1.Google Dataset Search – A search engine for datasets:

https://datasetsearch.research.google.com/

2. IBM's collection of datasets for enterprise applications:

https://developer.ibm.com/exchanges/data/

3. Kaggle Datasets:

https://www.kaggle.com/datasets

4. Huggingface Datasets – A Python library for loading NLP datasets:
https://github.com/huggingface/datasets

5. A large list organized by application domain:

https://github.com/awesomedata/awesome-public-datasets

6. Computer Vision Datasets (a really large list):
https://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm

7. Datasetlist – Datasets by domain:

https://www.datasetlist.com/

8. OpenML – A search engine for curated datasets and workflows:
https://www.openml.org/search?type=data

9. Papers with Code – Datasets with benchmarks:

https://www.paperswithcode.com/datasets

10. Penn Machine Learning Benchmarks:

https://github.com/EpistasisLab/pmlb/tree/master/datasets

11. UCI Machine Learning Repository:

https://archive.ics.uci.edu/ml/index.php

12. VisualDataDiscovery (for Computer Vision):

https://www.visualdata.io/discovery

13. Roboflow Public Datasets for computer vision:

https://public.roboflow.com/