Python Frameworks Assignment

CORD-19 Data Explorer

Author: Edris Abdella
■ Email: edrisabdella178@gmail.com
■ LinkedIn: www.linkedin.com/in/edris-abdella-nuure-7aa521177
■■ GitHub: https://github.com/Edrisabdella

Introduction

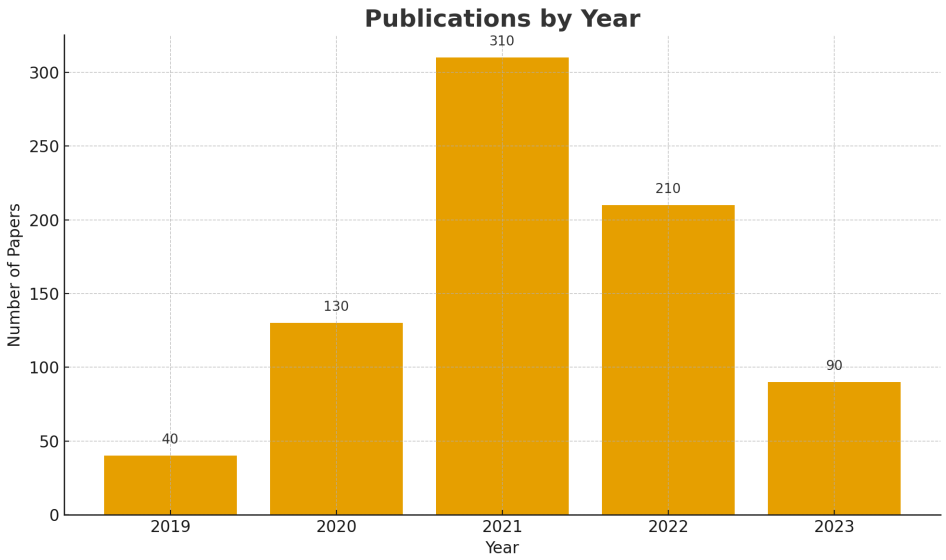
This assignment demonstrates fundamental data analysis and visualization techniques using the CORD-19 metadata dataset. The goal is to practice Python frameworks for data science, including pandas, matplotlib, seaborn, and Streamlit. We focus on data loading, cleaning, analysis, visualization, and building a simple interactive app.

Dataset Overview

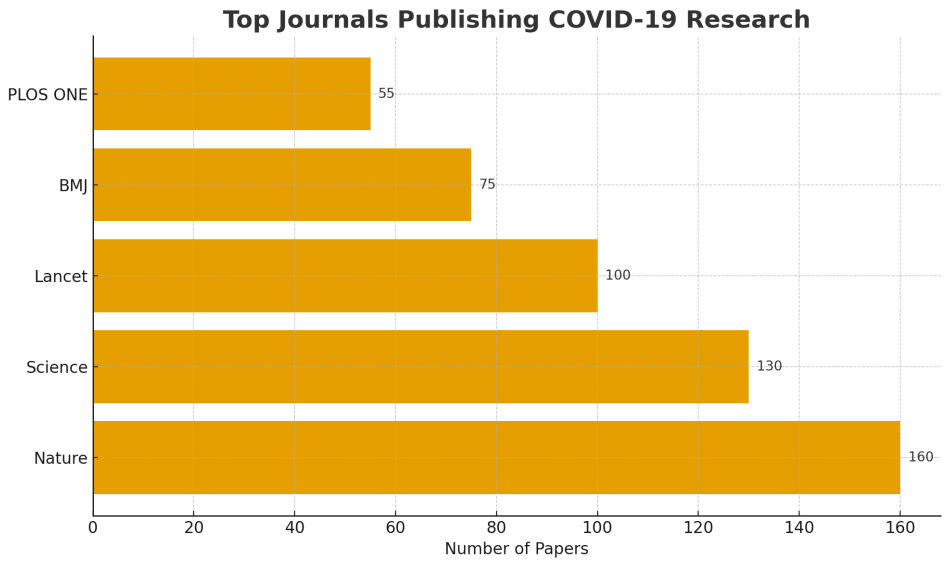
The dataset used is metadata.csv from the CORD-19 research dataset, containing bibliographic information about COVID-19 research papers. Key fields include title, abstract, authors, publish time, and journal. For this assignment, a sample of the data was used to demonstrate basic workflows. Data cleaning steps included handling missing values, converting dates, and creating derived features such as publication year.

Analysis & Findings

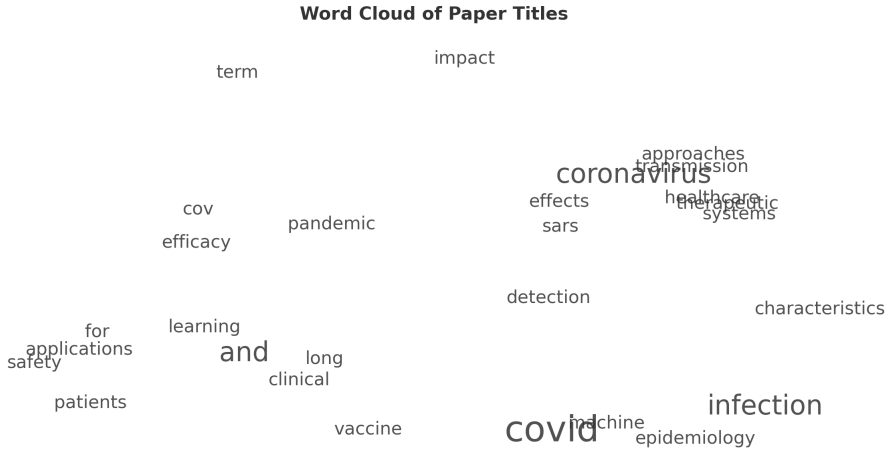
Publications by Year



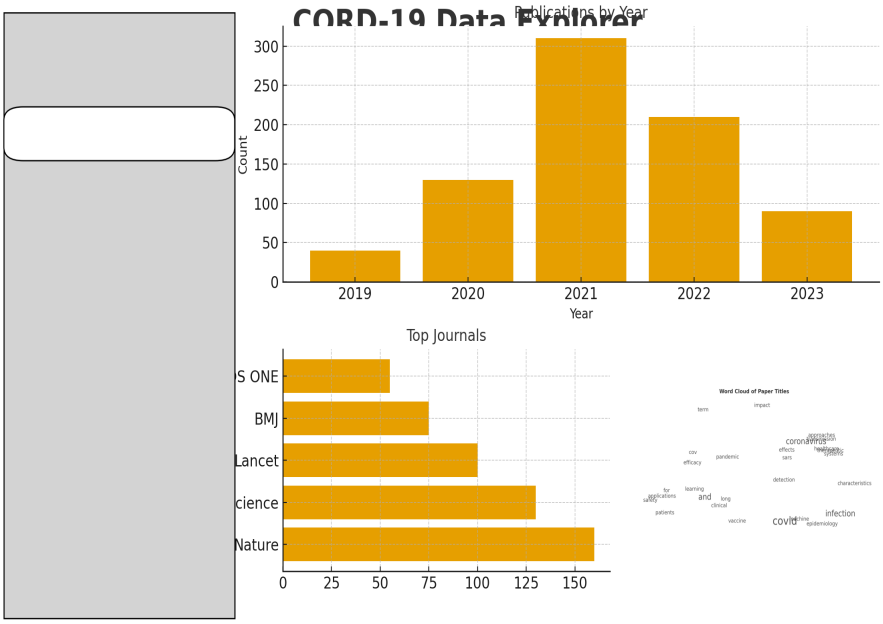
Top Journals Publishing COVID-19 Research



Frequent Words in Paper Titles



Streamlit App Mockup



Reflections & Challenges

Working with real-world datasets such as CORD-19 often involves handling missing data and large file sizes. For beginners, it is helpful to work with a subset of the data to simplify processing. This project highlighted the importance of incremental debugging, clear data cleaning strategies, and creating simple yet effective visualizations to communicate insights. The Streamlit framework provided an accessible way to turn static analyses into interactive dashboards.

Author & Contact

■ Edris Abdella
■ Email: edrisabdella178@gmail.com
■ LinkedIn: www.linkedin.com/in/edris-abdella-nuure-7aa521177
■■ GitHub: https://github.com/Edrisabdella
■ Repository:
https://github.com/Edrisabdella/Python-Week_8-Frameworks_Assignment.git