# Разведочный анализ данных абитуриентов ИИК

**Дано:** анонимизированная выгрузка данных о студентах, поступивших в ИИК.

**На выходе:** Jupyter Notebook, иллюстрирующий выполненное задание, с графиками и пояснениями.

**Первичная обработка**

1. Сделать преобразование данных в графе «пол» - М=мужской, Ж=женский
2. Вместо полей "информатика" и "физика" завести два поля: "3балл " и "3экз" - в одном балл по физике или информатике, в другом строка, в которой написан предмет, по которому сдан экзамен (физика или информатика).
3. Баллы за индивидуальные достижения трансформировать в два булевых столбца: "олимпиада" и "медаль"
4. "Школа" распарсить в поля "населённый пункт" и "школа", добиться как можно меньшего числа вариантов написания одной и той же школы

Можно п. 4 отложить на потом, если будет тягостно, сконцентрировавшись на анализе того, что хорошо парсится.

**Анализ датасета**  
Провести разведочный анализ данных с целью выявления неожиданных примеров корреляции признаков, желательно, касающихся выбора направления. Для основного анализа можно рассматривать только поступление на бюджет. Остальных можно вообще убрать из анализа.

Например, попробовать проанализировать:

* Есть ли изменения в портрете поступающего на ИБАС (или ИИК в целом) со временем? (в распределении баллов, гендере и пр.)
* Есть ли отличия студента, поступающего на ИБАС в сравнении с ПМИ или ФИИТ или ИВТ или ИИК в целом (школы, пропорции в баллах ЕГЭ, олимпиады, медали)?