Итоговый проект

**Цель:** применить полученные в ходе программы знания и навыки для решения комплексной задачи

**Описание задания:**

Вам необходимо разработать систему аналитики для учёта и планирования отпусков. Используйте для выполнения задания собственные данные.  
  
В таблице ниже перечислены 10 заданий. За правильное выполнение каждого из них вы получите баллы — они указаны в последнем столбце. Максимальное количество баллов за все задания — 100. Для зачёта достаточно набрать 60 баллов.

| **Номер** | **Описание задания** | **Баллы** |
| --- | --- | --- |
| 1 | Разработайте Google Форму, позволяющую планировать отпуск. Форма должна содержать следующие поля:   * страна * город * дата начала * дата окончания * планируемый бюджет   Результаты должны сохраняться в Google Таблицу | 10 |
| 2 | Добавьте в Google Таблицу столбцы для заполнения фактической информации об отпуске. Они должны содержать следующую информацию:   * стоимость билетов, * стоимость отеля, * траты на отдыхе, * общая оценка отпуска по шкале от 1 до 5, * столбец с итоговой стоимостью отпуска — сумма трех столбцов выше, * отклонение итоговой стоимости от плана, * стоимость одного дня отпуска.   Последние 3 столбца должны вычисляться с помощью формул | 10 |
| 3 | Внесите данные по 10 и более отпускам за последние 3 года. Данные могут быть вымышленные | 5 |
| 4 | Разработайте отчёт в Google Таблицах или Looker Studio со следующими показателями:   * вывести самый дорогой отпуск суммарно * среднее отклонение фактической цены отпуска от планируемой * расход бюджета: какую долю в среднем занимают авиабилеты, проживание и траты на месте ***! Важно:*** *необходимо создать столбчатую диаграмму, а не круговую, так как данные по разным категориям бюджета находятся в разных столбцах* * рейтинг стран по стоимости отеля за день * рейтинг стран по стоимости трат на месте за день   Отчёт должен содержать фильтр по датам и по оценке | 30 |
| 5 | Рассчитайте корреляцию между ценой отпуска за день и оценкой за отзыв.  Постройте точечный график с этими показателями | 5 |
| 6 | Проверьте статистическую гипотезу, что вы хорошо планируете отпуска: среднее отклонение планируемых трат от реальных равно 0 | 5 |
| 7 | ***!Важно:*** *01.02.2023 вышло обновление библиотеки sqlalchemy – версия 2.0.0, из-за чего могут не работать SQL-запросы из Python. Чтобы решить проблему, необходимо откатиться к предыдущей рабочей версии.*  *Для этого до выполнения задания:*  *1) Скопируйте и запустите в самой первой ячейке код –* ***!python -m pip install --upgrade 'sqlalchemy<2.0'***  *2) После этого вам будет предложено обновить среду выполнения – обновите.*  *Далее можно последовательно запускать все ячейки, начиная с import pandas as pd.*   * При помощи Python и SQL найдите, как называются ваши города в таблице City базы данных world.db. Строка подключения — postgresql://netology:NetoSQL2019@84.201.153.170:19001/world-db * Создайте соответствующий справочник на отдельном листе Google Таблиц * Добавьте колонку с английским названием города при помощи функции VLOOKUP (ВПР)   **Алгоритм выполнения задания:**   1. Подгрузить базу данных world.db в Python 2. Найти города 3. Копировать названия городов из базы вручную и вставить в свою гугл-таблицу на отдельный лист 4. С помощью функции VLOOKUP (ВПР) подтянуть города из справочника с английскими названиями в основную таблицу.   ***!Важно:*** *Название города в базе может быть записано не совсем стандартно или с нечитаемыми символами. Если вам не удалось найти город, попробуйте произвести поиск по части слова.*  *Например, для Moscow:* ***LIKE:WHERE name LIKE “%%osco%%”*** *Двойной знак процентов заменяет часть слова*.  *Если же и этот способ не помог, города, который вы хотите найти, просто нет в таблице. В таком случае найдите английский вариант названия города в Интернете и напишите комментарий преподавателю, что нашлись не все города – он в курсе этой особенности базы данных.* | 10 |
| 8 | Скачайте данные в CSV и откройте их в Python | 10 |
| 9 | При помощи SQL и Python получите датафрейм с названием города и его населением из таблицы City | 5 |
| 10 | При помощи Python соедините данные из выгруженного CSV файла (пункт 8) и таблицы с населением города.  Сгруппируйте итоговый датафрейм по странам и рассчитайте среднюю численность населения в городах, в которых вы отдыхали | 10 |

## В качестве решения необходимо прислать:

* ссылку на Google Форму,
* ссылку на Google Таблицу,
* ссылку на отчёт в Looker Studio при наличии,
* ссылку на Google Colaboratory.

*Перед отправкой проверьте доступ к файлам по ссылке. Для этого вы можете открыть ссылки в браузере в режиме инкогнито.*

[*Как запустить Chrome в режиме инкогнито*](https://support.google.com/chrome/answer/95464?co=GENIE.Platform%3DDesktop&hl=ru)*.*

[*Как запустить Safari в режиме инкогнито.*](https://support.apple.com/ru-ru/guide/safari/ibrw1069/mac)