Data Skills Workshop

Matthew Yeager
Volunteer
MatthewBrandonYeager@gmail.com

Chanse Hungerford
NA
Chanse.Hungerford@gmail.com

Netha Hussain
WikiWomen's User Group
Netha.Hussain@gmail.com

What is the event or activity you want to organize?

We will develop a series of video workshops to empower volunteers and community groups on any language Wikipedia with the data literacy skills to make better informed decisions. We will produce videos on the WikiLearn platform and Python notebooks to teach direct skills for improving community discussions, driving data-backed decisions, and establishing a common language for data questions and exploration.

We'll pivot away from a traditional approach that focuses on computer science fundamentals, to instead equip community groups with meaningful examples, concepts, and snippets which can improve their workflows and decisions today. In a pilot series of 6 videos, we will showcase approaching data wrangling, visualization, and analysis in simple, relevant examples. For instance, exploring the impact of a recent hack-a-thon, or determining which topic areas are currently under-represented within a project, or beginning to understand how an article has evolved over time through edits, editors, and citations.

We target technology-curious community volunteers who are looking to get stuff done, without needing to become familiar with computer science concepts. They've probably copied template code before and made some modifications to have it work for their project, without knowing exactly what is going on under the hood. Python notebooks come ready to access and plot data so that users can quickly start asking and answering their own questions. In this way, they are able to self-serve and collaborate more effectively, by asking more well informed questions when needing help from others.

What do you want to achieve by doing this community building activity?

We want to empower community volunteers with knowledgeable approaches to better access data and make more robust decisions in a community consensus space. Historically it has been difficult to directly empower Wikipedians at the research level, since project decisions are best left with those closest to the work.

In this way, we hope to lower the barrier to research methodologies and data by providing skills instead of solutions. We are looking to directly empower community volunteers and projects to capture how skills gained from our workshops would lead to improving at least 5 decisions or workflows they struggle with.

Additionally, we hope to broaden the backgrounds of users and projects that can now leverage simple data analysis and demonstrations, using Python as a tool.

When do you plan to host the event/activity? What is influencing the timing of the event?

We plan to release our first module in January of 2026, providing regular releases each 2-3 weeks. These video releases will be pre-recorded and come with copies of the Python notebook so that users can work through and build the examples themselves. We expect the final module to be released by April, 2026.

Releasing our first module in January is influenced by our time needed to develop and produce episodes, work alongside beta testers, and the seasonal reduction of activity that occurs during winter holiday months.

Our delivery plan includes the following high level allocation of time:

- July - September: Project Kickoff – Curriculum design, content creation tests, and WikiProject partnerships for beta testing.
- October - February: Build out Content – Python notebooks and videos. Release first video in January, following a release schedule of every 2-3 weeks.
- January - April: Release Schedule – Support module releases with outreach campaigns, and availability for Q&A office hours as needed.
- May - June: Impact – Capture usage analytics and WikiProject decision case studies.

Who will participate in your event/activity?

We'll aim for broad inclusion to engage both seasoned volunteers working directly in coordinating WikiProjects, as well as with everyday editors looking to gain self-service data skills. We expect to attract a wide range of technical comfortability as well as diverse backgrounds by engaging WikiProjects of all sizes. Our video material will include subtitles to ensure better accessibility for participants who are not native speakers.

We expect to reduce the technical prerequisite to about the same level as an entry level Excel user. They know how to sort and filter data, add in a chart (somewhere), and paste in formulas they found from googling their question online. Just like those users, we are looking to engage with editors looking to leverage the vast data sources Wikipedia has to offer, without the lift of needing to know implementation details.

This project is expected to create a diverse, data-literate community of Wikimedians who can use data, and in extension, technical tools to make informed decisions in their communities, thereby influencing the policies, practices, and workflows of the communities to which they belong.

Our team has two long-time Wikimedians who are active participants in several offline and online community spaces, including Wikimedia conferences such as WikiCredCon and Wikimania. We also have experience working together with community organizers in mapping knowledge gaps, customizing existing tools, and planning outreach events.

We also have extensive experience working with underrepresented communities such as WikiWomen's User group (and other movement

entities for advancing gender equity), geographical entities of the Global South such as Wikimedians of Kerala and thematic groups such as WikiProject: Women's Health. We will leverage our existing connections to further understand the needs of these and other organized communities, by inviting their first-look feedback to ensure the content remains accessible.

The program is targeted at a wide variety of tech-curious Wikimedians, particularly those working in building and maintaining WikiProjects, conducting outreach, and actively creating content. Being active participants in several Wikimedia community spaces ourselves, we will invite participants from targeted communities through talk pages/community noticeboards/social media/messaging platforms/mailing lists/personal invites to individuals and affiliates. In this way, we will ensure that we invite the global community, including underrepresented communities. Furthermore, we will participate in global Wikimedia conferences (Wikimania 2025, EduWiki 2025) and spread the word about the program.

We will also ensure that we have adequate representation in the beta testers of the project. As the skills workshops are an online, perennial resource with no accessibility barriers, we did not specifically invite formal commitment to participate from Wikimedia communities.

Where will the event/activity take place?

We will host the content on the WikiLearn platform, to allow for fully remote, asynchronous learning opportunities. In addition, we will have developed Python notebooks that allow users to not only follow along with the presentation, but also to explore on their own. All content will be made available under CC-BY-SA 4.0 or lower licenses.

Experience organizing similar events or activities

Chanse Hungerford, Ph.D., works in the semiconductor industry as a researcher, alongside researchers and hardware engineers. Over the past 18 months, Chanse has developed and led a weekly Python Skills Class to enable his colleagues to find better approaches to answering their own questions. These professionals, looking to get work done, don't have a computer science background and are not interested in obtaining one. They've tried learning Python before through eBooks and maybe even another course, but always disconnect due to an approach which focuses on computer science fundamentals. Chanse has been building curriculum and teaching with an approach that prioritizes improving people's day to day workflow over rote syntax memorization. His success comes from not only being able to break down complex projects into simple conceptual steps, but also in being part of the user group he is directly targeting.

Chanse will be working, in collaboration with Matthew and Netha, to lead developing our curriculum by taking common Wikipedia questions and workflows, and breaking them down into sizable bites in a "follow along with me" type of video.

Matthew Yeager has been a Wikipedia volunteer for several years working on a variety of community, technical, and content areas. With a background in software engineering, Matthew has worked alongside empowering everyday

users by democratising access to data and research techniques.

Matthew will lead this project to ensure quality deliverables, and a mindful collaboration between great teaching approaches and rich, data-backed examples showcasing Wikipedia workflows and questions.

Chanse & Matthew produced a prototype ebook chapter in 2019 which shares the guided approach starting with section "1.2 Part 2: Beginner's Guide to Plotting" located here: https://drive.google.com/file/d/1__-tstOmitiY7n0XNcuwXQe6arrhi14U/view?usp=sharing

Netha Hussain is a published researcher, and longtime Wikipedian, notable for her work in covering misinformation in the medical space and advancing gender equity in Wikimedia spaces. Netha has been collaborating with WikiProjects in medicine and gender, affiliates in South Asia and Sweden, as well as women's usergroups over the years to highlight knowledge gaps and pursue equity.

Netha has successfully funded grants for mapping knowledge gaps, creating Wikimedia driven educational platforms and conducting outreach events, as well as co-authoring and aggregating research on women in Wikimedia.

Netha will advise the team on meaningful examples and challenges facing Wikipedia community projects. Additionally, Netha will lead community outreach and partnership with beta testers.

Together, we've created a demo video showcasing a low fidelity mockup of what the screen recording and filming could look like: https://photos.app.goo.gl/csZSwsj3pBmrtN3h9

Specific outputs

We will deliver a series of 6 video workshops on to the WikiLearn platform, complete with closed captioning and accompanied with Python notebooks. These materials will directly target lowering the barrier for volunteers to access data and improve their community decision making processes by utilizing relevant Wikipedia data examples.

We will provide activity and financial reports to the Wikimedia Foundation summarizing our work and learnings at the midterm and end-of-project checkpoints. This will continue to facilitate the trust placed within us to deliver these workshops, as well as enable other researchers to learn from our approach and decision making.

In raising awareness of our new resource, we will develop posts and promotional content in Wikimedia spaces to explain our value proposition and invite people to join the course.

After the grant period, Matthew and Netha will continue to work as volunteers to maintain the resources, but with lower engagement levels. We will be available to answer video comments and maintain a list of candidate topics/questions that may be explored in the future.

We will use snapshots of robust datasets to ensure examples remain working throughout the years to come.

Evaluation

We'll have succeeded in meaningfully reaching our users when they leverage their skills to make better informed decisions within their projects. We will look to capture and quantify at least 5 occasions where community members who took our course, better contributed to their

projects and workflows through data retrieval and analysis that they would have otherwise not been able to obtain.

We will follow along with what users go on to make with their skills by reaching out and leveraging existing and new connections throughout the WikiProjects. These 5 occasions may be more comprehensive conversations within an RfC backed with data, or in building a dashboard with previously difficult to access or structure data, or in new proposals which explore knowledge gaps that deserve more of a project's time.

Additionally, we'll maintain shorter term metrics that allow us to better target and drive course outreach depending on user consumption. We'll seek feedback from the participants using a feedback form and conduct short interviews with selected participants who completed the course.

Budget

We are requesting $17,280 to produce a pilot series of 6 video modules and python notebooks. Our costs represent solely our staff's time and applied expertise in content creation and community engagement. We've considered software and hardware costs, and have instead stripped back to use existing skills and platforms in our pilot proposal.

An itemized breakdown is available following the requested template: https://docs.google.com/spreadsheets/d/1FpjSM7L4Z7SySgvvr46-OcwxcIjFuwLEglRTTMM1g24/edit?usp=sharing