Yujia Shen

yms5500@psu.edu | (+1)8142806883

CAREER OBJECTIVE

Aspiring Data Analyst skilled in Python, SQL, Tableau, and statistical modeling, seeking an internship to leverage data-driven insights for impactful decision-making.

EDUCATION

Wenzhou-Kean University (WKU) **Wenzhou, China**
Bachelor of Science, Major in Finance 09/2020 – 06/2024
GPA: 3.73/4.00
Awards: Dean's Scholarship 2020, The 3rd Innovation and Entrepreneurship Competition 2020
Core Coursework: Big Data Computing, Statistical Data Mining, Data Mining in Business Applications, Fundamentals and Applications of Business Information Systems

Pennsylvania State University (PSU) **State college, USA**
Master of Science, Major in Applied Statistics 09/2020 – Present
GPA: 3.89/4.00
Core Coursework: Regression Methods, Applied Data Mining & Statistical Learning, Analysis of Variance and Design of Experiments, Design and Analysis of Clinical Trials

PROFESSIONAL CERTIFICATE

Google Advanced Data Analytics **Google** 01/2025
- Developed expertise in statistical modeling, predictive analytics, and machine learning, utilizing Python libraries like scikit-learn, statsmodels, xgboost, TensorFlow, and PyTorch to analyze customer behaviors and forecast sales trends in three AI-driven commercial programs.
- Conducted data cleaning, transformation, and visualization for large-scale datasets using Python (pandas, numpy) and SQL, optimizing data workflows for over 10 million records and presenting actionable insights to stakeholders.

Data Visualization with Tableau Specialization **University of California, Davis** 01/2025
- Built 10 interactive Tableau dashboards as part of certification, leveraging advanced techniques like calculated fields, parameterized controls, and data blending to effectively communicate labor market and sales trends.
- Built 10 interactive Tableau dashboards as part of certification, leveraging advanced techniques like calculated fields, parameterized controls, and data blending to effectively communicate labor market and sales trends.

RESEARCH EXPERIENCE

The Demand for Fintech Skills in China: Evidences on Wages and Firms' Performance **Wenzhou, China**
Student Partnering with Faculty (SpF) Research Program (2022-2023) with ¥30,000 Grant 09/2023 – 01/2024
Principal Investigator (Led by Professor Fa-Hsiang Chang. Ph.D. Economics, University at Buffalo—SUNY)
- Automated data pipelines using Python with libraries like pandas, numpy, and re to process and clean **30 GB** of unstructured text data from web-scraped job descriptions, enhancing data quality and analysis readiness.
- Designed and optimized relational databases with SQL, applying indexing and normalization techniques to efficiently store and manage over **30 million records** of large-scale data. Developed automated integrity checks using SQL constraints and triggers, achieving **90%** data accuracy and reliability.
- Improved data usability by cleaning and mining data using sklearn and gensim method, conducting advanced natural language processing (NLP) tasks to structure raw text data for downstream analysis.

Labor Market Concentration and Its Effect on Local Labor Market in China **Wenzhou, China**
Student Partnering with Faculty (SpF) Research Program (2022-2023) with ¥30,000 Grant 09/2022 – 09/2023
Research Assistant (Led by Professor Fa-Hsiang Chang. Ph.D. Economics, University at Buffalo—SUNY)
- Managed more than **40 million records** in a Huawei RDS cloud database using SQL, securely uploading and maintaining data while ensuring role-based access control and efficient retrieval.
- Automated data analysis pipelines with Python scripts, leveraging libraries such as matplotlib and seaborn for visualizations, reducing manual effort by **80%** and accelerating data workflows.
- Developed **10+ interactive dashboards** in Tableau, integrating data directly from Huawei RDS to provide actionable insights on labor market trends, skill demands, and wage patterns.

WORKING PAPER

Y, Shen., W, Zeng., X, Yuan., & Chang, F.H. (2024). Skill Development in Finance Major: Are Quantitative Skills Really Important for Undergraduates?
- **Contribution:** Processed a **20-million-record** dataset using pandas, numpy, and SQL with indexing. Built regression models with scikit-learn, ensuring data accuracy and optimizing workflows. Demonstrated expertise in large-scale data analysis.

TCHNICAL SKILLS

Programming and Scripting Languages:
- **Advanced:** Python, SQL – Proficient in data cleaning, manipulation, and analysis; experienced in building and optimizing data workflows for large datasets, as well as automating data extraction and processing.
- **Intermediate:** Stata, R, SAS – Capable of conducting statistical analysis, data modeling, and hypothesis testing for research projects and complex data tasks.

Data Visualization and Reporting Tools:
- **Intermediate:** Tableau – Skilled in creating interactive and insightful dashboards to communicate complex data trends and findings effectively.