

A red parallelogram tilted to the right, containing the text "Projects ada" in white.

Projects
ada

EPFL

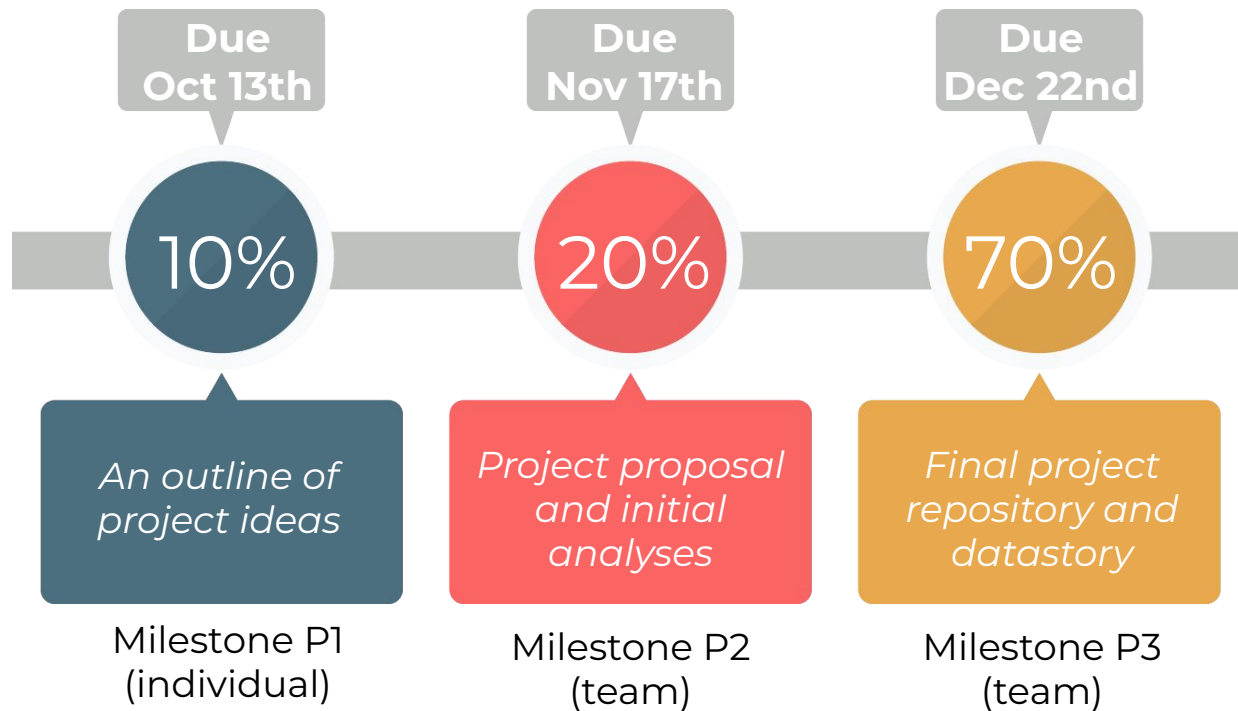
Project

Weight: 25%

All info available here!

Overview

Three milestones: first glimpse, proposal, execution



All the dates have deadline at 23:59

Datasets

Provided datasets are available here:

https://go.epfl.ch/ada23_datasets

CMU Movie Summary Corpus

42,306 movie plot summaries extracted from Wikipedia

Aligned metadata extracted from Freebase, including:

- revenue, genre, release date, runtime, and language
- Character names and info about the actors who portray them

Modalities: Text / Graphs / Numeric data

Can be combined with other sources (e.g. IMDb)

Beer reviews

Beer reviews from two websites: BeerAdvocate and RateBeer

Collected for a period of 17 years (2001 – 2017) and includes:

- Metadata about beers, breweries, and users
- Ratings across diff. aspects: appearance, aroma, palate, taste
- Textual review

Modalities: Text / Numeric data

Can be used for observational studies (matched data available)

Wikispeedia

Human navigation behavior data on a subset of Wikipedia, including:

- Human navigation paths
- Wikipedia hyperlink graph
- Plain-text content of Wikipedia articles

Modalities: Text / Graph / Navigation behavior

Generic and versatile

Wikipedia Requests for Adminship

Votes and outcome of Wikipedia editor adminship elections

Collected for 10 years from 2003 – 2013, including:

- Voter/votee graph
- Outcome of each election
- Textual comment about the vote

Modalities: Text / Graph

YouNiverse*

Metadata about Youtube channels and videos

Collected for 14 years from 2005 – 2019, including:

- Channel-level time series: weekly subscriber and view counts
- Co-commenting graph

Modalities: Timeseries / Graph

[Helpful dataset overview](#)

Coronawiki

Pageview statistics for 12 Wikipedia language versions

Collected during the COVID-19 pandemic, includes

- Monthly pageview statistics
- Mobility reports published by Apple and Google

Modalities: Timeseries

Can be used for observational studies

Milestone 1

To be done individually, due 13th Oct 23:59

1. Read the paper(s) relevant (Column G) to the dataset (fine if you don't fully grasp the technical details about the method)
2. Play around with the dataset
3. Propose at least three bold and creative ideas (non-binding)

Deliverable: Outline of project ideas (up to 500 words)

Interesting examples from 2022

Milestone 2

To be done as a team, due 17th Nov 23:59

1. A detailed project proposal (up to 1000 words)
2. A well-commented notebook with initial analyses and data handling pipelines

Milestone 3

To be done as a team, due 22nd Dec 23:59

1. The final notebook: complete with all the analysis and findings
2. A data story
 - 2.1. Not a loose collection of results
 - 2.2. Examples from [2017](#), [2018](#), [2019](#), [2020](#), [2021](#), [2022](#)

Thanks for your attention!