1. **Описать основные бизнес-отчеты (2-3 штуки), которые мы хотим видеть по нашему бизнесу**

Персонализация рекомендаций:

* рекомендовать фильмы, которые понравились пользователям с похожим набором фильмов (жанры, актеры…)
* после завершения просмотра и выставление положительной оценки фильму рекомендовать его продолжение или похожий фильм
* рекомендовать фильмы с актерами, которые часто встречались в просмотренных пользователем фильмах (аналогично с режиссерами)

Доработка поиска:

* добавить возможность поиска по ситуационным тегам, например, «осенняя депрессия под пледом», «зимнее настроение без выхода на улицу в минус 30», «веселые фильмы для сильно веселой компании» и т.д.
* добавить возможность поиска еще не вышедших в прокат фильмов и выставления напоминаний об их выходе

Привлечение давно не заходивших пользователей:

* направление еженедельных уведомлений: популярные фильмы за неделю, похожие на просмотренные ранее
* предоставление скидки на подписку (размер скидки зависит от количества просмотренных фильмов и даты его регистрации)

1. **Описать основные имеющиеся данные и источники их поступления**

Основной источник - сайт онлайн-кинотеатра.

Информация о пользователях:

* Основные данные (Id, имя, фамилия, дата рождения….)
* Дата регистрации
* Дата последнего входа
* Дата последнего просмотренного фильма

Информация о фильмах:

* Основные данные (Id, наименование, дата выхода, продолжительность, описание, страна…)
* Список актеров, режиссер
* Список жанров
* Рейтинг

Информация о просмотрах:

* По фильмам и пользователям
* Дата просмотра
* Продолжительность просмотра
* Поставленная оценка

1. **Описать основные сущности в хранилище данных (схема звезда) и процесс заливки данных**

Описание во вложенных файлах:

* er-diagram.jpg
* таблица description (лист «fields»)

1. **Описать основные проверки на качество данных (10 штук), которыми будем пользоваться при заливке**

**Таблица Views:**

* Значение **rating** должно быть целым числом в интервале от 1 до 10.
* При повторном просмотре пользователем фильма необходимо **вносить изменения в уже существующую запись с user\_id пользователя и movie\_id фильма**, а не создавать новую запись.
* **Новая запись** в таблице создается только после начала воспроизведения соответствующего фильма, а значение **view\_date** заполняется автоматически датой и временем начала воспроизведения.

**Таблица Users:**

* Значение поля **login** должно быть непустым и уникальным.
* Значение поля **password** должно содержать не менее 8 и не более 50 символов, должны использоваться только латинские буквы или арабские цифры. Обязательно наличие минимум одной цифры и заглавной буквы.
* Значение поля **email** должно быть непустым. Это необходимо для подтверждения учетной записи пользователя и изменения пароля.
* Поля **last\_login\_date** и **last\_movie\_date** заполняются автоматически при очередном входе пользователем в личный кабинет и в момент начала воспроизведения любого фильма, соответственно.

**Таблицы Subscriptions (categories, statuses):**

* Категория и статус подписки выбираются пользователем на этапе оформления подписки. Должна быть возможность выбрать только из имеющихся значений **category\_name** и **status\_name**.
* При оформлении пользователем новой подписки в таблицеSubscriptionsсоздается **новая запись**. При продлении изменения вносятся в уже имеющуюся запись подписки (изменяется subscription\_date).

**Таблица Movies:**

* Обязательно должны быть заполнены поля creator\_id, stars\_ids, genre\_ids, release\_date, title, runtime, country\_id, language\_id, certificate.
* Поля **creator\_id**, **stars\_ids**, **genre\_ids** могут содержать как одно, так и несколько значений в виде списка.
* Значение поля **rating** вычисляется автоматически на основании всех оценок фильма, изменяется при добавлении новой оценки. Формат numeric( ,2).

**Таблицы Creators, Languages, Genres, Libraries, Stars:**

* Поля **creator\_first\_name**, **language\_name**, **genre\_name**, **country\_name**, **star\_first\_name** в соотвествующих таблицах должны быть непустыми.

1. **Придумать Data-проект, который должен улучшить показатели Вашего бизнеса и расписать его по Crisp-DM**

Во вложенных файлах:

* таблица description (лист «Crisp-DM»)

1. **Описать требуемые роли в команде по работе с данными на этапах 4 и 5**

Во вложенных файлах:

* таблица description (лист «Crisp-DM»)

**Этап 4** (обучение модели, оценка результатов):

* DS-специалист

**Этап 5** (оценка результатов работы модели, оценка необходимости доработки модели):

* Аналитики
* DS-специалист