```
code
├── .gitignore
├── docker_services                                  # directory containing components instantiated via Docker
│   ├── .env                                         # file containing the environment variables used by the Planning Engine
│   ├── config
│   ├── dags                                         # required by Airflow
│   │   ├── airflowDAG.py                            # the Airflow DAG which defines the Planning Engine
│   │   ├── planning_utils.py                        # collection of utility methods used by the Planning Engine
│   ├── docker-compose.yml                           # define containers and their dependencies
│   ├── Dockerfile                                   # define instructions to build the FlowETL Docker image
│   ├── logs                                         # where the DAG runtime logs are written, required by Airflow
│   └── plugins                                      # requured by Airflow
├── etl_logs                                         # directory where the ETL worker execution logs are written to
├── etl_worker.py                                    # logic defining the ETL worker
├── evaluation                                       # folder containing the experiment and evaluation files
│   ├── bonobo_etl_experiment                        # folder containing the files for the Bonobo vs FlowETL experiment
│   │   ├── bonoboutils.py                           # utility methods used throughout the Bonobo evaluation
│   │   ├── bonoboworker.py                          # Bonobo ETL pipelines code
│   │   ├── polluter.py                              # logic used to artificially inject data wrangling issues within evaluation datasets
│   ├── flowetl_runtime_results
│   │   └── result.csv                               # results collected during the FlowETL evaluation experiment
│   ├── GT.md                                        # ground truth file outlining the trasformations to be applied to each dataset
│   ├── planning_engine_versions_comparison_experiment
│   │   ├── info.md                                  # outline of the methodology for this sub-experiment
│   │   ├── results.csv                              # results for the comparison between the two versions of the Planning Engine
│   │   ├── v1.json                                  # plans computed by the first version of the Planning Engine on the 13 evaluation
datasets
│   │   └── v2.json                                  # plans computed by the second version of the Planning Engine on the 13 evaluation
datasets
│   ├── sampling_percentage_experiment
│   │   └── results.csv                              # results for the experiment assessing the planning engine for varying sample sizes
│   └── schema_inference_experiment
│       └── results.csv                              # results for the experiment assessing the LLM vs algorithmic schema inference
├── evaluation_datasets
│   ├── source                                       # source datasets required for the evaluation task
│   │   ├── csv
│   │   │   ├── amazon_stock_data_source.csv
│   │   │   ├── chess_games_source.csv
│   │   │   ├── ecommerce_transactions_source.csv
│   │   │   ├── financial_compliance_source.csv
│   │   │   ├── netflix_users_source.csv
│   │   │   ├── pixar_films_source.csv
│   │   │   └── smartwatch_health_data_source.csv
│   │   └── json
│   │       ├── amazon_reviews_source.json
│   │       ├── flight_routes_source.json
│   │       ├── news_categories_source.json
│   │       ├── recipes_source.json
│   │       ├── social_media_posts_source.json
│   │       └── students_grades_source.json
│   └── target                                       # target datasets required for the evaluation tasks, corresponds to source datasets
after applying the GT
│       ├── csv
│       │   ├── amazon_stock_data_target.csv
│       │   ├── chess_games_target.csv
│       │   ├── ecommerce_transactions_target.csv
│       │   ├── financial_compliance_target.csv
│       │   ├── nextflix_users_target.csv
│       │   ├── pixar_films_target.csv
│       │   └── smartwatch_health_data_target.csv
│       └── json
│           ├── amazon_reviews_target.json
│           ├── flight_routes_target.json
│           ├── news_categories_target.json
│           ├── recipes_target.json
│           ├── social_media_posts_target.json
│           └── students_grades_target.json
├── input
│   ├── source                                       # extraction location for the source dataset
│   └── target                                       # extraction location for the target dataset
├── logs                                             # FlowETL runtime logs
├── observers
│   ├── driver.py                                    # driver code to manage both Observers
│   ├── logs                                         # Observers runtime logs
│   ├── observer_utils.py                            # utility methods and class definition for the Observers
├── output                                           # load location for the transformed source dataset
├── reporter.py                                      # logic defining the Reporting Engine
├── requirements.txt                                 # list of python packages to be installed before running the project
├── testvenv                                         # virtual environment to run Bonobo evaluation
├── unit_tests.py                                    # FlowETL unit tests
└── venv                                             # virtual environment to be activated before using FlowETL
```