**Проектная работа**

**Дашборд для Дзена**

**Что такое проект**

Поздравляем! Вы прошли курс в тренажёре. Самое время проверить полученные знания на практике и решить аналитический кейс. Выполнять работу вы будете самостоятельно.

Как закончите работу над проектом, отправьте его на проверку. Ожидайте комментарии в течение суток. Их нужно учесть; доработать проект и вернуть обновлённый вариант.

Скорее всего, вы снова получите комментарии по кейсу. Это нормально. Процесс доработки может проходить в несколько этапов.

Проект завершён, когда будут одобрены все доработки.

**Как отправить работу на проверку**

Выполните самостоятельную работу локально, на вашем компьютере. Подготовьте решение:

* ссылку на дашборд на сайте Tableau Public;
* файл с презентацией.

Также вы можете добавить файлы, необходимые для корректной работы.

Затем соберите все файлы в папку. Её размер *до архивации* не должен превышать 9 мб.

Сделайте из этой папки zip-архив и отправьте на проверку ревьюеру.

После проверки вы также получите zip-архив. Комментарии к работе вы найдёте внутри каждого из файлов.

**Часть 1. Составление технического задания**

Вы работаете аналитиком в [Дзене](https://zen.yandex.ru/). Почти всё ваше время занимает анализ пользовательского взаимодействия с карточками статей.

Каждую карточку определяют её **тема** и **источник** (у него тоже есть тема). Примеры тем: «Красота и здоровье», «Россия», «Путешествия».

Пользователей системы характеризует возрастная категория. Скажем, «26-30» или «45+».

Есть три способа взаимодействия пользователей с системой:

* Карточка отображена для пользователя (show);
* Пользователь кликнул на карточку (click);
* Пользователь просмотрел статью карточки (view).

Каждую неделю начинающие менеджеры Денис и Валерия задают вам одни и те же вопросы:

* Сколько взаимодействий пользователей с карточками происходит в системе с разбивкой по темам карточек?
* Как много карточек генерируют источники с разными темами?
* Как соотносятся темы карточек и темы источников?

На шестую неделю работы вы решаете, что процесс пора автоматизировать. Для Дениса и Валерии нужно сделать дашборд.

Дашборд будет основываться на пайплайне, который будет брать данные из таблицы, в которых хранятся сырые данные, трансформировать данные и укладывать их в агрегирующую таблицу. Пайплайн будет разработан для вас дата-инженерами.

МИНИ-ТЕСТ:

Какую последовательность шагов нужно выполнить для проектирования и реализации дашборда?

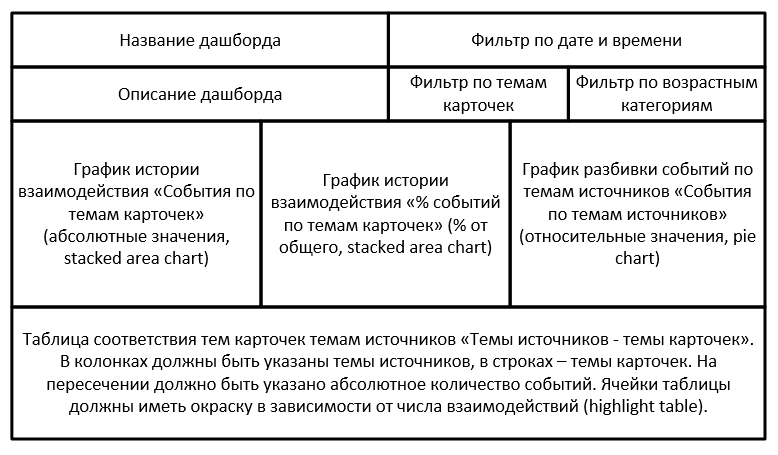
Правильный ответ

Нужно подробно обсудить с Валерией и Денисом состав дашборда, его внешний вид и набор отображаемых данных. После этого пообщаться с администраторами БД и выяснить, куда и как собираются нужные данные. С ними же нужно решить, где хранить агрегирующие таблицы. В финале приступить к разработке пайплайна и дашборда.

Верно! Идём к менеджерам обсуждать макет дашборда и писать ТЗ. Потом к админам, и только в конце писать код.

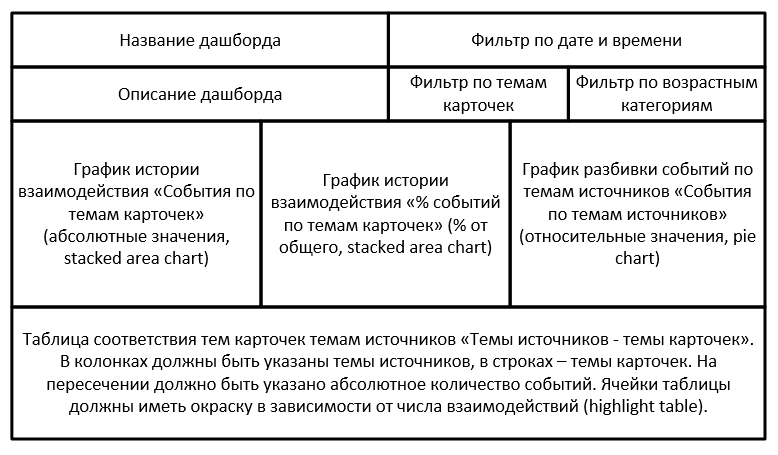
Пообщавшись с менеджерами и администраторами баз данных, вы написали краткое ТЗ:

* Бизнес-задача: анализ взаимодействия пользователей с карточками Дзена;
* Насколько часто предполагается пользоваться дашбордом: не реже, чем раз в неделю;
* Кто будет основным пользователем дашборда: менеджеры по анализу контента;
* Состав данных для дашборда:
  + История событий по темам карточек (два графика - абсолютные числа и процентное соотношение);
  + Разбивка событий по темам источников;
  + Таблица соответствия тем источников темам карточек;
* По каким параметрам данные должны группироваться:
  + Дата и время;
  + Тема карточки;
  + Тема источника;
  + Возрастная группа;
* Характер данных:
  + История событий по темам карточек — абсолютные величины с разбивкой по минутам;
  + Разбивка событий по темам источников — относительные величины (% событий);
  + Соответствия тем источников темам карточек - абсолютные величины;
* Важность: все графики имеют равную важность;
* Источники данных для дашборда: дата-инженеры обещали подготовить для вас агрегирующую таблицу dash\_visits. Вот её структура:
  + record\_id — первичный ключ,
  + item\_topic — тема карточки,
  + source\_topic — тема источника,
  + age\_segment — возрастной сегмент,
  + dt — дата и время,
  + visits — количество событий.
* Таблица хранится в специально подготовленной для вас базе данных zen;
* Частота обновления данных: один раз в сутки, в полночь по UTC;
* Какие графики должны отображаться и в каком порядке, какие элементы управления должны быть на дашборде (макет дашборда):



**Часть 2. Создание дашборда**

Ещё раз изучим макет дашборда:



Для создания дашборда выполните вот эти шаги:

1. Скачайте файл <https://---------------------.pem> в папку с файлом .ipynb, в котором вы будете работать с базой
2. Создайте коннекцию к базе, скопировав код ниже:

*# импортируем библиотеки*

import pandas as pd

from sqlalchemy import create\_engine

db\_config = {'user': 'praktikum\_student', *# имя пользователя*

'pwd': '\*\*\*\*\*\*\*\*\*\*\*\*\*', *# пароль*

'host': '---------------------',

'port': 64--, *# порт подключения*

'db': 'zen-project'} *# название базы данных*

connection\_string = 'postgresql://{}:{}@{}:{}/{}'.format(db\_config['user'],

db\_config['pwd'],

db\_config['host'],

db\_config['port'],

db\_config['db'])

engine = create\_engine(connection\_string)

1. Коннекция к базе хранится в переменной engine. Выполнить sql-запрос можно вот так:

Скопировать кодPYTHON

pd.io.sql.read\_sql(query, con = engine)

1. В Юпитере, напишите код для выгрузки данных из таблицы dash\_visits в файл dash\_visits.csv;
2. В Tableau Public на основе файла dash\_visits.csv сформируйте дашборд в соответствии с макетом;
3. Опубликуйте дашборд на сайте Tableau Public. Убедитесь, что он доступен для всего интернета - для этого можно открыть дашборд в нескольких браузерах. Если дашборд не будет доступен всему интернету, ревьюер не сможет его проверить.
4. Ответьте на вопросы менеджеров, используя дашборд:
   * Cколько взаимодействий пользователей с карточками происходит в системе с разбивкой по темам карточек?
   * Как много карточек генерируют источники с разными темами?
   * Как соотносятся темы карточек и темы источников?

Подготовьте небольшую презентацию с отчётом — ответы на вопросы и графики.

РЕШЕНИЕ СНИЗУ

Презентация в формате Гугл презентаций:

<https://docs.google.com/presentation/d/1niGMSkNQcOxBGRNE59FfsuHbrBTh3D06lwNOSWgdEfs/edit?usp=sharing>

Презентация в формате PDF:

<https://docs.google.com/presentation/d/1niGMSkNQcOxBGRNE59FfsuHbrBTh3D06lwNOSWgdEfs/edit?usp=drive_link>

Ссылка на Tableau-Public:

<https://public.tableau.com/views/__16763096990910/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link>