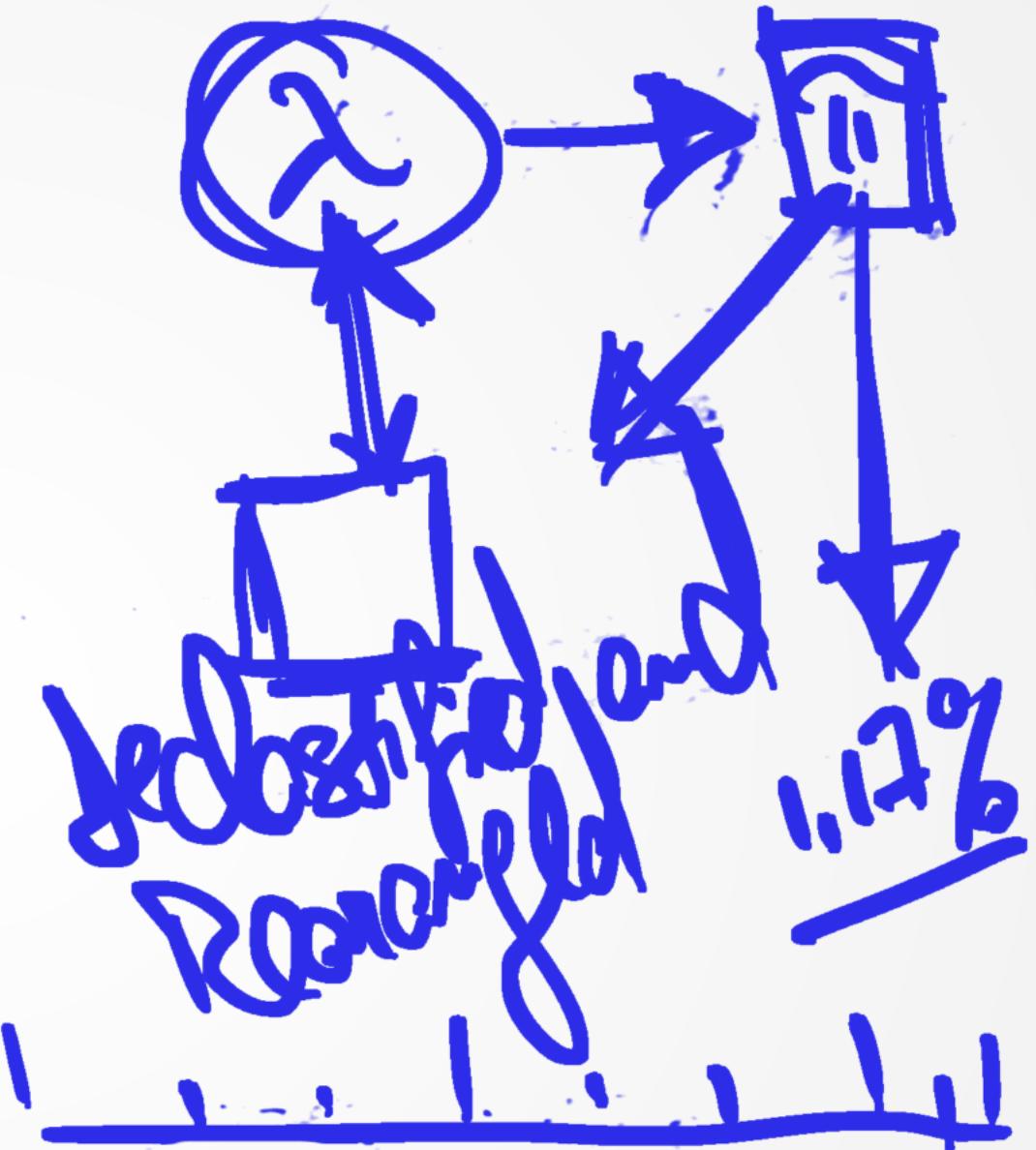
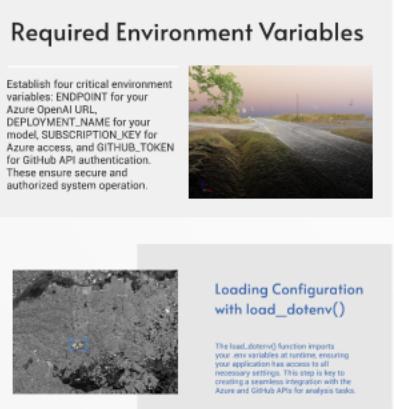


User Guide: Government GitHub Analysis System

Step-by-step Instructions for Environment Setup and Data Analysis

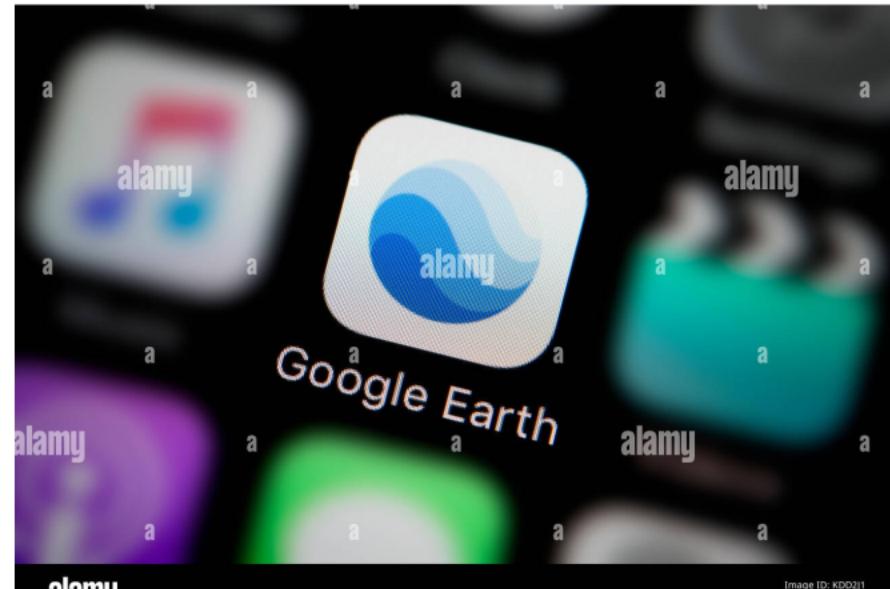


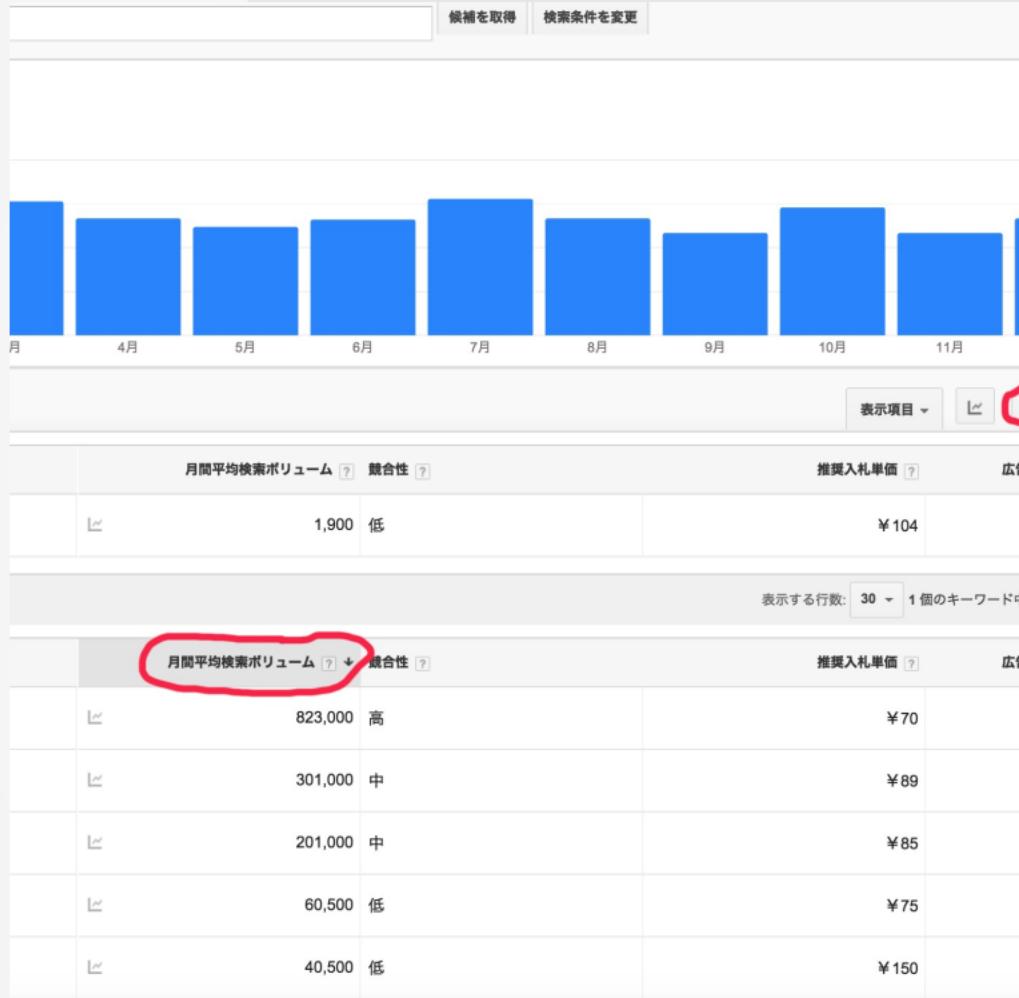
Environment Setup



Google Colab vs Jupyter Notebook

Google Colab is a cloud-based interactive notebook environment that facilitates sharing and collaboration, while Jupyter Notebook runs locally or on a server, giving users more control over their environment. Choose based on your specific needs for accessibility and resource management.





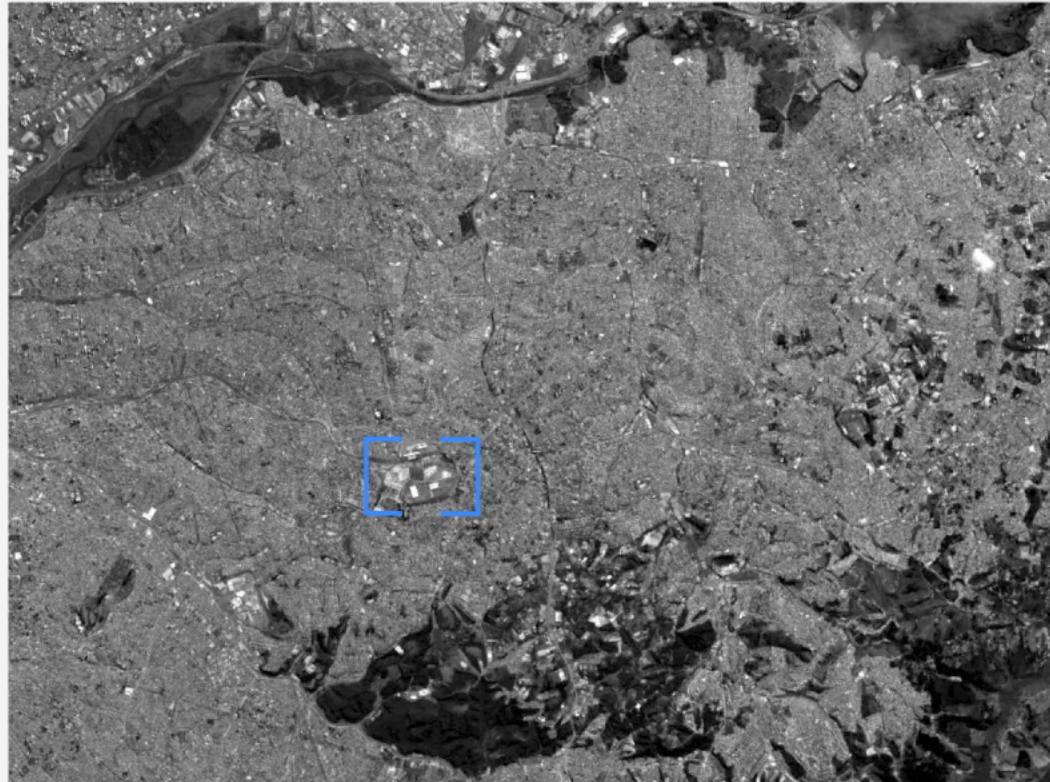
.env File Configuration

.env files are essential for storing configuration variables without hardcoding them into your code. This approach enhances security and allows for easier updates to your environment settings.

Required Environment Variables

Establish four critical environment variables: ENDPOINT for your Azure OpenAI URL, DEPLOYMENT_NAME for your model, SUBSCRIPTION_KEY for Azure access, and GITHUB_TOKEN for GitHub API authentication. These ensure secure and authorized system operation.





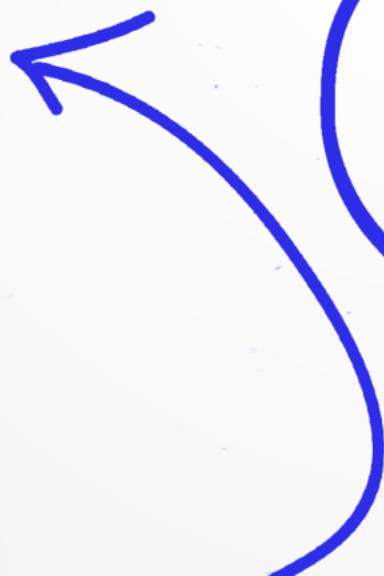
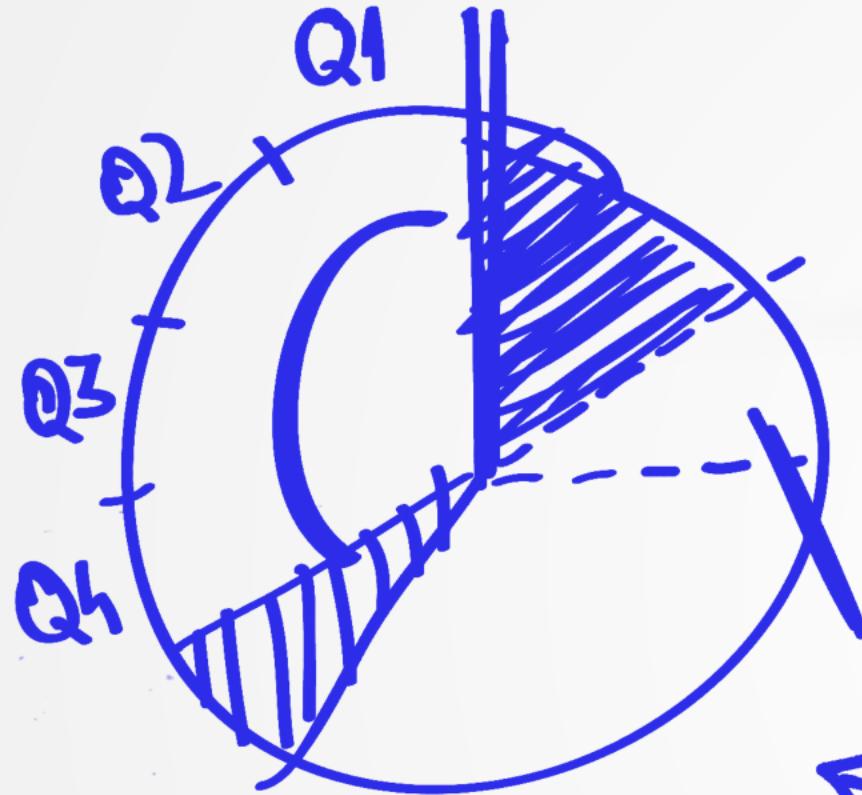
Loading Configuration with `load_dotenv()`

The `load_dotenv()` function imports your `.env` variables at runtime, ensuring your application has access to all necessary settings. This step is key to creating a seamless integration with the Azure and GitHub APIs for analysis tasks.

User Guide: Government GitHub Analysis System

Step-by-step Instructions for Environment Setup and Data Analysis



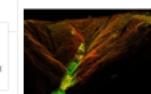


0,2143; 0,120013, 17,86

Fetching GitHub Government Accounts

Automatic YAML File Retrieval

The system automatically fetches a YAML file from GitHub containing harvested data for all GitHub accounts globally. This file is maintained by GitHub and serves as the foundational dataset for subsequent analysis.



Analyzing All Accounts vs. Filtered Accounts

This system offers the option to analyze all accounts by default, or narrow down to specific high-impact organizations. This flexibility enables targeted research, facilitating focused analysis on significant contributors.

High-impact Organizations for Filtering

Utilizing filters allows for a more precise analysis of key government entities such as GSA, alphagov, and opengovsg. This targeted approach enhances the relevance of the analysis results for stakeholders.



GSA - General Services Administration

The GSA plays a pivotal role in managing government procurement, providing a wide range of services. Analyzing their GitHub accounts can uncover innovative technology implementations in public sectors.



alphagov - UK Government Digital Service

alphagov represents the digital service for the UK government, focusing on making public services accessible online. Their GitHub contributions inform best practices in public service delivery and user-centered design.



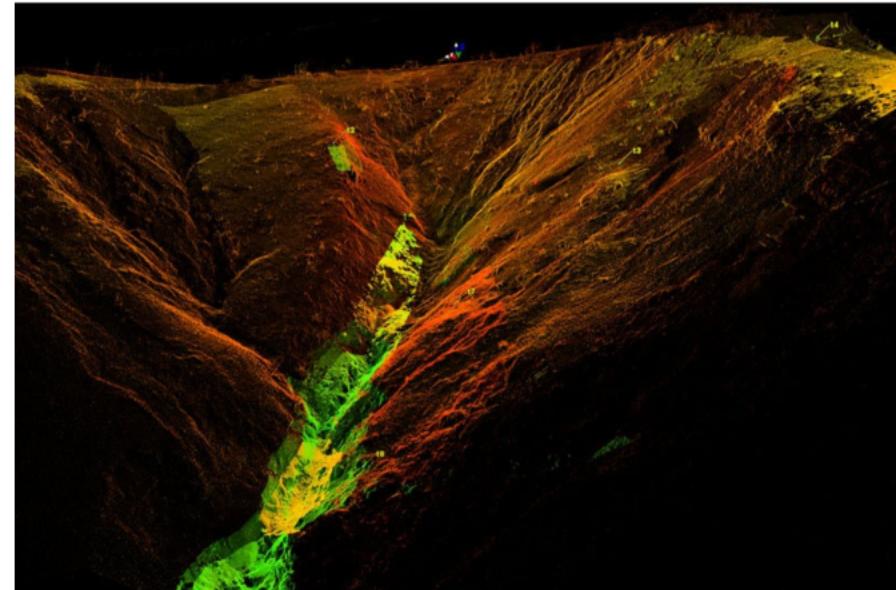
opengovsg - Singapore's Open Government

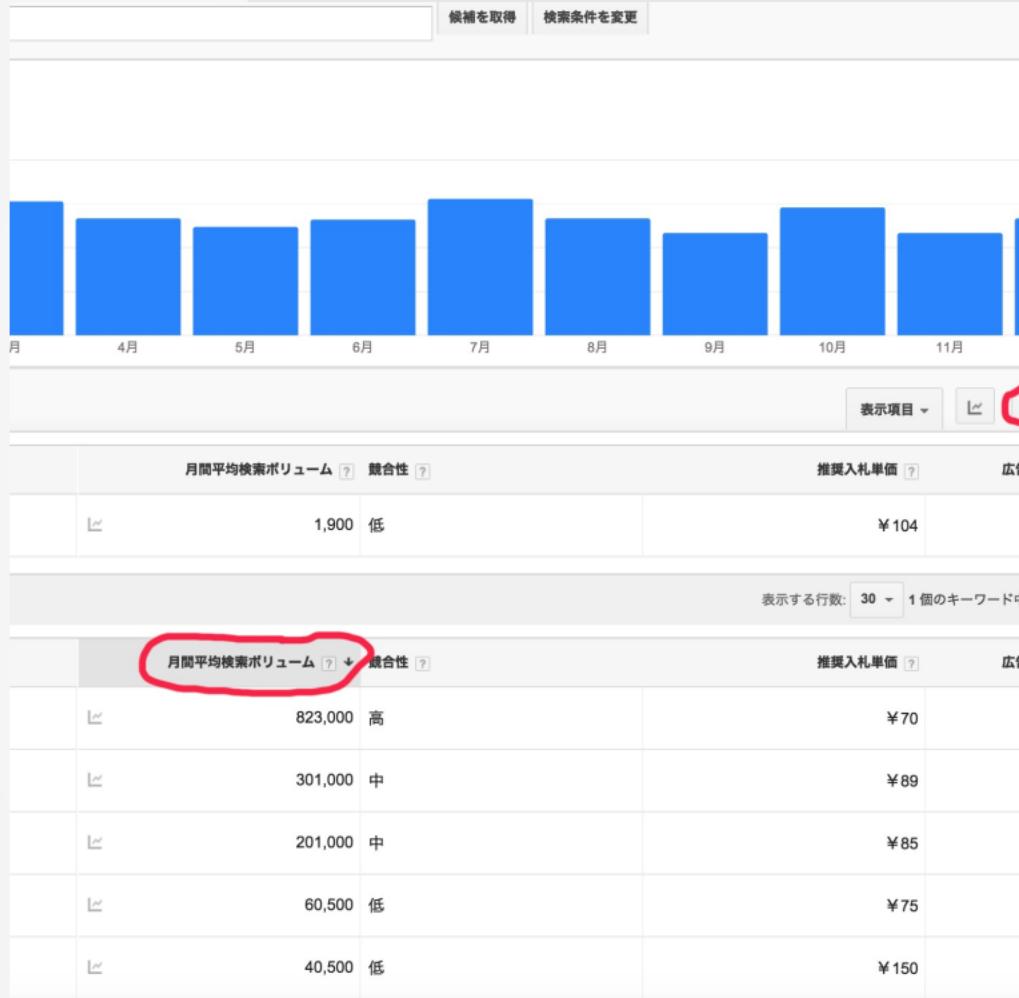
The opengovsg initiative aims to enhance transparency and citizen engagement through open data. Analyzing their GitHub repositories can provide insights into effective civic tech solutions and collaborations.



Automatic YAML File Retrieval

The system automatically fetches a YAML file from GitHub containing hundreds of official government GitHub accounts globally. This file is maintained by GitHub and serves as the foundational dataset for subsequent analysis.





Analyzing All Accounts vs. Filtered Accounts

The system offers the option to analyze all accounts by default, or narrow down to specific high-impact organizations. This flexibility enables targeted research, facilitating focused analysis on significant contributors.

High-impact Organizations for Filtering

Utilizing filters allows for a more precise analysis of key government entities such as GSA, alphagov, and opengovsg. This targeted approach enhances the relevance of the analysis results for stakeholders.



GSA - General Services Administration

The GSA plays a pivotal role in managing government procurement, providing a wide range of services. Analyzing their GitHub accounts can uncover innovative technology implementations in public sectors.



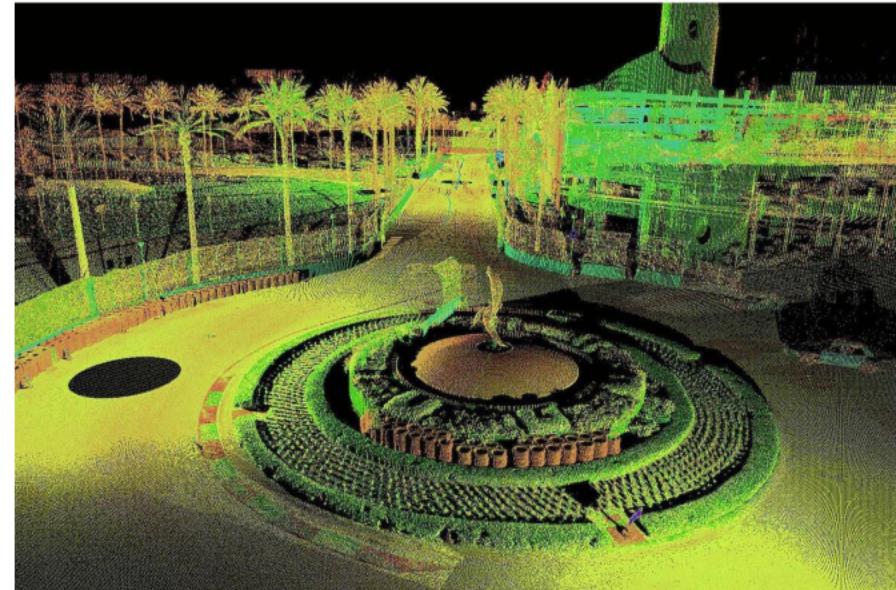
alphagov - UK Government Digital Service

alphagov represents the digital service for the UK government, focusing on making public services accessible online. Their GitHub contributions inform best practices in public service delivery and user-centered design.



opengovsg - Singapore's Open Government

The opengovsg initiative aims to enhance transparency and citizen engagement through technology. Analyzing their GitHub repositories can provide insights into effective civic tech solutions and collaborations.

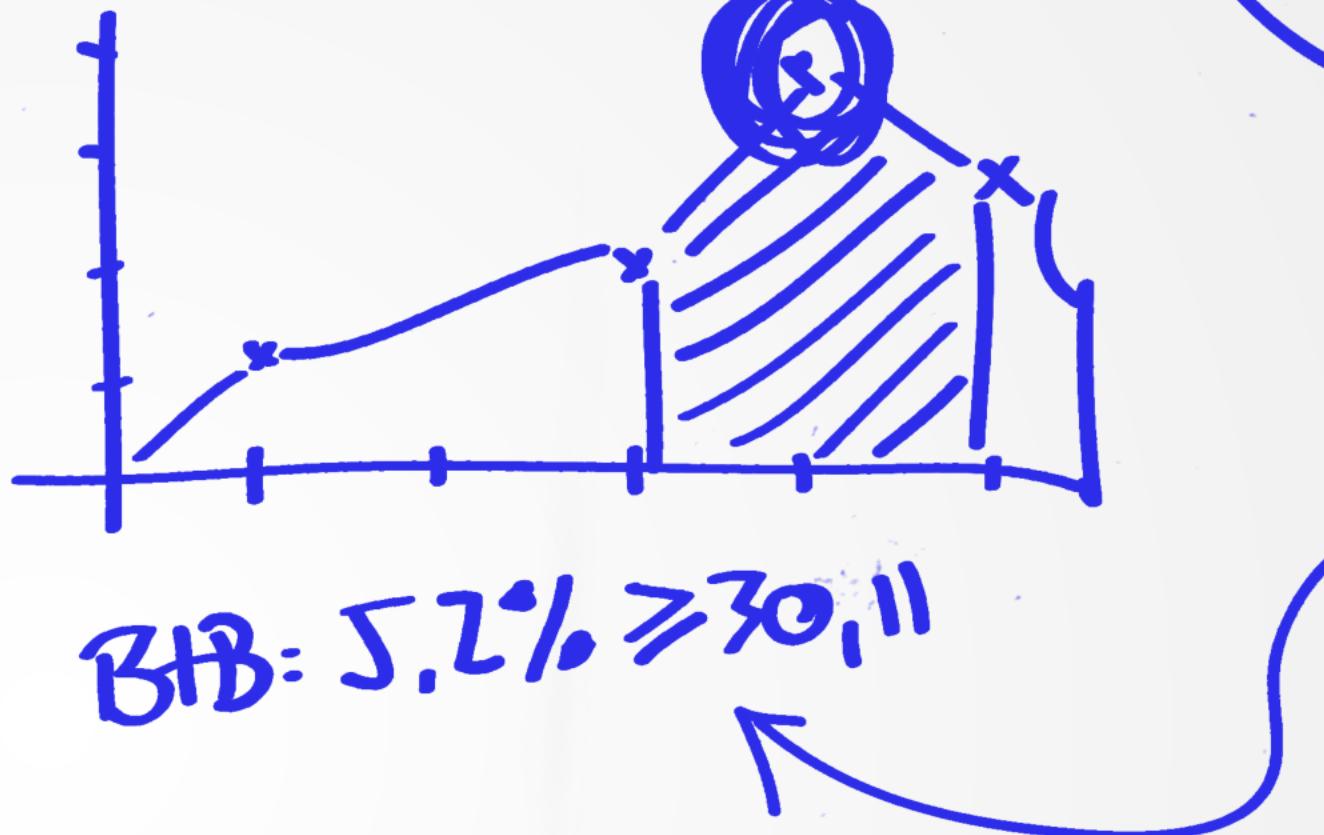
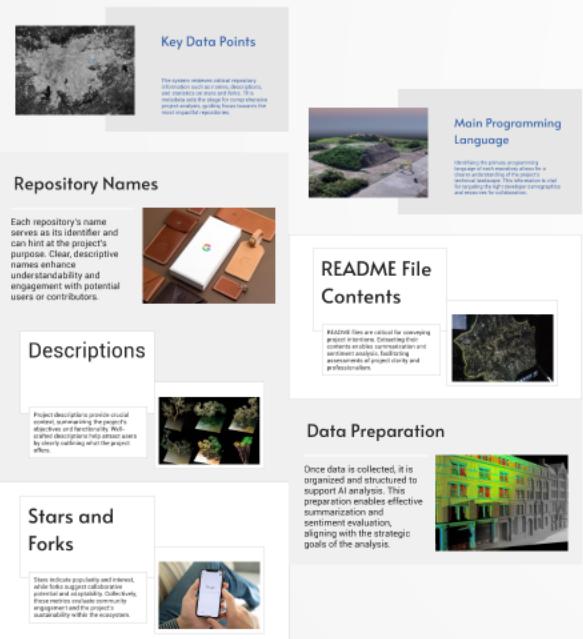


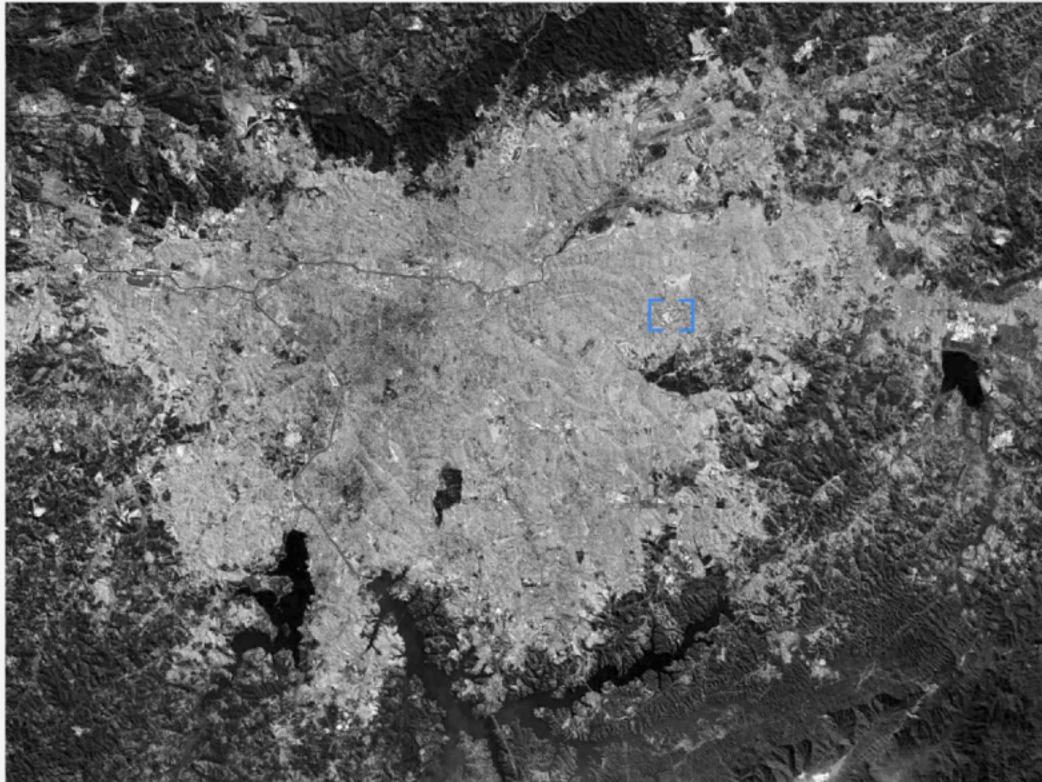
User Guide: Government GitHub Analysis System

Step-by-step Instructions for Environment Setup and Data Analysis



Repository Metadata Extraction





Key Data Points

The system retrieves critical repository information such as names, descriptions, and statistics on stars and forks. This metadata sets the stage for comprehensive project analysis, guiding focus towards the most impactful repositories.

Repository Names

Each repository's name serves as its identifier and can hint at the project's purpose. Clear, descriptive names enhance understandability and engagement with potential users or contributors.



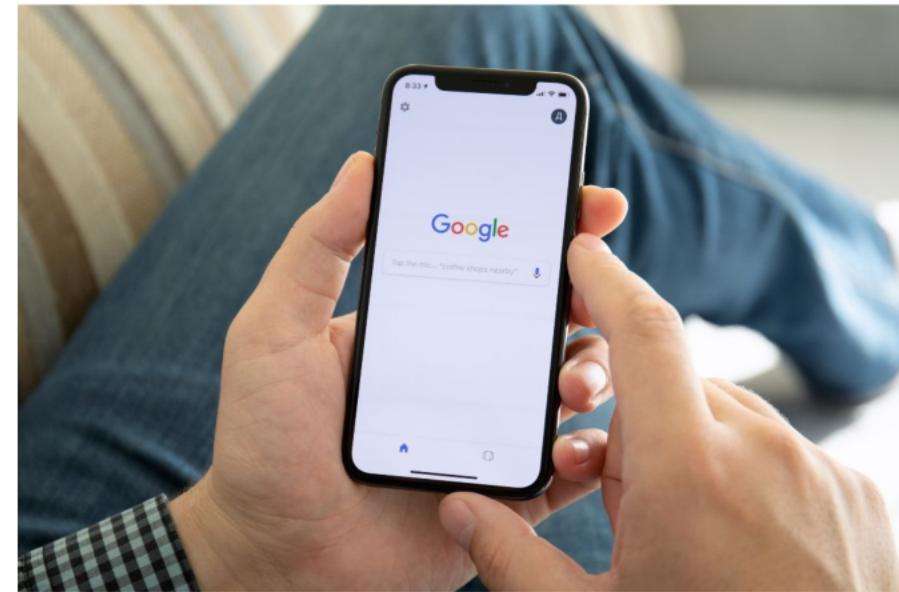
Descriptions

Project descriptions provide crucial context, summarizing the project's objectives and functionality. Well-crafted descriptions help attract users by clearly outlining what the project offers.



Stars and Forks

Stars indicate popularity and interest, while forks suggest collaborative potential and adaptability. Collectively, these metrics evaluate community engagement and the project's sustainability within the ecosystem.





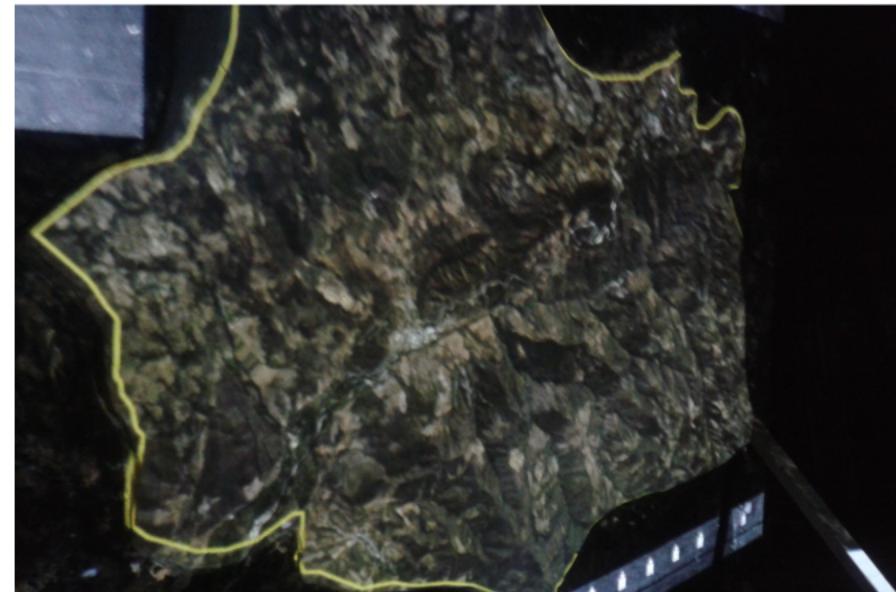
Main Programming Language

Identifying the primary programming language of each repository allows for a clearer understanding of the project's technical landscape. This information is vital for targeting the right developer demographics and resources for collaboration.

README File

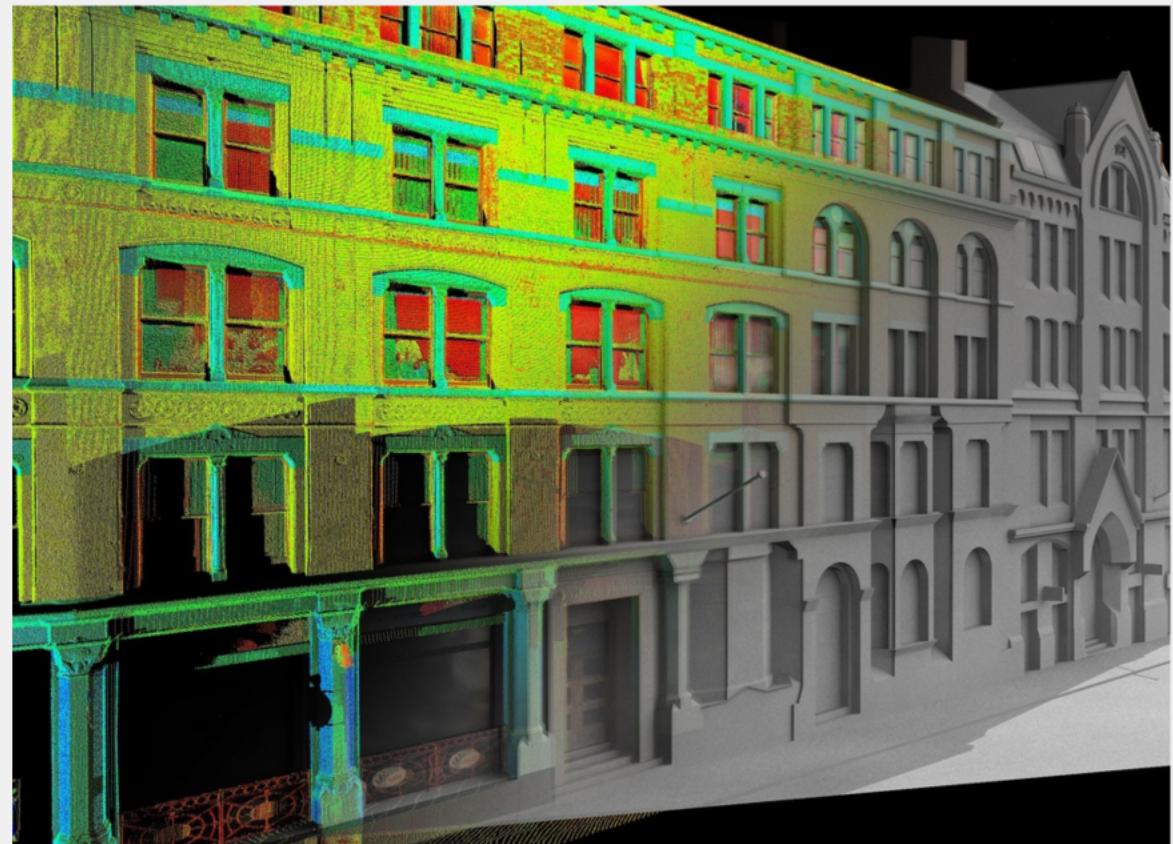
Contents

README files are critical for conveying project intentions. Extracting their contents enables summarization and sentiment analysis, facilitating assessments of project clarity and professionalism.



Data Preparation

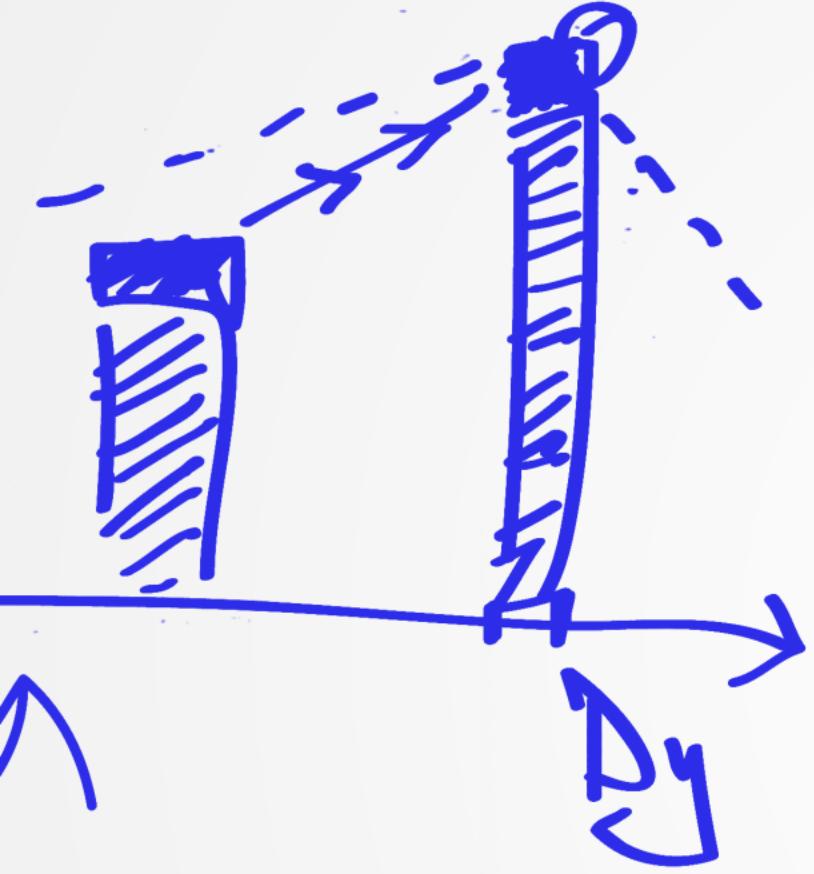
Once data is collected, it is organized and structured to support AI analysis. This preparation enables effective summarization and sentiment evaluation, aligning with the strategic goals of the analysis.



User Guide: Government GitHub Analysis System

Step-by-step Instructions for Environment Setup and Data Analysis

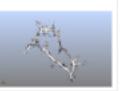




AI-Powered Analysis and Export

Summarization and Sentiment Analysis Process

The system utilizes GPT-4 for generating concise summaries of each project. Initially, a sentiment score assesses the overall tone, context, and language used in the project's presentation, thereby reflecting project presentation quality.



Generated Output Files

Results from the analysis are stored in multiple formats, including PDFs and JSON files, for downstream applications. These files facilitate easier integration with other systems and provide a structured way to review and analyze the results of the AI-powered sentiment analysis.



Summarized Project Overviews

This summarized project overview provides a clear, concise description of each open-source or commercial project. It highlights key features and clearly understands core objectives. This summary is invaluable for stakeholders across various sectors, including project relevance and potential impact.



Sentiment Evaluation

Sentiment analysis not only evaluates the clarity and enthusiasm in text-based data but also considers the tone and quality of project descriptions. By combining tone and language analysis, this feature ensures a more holistic gauge of project professionalism and engagement.



Trend Analysis

Analyzing sentiment data helps identify geographic or organizational trends in project reception. Policymakers can target specific regions or industries for further support or resources based on the most positively received open-source projects.



Opportunity Identification

The analysis highlights standout projects worthy of further funding or integration into larger initiatives. Notable examples, such as Singapore's A3DGov, showcase innovation and potential for replication, aiding in effective public sector investment strategies.



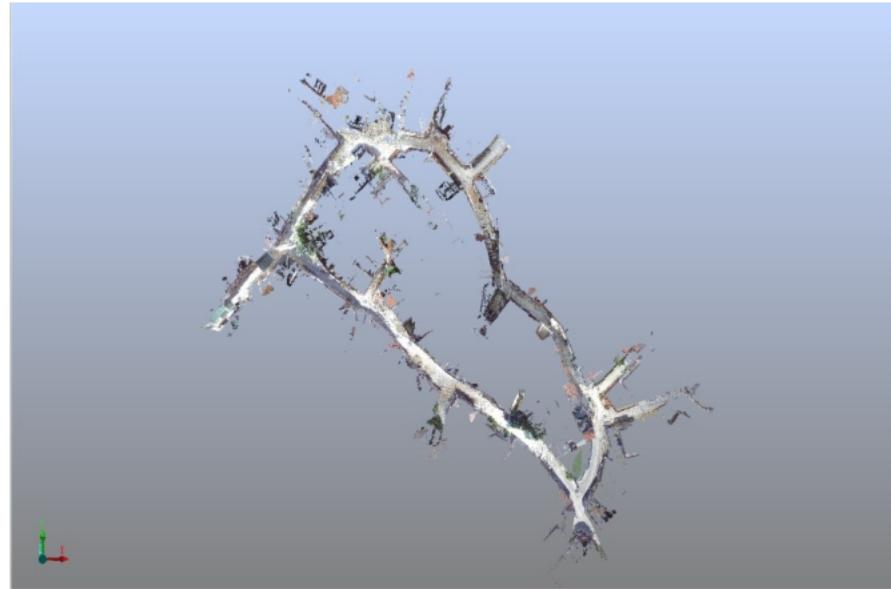
Data-Driven Insights

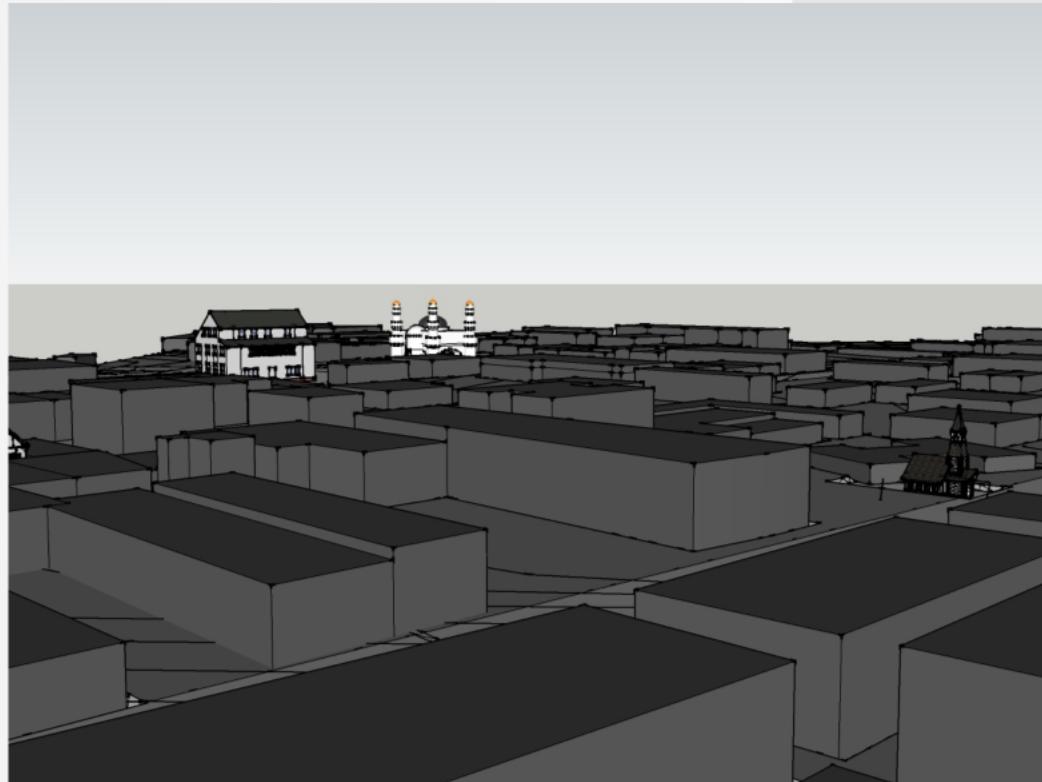
Employing AI-supported, reproducible insights, an actual project can benefit from the use of machine learning models to derive deeper insights and reduce reliance on subjective judgment, enabling a more informed decision-making process in public-sector technology investments.



Summarization and Sentiment Analysis Process

The system utilizes GPT-4o for generating concise summaries of each project's README, capturing key objectives. Additionally, a sentiment score assesses the overall tone, categorizing it as positive, neutral, or negative, thereby reflecting project presentation quality.





Generated Output Files

Results from the analysis are saved in multiple formats for easy access: `readme_analysis.csv` for summaries and sentiments; `readme_analysis.json` for structured data; `sentiment_by_country.csv` aggregates sentiment by country; and `sentiment_by_org.csv` provides organizational sentiment insights.

Summarized Project Overviews

The summarized project overviews provide a clear, concise description of each open-source government initiative, allowing stakeholders to quickly understand core objectives. This format is particularly beneficial for non-technical reviewers evaluating project relevance and potential impact.



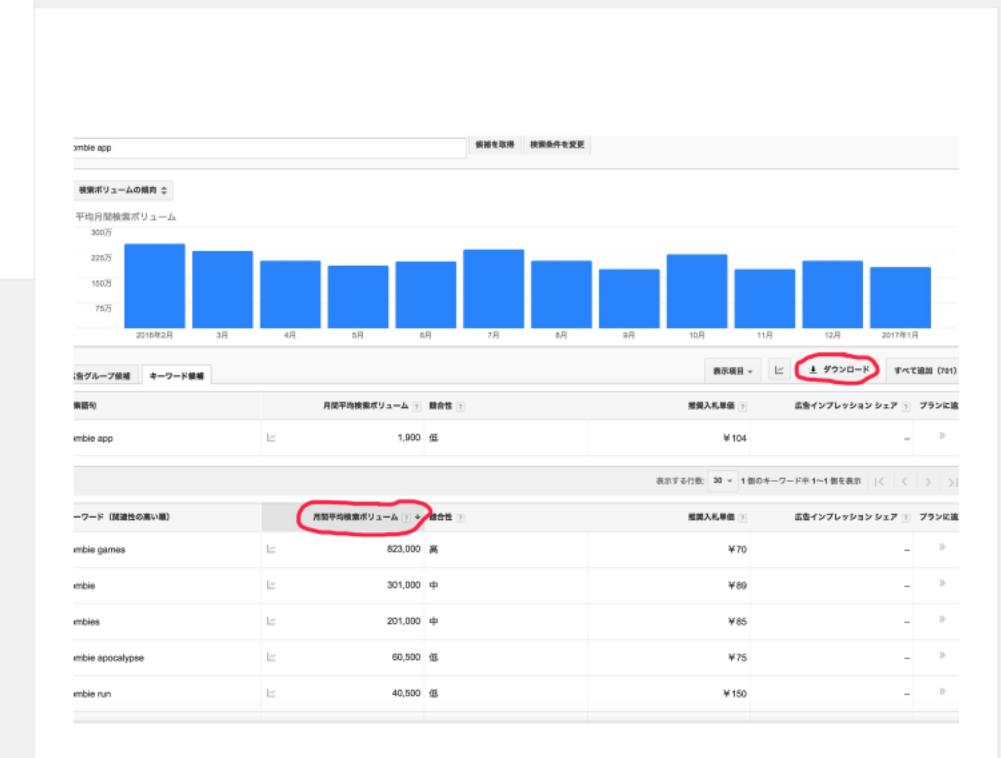
Sentiment Evaluation

Sentiment analysis not only evaluates the clarity and enthusiasm in README files but also acts as an indirect measure of project quality. By understanding tone and presentation style, stakeholders can gauge overall project professionalism and confidence.



Trend Analysis

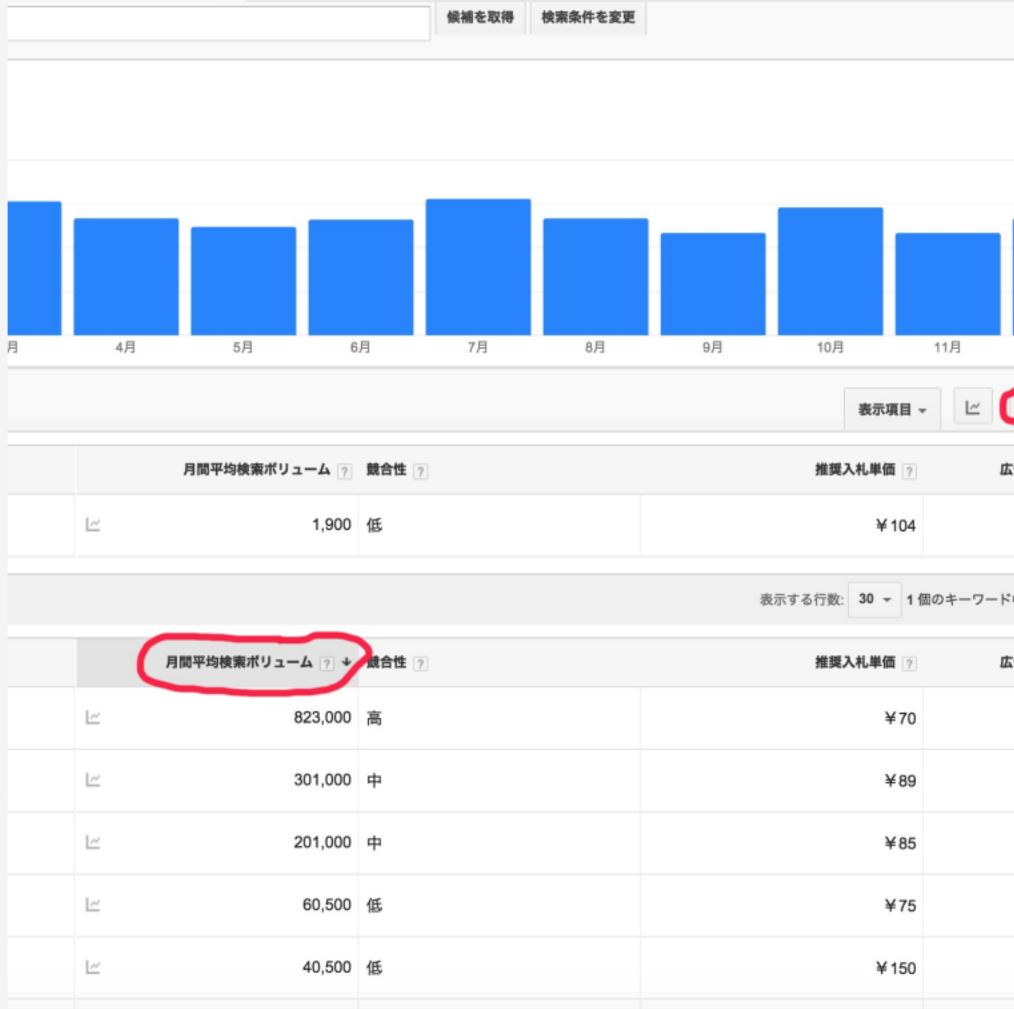
Analyzing sentiment data helps identify geographic or organizational trends in project reception. Policymakers can target investment areas by recognizing which countries or organizations boast the most positively received open-source projects.



Opportunity Identification

The analysis highlights standout projects worthy of further funding or integration into larger initiatives. Notable examples, such as Singapore's AskGov, showcase innovation and potential for replication, aiding in effective public sector investment strategies.





Data-Driven Insights

Employing AI-supported, reproducible insights based on actual project content reinforces data-driven decision-making. This reduces reliance on subjective judgment, enabling a more systematic approach to evaluating public-sector technology investments.

User Guide: Government GitHub Analysis System

Step-by-step Instructions for Environment Setup and Data Analysis

