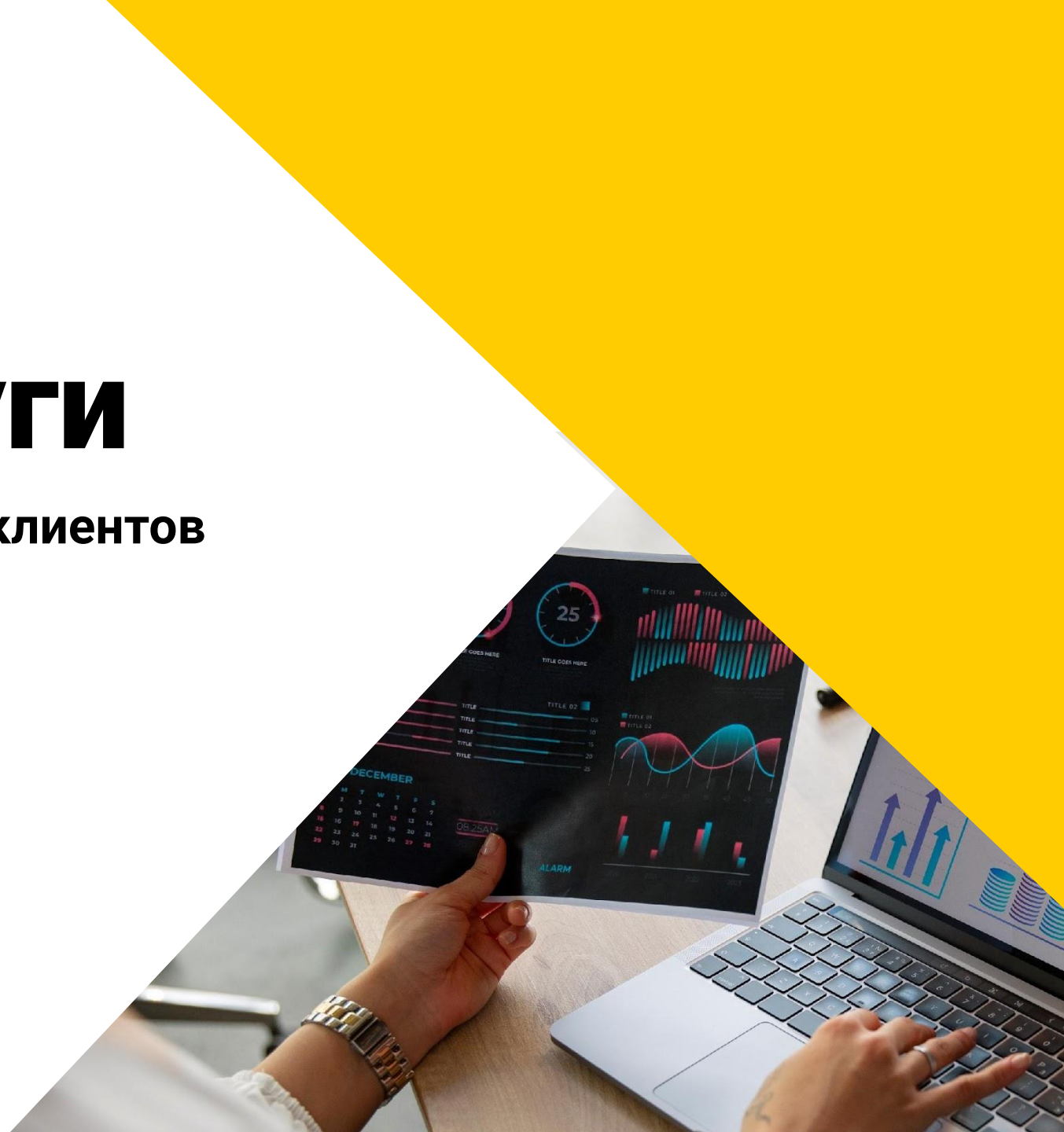
Бытовые услуги

Исследование удовлетворенности клиентов

Подготовили:

Елфимов Константин

Иванов Вячеслав



Цели и задачи проекта

● Цель ●

Выяснить, в какой тип климатического оборудования следует инвестировать в различных регионах.

1 Очистка данных

Обработка выбросов и пропусков, удаление не значимых для анализа колонок.

2 Проверка гипотез

Проверка гипотез о влиянии различных факторов на оценку комфорта респондентов.

3 Регрессионный анализ

Построение модели линейной регрессии для предсказания температуры в помещении на основе выбранных факторов.

Предобработка данных

	Тип	% пропусков	Кол-во выбросов	Кол-во уникальных значений
год	datetime64[ns]	0.000000	NaN	3
время_года	object	0.000000	NaN	4
климат	object	0.000000	NaN	5
город	object	0.000000	NaN	8
страна	object	0.000000	NaN	3
способ_охлаждения	object	0.000000	NaN	3
режим_при_смешанном_типе_охлаждения	object	30.497925	NaN	2
способ_обогрева	object	53.941909	NaN	1
возраст	float64	14.315353	12.0	29
пол	object	14.937759	NaN	2
ощущение_температуры	float64	0.000000	47.0	22
ощущение_температуры_(bool)	float64	0.000000	102.0	2
предпочтительное_изменение_температуры	object	0.000000	NaN	3
ощущение_движения_воздуха_(bool)	float64	58.921162	49.0	2
предпочтительное_изменение_движения_воздуха	object	0.000000	NaN	3
оценка_комфорта	float64	55.394191	0.0	6
утепление	float64	0.000000	23.0	83
температура_воздуха_в_помещении	float64	0.000000	54.0	115
температура_воздуха_на_улице	float64	85.062241	5.0	25
rh	float64	0.000000	0.0	289
скорость_воздуха	float64	0.000000	49.0	56
рост	float64	70.331950	1.0	29
вес	float64	56.016598	4.0	47
занавески	float64	85.062241	0.0	2
вентилятор	float64	69.709544	0.0	2
окно	float64	42.531120	69.0	2
двери	float64	67.842324	29.0	2
отопление	float64	83.609959	17.0	2
среднемесячная_температура_на_улице	float64	0.000000	14.0	20
количество_рекламаций	int64	0.000000	34.0	4

- 1) Обработка ошибок в определении типов данных, удаление дубликатов;
- 2) Заполнение пропусков в отдельных колонках (заполнение медианой/средним по группе, на основе зависимостей от других колонок, константой, методами из ML);
- 3) Анализ выбросов в отдельных колонках (тройной интерквартильный размах), что требует дополнительного внимания к этим колонкам (например, в США температура в некоторых колонках была в фаренгейтах, были ошибки со знаками после запятой);
- 4) Исправление опечаток в значениях климат и предпочтительное изменение температуры;
- 5) Специфические пропуски. Например, только в США нет данных по полу, только в Австралии нет данных по оценке комфорта и т.д.;
- 6) Оценка комфорта: была попытка применить различные алгоритмы ML для заполнения пропусков, однако, из-за малого количества данных и несбалансированности классов все модели крайне переобучались.

Общие сведения

482

респондента **приняли участие** в исследовании.

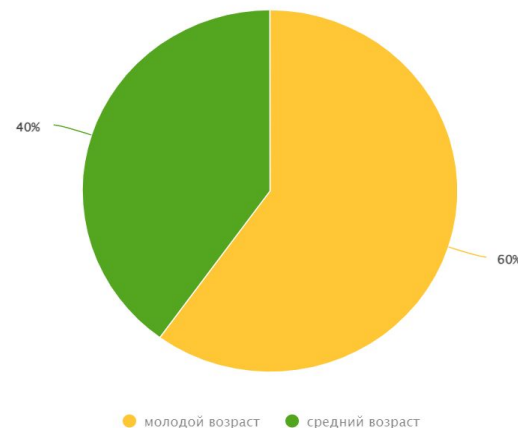
78.8%

опрошенных **довольны** температурным режимом.

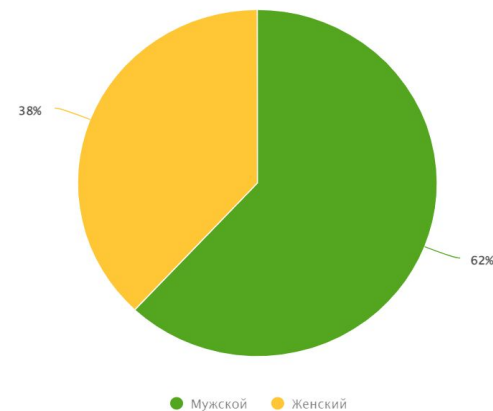
В исследовании преимущественно приняли участники **молодого возраста** (до 44 лет), **мужского пола** (данные о поле в США отсутствуют).

В **США** можно наблюдать **самую низкую удовлетворенность** температурным режимом.

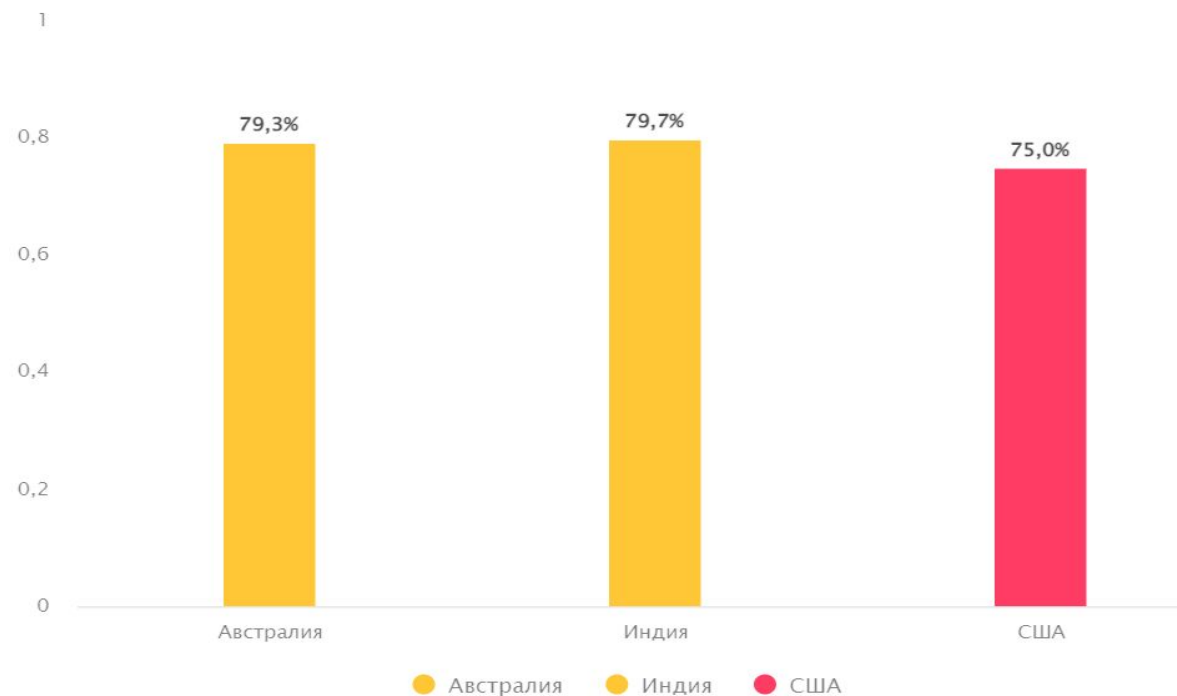
Возрастной состав



Половой состав



Процент удовлетворенных температурой респондентов по странам



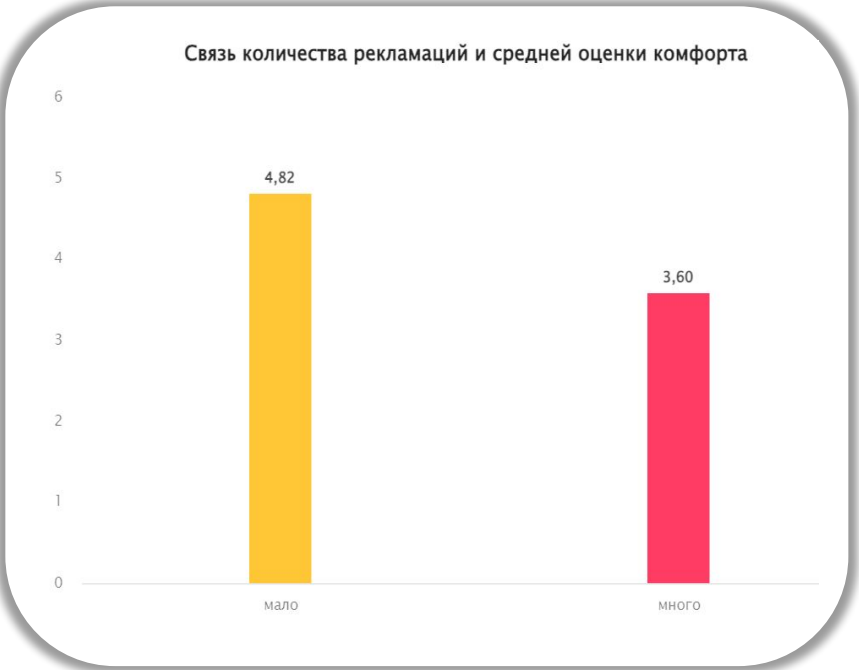
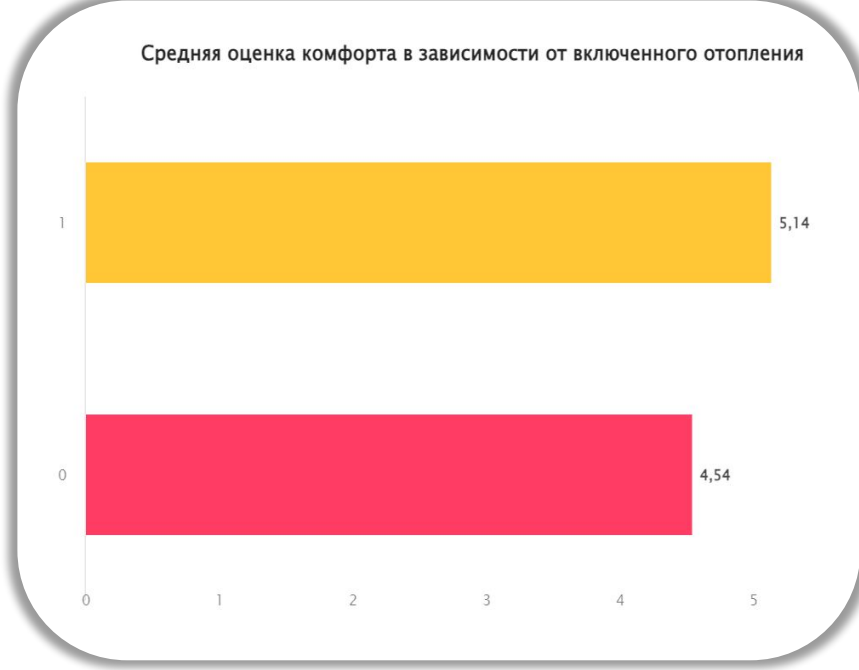
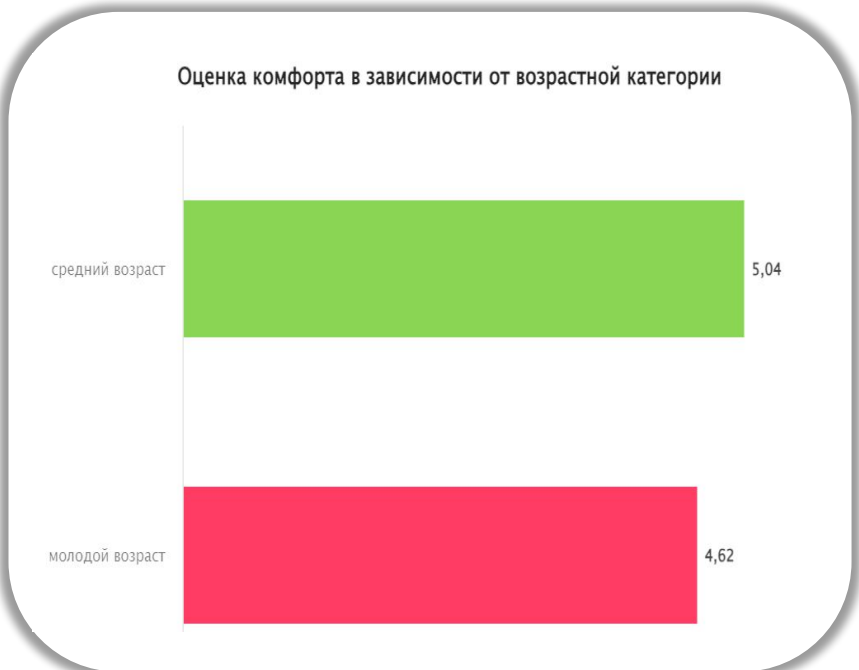
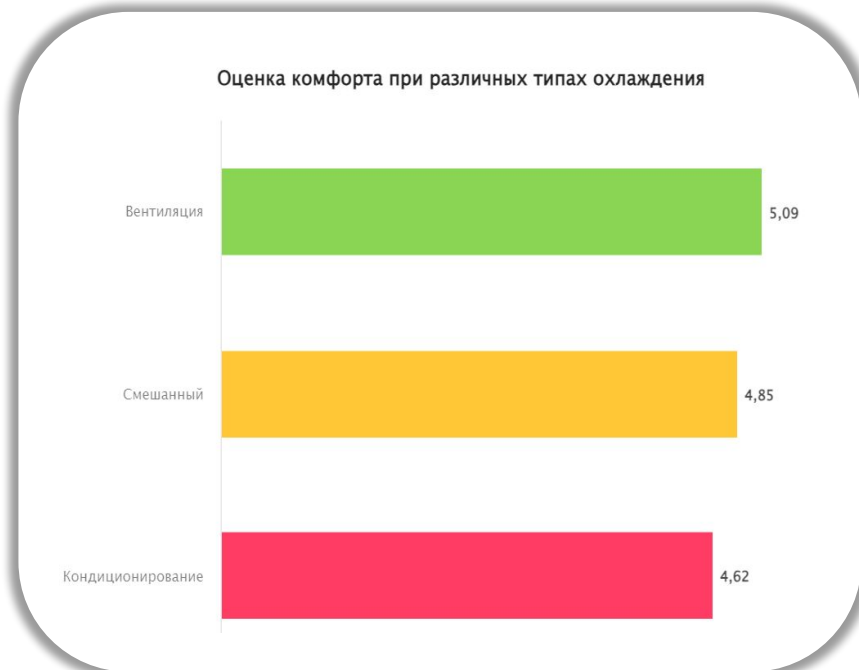
Факторы, влияющие на оценку комфорта

Отопление значимо влияет на оценку комфорта (критерий Манна-Уитни, $p: 0.001$). Для дальнейшей проверки гипотез использовались **только данные респондентов с выключенным отоплением**, т.к. мы ориентируемся на оценку комфорта для систем **охлаждения**, которые используются обычно в **жаркую погоду**.

Тип охлаждения значимо влияет на оценку комфорта (критерий Краскела-Уоллеса, $H: 15.014$, $p: 0.001$).

Возрастная группа значимо влияет на оценку комфорта (критерий Манна-Уитни, $p: 0.002$).

Количество рекламаций значимо влияет на оценку комфорта (корреляция: -0.6 , $p: 0.0498$).



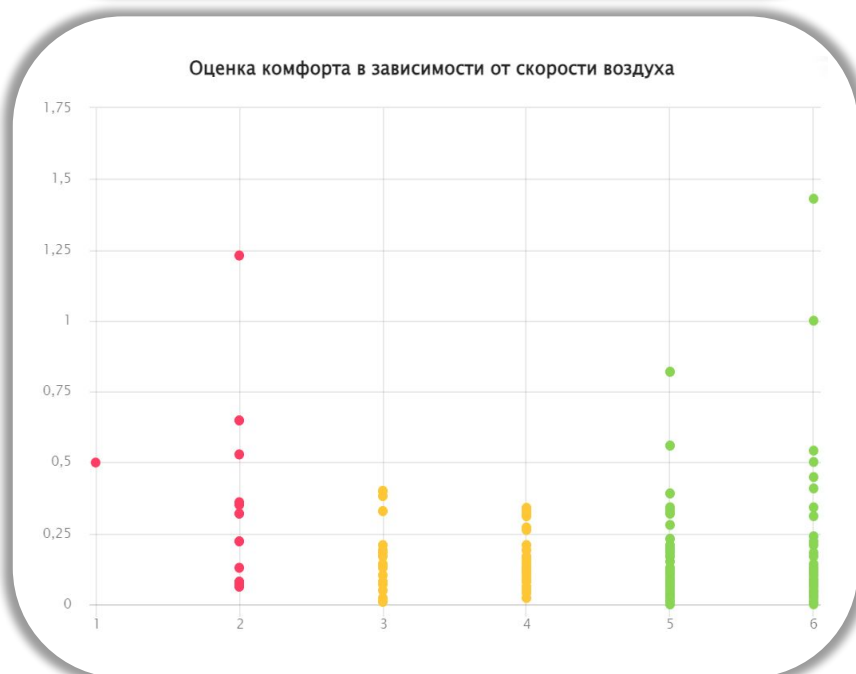
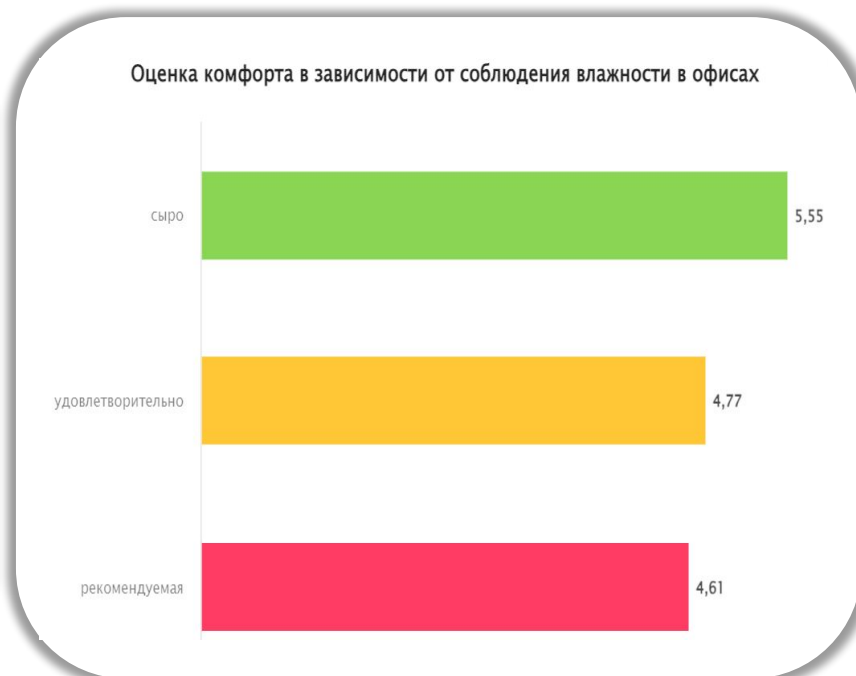
Факторы, влияющие на оценку комфорта

В США самая низкая оценка комфорта из-за доминирования кондиционирования.

Страна значительно влияет на оценку комфорта (критерий Манна-Уитни, $p: 0.001$).

Степень соблюдения влажности в офисах значительно влияет на оценку комфорта (критерий Краскела-Уоллеса, $H: 8.85$, $p: 0.012$).

Скорость воздуха значительно влияет на оценку комфорта (Корреляция: -0.29 , $p: 0.001$).



Дополнительная проверка гипотез

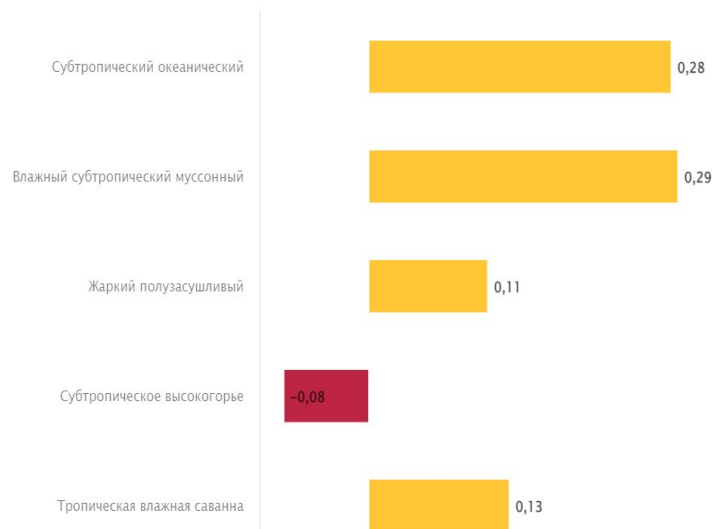
Климат значительно **влияет на ощущение температуры** (критерий Краскела-Уоллеса, $N: 10.894$ $p: 0.027$).

Климат также **влияет на оценку комфорта** (критерий Краскела-Уоллеса, $N: 31.586$ $p: 0.0002$).

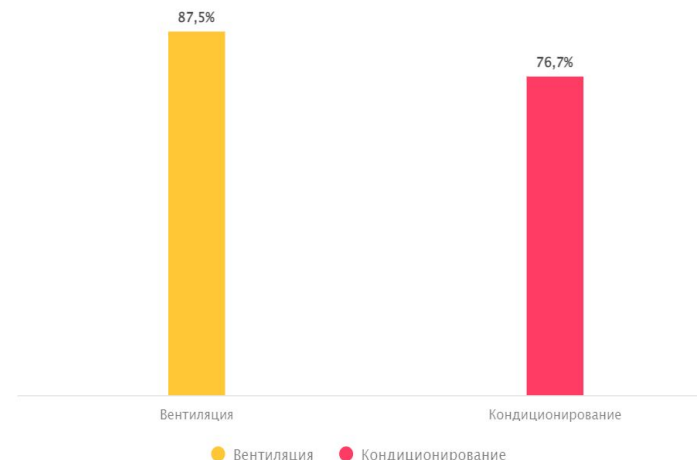
Отдельный анализ по Австралии показал, что в целом **Кондиционирование** показывает **меньший процент удовлетворенности** клиентов, но результат статистически незначим (коэффициент Крамера: 0.13 , $p: 0.168$).

Также было показано, что **ощущение температуры** значительно **связано с оценкой комфорта** (критерий Манна-Уитни, $p: 7.597e-09$).

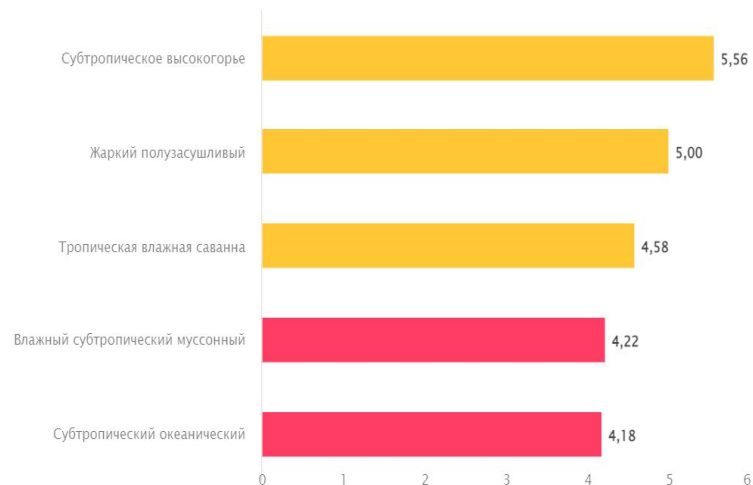
Ощущение температуры в зависимости от климата



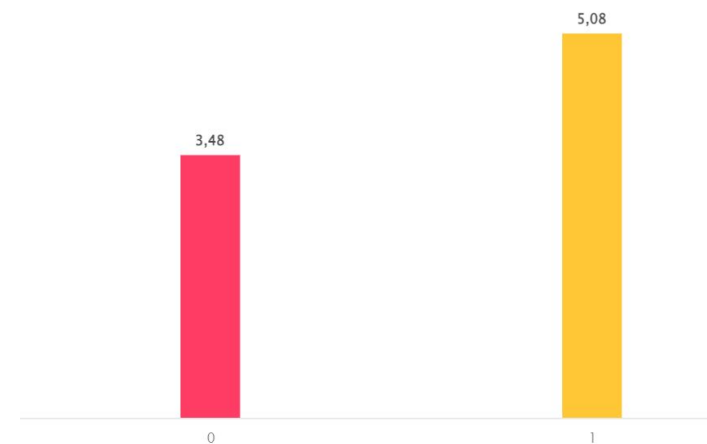
Процент удовлетворенных клиентов в зависимости от режима охлаждения (Австралия)



Оценка комфорта в зависимости от климата



Средняя оценка комфорта в зависимости от ощущения температуры



Регрессионное моделирование

По итогам регрессионного анализа нам удалось достичь значения $R^2=60.9\%$ и $MARE=4.9\%$. Отбор факторов происходил на основе предварительного корреляционного анализа.

Коэффициент детерминации: 60.9%
Средняя абсолютная ошибка: 4.9%

Переменная	Коэффиц иент	t- значение	p-value
const	20,78	38,02	0,00
ощущение_температуры_(bool)	-0,25	-2,97	0,00
скорость_воздуха	0,33	3,01	0,00
температура_воздуха_на_улице	0,98	3,36	0,00
время_года_Зима	0,65	1,13	0,26
время_года_Лето	1,22	2,04	0,04
время_года_Осень	0,26	0,42	0,68
климат_Влажный_субтропический_муссонный	5,39	15,38	0,00
климат_Жаркий_полусухой	6,38	16,55	0,00
климат_Субтропическое_высокогорье	1,87	5,12	0,00
климат_Тропическая_влажная_саванна	2,56	5,53	0,00
город_Бангалор	1,01	2,11	0,04
город_Дели	5,39	15,38	0,00
город_Сидней	3,51	7,31	0,00
город_Техас	1,07	4,97	0,00
город_Хайдарабад	1,34	2,50	0,01
город_Ченнай	1,56	3,34	0,00
город_Шимла	1,87	5,12	0,00
влажность_в_офисе_категория_сыро	-0,23	-0,83	0,41
влажность_в_офисе_категория_удовлетворительно	0,03	0,14	0,89
используемый_тип_охлаждения_Кондиционирование	1,07	4,97	0,00
используемый_тип_охлаждения_Смешанный_Вентиляция	-1,39	-2,55	0,01
используемый_тип_охлаждения_Смешанный_Кондиционирование	-0,64	-1,17	0,25

Подведем итоги



Вентиляция – наш выбор!

По итогам анализа вентиляция показалась клиентам более комфортным способом охлаждения во всех регионах (кроме США!). Возможно, это связано с техническими особенностями работы (например, скорость воздуха в кондиционерах). Однако, нужно учитывать множество других факторов (цены, логистика и др.).



Комфорт - разнообразен

Оценка комфорта складывается из многих параметров, которые необходимо учитывать при установке оборудования: тип климата, влажность, скорость воздуха в оборудовании и т.д.



Температуру в помещении нам не угадать

Показатели регрессионной модели дают неплохие результаты для предсказания температуры в помещении, но требуется намного больше данных, а также контроль качества за собранными данными.

Дашборд

Анализ удовлетворенности клиентов системами охлаждения

Количество респондентов

482

Процент респондентов, удовлетворенных температурой

78,8%

Процент удовлетворенных температурой респондентов по странам



Возрастной состав

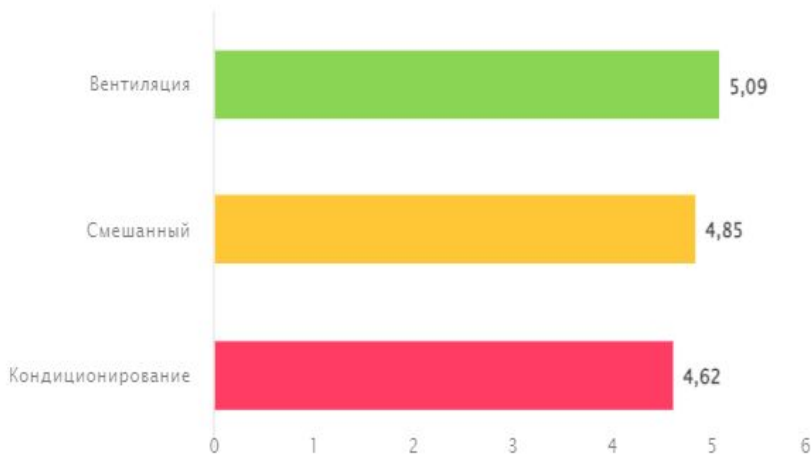


Половой состав



Анализ факторов, влияющих на оценку комфорта респондентами

Оценка комфорта при различных типах охлаждения



Типы охлаждения в различных странах



↑ ↓ / Страна / Климат

Влияние различных факторов на оценку комфорта

