

Практическое задание 6. SQL для подготовки данных

Использование внешнего набора данных для выявления тенденций продаж



Рассмотреть данные переписи населения США об использовании общественного транспорта относительно ZIP-кода.

Требуется проанализировать, имеется ли корреляция между уровнем использования общественного транспорта с продажами в организации в определенном регионе.

Использование внешнего набора данных для выявления тенденций продаж



Этот набор данных содержит три столбца:

zip_code: это пятизначный почтовый индекс США, который используется для идентификации региона.

public_transportation_pct: процент населения ZIP-кода, которое, как было установлено, использует общественный транспорт для поездок на работу.

public_transportation_population: необработанное количество людей ZIP-кода, которые пользуются общественным транспортом, чтобы добраться на работу.

Использование внешнего набора данных для выявления тенденций продаж



1. Загрузите набор данных **public_transportation_statistics_by_zip_code.csv** с GitHub:
https://github.com/BosenkoTM/SQL-for-Begginer-Data-Analytics/blob/main/Datasets/public_transportation_statistics_by_zip_code.csv
2. Скопируйте данные **public_transportation_statistics_by_zip_code.csv** в базу данных клиентов организации, создав для них таблицу в наборе данных.
3. Найдите максимальное и минимальное процентное соотношение в данных. Значения ниже 0 принять как пустое значение.
4. Рассчитайте средний объем продаж для клиентов, проживающих в регионах с высоким уровнем использования общественного транспорта (более 10%), а также с низким уровнем использования общественного транспорта (менее или равным 10%).

Использование внешнего набора данных для выявления тенденций продаж



5. Загрузить данные в pandas и построить гистограмму распределения (использовать `my_data.plot.hist(y='public_transportation_pct')` для построения гистограммы).

6. Сгруппируйте клиентов по их `zip_code`, и посмотрите на среднее количество транзакций на одного клиента. Экпортируйте эти данные в Excel и создайте диаграмму рассеяния, чтобы лучше понять взаимосвязь между использованием общественного транспорта и продажами.

7. На основании этого анализа, какие рекомендации вы бы дали руководству компании при рассмотрении возможностей расширения?