# **Анализ удовлетворенности клиентов. Исследование комфорта при работе климатического оборудования.**

**Данные для проекта:**

https://drive.google.com/file/d/1eT\_a2KPPL9JvPUS0FVrW0Ofd4db3gQdd/view?usp=sharing

**Описание проекта:**

Вы работаете аналитиком в международной компании, занимающейся производством и продажами климатического оборудования: систем кондиционирования и вентиляции. На одной из выставок, посвященной климатическому оборудованию, группа инженеров из дружественной компании предоставила результаты своего исследования: данные по использованию климатических систем в офисах разных стран, городов, а также оценочные данные от респондентов по тому, насколько им комфортно было находиться в заданных условиях. Руководство поставило вам задачу проанализировать полученные данные и предоставить выводы о том, в какой тип климатического оборудования стоит инвестировать больше, чтобы повысить удовлетворенность клиентов созданными условиями в зависимости от региона? И на что стоит обратить внимание при развитии бизнеса в разных регионах?

* **Описание данных  
  Год** - Год проведения исследования   
  **Время года** - Весна, лето, осень, зима   
  **Климат** - Название типа климата   
  **Город** - Город, где было проведено исследование   
  **Страна** - Страна, где было проведено исследование   
  **Способ охлаждения:** -*Кондиционирование* = может быть разных типов, окна закрыты.  
   - *Вентиляция* = механические системы вентиляции и естественное охлаждение.  
   - *Смешанный* = механическое охлаждение, окна могут быть как открыты, так и закрыты.  
   - *NA* = не применимо, например исследование холодной зимой, когда включено отопление.

**Режим при смешанном типе охлаждения:** -*NA* = здания, не относящиеся к смешанному типу.  
 - *Кондиционер* = кондиционер включен, окна закрыты.  
 - *Вентиляция* = кондиционер выключен, окна открыты/механическая  
 вентиляция.

**Способ обогрева:** -*Механическое отопление* = в здании имеется механическая система отопления (переносные обогреватели не считаются механическим отоплением на уровне здания).  
 - *NA* = отсутствие механического отопления, неизвестно или исследование только в жаркую летнюю погоду.

**Возраст -** Возраст респондента   
**Пол** - Пол респондента   
**Ощущение температуры** - Оценка температуры по ощущениям: от -3 (холодно) до +3 (горячо)   
**Ощущение температуры (bool)** - 0 = неприемлемо, 1 = приемлемо   
**Предпочтительное изменение температуры** - Без изменений, теплее, холоднее   
**Ощущение движения воздуха (bool)** - 0 = неприемлемо, 1 = приемлемо   
**Предпочтительное изменение движения воздуха** - Меньше, без изменений, больше   
**Оценка комфорта** - От 1 (очень неудобно) до 6 (очень комфортно)   
**Утепление** - Фактор связанный с одеждой респондента   
**Температура воздуха в помещении** - Измеренная температура в помещении   
**Температура воздуха на улице** - Измеренная температура на улице   
**RH** - Относительная влажность   
**Скорость воздуха** - Скорость воздуха, м/с   
**Рост** - Рост респондента  
**Вес** - Вес респондента   
**Занавески** - 0 = открыто, 1 = закрыто; если неизвестно - нет данных   
**Вентилятор** - Режим вентилятора: 0 = выключен, 1 = включен; если неизвестно - нет данных   
**Окно** - 0 = открыто, 1 = закрыто; если неизвестно - нет данных   
**Двери** - 0 = открыты, 1 = закрыты; если неизвестно - нет данных  
**Отопление** - Режим обогрева: 0 -выключен, 1 - включен; если неизвестно - нет данных   
**Среднемесячная температура на улице** - Среднемесячная температура наружного воздуха на момент исследования, C.  
**Количество рекламаций** - Количество жалоб на установленное оборудование, полученных от клиента, за прошедший период

* **Исследование данных на качество и предобработка**Для проведения исследования необходимо:
  + Названия столбцов привести в соответствии с PEP8
  + Исследовать типы данных, привести данные к правильному типу, где это необходимо
  + Проверить данные на пропуски и заполнить их при необходимости. Обосновать решение.
  + Проверить данные на явные и неявные (в категориальных столбцах) дубликаты. Обработать дубликаты.
  + Провести аналитический и графический анализ данных. Сделать выводы.
  + Проверить данные на выбросы и адекватность данных. Предположить причину возникновения выбросов и обработать такие значения.

* **Проведение расчётов и исследований**
  + Создать дополнительный категориальный столбец, который покажет количество рекламаций по каждому клиенту отдельно: мало (≤1), средне (=2), много (>2).
  + Рассчитать средний возраст респондентов по полу и стране
  + Добавить категориальный столбец, разделив респондентов на три категории по возрасту: до 44 лет - молодой возраст, 45 - 59 лет - средний возраст, 60 лет и больше - пожилой возраст.
  + Рассчитать среднюю комфортную температуру в зависимости от возрастной категории
  + Рассчитать медианное значение температуры и влажности для каждого типа охлаждения
  + Добавить столбец с процентом удовлетворенных респондентов температурой воздуха в помещении относительно всех респондентов одной страны и одного пола
  + Составить сводную таблицу, в которой данные будут сгруппированы по стране, полу, возрастной группе и подсчитаны средняя температура воздуха в помещении, на улице и средняя относительная влажность для каждой из этих групп.
  + Самостоятельно найти информацию по стандартной комфортной относительной влажности в помещениях (в нашем случае - офисы). Создать новый категориальный столбец, в котором будет показано, удовлетворяет ли зафиксированная относительная влажность стандарту или нет. Категории назначить самостоятельно.
  + Исследовать корреляцию между параметрами. Объяснить наиболее коррелирующие параметры, имеет ли это реальный смысл или просто особенность данных?
* **Проверка гипотез**
  + Влияет ли способ охлаждения на оценку комфорта?
  + Влияет ли пол на оценку комфорта?
  + Влияет ли возрастная группа на оценку комфорта?
  + Есть ли взаимосвязь между количеством рекламаций и оценкой комфорта?
  + Проверить гипотезу о том, что средняя оценка комфорта отличается в зависимости от страны
  + Дополнительно выдвинуть и проверить не менее 3 гипотез по влиянию разных факторов на оценку комфорта/ощущение температуры/ощущение движения воздуха.
* **Регрессионное моделирование**Построить регрессионную модель температуры воздуха в помещении в зависимости от типа охлаждения и других значимых факторов. Выбор факторов - на усмотрение исполнителя.
* **Дашборд**На основе результатов выполненных расчетов и исследований составить информативный дашборд: показать наиболее интересные найденные зависимости, результаты проверки гипотез, основные статистики/распределения предоставленных данных: то, что может быть интересно и полезно заказчику для развития его бизнеса.
* **Выводы**Оформите выводы и дашборд в презентацию для руководства вашей компании (не более 7 слайдов).  
  При составлении помните про основной вопрос заказчика: “В какой тип климатического оборудования стоит инвестировать больше, чтобы повысить удовлетворенность клиентов созданными условиями в зависимости от региона? И на что стоит обратить внимание при развитии бизнеса в разных регионах?”