Ivy Zhou

Email: ivyzhou752@gmail.com | Phone: +1 (253) 279-5327 | LinkedIn Profile

(Available to Relocate Nationwide)

EDUCATION

Northeastern University   *Boston, MA/Portland, ME   01/2022 - 12/2023*
Master's in Data Analytics - Applied Machine Learning (GPA: 3.96/4.0)

University of Washington, Seattle   *Seattle, WA   09/2017 - 08/2020*
Bachelor of Science in Mathematics

SKILLS

Programming Languages: Python, SQL, R, Java
Technical Tools and Skills: Tableau, Git, Excel, AWS, Snowflake, Microsoft Azure, Power Apps, Google Cloud, PyTorch, Apache Hadoop, Apache Spark, Machine Learning, Deep Learning, Data Handling and Processing, Infrastructure and Data pipeline.

Analytical Skills: Predictive Analytics and Forecasting, Customer Insights and Market Understanding, Optimization and Efficiency Models, Data Processing and Natural Language Processing, Data Exploration and Statistical Analysis, Advanced Data Mining.

PROFESSIONAL EXPERIENCE

DreamStudio, Data Scientist and Full Stack Developer *(https://dreamstudio.com/)*   *Remote 06/2024 - Present*
- Developed a scalable pipeline for generating prompt-driven narratives using community datasets, leveraging various Stable Diffusion AI models via APIs for automated image generation.
- Customized Odoo ERP and CRM modules, and built a helpdesk ticketing system hosted on a private Google Cloud server.
- Integrated AI-driven products into GitHub's CI/CD pipeline, streamlining the deployment of the main website's storyboard gallery and boosting user engagement with interactive content.
- Collaborated with cross-functional teams to implement solutions across three GitHub projects, optimizing workflows, reducing operational costs, and enhancing scalability.

Model Earth, Data Scientist   *Remote 01/2024 - 09/2024*
- Developed and implemented statistical and predictive models to assess industrial development trends and their environmental impact across geological regions, utilizing Python-based machine learning frameworks (Scikit-learn, XGBoost). Delivered insights through interactive data visualizations and real-time predictions via a custom-built Streamlit app, RealityStream.
- Analyzed five USDA datasets to evaluate the performance of multiple statistical and classification models (LR, RFC, XGBoost, SVM), generating detailed performance/importance reports and identifying key factors affecting outcomes.
- Improved and standardized the training processes for predictive models, enabling real-time model training with customized datasets on Streamlit. Achieved 71% precision and a 78% ROC AUC with XGBoost.
- Managed and regularly updated dashboards to incorporate new geographical features, ensuring stakeholders had access to up-to-date, actionable insights.

IDEXX Laboratories, Inc, People Data Analyst   *Westbrook, ME 01/2022 - 08/2023*
- Developed a structured approach for analyzing quarterly employee surveys, using NLP-based sentiment analysis to extract insights from unstructured data. This analysis achieved 74% accuracy, leveraging AWS Comprehend and Python libraries (NLTK, SpaCy) to guide HR decision-making.
- Designed an ad-hoc text analysis filter to assist management executives in identifying potential workplace safety concerns within employee reports, employing fuzzy word-matching techniques using Levenshtein distance with Python (FuzzyWuzzy).
- Performed data preparation and exploratory analysis using Python (NumPy, Pandas, Seaborn, Matplotlib) to ensure the integrity and quality of stored source data prior to saving, enabling reliable insights for informed decision-making.
- Updated and maintained ad-hoc Tableau dashboards with quarterly survey data, providing HR leadership with clear and actionable insights to track employee sentiment and workplace trends.
- Contributed to building an ETL process for the HR department to enhance operational efficiency and reduce data transfer errors. This involved extracting data from various HR management platforms, transforming it within AWS EC2, and loading it into a Snowflake cloud SQL data warehouse.

1st Security Bank of Washington, Quality Assurance Analyst   *Seattle, WA 03/2021 - 11/2021*
- Ensured data integrity for the Home Lending Department by conducting data mining and analysis from multiple sources to boost operational accuracy.
- Reviewed and processed client documents for mortgage loan closings, ensuring alignment with legal requirements to streamline the closing process.
- Managed monthly data collection and analysis for loan profiles, significantly improving data accuracy and providing critical insights for the compliance team.
- Analyzed credit report datasets generated by the Encompass system daily, reducing the risk of incomplete case closures and enhancing decision-making for loan approvals.

PROFESSIONAL PROJECTS

The Dempsey Cancer Care Center Data Engineering and Analytics – Python, Microsoft Azure, MySQL, Tableau
- Conducted exploratory data analysis to assess how the quality of the client's current database structure significantly affects operational efficiency and leads to resource wastage.
- Performed predictive analysis using forecasting techniques to optimize resource allocation with limited funding and personnel, aiming to increase service demand.
- Engineered a comprehensive data pipeline for data centralization, incorporating automated data extraction, Python-based data preparation, and ETL processes in a relational database hosted on Microsoft Azure MySQL.
- Implemented a file storage solution in Azure for processed Mindbody source data and developed an Azure Data Factory (ADF) pipeline for efficient data transfer, greatly enhancing the source data loading process.
- Created dynamic operational dashboards on Tableau Cloud, leveraging advanced data visualization techniques to provide actionable recommendations for optimizing service scheduling.
- Proposed future project developments to incorporate linear and integer programming, aiming to further reduce resource wastage in operational processes.

VIP Tires and Service Analytics - R-Studio, Python, Tableau, Data Mining, Statistical Analysis
- Conducted a comprehensive customer behavior analysis in R (arules and arulesViz) for market basket analysis to refine and enhance marketing strategies.
- Created interactive visualizations and detailed numerical reports to identify customer buying patterns and service correlations, improving customer retention by aligning strategies with evolving market dynamics.
- Processed and prepared a large-scale dataset from 65 stores using R libraries (dplyr, tidyr, tidyverse, mice), and performed exploratory data analysis to uncover trends in purchases by time, location, and service types.
- Proposed a future project to evaluate the newly implemented customer loyalty program, which includes inferential analysis, normality tests, hypothesis testing, and structural modeling (mediation analysis) using existing data. The project also involves developing a new customer survey to gather demographic information for future behavioral analysis.