



## ИТОГОВОЕ ЗАДАНИЕ (ПРОДВИНУТЫЙ УРОВЕНЬ)

По индивидуальному заданию (выбранный Вами кейс):

<https://drive.google.com/drive/folders/1b3pGfEhAJpq8yuI6Eua1eOTefiOg4kel?usp=sharing>

1. *Используя платформу Data Science and Machine Learning* (любую на выбор: **Knime**, RapidMiner, Trifacta, **Loginom**) выполнить:

- консолидацию данных кейса;
- провести очистку;
- трансформацию данных (ETL);
- выполнить preprocessing (предварительная обработка данных) (<https://www.sv-europe.com/crisp-dm-methodology/>);
- обогатить данные кейса – провести ABC-XYZ, RFM-анализ;
- обучить и оценить по метрикам качества модель машинного обучения (решение одной из задач: классификации и/или регрессии и/или кластеризация и/или анализ временных рядов – одна или несколько на выбор исходя из данных кейса).
- экспортировать подготовленный датасет в BigQuery (или в формате .csv в Google Drive).

2. *В Colab* (подключившись к BigQuery или используя подготовленный датасет .csv в Google Drive):

- провести EDA с использованием любой библиотеки Python (pandas-profiling, dataprep, bambolib, D-tale или другую)
- обучить модель машинного обучения с использованием любого фреймворка: scikit-learn (или LightGBM или XGBoost или CatBoost или h2o- 3 или AutoGluon или PyCaret).
- провести прогнозирование ответа лучшей модели на новых данных (новые, не видимые ранее моделью) данные до обучения взять из исходных или сгенерировать самостоятельно используя исходные данные).
- сравнить результаты, полученные в пункте 1.

### Платформы BI:

**На выбор любой** из 3 4, 5, 6 пунктов

В любой из BI-платформ (можно, конечно, во всех):

3. Подключившись к BigQuery (или использовать уже подготовленный датасет) разработать несколько отчетов в Power BI. Выявить инсайты, оформить их в “закладках”. Внедрить интерактивные отчеты в Colab.

4. В Tableau Public использовать подготовленный датасет построить дашборды, выявить инсайты, оформить истории (Story), опубликовать в своем аккаунте в Tableau Public, внедрить интерактивные отчеты в Colab.

5. В DataLens импортировать подготовленный датасет построить дашборды, выявить инсайты, оформить истории (Story), оформить отчет в Word

(используя скрины)

6. В BigQuery создать, обучить и оценить по метрикам качества модели машинного обучения используя SQL; создать интерактивный отчет, включающий исходные данные и результаты (прогнозы) машинного обучения (можно на другой странице) в Data Studio и внедрить в Colab:

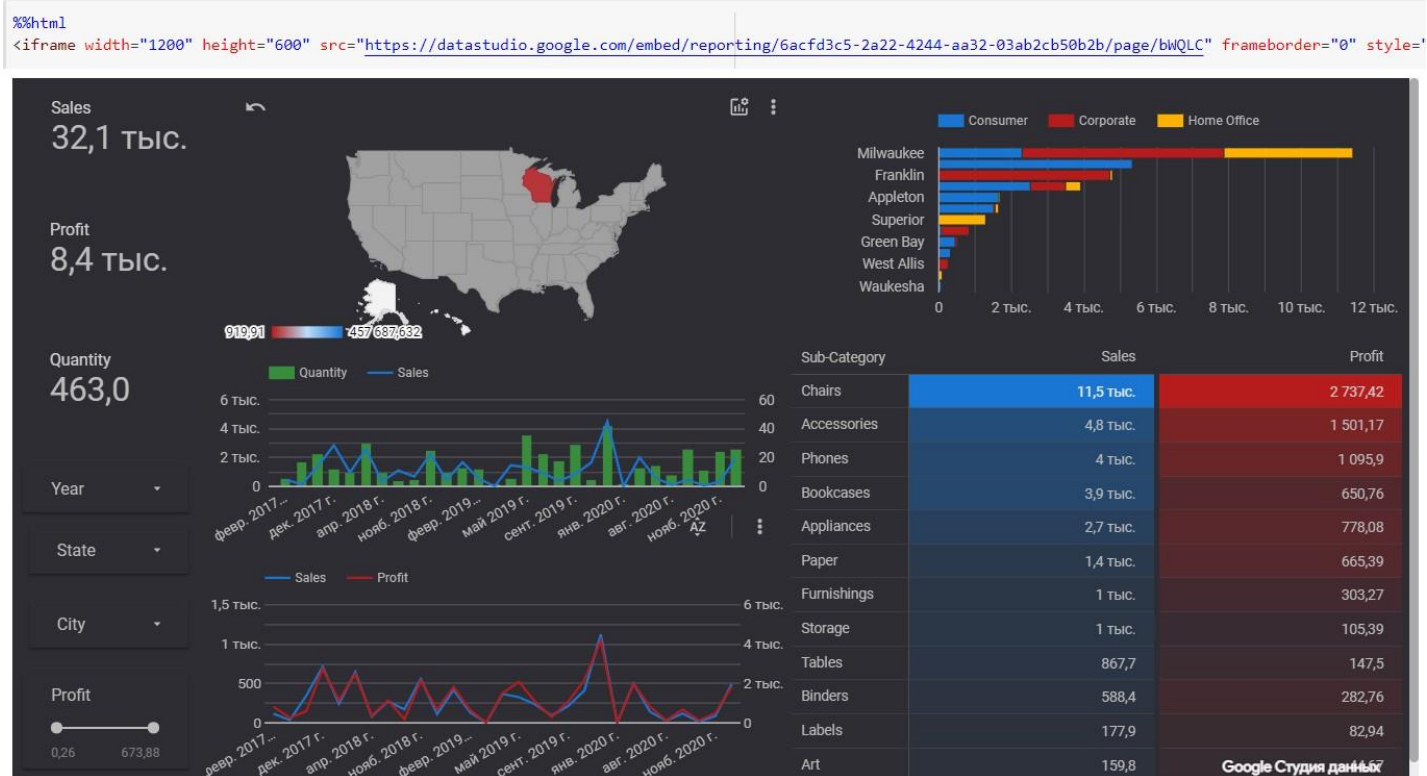


Рисунок 1. Пример визуализации внедренной в Colab

