WILLIAM ZHU

wzhu4@outlook.com | Arlington, VA 22201 | 610-517-3106 | LinkedIn | GitHub Website

Professional Summary

Data Scientist with 3 years of experience in hospitality, meal subscription and survey program evaluation.

Tools: SQL, Python (pandas, numpy, statsmodel, sklearn, matplotlib, seaborn, networkx, nltk, spacy, gensim, geopandas) AWS (Sagemaker, S3, Redshift, Glue, Airflow), R, Tableau, Git/Github, HTML/CSS

Machine Learning: Linear and Logistic Regressions, Random Forest, Boosting, KNN, Model Selection, PCA, Clustering
Analytical Methods: AB Testing, Cohort Analysis, Network Analysis, Text Analysis, Parallel Computing, Time Series Forecasting

Education

The University of Chicago (Chicago, IL)

MA Computational Social Science (STEM), Quadrangle Scholar   *June 2022*

Courses in Python Programming, Machine Learning, Deep Learning, Natural Language Processing, Social Network Analysis, Algorithms, Discrete Math, Databases, Large Scale Computing
Marketing Concentration at Chicago Booth School of Business

Courses in Competitive Strategy, Marketing Strategy, Consumer Behavior, Experimental Marketing, Applied Regression Analysis, Big Data, Data-Driven Marketing, Macroeconomics

Haverford College (Haverford, PA)

BA Sociology, Minor in Statistics, GRE: Verbal: 165/170, Quantitative: 167/170, Writing: 5/6   *May 2019*

Work Experience

Choice Hotels International (Rockville, MD)

Data Scientist   *February 2022 – Present*

- Building and productionizing ML models to improve loyalty program membership retention, drive value contribution, and detect membership enrollment fraud behaviors (Python and SQL AWS Sagemaker, Redshift, S3, Glue, Airflow)
- Delivered insights to the leadership team on the overlap and customer profiling between Choice and Radisson customers
- Identified customer demographics and behaviors by hotel brands using third-party customer data platform

RealEats (Geneva, NY)

Data Analytics Consultant   *June 2021 – December 2021*

- Drove analytics insights that led to $16 million Series A funding by using Python to automate monthly customer retention reports for CEO, CFO, investors, and board of directors (Python and SQL on AWS Sagemaker and Redshift)
- Improved 2-month customer lifetime value (LTV) by 7% by identifying and implementing promo codes associated with high Average Order Value (AOV) and low churn rate (Python statsmodel, SQL, Google Data Studio)
- Solidified brand positioning strategy for the CMO by identifying top customers persona using logistic regressions and decision trees; compiled customer database by merging sales records, Google Analytics API, U.S. Census, Zillow home value index, and various third-party databases (Python and SQL on AWS Sagemaker and Redshift)

Westat (Rockville, MD)

Research Assistant   *July 2019 – June 2020*

- Received the highest performance rating for all 8 evaluation projects by performing survey data analysis and report writings for clients including the U.S. Department of Education, Verizon, and National Science Foundation (R, Excel)
- Co-authored an evaluation report that led to $260K in funding for Racial Equity Action Leadership Program by showing its effectiveness at guiding 40+ participants to craft company-wide racial equity plan through focus groups and surveys
- Co-authored a peer-reviewed journal article titled "Using state data sets and meta-analysis of low-powered studies to evaluate a school-based dropout prevention program for students with disabilities." in *Studies in Educational Evaluation*

Projects

Exploring Public Bike Share Trip Patterns [5-min presentation, blog post]   *March 2021 – Dec 2021*

- Discovered associations between Chicago's public bike stations' usage volume and crime rates, socio-economic status, and demography by merging and analyzing 300K public bike trip records, 210K Chicago crime records, 11K bus stops records, U.S. Census, and Zillow home value index (Python pandas, networkx, statsmodels) [github repository]
- Achieved a prediction accuracy of 70% (AUC=0.76) by using Random Forest to predict whether e-bike trips cross zip codes; implemented parallel computing on 2.1 Million bike trip data (AWS S3 and EMR, PySpark, Dask) [github repository]

Uncovering Links between Cultural Values and Company Directions [5-min presentation]   *March 2021 – June 2021*

- Showed that firms with an agile culture are more likely to receive M&A and investment offers by analyzing keyword frequency in 4 Million Glassdoor employee text reviews from 6K firms in the U.S. (Python pandas, statsmodel, nltk, spacy, fuzzywuzzy)