Data governance for ML (DVC)

- Task description:
    1. Make initial setup using `Data version control` tool. Add dataset, so another could obtain it via `dvc pull`
    2. Define 2-3 pipelines that would preprocess data in different ways (basic cleaning, scaling, aggregations, etc.). Each pipeline should be reproducible using `dvc repro` .
    3. Use some existing solution for your dataset, run experiment on the data using development environment from previous step and save metrics using `dvc metrics.`
- Criteria:
    1. Pipelines defined in a simple, reproducible manner
    2. Following DVC best practices
    3. Code style / code quality tools used
    4. There is an existing remote from which one could pull data (use free tier of AWS/GCP, Google Drive, or any other that would be easy to share)
- Materials
    1. https://dvc.org/
    2. https://www.youtube.com/watch?v=kZKAuShWF0s
    3. https://www.youtube.com/watch?v=xPncjKH6SPk&t=2s
    4. https://www.youtube.com/watch?v=kLKBcPonMYw&t=1s