

PipeLine

- Пайплайн(PipeLine) для нашей задачи представляет из себя алгоритм для обработки данных, полученных из A/B – теста. Для нашей задачи выделим четыре этапа:
 1. Извлечение
 2. Очистка данных
 3. Разведочный анализ данных(EDA)
 4. Заключение.

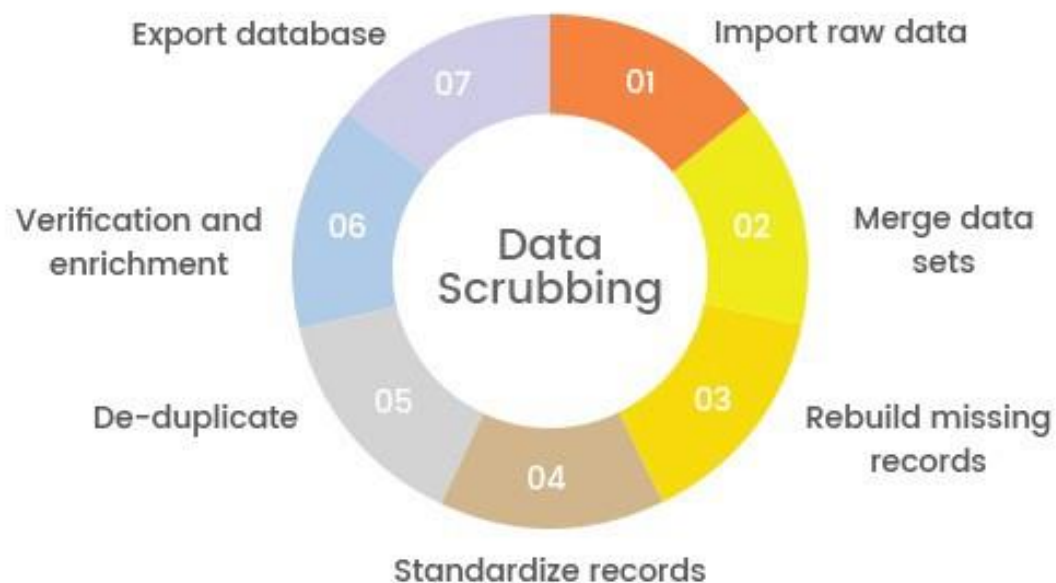
Извлечение

- Этот этап подразумевает сбор данных для анализа по двум тестируемым группам и их последующая конвертация в определенные форматы, в нашем случае – таблица csv с данными о взаимодействии пользователей с web - страницей обучающего курса.



Очистка данных

- После извлечения данные, скорее всего, не пригодны для анализа в первоизданном виде и для дальнейшего анализа они должны пройти следующие преобразования:
 - Очистка от аномалий;
 - Заполнение/удаление пропусков;
 - Удаление дубликатов;
 - Преобразование в единый набор данных, если таблиц несколько.

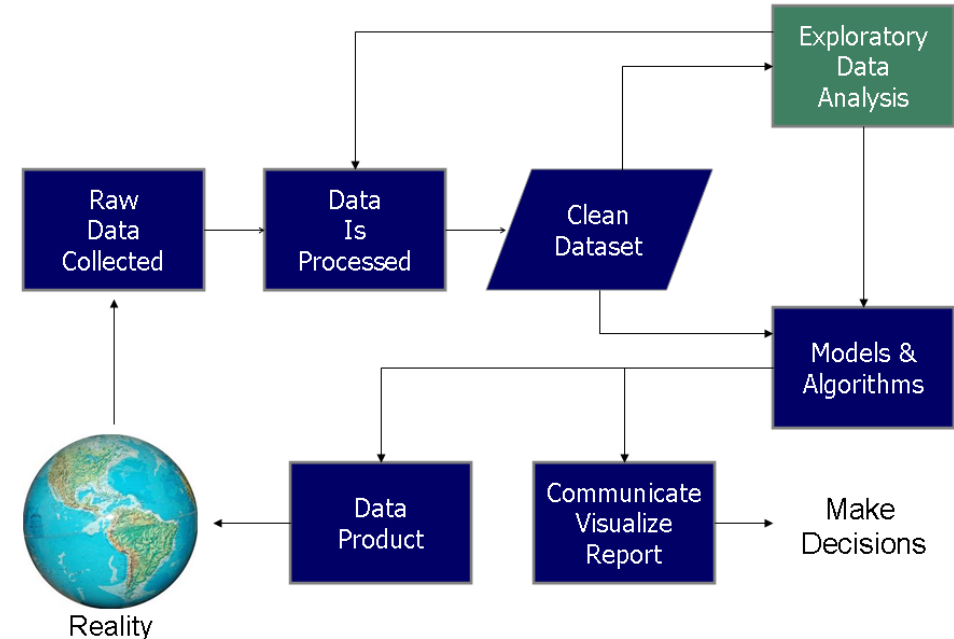


Разведочный анализ данных(EDA)

- Один из важнейших этапов А/В – тестирования, в котором предстоит выбрать метрики для оценивания значимости влияния изменений на тестовую выборку пользователей, выбрать наиболее релевантные переменные, создание новых признаков, визуализация.

В нашем случае данные метрики:

- 1) Количество уникальных кликов на странице за день;
- 2) Количество уникальных просмотров страницы за день;
- 3) Вероятность перехода по клику, то есть первое делим на второе;
- 4) Валовая конверсия - являющаяся отношением количества регистраций к количеству уникальных кликов;
- 5) Чистая конверсия - являющаяся отношением количества платежей к количеству уникальных кликов;
- 6) Удержание – отношение количества платежей к количеству регистраций.



Заключение

- На данном этапе, на основе метрик проверяется гипотеза о значимости результатов тестирования. И если результаты значимые, то изменения в интерфейсе для тестовой группы можно применить для всех пользователей. Иначе тестируемые изменения убираются из интерфейса вовсе.