

Итоговое задание (аналитик данных)

ETL: подготовка (preprocessing), исследовательский анализ данных

1. Выбрать индивидуальное задание:
 - можно использовать “свои” данные
 - использовать подготовленные датасеты (<https://drive.google.com/drive/folders/1b3pGfEhAJpq8yuI6Eua1eOTefiOg4kel?usp=sharing>)
 - или использовать датасеты с Kaggle (<https://www.kaggle.com/datasets?fileType=csv>).
2. Используя любую аналитическую платформу Loginom, Knime, RapidMiner (любую на выбор или в Colab) выполнить консолидацию данных кейса (если кейс состоит из нескольких таблиц), провести очистку, трансформацию данных (ETL), выполнить preprocessing данных (<https://www.sv-europe.com/crisp-dm-methodology/>), провести исследовательский анализ данных (EDA) путем создания различных визуализаций и описательных статистик.
3. **Желательно (но не обязательно)**: обогатить данные кейса: провести ABC-XYZ, RFM-анализ, дополнить данные результатами решения задачи классификации, регрессии, кластеризации (одна или все на выбор исходя из данных кейса).
4. **Желательно (но не обязательно)**: обучить и оценить по метрикам качества модели машинного обучения (использовать различные алгоритмы); сравнить полученные результаты.
5. **Желательно (но не обязательно)**: экспортировать подготовленный датасет в BigQuery (или сохранить в формате .csv в Google Drive).
6. **Желательно (но не обязательно)**: обучить и сравнивать модели машинного обучения с использованием библиотеки OmniXAI. Провести объяснение модели (оценить важность признаков на основе различных методов SHAP, LIME, PDP и др.
7. **Желательно (но не обязательно)**: сравнить результаты, полученные в платформах Data Science and Machine Learning.

Платформы BI:

На выбор 8, 9, 10 пункты: в любой из BI-платформ:

8. 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 интерактивных отчетов в Power BI Desktop.

9. В Tableau Public (используя подготовленный датасет из предыдущего задания) построить дашборды, выявить инсайты, оформить истории (Story), опубликовать в своем аккаунте в Tableau Public), поделиться аналитическим приложением.
10. Создать интерактивный отчет в Looker Studio <https://lookerstudio.google.com/navigation/reporting> (используя подготовленный датасет из предыдущего задания), поделиться аналитическим приложением.