

Jeffrey Pan

pan.je@northeastern.edu | github.com/jpan1234 | jeffreypan.com | linkedin.com/in/JeffreyYPan | 781.439.9298

EDUCATION

Northeastern University - Khoury College of Computer Sciences, Boston, MA GPA: 3.81 Sept 2020 - Present

Candidate for a Bachelor of Science in Data Science and Environmental/Sustainability Sciences

Achievements and Certificates: Dean's List 2020-2024 (Fall) and Google Data Analytics Certification (Ongoing)

Coursework: Advanced Programming with Data, Database Design, Information Visualization, Machine Learning, Large Scale Information Storage and Retrieval, Statistics, Foundations of Data Science

TECHNICAL KNOWLEDGE

Programming Languages: Python, SQL, R

Software: NoSQL, Redis, SkLearn (LDA, Vectorizer, KNN, KMeans), Plotly (Dash, Express), Gensim (Coherence), PyMySQL, SQLAlchemy, Pandas, Numpy, GeoPandas, BeautifulSoup, Seaborn, Altair, MySQL, Microsoft SQL, BigQuery, SQLite, Google Workspace (slides, docs, sheets, CoLab), Tableau, Microsoft Suite (Excel, Word, Access, PowerBI), Docker, Appsmith, and RStudio

WORK EXPERIENCE

Northeastern University - *Teaching Assistant: Introduction to Programming DS 2000* Jan 2024 - Present

- Reinforce the development of technical skills with Python programming, introducing topics such as data analysis with Pandas and basic machine learning methodology such as linear regression
- Facilitate and manage a learning environment for 400+ incoming and prospective Data Science students
- Ensured swift and efficient support for student inquiries by maintaining a commitment to responding within a timeframe of under 5 minutes

Veolia North America - *Energy Data Analyst*, Boston, MA Jan 2022 - Sep 2023

- Revamped and streamlined manual data cleaning processes using Python and Excel automation, achieving an efficiency boost from hour-long tasks to just 5 minutes a month
- Developed and designed automated tax form populator with SQL - Python connections (using SQLAlchemy), saving \$4000 monthly for clients in tax remissions
- Performed quality analysis and assurance on large-scale (10,000+ rows) utility datasets for multiple clients
- Scripted queries for further data analysis with client utility usage using Microsoft SQL Server and reported findings to clients weekly
- Designed and constructed invoice data reports using Tableau for KPI tracking of client projects

DATA PROJECTS (available in Github repo)

HuskyHealth Application - Northeastern University Nov 2023 - Dec 2023

- Point project manager of the development of RestAPI connections on Appsmith using Python and Docker backend connections to MYSQL database
- Demonstrated expertise in designing and implementing the MYSQL database structure for a healthcare application, contributing significantly to the robust foundation of the system

Tenax Strategies Cannabis Markets Dashboard - Northeastern University Nov 2023 - Dec 2023

- Architected and constructed a dynamic Tableau dashboard to monitor and analyze Cannabis prices at Tenax Strategies
- Pioneered the implementation of an automated web-scraping system from the Massachusetts Cannabis Control Commission, ensuring timely and accurate weekly updates for client decision-making processes

Billboard Top 100 Songs Analysis Dashboard - Northeastern University April 2023

- Implemented Python's Dash package to create a dashboard analyzing the Billboard Top 100 songs
- Utilized various APIs from Spotify, Genius, and last.fm to obtain song lyric data and its features, which were then used for sentiment analysis and visualization
- Applied advanced techniques such as cluster analysis and k-nearest neighbors (KNN) to provide similar song recommendations based on unique song metrics (danceability, cadence, etc.)

RESEARCH / LAB EXPERIENCE

Haelewaters Lab, Ghent University - *Research Assistant & Web Designer*, Boston, MA Nov 2022 - Present

- Research Assistant and co-author for *beetlehangars.org* about a new fungi species: *Hesperomyces harmoniae*
- Cleaned large datasets (10,000+) using Python and Excel - implemented a combination of GeoPandas and Plotly to effectively and efficiently determine fungal species coordinates for graphing purposes
- Developed web-scraping techniques for automated online data collection for samples from Flickr

NEU Sustainability and Social Change Lab - *Research Assistant*, Boston, MA Dec 2022 - Dec 2023

- Assisting in experimental design, data analysis, content analysis, literature reviews, and running studies using Python, R, and Qualtrics for Northeastern University's SSC Lab
- Scraping and Analyzing Labor Union Twitter and Speech data from various sources on Climate Change and performing sentiment analysis, natural language processing, and topic modeling using Python