

Бытовые услуги. Исследование удовлетворенности клиента

Выполнили:

Кацук Ирина

Дубровин Иван

Шкаев Максим

Лесных Виктория

2023

Выводы

- Количество данных достаточно мало, много пропусков, поэтому сделать выводы о направлении развития компании на таких данных не очень разумно, следует запросить больше качественных данных
- Компания получила данные из трех стран: Австралии, США и Индии. В отчете Австралия и США находятся в одной климатической зоне. Индия представлена различными климатическими зонами.
- В Индии клиенты с низким уровнем комфорта желали бы сделать потеплее или оставить температуру без изменений. Недовольные клиенты пользовались способом охлаждения - Вентиляция или Смешанный. Следовательно, в данном регионе следует развивать способ охлаждения Кондиционирование, на который жалоб нет.
- В США 50% недовольных хотели бы похолоднее, а 50% - потеплее. Рекомендация для США: продолжать развивать Кондиционирование, обратить внимание на условия работы в офисах, проводить семинары о важности нормальной влажности в офисах.
- В Австралии большинство недовольных хотели бы похолоднее. Недовольных там немного, поэтому можно развивать Смешанный тип охлаждения с увеличенной мощностью для охлаждения.

Исследование данных на качество и предобработка

После загрузки данных обнаружилось:

- Количество пропусков: 3817
- Доля пропусков: 27.2%
- Доля строк, содержащих пропуски: 100.0%
- Количество полных дубликатов: 2
- В столбцах “Предпочтительное изменение температуры” и “Климат” название категорий требуют корректировок: Тепле, Холодн, Субтроп океанич
- Есть выбросы: Температура воздуха в помещении (похоже, часть данных в шкале Фаренгейта), Скорость воздуха, Рост, Среднемесячная температура воздуха на улице
- Данные распределены, в основном, ненормально

Подготовка данных

- Обработка отсутствующих значений
 1. Удалили строки с отсутствующим возрастом, так как в этих строках также отсутствует важный параметр “Оценка комфорта”
 2. Для отсутствующих значений по полу ввели строку “Не указан”
 3. Отсутствующие значения оценки комфорта заполнили средними значениями, сгруппированными по способу охлаждения
 4. Пропуски в температуре воздуха на улице заполнили средними значениями среднемесячной температуры на улице, сгруппированной по городу и времени года
 5. Удалили колонки рост и вес, так как в них слишком много пропусков
 6. Значение колонки "Отопление" зависит от типа охлаждения. Заполнили пропуски значениями, в зависимости от типа охлаждения
- Обработка выбросов
 1. Преобразовали температуру воздуха в помещении из шкалы Фаренгейта в шкалу Цельсия
 2. Заменяли выбросы в колонке “Скорость воздуха” на медианные значения
 3. Преобразовали среднемесячную температуру на улице. Явные выбросы заменяем на средние значения, сгруппированные по городу, остальные перевели из шкалы Фаренгейта в шкалу Цельсия

Проведение расчетов и исследований

- Сильная связь между страной(городом) и годом, объясняется тем, что данные по странам собраны в разные периоды времени.
- Обнаружена сильная связь между городом и способом обогрева, а также между городом и способом охлаждения. Это связано с тем, что в разных городах используются разные способы обогрева/охлаждения
- Сильную связь между страной(городом) и занавесками; вентилятором и занавесками можно объяснить тем, что данную информацию указывали только в определенных городах
- Связь между городом и оценкой комфорта(категориальная) можно объяснить тем, что в некоторых городах система охлаждения настроена более оптимально
- Связь “способ_охлаждения” и “занавески”, видимо, можно объяснить тем, что при определенных способах охлаждения занавески закрывают или открывают.
- Обнаружена связь между “ощущение_движения_воздуха(bool)” и вентилятором
- Заметная связь между “способ_обогрева” и “оценка_комфорта_кат”.
- Заметная связь между “вентилятор” и “оценка_комфорта_кат”.
- Обнаружена сильная связь между утеплением и годом. Возможно, это связано со страной, так как измерения были проведены в разных странах в разные годы

Проверка гипотез

- Исследовано 9 гипотез о влиянии факторов
- Обнаружено влияние фактора “способ охлаждения” на фактор “оценка комфорта”
- Обнаружено влияние фактора “пол” на фактор “оценка комфорта”
- Обнаружено влияние фактора “количество рекламаций” на фактор “оценка комфорта”
- Обнаружено влияние фактора “страна” на фактор “оценка комфорта”
- Обнаружено влияние фактора “скорость воздуха” на фактор “оценка комфорта”
- Обнаружено влияние фактора “утепление” на фактор “оценка комфорта”

Регрессионное моделирование

- Построена модель линейной регрессии для фактора "температура воздуха в помещении"
- Для построения модели использовались факторы: "температура_воздуха_на_улице", "утепление", "rh", "способ_охлаждения", "отопление", "вентилятор", "занавески", "климат", "способ_обогрева", "двери", "окно"
- Из анализа коэффициентов линейной регрессии наибольшее влияние на модель оказывают факторы "способ охлаждения", "занавески", "климат", "отопление", "температура воздуха на улице"
- В результате работы модели на тестовых данных r^2 (коэффициент детерминации, изменяется от 0 до 1) равен 0.67, а метрика MAPE (относительная ошибка в процентах - от 0 до 100) равна 4.45. Таким образом, выбранные факторы на 67% влияют на модель, а ошибка 4.45% является достаточно низкой

Дашборд

Климатическое оборудование

Количество стран
3

Количество городов
7

Количество отзывов
413

Средняя оценка комфорта
4,81

Среднее количество рекламаций
0,12

пол Нет выбранных значений

страна Нет выбранных значений

возраст Нет выбранных значений

город Нет выбранных значений

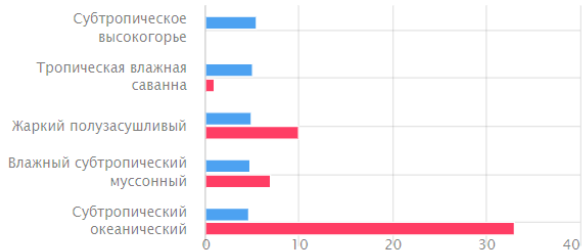
климат Нет выбранных значений

Круговая диаграмма полов



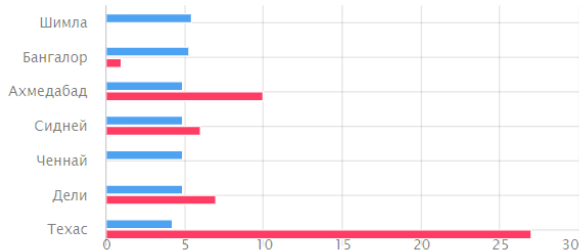
● Мужской ● Женский
● не указан

Диаграмма зависимости климат-комфорт



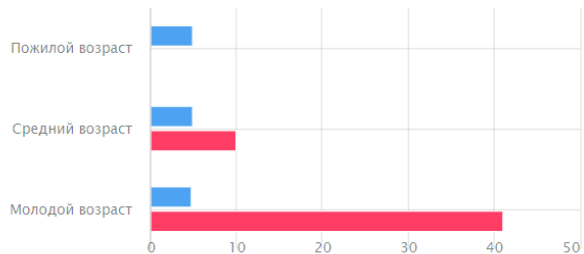
● Средняя оценка комфорта ● Количество рекламаций

Диаграмма зависимости город-комфорт



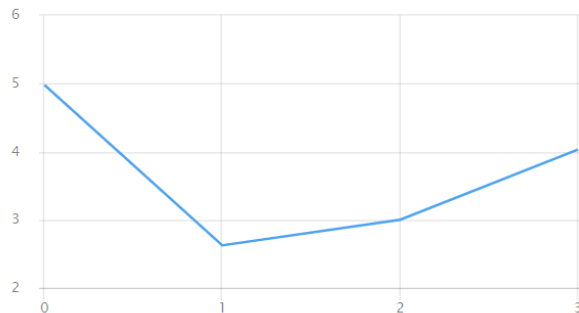
● Средняя оценка комфорта ● Количество рекламаций

Диаграмма зависимости возраст-комфорт



● Средняя оценка комфорта ● Количество рекламаций

Диаграмма зависимости рекламации-комфорт



А теперь ваши вопросы...

