



**Пояснительная записка к
тестовому заданию
на позицию младшего аналитика
сервиса «Mego.travel»**

Выполнил:

кандидат на позицию младшего аналитика
Хильчук Максим Игоревич

Описание задания

| Номер задания | Описание данных | Задача |
|---------------|---|---|
| №1 L2B | Направления - это пары городов, где каждый город представлен в виде трехбуквенного кода. Поле L2B рассчитывается как Поиски разделить на Сегменты. Округление вверх до целого. При этом если Сегменты равны 0, то L2B = Поиски. | 1.1 Найти показатель L2B для всех направлений, в которых встречается код MOW. При расчете игнорировать направления без Сегментов. |
| | | 1.2 Определить топ-3 городов с наибольшим суммарным значением L2B по всем направлениям, где они представлены. |
| №2 Маршруты | Откуда и Куда представлено 3-х буквенными кодами. День - это день недели, от 1 до 7. | Для каждой уникальной связки Откуда-Куда нужно выбрать одного наиболее выгодного поставщика. Выгодный поставщик - это такой поставщик, который больше всего дней недели оказывался дешевле (то есть имел меньшую Стоимость) других поставщиков. |

Решение

Ниже будет описана последовательность действий, логика решения, а также приведены результаты заданий.

Задание 1.1

Задание решено в двух вариантах:

- с использованием MS Office Excel ([Задание1\(EXCEL\).xlsx](#));
- с использованием СУБД PostgreSQL ([Задание1\(SQL\).sql](#)).

Решение через MS Office Excel

Было выполнено преобразование таблицы к удобному формату для написания вспомогательных формул ([Лист «Задание 1.1 формат. таблицы»](#)).

1. Столбец А «Направление» был разделен на 2 столбца (В «Откуда» и С «Куда») при помощи функции ПСТР(). Таким образом, стало возможно проверить наличие направления «MOW» для отправления и прибытия рейсов.
2. В столбце G «Проверка на условия» была прописана функция ЕСЛИ(), проверяющая каждую строку на выполнение условий встречаемости кода «MOW» и равенства столбца сегментов нулю.
3. Далее, с использованием фильтров, стало возможно выбрать только ячейки с пометкой «ок», удовлетворяющие условиям задачи.
4. Результирующая таблица к заданию представлена на [Листе «Задание 1.1 ИТОГ»](#).

Решение через СУБД PostgreSQL

Для решения был использован клиент PostgreSQL «DBeaver».

Вначале исходная таблица с данными была подготовлена к импорту в SQL клиент:

- столбец А «Направление» был разделен на 2 столбца (В «From» и С «To»);
- таблица была переконвертирована в формат csv (Рисунок 1).

| | A | B | C |
|----|-------------------------------|---|---|
| 1 | From,To,Searches,Segments,L2B | | |
| 2 | MOW,AER,5458010,1886,2893 | | |
| 3 | MOW,LED,3728800,102,36556 | | |
| 4 | AER,MOW,2741115,673,4072 | | |
| 5 | MOW,SIP,2634666,663,3973 | | |
| 6 | MOW,IST,2588738,43,60203 | | |
| 7 | LED,MOW,2504775,169,14821 | | |
| 8 | IST,MOW,2030168,14,145012 | | |
| 9 | MOW,KRR,2024813,155,13063 | | |
| 10 | MOW,MRV,2021291,908,2226 | | |

Рисунок 1. Фрагмент таблицы, подготовленной для импорта в DBeaver

Для получения ответа по условиям задачи был составлен запрос:

```
SELECT n."From", n."To", n.l2b
FROM newtable n
WHERE ("From" LIKE 'MOW' OR "To" LIKE 'MOW') AND segments != 0
```

По результатам запроса получили таблицу (Задание1.1(SQL)_результат.csv).
Результат аналогичен решению через MS Office Excel.

Задание 1.2

Задание решено в двух вариантах:

- с использованием MS Office Excel (Задание1(EXCEL).xlsx);
- с использованием СУБД PostgreSQL (Задание1(SQL).sql).

Решение через MS Office Excel

На первом этапе выполнения задания была преобразована исходная таблица с данными.

На основании ранее проведенного деления столбца с маршрутами на отдельные столбцы «Откуда» и «Куда» данные были собраны по принципу «Направление» (столбец А), город (столбец В). Данные столбцов «Поиски», «Сегменты» и «L2B» были продублированы.

На следующем этапе на основании полученной таблицы с данными была составлена автоматически собираемая сводная таблица с суммой L2b для каждого города (Лист «Задание 1.2 сводная»). В сводной данные были отфильтрованы по убыванию величины L2b.

Ответ за задание приведен в таблице 1 и на листе «Задание 1.2 ИТОГ»

| № п/п | Город | Сумма по полю L2B |
|-------|-------|-------------------|
| 1 | MOW | 16724331 |
| 2 | LED | 7399457 |
| 3 | KRR | 4679392 |

Таблица 1. Итоговая таблица для задания 1.2

Решение через СУБД PostgreSQL

Для решения были использованы данные, ранее импортированные для решения задания 1.1 (рисунок 1).

Был использован запрос:

```
SELECT p.city, sum(l2b) AS "sum l2b"
FROM (SELECT n."From" AS city, l2b
      FROM newtable n
      UNION all
      SELECT n."To" AS city, l2b
      FROM newtable n) AS p
GROUP BY p.city
ORDER BY sum(l2b) DESC
LIMIT 3
```

В запросе использован один подзапрос объединяющий данные по вылету и прилету в единую таблицу, группировка по городу, фильтрация от большего к меньшему и ограничение на 3 позиции.

В результате получена таблица (Задание1.2(SQL)_результат.csv). Полученные данные идентичны таблице 1 для решения через MS Office Excel.

Задание 2

Задание решено при помощи MS Office Excel.

К исходной таблице (Лист «Задание №2 Маршруты») был добавлен столбец С «Маршрут», в котором выполнена функция СЦЕПИТЬ() для столбцов А («Откуда») и В («Куда»).

На основании полученной таблицы была собрана сводная таблица (Лист «Этап1_Сводная таблица») с суммой стоимости перевозок каждого из поставщиков в каждый из дней.

Полученная сводная была обработана вручную (Лист «Этап2_Обработка сводной») для дальнейшего извлечения итогов задания. В столбцах В, С, D с использованием функции ЕСЛИ() вычленены данные о маршрутах и днях по отдельности из исходного столбца А. В столбце L функция ЕСЛИ() помогла определить наиболее выгодного поставщика для каждого из дней каждого направления.

Далее с использованием фильтра была отобрана информация, необходимая для итогового подсчета (Лист «Этап3_Итоговый подсчет»).

Итоговый подсчет моды для данных наиболее выгодного поставщика на каждом из направлений выполнен в столбце К. Использована многоуровневая формула (Рисунок 1) с проверками:

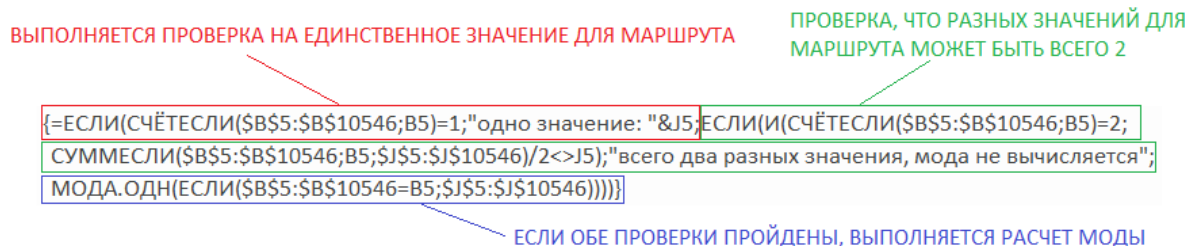


Рисунок 2. Формула для расчета наиболее выгодного поставщика для каждого из направлений (моды для каждого направления)

Функция ЕСЛИ() внутри формулы моды работает, когда формула введена "массивом" (через ctrl-shift-enter). В случаях, когда расчет моды невозможен по причине малой выборки данных внутри маршрута, в ячейках прописывается дополнительная информация.

Правее на Листе «Этап3_Итоговый подсчет» расположен блок формирования итоговой таблицы – ответа на поставленную задачу. Итоговая таблица собрана из данных при помощи функции ВПР().

Ответ на поставленное задание в итоговой таблице (без формул) представлен на Листе «ИТОГ».