Project pipeline

Outlines the high-level components of the project. More detail will be added to each section as progress is made.

# Data cleaning

## Description

Module converts the raw train.csv file to the cleaned train.csv file.

## Inputs

* train.csv

## Outputs

* train\_clean.csv

## Script

# generate\_pairs

## Description

## Inputs

## Outputs

## Script

# Generate modelling data

## Description

## Inputs

## Outputs

## Script

# Split modelling data

## Description

## Inputs

## Outputs

## Script

# Train the model

## Description

## Inputs

## Outputs

## Script

# Evaluate the model

## Description

## Inputs

## Outputs

## Script