

# YIXUAN FENG

Boston, MA | (857)-268-8315 | feng.yix@northeastern.edu | [linkedin.com/in/yixuan-feng](https://www.linkedin.com/in/yixuan-feng) | [github.com/yixuan-feng](https://github.com/yixuan-feng)

## EDUCATION

### Northeastern University

*Master of Science in Data Analytics Engineering*

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

**Boston, MA**

May 2022

### University of Washington

*Bachelor of Science in Chemical Engineering*

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

**Seattle, WA**

Jun 2019

## SKILLS

**Programming Skills:** R (ggplot2, dplyr), Python (Pandas, NumPy, Matplotlib), SQL (MySQL, Microsoft SQL Server), MATLAB, JAVA

**Machine Learning Techniques:** Exploratory Data Analysis, Regression, Classification, Random Forests, Decision Trees, Predictive Analytics, Data Mining, Data Cleaning, Statistical Modeling

**Tools:** MySQL Workbench, ER Studio, Tableau, Power BI, SSPS, Minitab, Microsoft Excel, Microsoft Word, Microsoft PowerPoint

## WORK EXPERIENCE

### MEMIC

*Data Analyst Co-Op*

**Portland, ME**

July 2021 – December 2021

- Built a point-to-point data mapping to assist migration of core business systems with increased data security and accessibility
- Collaborated with development team on ETL procedures to ensure data quality for data reporting using SQL Server and IBM Db2
- Analyzed insurance claims data to understand the characteristics of the population of claims, developed probabilities for making reserving decisions, built 82% accuracy machine learning models in R Studio for predicting losses on an incurred insurance
- Researched external data from public statistics to comprehend the potential causes of worker's compensation claims, documented API access techniques to enrich internal data
- Developed insightful visual presentation using Power BI to summarize aggregate claims data and to facilitate insurance loss control

### Northeastern University

*Web Design and Data Assistant*

**Boston, MA**

March 2021 – Present

- Designed and maintained academic websites, analyzed survey results, visualized to interpret patterns within questionnaire dataset
- *Teaching Assistant in Business Statistics*  
Sep 2020 – Dec 2020
- Mentored 60+ students in business statistics concepts, assisted professors with assignment submission and grading system

### Jinan Tata Network Technology Co., Ltd

*Data Analyst Intern*

**Hangzhou, China**

Jun 2019 – Aug 2019

- Built a marketing plan for eCommerce business in Python and Excel to increase product visibility and market performance
- Automated the process to reduce 40% manual task of data cleaning and data normalization, examined sales trends in Python
- Performed market research and competitive analysis; suggested clients with recommendations using Tableau dashboards
- Increased ROI by 20% while effectively managing \$10,000 in paid search campaigns to grow business online

## ACADEMIC PROJECTS

### E-Commerce Retail Data Warehouse Design

- Built a data warehouse to aggregate sales and inventory data automatically to support business intelligence processes
- Used star schema to design a dimensional model on the top of an OLTP database; enabled SCD to keep track of the changes of customer data and product inventory
- Constructed automatic ETL process using SSIS to populate and backfill 1 million records into 6 tables in SQL Server through 1 initial load and 8 batches processing
- Fed data into OLAP cube in SSAS and auto-refreshing dashboard in Tableau to empower self-service analysis on revenue and inventory sustainability

### Company Email Database Design

- Utilized MySQL to develop a database management system for companies to store, manipulate and analyze data of emails and files
- Designed company email data model and mapped to EER and Relation Model
- Wrangled data using R and implemented into MySQL; created triggers and views to test data integrity and database security

### Hotel Booking Status Data Analysis

- Built and trained classification models in R by developing machine learning algorithms on 119,391 hotel booking data to predict booking cancellation status
- Generated R shiny web application with integrated R and Tableau for presenting the change of hotel bookings and rates over time

### Analysis of California Elementary School Vaccination

- Conducted data analysis and trend interpretation to explore trends and correlation between vaccination rate and infection rate
- Utilized Pandas and NumPy to assess data quality and organize data, visualized dataset in Matplotlib to identify patterns and outliers