

# Курсовая работа – Аналитика внутренних авиарейсов с помощью методов Big Data.

## 1. Бизнес – анализ данных

Имеется датасет размером примерно 6 млн строк с данными по всем авиарейсам за 2015 год в США, в т.ч. отмененным и перенесенным.

Датасет представлен следующим набором файлов:

**Flights** – Данные о полетах авиакомпаний США за 2015 год.

**Airlines** – Справочник авиакомпаний США (коды IATA).

**Airports** – Данные аэропортов США.

**L\_AIRPORT** – Справочник аэропортов мира (коды IATA).

**L\_AIRPORT\_ID** – Справочник аэропортов мира (цифровые коды).

В качестве знакомства с датасетом проведем первичную аналитику:

- Построим график количества рейсов по месяцам и ответим на вопрос – на какие месяцы приходится пик перелетов и когда отмечается спад.
- Построим график количества рейсов по дням недели и ответим на вопрос – как распределено количество рейсов по дням недели в целом.
- Определим топ 10 городов по количеству рейсов - будут ли среди них заметно выделяющиеся хабы.
- Определим топ 10 авиакомпаний по количеству рейсов - это и будут самые популярные авиакомпании страны.

## 2. Исследование данных

С целью оптимизации текущих процессов в работе авиасообщения в США и тиражирования местных лучших практик необходимо найти ответ на следующие вопросы:

- Найти аэропорт с минимальной задержкой вылета
- Выявить самую пунктуальную авиакомпанию на прилет в \*
- Найти аэропорт, где самолёты проводят в среднем больше всего времени на рулении

\* - Порядковый номер IATA кода из файла airports для 211 группы = номер группы - порядковый номер студента в ведомости

\* - Порядковый номер IATA кода из файла airports для 212 группы = номер группы + порядковый номер студента в ведомости

Рассчитывая минимальную задержку вылета обратим внимание как на топ 10 аэропортов с минимальной задержкой вылета (средней), так и найдем самую среднюю задержку вылета для аэропортов из топ-10 по количеству рейсов.