# Основные тезисы

Данные по рекомендациям мы можем получить в трех категориях - Книги, Кружки, Мероприятия. Данные по пользователям есть только в Книгах и Кружках.

## 1) Книги

### Пользователи

По пользователям есть только данные о возрасте (дата рождения)

### Рекомендовать можно:

Книгу, Автора

### Комментарии:

Вся полезная информация в таблице **Каталог.xlsx**, по возрастам пользователей в **Читатели.xlsx**. Таблицы сопоставлений-связи (**Выдача\_1.xlsx**, **Выдача\_2.xlsx**) - эти таблицы не содержат полезной информации - только для сопоставления книги и пользователя.

### Методы решения:

Предполагается сопоставление пользователей по различным признакам (кластеризация) или другие методы, исходя из этого рекомендовать те книги и тех авторов, которые характерны для данного множества.

## 2) Кружки

### Пользователи

По пользователям есть только данные о возрасте (сколько лет) и поле.

### Рекомендовать можно:

Услугу (конкретный кружок), Организацию (где можно записаться на разные активности), возможно тему (Архитектура, Музыка итд)

### Комментарии:

Теоретически, организация кружка куда ходит пользователь предоставляет нам информацию о территориальном расположении пользователя, исходя из этого мы просто можем рекомендовать ближайшие организации и услуги по территориальному признаку то есть организация - сильный признак. Таблица связи-сопоставления - **MegaRelation\_hackaton.csv** здесь есть сопоставления пользователя, услуги, организации, итд. Таблица **request\_hackaton.csv** не несет никакой полезной информации так как в таблице связи **MegaRelation\_hackaton.csv** зафиксировано что пользователю интересна определенная услуга, то смог ли пользователь сдать экзамены, закончить обучение итд нам не важно.

### Методы решения:

Как и в случае с книгами сопоставление пользователей по похожести (кластеризация) и рекомендация характерных для данной группы услуг и организаций. Можно попробовать провести простую классификацию на основании данных по организации, услуге, теме, возрасте, поле пользователя, в качестве целевых признаков используя организацию или тему, возможно конкретно услугу.

## 3) Мероприятия

Указаны только названия, тип мероприятия, возрастная категория где проходит но рекомендацию можно выдать только на основе возраста.

### Пользователи

Нет никакой информации

### Рекомендовать можно:

Мероприятие

### Комментарии:

Просто получаем список будущих мероприятий и на основании возраста рекомендуем пользователям из Книг и Кружков определенные мероприятия

### Методы решения:

Просто сопоставление возраста и возрастной категории мероприятия

## Таблицы

|  |  |  |
| --- | --- | --- |
| **Книги** | | |
| **Таблица** | **Описание** | **Комментарий** |
| Каталог.xlsx | Справочник книг: название, автор, издательство | Содержит главную информацию по книгам |
| Читатели.xlsx | Данные по пользователям - указан только возраст | Информация по пользователям |
| Экземпляры.xlsx, Экземпляры\_2.xlsx | Таблицы сопоставлений конкретных экземпляров и названий книг из каталога | Не содержит полезной информации |
| Выдача\_1.xlsx, Выдача\_2.xlsx | Таблицы сопоставлений пользователей и экземпляров книг | Не содержит полезной информации |
| **Кружки** | | |
| Classificator\_hachaton.csv | Классификатор кружков: id классификатора и название темы | Содержит классификацию типов кружков |
| MegaRelation\_hackaton.csv | Таблица сопоставлений пользователей, организаций, услуг | Не содержит полезной информации |
| org\_hackaton.csv | Данные по организациям и их географии (метро) | Информация по организациям |
| Pupil\_hackaton.csv | Данные по пользователям - указан возраст и пол | Информация по пользователям (в основном дети) |
| request\_hackaton.csv | Данные о заявлениях на запись - зачислен или нет | Не содержит полезной информации |
| services\_hackaton.csv | Данные по услугам кружков - id услуги, тема (классификатор), организация | Информация по услугам (кружки) |
| **Мероприятия** | | |
| Мероприятия.xlsx | Данные по прошедшим и будущим мероприятиям | Информация о мероприятиях и возрастной категории |