

OPPORTUNITY CUP 2022

Подозрительный кейс:

Разработайте алгоритм
для определения мошеннических
банковских операций



Добро пожаловать!



Дорогие команды, мы рады видеть вас среди участников кейс-чемпионата Opportunity Cup 2022!



В рамках ИТ-трека вам предстоит решить кейс от Сбера по обнаружению мошеннических операций среди всех транзакций банка.

В тексте кейса вы найдете:

- ☐ Вводную информацию о работе с базами данных;
- ☐ Новые термины, такие как “fraud” и “ETL-процесс”;
- ☐ Критерии оценки решений отборочного этапа.

Для помощи в подготовке вашего решения мы проведем:

29.09, 18:30

Вводный онлайн-тренинг, где опытный кейсер **Максим Комяков**, выигравший не один кейс-чемпионат, поделится лайфхаками по успешному решению кейсов в рамках различных чемпионатов!

01.10, 14:00

Вводный онлайн-брифинг и QA-сессия с экспертами-составителями кейса:

Наталья Носенко – руководитель направления, департамент SberData;

Алексей Сазанов – исполнительный директор, лидер продукта.





В банковской сфере **Фрод (Fraud)** — это проведение мошеннических (неправомерных) операций с использованием банковских карт без ведома их владельцев.

В таких случаях убытки несет и покупатель, и продавец, так как теряет товары, купленные фродером и сумму за комиссию при переводе платежа. Дополнительно, продавцу придется платить комиссию за чарджбек (возврат денег) и компенсировать расходы на расследование.

Как обнаружить фрод?

Выявление мошеннических транзакций возможно путем анализа сохранённых записей по этим транзакциям, а именно поиск отклонений от нормы.

Пример признаков фрода:

- Очевидно шаблонное поведение (например, равные временные промежутки между кликами)
- Много кликов с одного IP/ID
- Подозрительная активность в ночные часы

В данном кейсе вам предстоит поработать с форматом **JSON**.

Это простой, удобный для чтения как человеком, так и компьютером текстовый формат обмена данными. Он представляет собой текст в виде пар параметр-значение. **JSON** позволяет передавать не только сами данные, но и их структуру. Встроенная поддержка **JSON** есть во многих языках программирования (Python, C#, C++, PHP ...).

В СУБД PostgreSQL Server, начиная с 2016 версии, существует возможность встроенной работы с форматом данных JSON.



Сбер не единственная компания в мире, которая страдает от фрода и от мошенников. Это настолько большая проблема, что ей занимаются целые государственные департаменты. В самой основе современной финансовой индустрии лежат политики противодействия мошенничеству, отмыванию денег и финансированию терроризма.

Постановка задачи



Новое сообщение

Кому: Команда SberData

Тема: Цели проекта

ЦЕЛЬ: Используя полученный JSON - файл с банковскими транзакциями реализовать алгоритм выявления мошеннических операций:

1. Подобрать техническое средство для обработки файла;
2. Подготовить блок-схему ETL-процесса с описанием функциональных компонентов, этапов преобразования данных и использованных технических средств;
3. Выделить не менее 3 fraud-паттернов с их подробным описанием;
4. Получить список мошеннических операций на основе выявленных паттернов.

Решение необходимо подготовить в виде двух файлов:

1. Excel - отчет* с перечнем паттернов мошеннических операций и списком подозрительных транзакций с разметкой по паттернам;
2. Презентация с подробным описанием подхода к решению, используемых средств и полученных результатов.

*Отчет необходимо оформить в приложенном шаблоне Excel-отчета

Переслать

Ответить



Мы ждем ваши решения до 6:00 МСК 7 октября!



Присылайте их на почту Cases_opportunity-cup.ru

.PDF

Презентацию в формате .pdf или .pptx

.XLSX

Excel-отчет в формате .xlsx



В презентацию обязательно включите слайд с резюме команды (имена, вузы, опыт)



Презентация объемом не более 8 слайдов (не считая титульный и приложения)

Ваше решение должно содержать **2 файла**:

- Презентация в формате .pdf с отчетом о проделанной работе: объем не более 8 слайдов, включая титульный слайд, резюме решения и резюме команды.
- Excel-файл, заполненный в соответствии с полученным шаблоном. В файле оценивается:
 - Количество выявленных fraud-паттернов
 - Процент покрытия выявленных мошеннических транзакций

Критерии оценки презентаций с отчетом:

Качество анализа

- Качество проведенного анализа входных данных;
- Описание способа обработки данных (используемый подход);
- Уровень бизнес-логики формирования паттернов;
- Качество и уровень описания принципиальной схемы работы с данными;
- Ссылки на аналитические и информационные источники;
- Каждое действие подкреплено сильной аргументацией;
- Все факты и цифры подтверждены расчетами и ссылками на источники.

Структура и логика

- Решение следует единой логике, нет необоснованных переходов между темами;
- Отслеживается четкая структура решения (связи между элементами решения);
- Решение целостно – отсутствуют пробелы в аргументации.

Соответствие поставленной задаче

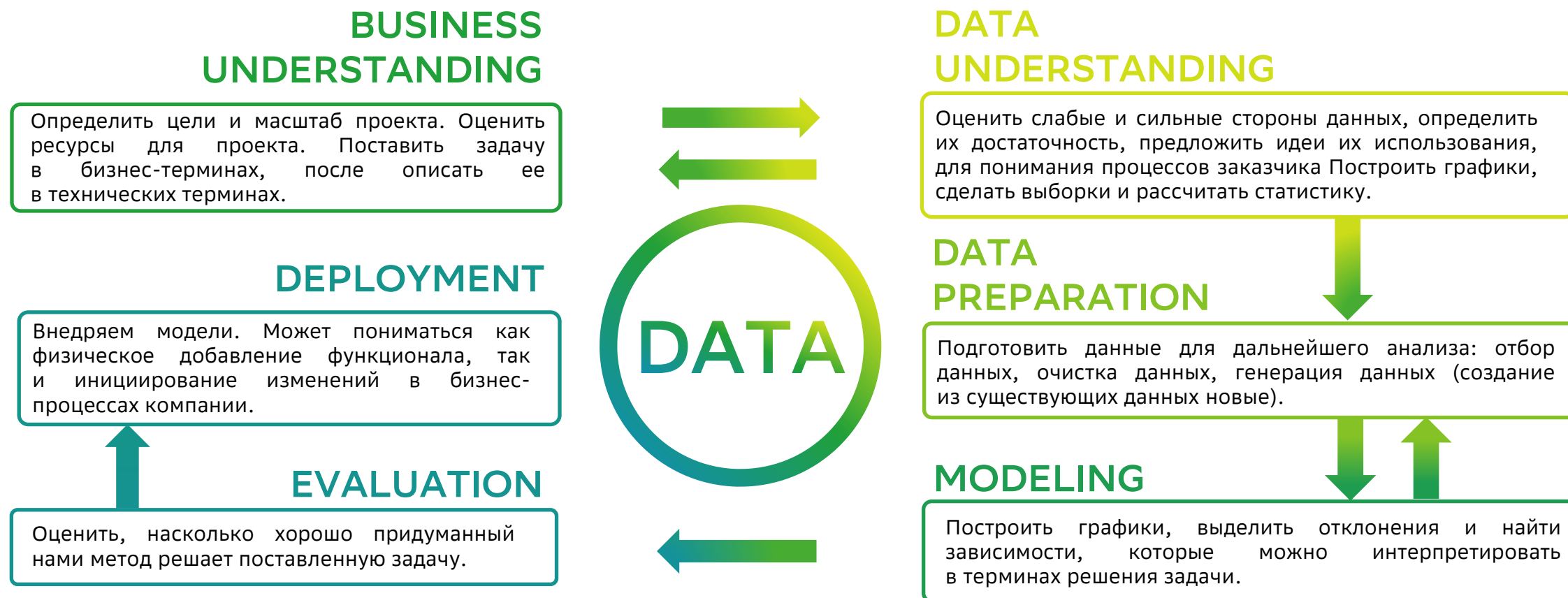
- Даны точные ответы на вопросы кейса, а не собственную постановку задачи;
- Решение адекватно и принципиально реализуемо;
- Использован оригинальный, но соответствующий задаче подход к решению или идея.

Качество презентации

- Соблюдены формальные требования: объем, резюме команды и решения;
- Все слайды имеют четкую структуру: информация читается в логичном порядке, понятны акценты;
- Использована корректная цветовая палитра;
- Использован деловой стиль оформления: в тексте и в картинках;
- Нет технических ошибок, опечаток (рекомендуем отправлять решение в формате .pdf, предварительно проверив).

Процесс исследования данных (CRISP-DM)

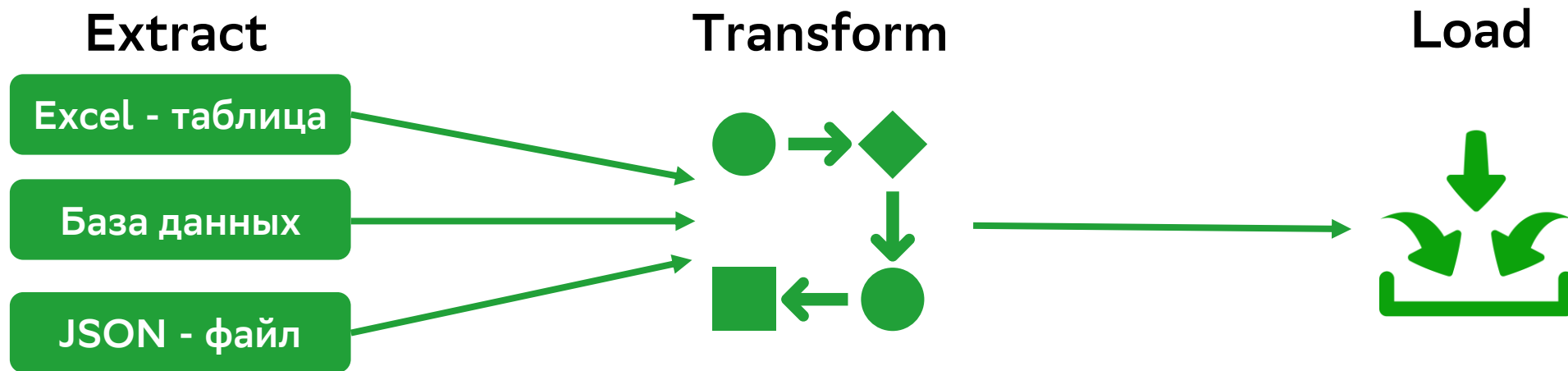
После того, как мы получим данные из **JSON** - файла, можем приступить к их подготовке и выделению нужных паттернов. При построении вашего решения, вы можете опираться на методологию **CRISP-DM** или любую другую, удобную для вас.



Описание ETL-процесса

Что такое **ETL**?

Аналитик данных работает с информацией из множества разных источников. Чтобы анализировать разнородную информацию, нужен универсальный процесс, оптимизирующий работу с данными. **ETL** — это процесс транспортировки данных, при котором информацию из разных мест преобразуют и кладут в новое место. ETL расшифровывается как extract, transform, load, то есть «извлечь, трансформировать, загрузить».



Extract. Система берет данные из источников и перемещает в промежуточный буфер для дальнейшей обработки. Также может проводиться валидация, проверка данных на соответствие тем или иным критериям. Система проверяет, можно ли загрузить их без потерь в новое хранилище.

Transform. Это «сердце» любого ETL, этап, когда мы применяем бизнес-логику и делаем фильтрацию, группировку и агрегирование, чтобы преобразовать сырые данные в готовый к анализу датасет. Эта процедура требует понимания бизнес-задач и наличия базовых знаний в области.

Load. Наконец, мы загружаем обработанные данные и отправляем их в место конечного использования. Полученный набор данных может быть использован конечными пользователями, а может являться входным потоком к еще одному ETL.

Приложение: карьерные возможности в Сбере



SBERSEASONS – это оплачиваемая программа стажировок для студентов IT, математических, экономических и технических специальностей.

Какие преимущества есть у стажёров?



Поддержка сообщества,
кураторы



Боевые задачи



Совмещай работу
и учёбу
Нужно быть студентом 3+
курса очной формы обучения



Наставник
и обучение



Миллионы пользователей
продукта

Кроме **Sberseasons** на портале
SberGraduate можно узнать:

- ❑ О том как стать Амбассадором Сбера в своем университете
- ❑ О том как пройти практику и зачесть ее в рамках своей образовательной программы

В рамках **SberSeasons** открыт набор по 5 направлениям:

- ❑ Анализ данных
- ❑ Разработка
- ❑ Кибербезопасность
- ❑ Бизнес
- ❑ Дизайн

Подробный список направлений
стажировки

Подпишись
на рассылку, чтобы
не пропустить
важные анонсы!

Прокачивайся
вместе со Сбером
на образовательных
программах!



OPPORTUNITY



CUP 2022



НИЯУ
МИФИ

Команда Карьерного клуба «Старт Карьеры» НИЯУ МИФИ подготовила данный кейс исключительно для использования в образовательных целях в рамках кейс-чемпионата «Opportunity Cup 2022». Авторы не намереваются иллюстрировать как эффективное, так и неэффективное решение проблемы. Данные, описанные в кейсе, не обязательно являются реальными и также могли быть изменены или выдуманы, с целью сохранения конфиденциальности данных.

Cases@opportunity-cup.ru
vk.com/cmephi

