

Теоретическая часть

1. Основные бизнес-отчеты по объему просмотров в интернет-кинотеатре.

- Анализ пользователей (в разрезе пола, возраста, суммы трат, предпочтения по жанрам)
- Анализ продаж (общие суммы и кол-ва по подпискам и единоразовым просмотрам, по скачиваниям фильмов, так же и в разрезе жанров)
- Анализ просмотров рекламы (кол-ва просмотров, переходов).

2. Описать основные имеющиеся данные и источники их поступления.

Данные	Источник
Информация по пользователям (имя, имейл, возраст, пол, и тд)	Анкета при регистрации
Кол-во единоразовых просмотров и суммы от аренды и покупки фильма	Статистика по сайту
Кол-во и суммы по ежемесячным подпискам	Статистика по сайту
Кол-во просмотров рекламы	Статистика по сайту
Кол-во загрузок фильмов	Статистика по сайту
Данные по рейтингам	Статистика по сайту
Информация по фильмам (название, год, жанр)	Список фильмов

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

Модель данных состоит из двух типов таблиц:

1. Таблица фактов FactService — факт оказания услуги (подписка, аренда, покупка фильма).
2. Таблицы измерений:
 - DimCustomer - список пользователей с информацией о них;
 - DimService - каталог услуг с ценами;
 - DimMovie - каталог фильмов с информацией о них;
 - DimGenre - список жанров;
 - DimRating - данные по рейтингам.

Схема Звезды:

	release date date	not null
	genre_id int FK	not null
DimRating	customer_id int PK FK	not null, выбираются из соответствующей таблицы (для которой являются PK)
	movie_id int PK FK	not null, выбираются из соответствующей таблицы (для которой являются PK)
	rating int	not null, check between 1-5
	date timestamp	not null, автоматически текущее время
Вся база данных		Ограничения внешнего ключа restrict

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

Проект - разработать модель рекомендации фильмов пользователю на основе его предыдущих просмотров и проставления рейтингов или на основе рейтингов от других пользователей из похожей целевой аудитории (по полу/возрасту).

Фаза	Действия	Специалисты
Business Understanding	Анализ информации по пользователям и истории их просмотров, анализ данных по рейтингам (средний по фильму, средний по жанру, по кол-ву оценок), определение и расчет метрик (покупки, рейтинги и т.п.), оценка ситуации, анализ затрат/выгод и ожидаемых результатов от внедрения проекта, создание плана проекта и целей для DS.	Аналитик/BI, владелец продукта
Data understanding	Сбор необходимых данных, исследования, проверка на качество.	Аналитик
Data preparation	Создание обучающей и тестовой выборки, очистка и расширение данных, создание data frame для обучения модели, документирование всех выполненных шагов по подготовке данных.	DS

Modeling	Построение модели, обучение модели, оценка качества, описание результата.	DS
Evaluation	Оценка результатов проекта, анализ с точки зрения внедрения в бизнес процесс.	Аналитик
Deployment	Создание плана внедрения, планирование мониторинга и тех.поддержки, внедрение проекта и заключительная отчетность.	DS, разработчик, аналитик, владелец продукта.

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

Роли	Описание
Специалисты по БД	<ul style="list-style-type: none"> ● Проектирование и оптимизация БД ● Создание и оптимизация запросов ● Контроль безопасности БД
Аналитики	<ul style="list-style-type: none"> ● Сбор бизнес-данных, анализ, отчётность. ● Составление, валидация, оценка метрик ● Прогнозирование и моделирование исхода различных вариантов действий ● Отслеживание результатов принятия решений
DS специалисты	<ul style="list-style-type: none"> ● Анализ данных, отчётность. ● Построение и валидация моделей ● Работа с большими данными
Разработчики	<ul style="list-style-type: none"> ● Разработка API и интеграция ● Настройка получения требуемых данных ● Внедрение модели и тестирование
Владелец продукта	<ul style="list-style-type: none"> ● Менеджмент ● Создание плана проекта и целей для DS. ● Контроль, внедрение и продвижение проекта.