Finding Machine Learning Datasets Beyond Kaggle

Accessing diverse and quality datasets is essential for successful ML projects. Here are several excellent alternatives to Kaggle:

1. UCI Machine Learning Repository
   - Classic resource with various datasets.
   - [UCI ML Repository](http://archive.ics.uci.edu/ml/index.php)

2. Google Dataset Search
   - Aggregates datasets from across the web.
   - [Google Dataset Search](https://datasetsearch.research.google.com/)

3. AWS Open Data Registry
   - Free access to public datasets on AWS.
   - [AWS Open Data Registry](https://registry.opendata.aws/)

4. Data.gov
   - U.S. government's open data portal.
   - [Data.gov](https://www.data.gov/)

5. Microsoft Azure Open Datasets
   - Pre-processed datasets for ML tasks.
   - [Azure Open Datasets](https://azure.microsoft.com/en-us/services/open-datasets/)

6. KDnuggets Datasets
   - Curated list of datasets for various analyses.
   - [KDnuggets Datasets](https://www.kdnuggets.com/datasets/index.html)

7. Google BigQuery Public Datasets
   - Easily query large-scale public datasets.
   - [BigQuery Public Datasets](https://cloud.google.com/bigquery/public-data)

8. Harvard Dataverse
   - Research data from various academic disciplines.
   - [Harvard Dataverse](https://dataverse.harvard.edu/)

9. OpenML
   - Platform for sharing datasets and ML experiments.
   - [OpenML](https://www.openml.org/)

10. Awesome Public Datasets on GitHub
    - Community-driven collection of public datasets.
    - [Awesome Public Datasets](https://github.com/awesomedata/awesome-public-datasets)

11. World Bank Open Data
    - Comprehensive global development data.
    - [World Bank Open Data](https://data.worldbank.org/)

12. NCBI
    - Biological and genomic datasets.
    - [NCBI](https://www.ncbi.nlm.nih.gov/)

Explore these resources to find the perfect dataset for your ML projects. Each platform offers unique datasets suited to various domains and research interests.