

Xuan Wang

Austin, Texas | 512-200-6763 | xxxuan0213@gmail.com | linkedin.com/in/xxuan-wang/ | xuanwangg.github.io/

SKILLS

Programming Languages: Python, SQL, PostgreSQL, Apache Spark, Git, R; Scikit-learn, Pandas, NumPy, Tensor-flow, Keras

Tools: Tableau, AWS (S3, DynamoDB, EC2), GCP BigQuery, Salesforce, Jupyter, Kubernetes, Azure, Docker, Excel

EDUCATION

MS In Information Studies - The University of Texas at Austin | GPA: 3.45/4.00 Aug 2019 - May 2021

BS In Electronic Commerce - Dalian University of Technology | GPA: 3.40/4.00 Sep 2015 - Jun 2019

PROFESSIONAL EXPERIENCE

Data Analyst - Angle Health Inc. Apr 2023 - Aug 2023

- Developed regression models to assess and track risk scores, enabling prompt monitoring and handling of high-cost members.
- Cultivated strong stakeholder-tech cohesion, assisting TPMs in translating business requirements into feasible tasks like PostgreSQL data integration and service alignment with timely progress updates.
- Utilized AWS Lambda and S3 buckets to architect, deploy efficient ETL pipelines and design Retool views for clients. Successfully streamlined data extraction processes, resulting in an impressive 60% improvement in data retrieval speed.

Data Analyst (AI/ML) - Movable Ink Mar 2022 - Jan 2023

- Developed machine learning models to pinpoint high-converting products and proactively communicated insights and proposed strategic actions to stakeholders across CX-Eng teams to facilitate well-informed decisions.
- Collaborated closely with engineers to customize GitHub Actions for deploying machine learning models through CI/CD pipelines; Assisted in troubleshooting, resolving, and conducting data validation to elevate data quality.
- Spearheaded the compilation and optimization of impactful SQL queries on BigQuery, aligning precisely with business imperatives to support the modeling team through interactive Tableau dashboards.
- Championed the development and implementation of a Python-based reporting automation tool on Cloud Run, driving a fivefold speed increase, significantly reducing manual input, and minimizing human errors.
- Performed EDA to process TB of inbound data from 5+ sources using PySpark, making data ready for downstream purposes.

Data Science & Analysis Intern - Renzoe Box, Inc. Sep 2021 - Mar 2022

- Utilized NLP to perform multi-label classification of textual data, resulting in the prediction of 64 tags for 28k+ beauty products.
- Created the first content-based ML solution in Spark to generate makeup recommendations by wrangling 500+ surveys.
- Modeled structured and unstructured data, deployed ML models with Flask on AWS EC2 to visualize the matching results.
- **Impact:** Implemented methodologies utilizing neural networks and exhaustive search to create personalized makeup recommendations based on individual preferences. Established relevant metrics to support decision-making process.

Data Research Assistant - Texas Department of Information Resources Dec 2020 - June 2021

- Automated web scraping with using requests, BeautifulSoup, and selenium, to fetch 12k+ vendor records through Salesforce API and feed as inputs to 15 websites for targeted file collections.
- **Impact:** Developed a multi-threaded Python application with optimized procedures and functions to fulfill extensive data requirements, leading to a time efficiency >10X, and significantly reduced oversight by contract managers by 90%.

Cloud Computing TA - UT Austin, Department of Computer Science Jan 2021 - May 2021

- Designed Git-based automated Python grading scripts with seamless integration, reduced human supervision by 70%.
- Developed S3 and DynamoDB handlers to facilitate distributed data manipulation with AWS Python SDK in virtual environments.
- Led a group of 60+ students in container exploration and successfully deployed of Helm Charts on a GKE single-node cluster.

Big Data & Distributed Programming TA - UT Austin, McCombs School of Business Aug 2020 - Jan 2021

- Guided students in leveraging EC2 GPUs to experience quicker execution capabilities of TensorFlow/Keras deep learning models.
- Illustrated the essential operations of RDDs on distributed datasets exceeding 1TB in size, employing Apache Spark to implement a collaborative-filtering movie recommendation system.