

# YIXUAN FENG

+1 (857) 268 8315 | feng.yix@northeastern.edu | linkedin.com/in/yixuan-feng

## Education

### Northeastern University | Boston, MA

May 2022

Master of Science in Data Analytics Engineering

Relevant Coursework: Computation & Visualization, Data Mining, Statistics Methods, Data Management & Database Design

### University of Washington | Seattle, WA

Jun 2019

Bachelor of Science in Chemical Engineering

Relevant Coursework: Programming in Data Science, Probability & Statistics for Engineers, Computer Programming

## Skills

Programming Skills: R (ggplot2, caret, dplyr), Python (Pandas, NumPy, Matplotlib, Seaborn), SQL (MySQL, Microsoft SQL Server)

Statistics & Machine Learning: Scikit-Learn, TensorFlow, NLTK, A/B Testing, SPSS, Regression, Classification, Clustering

Tools: Looker, Tableau, Power BI, MySQL Workbench, Spark, SSPS, Minitab, AWS, MS SSIS, MS SSAS, MATLAB

## Work Experience

### Analyst, Credit Risk | Avant | Chicago, IL

Jul 2022 - Present

- Produce credit card risk monitoring deck in Python and SQL to review delinquency performance and analyze risk strategies
- Track and report on Avant's portfolio and industry trends to stay ahead of an always-evolving consumer lending environment by identifying & implementing improvements, and suggesting additional analysis
- Collaborate with the product team & data team on new initiatives and projects by providing insights on credit risk management
- Create Looker dashboards to provide frequent updates and keep management better informed of current risk trends
- Streamline the process flow of projects and improve accuracy in JupyterLab to reduce compilation time and complexity by 50%

### Data Analyst Co-op | MEMIC | Portland, ME

Jul 2021 - Dec 2021

- Enhanced data security by establishing a P2P data mapping for 20 databases to assist with the migration of core business systems
- Provided real-time insights by creating data dashboards and reports by integrating SQL with Power BI
- Cooperated with DE team to develop ETL procedures using SQL Server & IBM Db2 which improved data quality by 60%
- Analyzed national insurance statistics and used R to build machine learning models to predict possible claims with 82% accuracy
- Researched external data, documented API access techniques from 45 public statistics to comprehend potential causes of incident

### Web Design and Data Assistant | Northeastern University | Boston, MA

Mar 2021 - Jul 2021

- Designed and maintained academic websites, assessed survey results, visualized to interpret patterns in questionnaire dataset

### Business Statistics TA | Northeastern University | Boston, MA

Sep 2020 - Dec 2020

- Mentored a batch of students in business statistics and hypothesis testing concepts, assisted professors with the classwork

### Data Analyst Intern | Jinan Tata Network Technology Co., Ltd | China

Jun 2019 - Aug 2019

- Collaborated with cross-functional teams and analyzed BI metrics to deliver time-sensitive reports for clients which improved channel sales pipelines and boosted brand awareness by exceeding channel sales target by \$20K
- Performed market research on website traffic to identify target customers; built a Tableau dashboard to present insights to clients
- Analyzed historical sales in Python to construct an E-commerce marketing plan and showcased marketing KPI in Excel
- Examined sales trends with Pandas and automated a process to reduce 40% of manual tasks of data cleaning & data normalization

## Academic Projects

### E-Commerce Retail Data Warehouse Design

- Built a data warehouse in SQL Server to automatically aggregate sales data and support business intelligence processes
- Deployed star schema to design a dimensional model with dimension and fact tables on top of an OLTP database
- Constructed ETL process using SSIS for automation of transforming and loading 1 M records for 8 batch schedule processes

### Company E-Mail Database Design

- Developed a management system in MySQL for companies to store, manipulate and review data of emails and project files
- Assisted with overall project progress and decision-making by wrangling 30,000 records in R and empowering self-service analysis on revenue to feed data into OLAP cube in SSAS which enables Tableau dashboard auto-refreshing

### Hotel Booking Status Data Analysis

- Predicted future hotel reservation cancellation status with 80% accuracy which enabled hotel owners to prevent losses in advance
- Explored 119K hotel booking data in R; trained classifier using algorithms including Logistic Regression, K-NN, Naive Bayes, Decision Trees; evaluated performance using accuracy, confusion matrix, and ROC curve

### Sentiment Analysis on Yelp Reviews

- Processed 10,000 Yelp restaurant reviews to predict customer ratings by identifying key features from text sentiment in Python
- Generated a pipeline to clean and review texts using the NLTK library in Python which increased efficiency by 60%