



## ИТОГОВОЕ ЗАДАНИЕ (НАЧАЛЬНЫЙ УРОВЕНЬ)

1. Выбрать индивидуальное задание: можно использовать “свои” данные или использовать датасеты с Kaggle <https://www.kaggle.com/datasets?fileType=csv> или взять любой датасет из PyCaret (<https://colab.research.google.com/drive/1kXxUEBiG3Zc7fM73vLc34-rAk0vsGjij?usp=sharing> , <https://pycaret.org/get-data/>)
2. В Colab (или используя платформу Data Science and Machine Learning - любую на выбор: **Loginom** (легко использовать визуализаторы статистика, качество данных), Knime, RapidMiner) выполнить консолидацию данных кейса, если кейс состоит из нескольких таблиц (пример на pandas: [https://colab.research.google.com/drive/1dfFXMdk\\_X0JY6-JbkMtaHkQ8jqwe8eSz?usp=sharing](https://colab.research.google.com/drive/1dfFXMdk_X0JY6-JbkMtaHkQ8jqwe8eSz?usp=sharing), ссылка на датасет с дубликатами и пропущенными значениями [https://docs.google.com/spreadsheets/d/1jjbT6CZGkj3OduClesmnqHHolxMTB1Er/edit?usp=drive\\_link&oid=115645456076953951741&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1jjbT6CZGkj3OduClesmnqHHolxMTB1Er/edit?usp=drive_link&oid=115645456076953951741&rtpof=true&sd=true), очищенный датасет [https://docs.google.com/spreadsheets/d/1jjbT6CZGkj3OduClesmnqHHolxMTB1Er/edit?usp=drive\\_link&oid=115645456076953951741&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1jjbT6CZGkj3OduClesmnqHHolxMTB1Er/edit?usp=drive_link&oid=115645456076953951741&rtpof=true&sd=true)); провести очистку, трансформацию данных (ETL), выполнить preprocessing данных (<https://www.sv-europe.com/crisp-dm-methodology/>), провести EDA с использованием любой библиотеки python (легко использовать библиотеки python автоматизации EDA Dataprep, Autoviz: пример использования [https://colab.research.google.com/drive/1PIgDvYLYxRVYzcBkcJ\\_ZQDqF4cLBWzhf?usp=sharing](https://colab.research.google.com/drive/1PIgDvYLYxRVYzcBkcJ_ZQDqF4cLBWzhf?usp=sharing)), библиотеки bamboolib, mitosheet, d-tale пример использования <https://disk.yandex.ru/i/nAWgDzty2n1zZw>).

### Платформы BI:

**На выбор любой** из 1 2, 3 пунктов.

В любой из BI-платформ:

1. Подключившись к BigQuery (или использовать подготовленный датасет или работать с исходными “сырыми” данными (обработать их в Power Query), создать в Power BI Desktop выражения KPI с использованием выражений анализа данных (DAX), разработать несколько отчетов в Power BI, опубликовать его в корпоративном облачном сервисе Power BI.
2. В Tableau Public (используя подготовленный датасет из предыдущего задания) построить дашборды, выявить инсайты, оформить истории (Story), опубликовать в своем аккаунте в Tableau Public, внедрить интерактивные отчеты в Colab.
3. Создать интерактивный отчет в Looker Studio <https://lookerstudio.google.com/navigation/reporting> (используя подготовленный датасет из предыдущего задания) и внедрить в Colab:

%%html  
<iframe width="1200" height="600" src="https://datastudio.google.com/embed/reporting/6acfd3c5-2a22-4244-aa32-03ab2cb50b2b/page/bWQLC" frameborder="0" style="width: 100%; height: 100%; border: none;">

