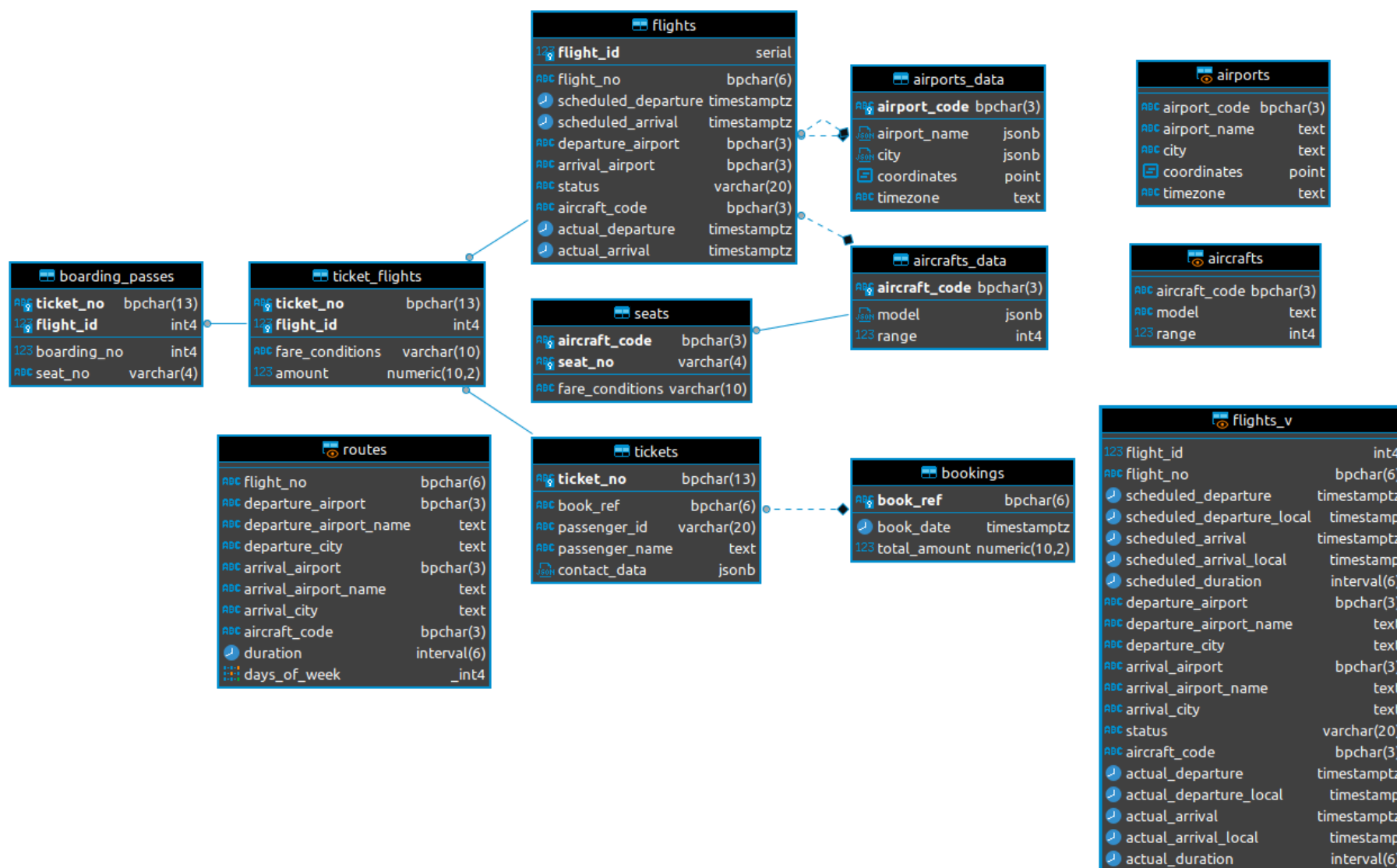


Структура датасета для анализа данных по проекту «Авиарейсы без потерь»

Михаил Пустовалов, DSPR-8

Начальные данные

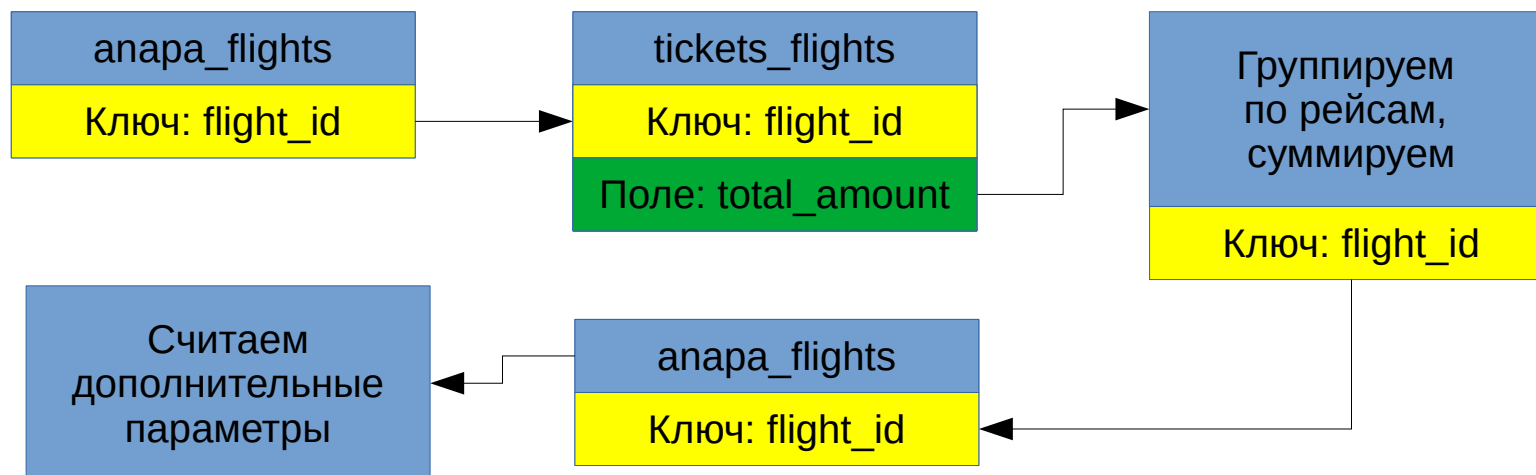


Принципы подбора данных для датасета

- Для анализа прибыльности рейса необходимы данные по «доходной» и «расходной» части.
- «Доходную» часть можно получить из данных по бронированиям и билетам.
 - Данные по выручке с каждого рейса
 - Данные по заполненности каждого рейса
 - Показатели для рейсов по стране для оценки
- «Расходную» часть оценить затруднительно, т. к. требуется большое количество дополнительных данных:
 - Расходы на топливо
 - Платежи за лизинг самолёта\отчисления в амортизационный фонд
 - Заработная плата лётного состава и бортпроводников
 - Расходы по наземному техническому обслуживанию
 - Аэропортовые расходы на слот, обслуживание пассажиров и багажа
 - Прочие сборы

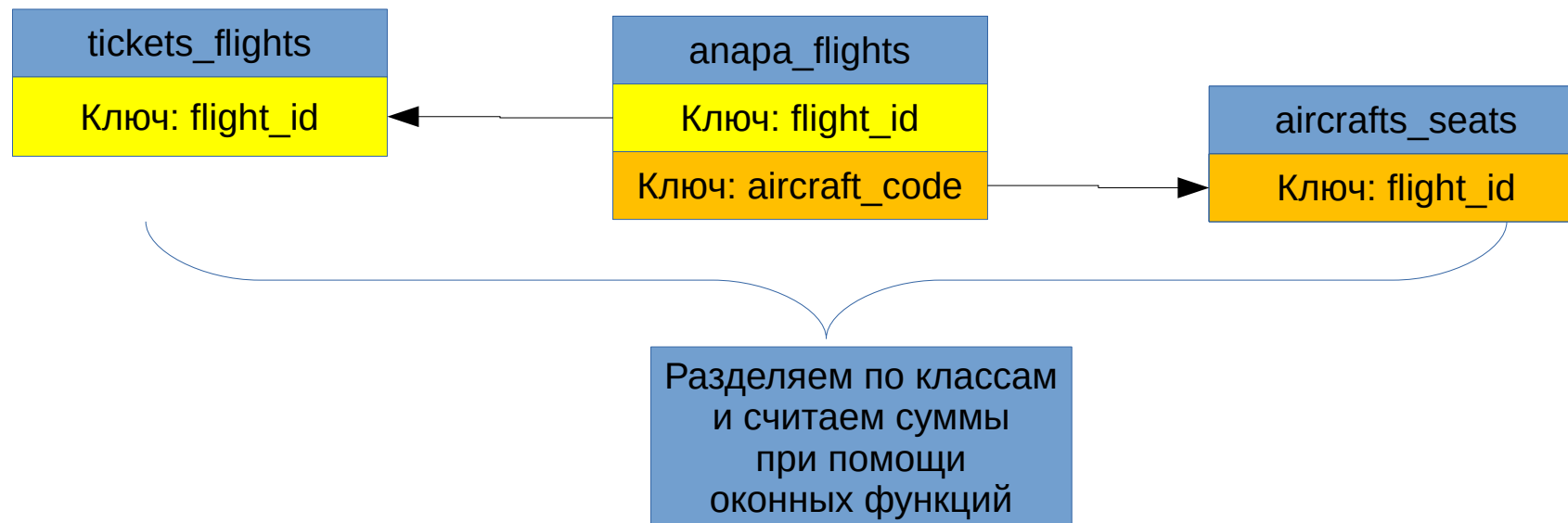
Промежуточные расчёты

- Логика расчётов выручки следующая:
 1. Выделяем во временные таблицы «зимние» рейсы из Анапы и из других городов (anapa_flights и other_flights).
 2. По таблице seats подсчитываем количество мест в каждом самолёте для разных классов обслуживания.
 3. Подсчитываем выручку по следующей схеме (для Анапы, аналогично для остальных рейсов):



Промежуточные расчёты

- Логика расчётов заполненности следующая:
 - 1) Подсчитываем количество мест в каждом самолёте суммарное и по классам, объединив таблицы aircrafts и seats, и сгруппировав по типу судна.
 - 2) Дальше по схеме:



- 3) Заполненность считается как отношение числа купленных билетов к числу мест в самолёте.

Результирующие датасеты

| anapa_dataset | | | |
|---------------|-------------------|------------|--|
| 123 | flight_id | int4 | Идентификатор рейса |
| ABC | flight_no | bpchar(6) | Номер (код) рейса |
| 🕒 | actual_arrival | timestampz | Фактическое время прибытия |
| ABC | arrival_airport | bpchar(3) | Аэропорт назначения |
| ABC | aircraft_code | bpchar(3) | Код судна |
| 123 | revenue | numeric | Выручка с рейса |
| 123 | flight_duration | numeric | Продолжительность рейса |
| 123 | anapa_rph | numeric | Выручка, отнесённая к продолжительности |
| 123 | total_tickets | int8 | Количество проданных билетов по всем классам |
| 123 | economy_tickets | int8 | Количество проданных билетов эконом-класса |
| 123 | business_tickets | int8 | Количество проданных билетов бизнес-класса |
| 123 | total_capacity | int8 | Общая вместимость самолёта |
| 123 | economy_capacity | int8 | Общее количество мест в эконом-классе |
| 123 | business_capacity | int8 | Общее количество мест в бизнес-классе |
| 123 | total_fill | numeric | Заполненность* по всем классам |
| 123 | economy_fill | numeric | Заполненность по эконом-классу |
| 123 | business_fill | numeric | Заполненность по бизнес-классу |

* Заполненность — отношение количества проданных билетов к вместимости самолёта.

Результирующие датасеты

| other_dataset | |
|-----------------------|------------|
| 123 flight_id | int4 |
| ABC flight_no | bpchar(6) |
| 🕒 actual_arrival | timestampz |
| ABC departure_airport | bpchar(3) |
| ABC arrival_airport | bpchar(3) |
| ABC aircraft_code | bpchar(3) |
| 123 revenue | numeric |
| 123 flight_duration | numeric |
| 123 other_rph | numeric |
| 123 total_tickets | int8 |
| 123 economy_tickets | int8 |
| 123 business_tickets | int8 |
| 123 total_capacity | int8 |
| 123 economy_capacity | int8 |
| 123 business_capacity | int8 |
| 123 total_fill | numeric |
| 123 economy_fill | numeric |
| 123 business_fill | numeric |

Идентификатор рейса

Номер (код) рейса

Фактическое время прибытия

Аэропорт вылета

Аэропорт назначения

Код судна

Выручка с рейса

Продолжительность рейса

Выручка, отнесённая к продолжительности

Количество проданных билетов по всем классам

Количество проданных билетов эконом-класса

Количество проданных билетов бизнес-класса

Общая вместимость самолёта

Общее количество мест в эконом-классе

Общее количество мест в бизнес-классе

Заполненность* по всем классам

Заполненность по эконом-классу

Заполненность по бизнес-классу

* Заполненность — отношение количества проданных билетов к вместимости самолёта.

Результаты

- Датасет подготовлен.
- В данных по выручке для всех рейсов в Новокузнецк — 0, данных по проданным билетам нет. Возможно, это отсутствующие данные (что наиболее вероятно). В противном случае получается, что в Новокузнецк из Анапы никто не летает, а значит рейс надо отменять. В любом случае требуется перепроверка этих данных.
- Для оценки «доходных» показателей предлагается использовать выручку с каждого рейса из Анапы, выручку, отнесённую к часу полёта, а также сравнение выручки с другими рейсами.
- Также можно использовать показатели «заполненности» каждого рейса.
- Для оценки «расходных» показателей необходимо большое количество дополнительных данных:
 - Расходы на топливо (расчёт количества топлива на рейс — нетривиальная задача, т. к. в разных фазах полёта расход различен, а также зависит от ветра\эшелона и т.д.)
 - Платежи за лизинг самолёта\отчисления в амортизационный фонд
 - Заработная плата лётного состава и бортпроводников
 - Расходы по наземному техническому обслуживанию
 - Аэропортовые расходы на слот, обслуживание пассажиров и багажа
 - Аэронавигационные сборы