

# Проект 4

## Авиарейсы без потерь.

Целиков Сергей  
Группа DST-PRO-11

# Структура датасета:

- **flight\_id** - Идентификатор рейса. Характеризует конкретный рейс. Уникальное поле
- **flight\_no** - Номер рейса. Определяет авиакомпанию. Необходим для дальнейшей возможной группировки
- **departure\_airport** – IATA код аэропорта вылета. Информационное поле. Для данного датасета одинаковое для всех строк. Можно удалить при постобработке.
- **departure\_airport\_name** – Название аэропорта вылета. Информационное поле. Для данного датасета одинаковое для всех строк. Можно удалить при постобработке.
- **departure\_city** – Название города вылета. Информационное поле. Для данного датасета одинаковое для всех строк. Можно удалить при постобработке
- **departure\_latitude** – Географическая широта аэропорта вылета. Может использоваться в дальнейшем для определения дальности перелета
- **departure\_longitude** – Географическая долгота аэропорта вылета. Может использоваться в дальнейшем для определения дальности перелета
- **arrival\_airport** - IATA код аэропорта прилета. Информационное поле. Может использоваться для группировки данных при постобработке.
- **arrival\_airport\_name** - Название аэропорта прилета. Информационное поле. Может понадобиться для связи с внешними данными или его можно удалить при постобработке.
- **arrival\_city** – Название города прилета. Информационное поле. Может понадобиться для связи с внешними данными или его можно удалить при постобработке

# Структура датасета (продолжение):

- **arrival\_latitude** - Географическая широта аэропорта прилета. Может использоваться в дальнейшем для определения дальности перелета
- **arrival\_longitude** - Географическая долгота аэропорта прилета. Может использоваться в дальнейшем для определения дальности перелета
- **scheduled\_departure** – Плановое время вылета. Может использоваться в постобработке.
- **scheduled\_arrival** – Плановое время прилета. Может использоваться в постобработке.
- **actual\_departure** – Реальное время вылета рейса. Может использоваться в постобработке.
- **actual\_arrival** – Реальное время прилета рейса. Может использоваться в постобработке
- **dow** – День недели (0 – воскресенье, 6 – суббота). Может использоваться для оценки прибыльности рейсов по дням недели
- **scheduled\_len** – Плановая длительность рейса в минутах. Может использоваться для оценки прибыльности рейсов
- **actual\_len** – Реальная длительность рейса в минутах. Может использоваться для оценки прибыльности рейсов
- **departure\_delay** – Задержка отправления рейса в минутах относительно плановой. Может использоваться для оценки прибыльности рейсов
- **arrival\_delay** – Задержка прибытия рейса в минутах относительно плановой. Может использоваться для оценки прибыльности рейсов

# Структура датасета (продолжение):

- **aircraft\_code** – Код модели воздушного судна. Может использоваться для группировки при анализе
- **model** – Название модели воздушного судна. Информационное поле. Может понадобиться для связи с внешними данными
- **range** – Максимальная дальность полета воздушного судна в км. Может понадобиться для оценки рентабельности использования данного типа воздушного судна на рейсе.
- **economy\_seats** – Общее количество мест эконом-класса для данного судна. Необходимо для анализа рентабельности рейса.
- **business\_seats** – Общее количество мест бизнес-класса для данного судна. Необходимо для анализа рентабельности рейса.
- **economy\_seats\_sold** - Количество проданных мест эконом-класса для данного рейса. Необходимо для анализа рентабельности рейса.
- **economy\_revenue** – Выручка от продажи мест эконом-класса для данного рейса. Необходимо для анализа рентабельности рейса.
- **business\_seats\_sold** - Количество проданных мест бизнес-класса для данного рейса. Необходимо для анализа рентабельности рейса.
- **business\_revenue** – Выручка от продажи мест бизнес-класса для данного рейса. Необходимо для анализа рентабельности рейса.

# Данные, которые можно добавить

При проведении исследования базы данных можно было бы получить выборку из таблицы билетов и полетов (Ticket\_flights) в привязке к таблицам посадочных талонов (Boarding\_passes), бронирований (Bookings) и перелетов (Flights). Это позволило бы получить список всех билетов и бронирований для исследуемых рейсов.

Подобная выборка могла бы дать следующую дополнительную информацию:

- За сколько дней до вылета осуществляется бронирование для конкретных пассажиров (является ли полет запланированным или спонтанным).
- Количество человек в бронировании (летит ли человек один или с семьёй/друзьями/коллегами)
- Сравнив проданные билеты и посадочные талоны, можно посчитать пассажиров, которые выкупили билеты, но не явились на посадку. Скорее всего, для этих пассажиров будут произведены возвраты средств и это уменьшит выручку от продажи билетов на рейс.
- Можно также провести из базы выборку обратных рейсов (с прилетом в город Анапа) в указанный интервал времени (+- один месяц), оценить также экономику этих рейсов и посчитать количество пассажиров, воспользовавшихся рейсами в обе стороны.
- Можно посчитать пассажиров, которые летают часто. Возможно, они пользовались данным направлением несколько раз за указанный период? Или за более широкий период времени? Какие другие направления они предпочитают?
- По базе посадочных талонов можно провести анализ предпочитаемых посадочных мест для разных пассажиров.
- Существуют ли альтернативные (стыковочные) варианты перелетов между городами. Какая их длительность/стоимость?

К сожалению, данный анализ практически невозможен из-за низкой производительности сервера, что делает затруднительным получение выборки в разумный период времени.

# Данные, которые можно добавить из ВНЕШНИХ ИСТОЧНИКОВ

- Расстояние между аэропортами (можно посчитать используя географические координаты)
- Стоимость горючего для каждого типа самолета и его потребление на час полета.
- Аэропортовые сборы с авиакомпаний и пассажиров
- Стоимость лизинга данного типа самолета.
- Количество членов экипажа, расходы авиакомпании в ФОТ для экипажей.
- Стоимость проживания экипажей в гостиницах (если обратный рейс приходится на следующие сутки)
- Возможность выбора альтернативных типов воздушных судов для рейса.