

Week 2 : Start Working With Dataset

Resources:

For Week #2, we'd like you to learn more about WikiProjects:

- **WikiProject Overview:** <https://en.wikipedia.org/wiki/Wikipedia:WikiProject>
- **WikiProject Council Guide:** https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Council/Guide#What_is_a_WikiProject, especially these sections:
 - **What is a WikiProject?**
 - **General principles**
 - **Tools for WikiProjects**
- "Using WikiProjects - Wikipedia For Beginners" [video] https://www.youtube.com/watch?v=y-_JpAdUBbI

To follow up on that, here are a few resources you could look at related to exploratory data analysis:

https://en.wikipedia.org/wiki/Exploratory_data_analysis

<https://lib.guides.umd.edu/datavisualization/explore>

Here are some resources that overview user stories:

https://en.wikipedia.org/wiki/User_story

<https://link.springer.com/article/10.1007/s00766-016-0250-x>

1 Introduction

3.1 Quality of an individual user story

<https://xtools.wmcloud.org/api#/>

Exploring Datasets:

 **Week2.ipynb** : Contains methods to query the active or inactive status of various WikiProjects and to find overlapping or multiple projects per article.

 **Week2-2nd.ipynb** : In this notebook, I calculated the quality score using a model and tracked its evolution over time. I also analyzed variables like edit counts and the number of unique editors. I intended to compare the results for two languages, so I applied the same analysis to French articles as well.

 **Week2-3rd.ipynb** : In this notebook, I queried the number of changes section-wise to determine if edits in specific sections lead to greater changes in the quality score.

 **USER STORIES**

Notes:

WikiProjects & WikiProjects Councils

General principles

Tagging pages with WikiProject banners

WikiProject talk page banners play a vital role in helping projects organize their efforts and direct editors to areas where they can contribute. These banners should be used responsibly to indicate active support for the tagged pages, recognizing that multiple projects may have an interest in the same articles. Importantly, placing a banner does not equate to ownership, and decisions about the content remain subject to broader community consensus.

WikiProjects define their scopes

WikiProjects on Wikipedia use talk page banners to organize and track articles within their defined scopes. Editors are encouraged to place banners judiciously and avoid over-tagging articles. When in doubt about the relevance of an article to a particular WikiProject, it's best to communicate with the project members rather than adding the banner directly. Respect the decisions of WikiProject participants regarding the scope of their projects and avoid edit wars over banner placements. This collaborative approach helps maintain a productive and harmonious editing environment on Wikipedia.

Advice pages

WikiProject advice pages are valuable tools for providing subject-specific guidance on applying Wikipedia's general policies and guidelines. They help organize information, provide clear explanations, and offer practical examples. However, it's essential to ensure these pages do not conflict with site-wide policies, avoid asserting ownership over articles, and remain consistent with the broader community consensus. When advice pages evolve into widely accepted guidelines, they transition from being project-specific to community-wide, with all editors participating equally in their development and maintenance.

Tools for WikiProjects

Bambots are a collection of Wikipedia bots/tools: [Bambot](https://bambots.brucemyers.com/)
<https://bambots.brucemyers.com/> (Important)

Here's a quick summary of the tools:

1. Recruiting Editors:

- Add to directories and categories.
- Use the {{WikiProject status}} template.
- Utilize Reports bot for finding active editors.

2. General Tools:

- WP 1.0 bot for article statistics[<https://wp1.openzim.org/#/>]

- InceptionBot for new page tracking.
[<https://en.wikipedia.org/wiki/User:InceptionBot>]
- Create a manual project watchlist.
- Enable Article Alerts.
[https://en.wikipedia.org/wiki/Wikipedia:Article_alerts/Subscribing]
- Use Popular Pages bot.
[https://en.wikipedia.org/wiki/Wikipedia:Popular_pages]
- Install Evad37's rater script.[<https://en.wikipedia.org/wiki/User:Evad37/rater>]

3. Tracking Progress:

- CleanupWorklistBot for maintenance tracking.[<https://bambots.brucemyers.com/CleanupWorklistBot.php>]
- JL-Bot for listing recognized content.[<https://github.com/JLaTondre/JL-Bot>]

These tools collectively help in organizing efforts, improving article quality, and fostering community involvement within WikiProjects.

(Within the data for querying "Inactive" status, see if any other variables seem important to include when building our tool):

Maybe we can check for vandalism.

To check if an article has been vandalized based on existing data on Wikipedia, we can query the revision history of the article to look for patterns or specific indicators of vandalism.

Alternatively, we could train a model to detect vandalism. I believe there is already a bot for this purpose, but I only found its GitHub link without instructions on how to use it.

ClueBot NG

The most well-known "bot" that fights vandalism is ClueBot NG. The bot was created by Wikipedia users Christopher Breneman and Cobi Carter in 2010 (succeeding the original ClueBot created in 2007; NG stands for Next Generation) and uses [machine learning](#) and [Bayesian statistics](#) to determine if an edit is vandalism.

While the bot has been effective in helping keep Wikipedia clean, some claim the bot is hostile to new users by not being able to apply a human brain's knowledge to the edit, and leaving impersonal [template messages](#).