Yijiao Zuo

PN:917-499-6058 | **EM**:yz3940@columbia.edu | **LI**: linkedin.com/in/yijiao-zuo-0b4551164

EDUCATION

Columbia University — **New York, NY**
Master of Quantitative Methods in Social Science – Data Science Track *(STEM)* — Jan 2021 – Oct 2022

University of California, San Diego — **San Diego, CA**
Bachelor of Arts of Economics / Minor in Mathematics — Sep 2016 – Mar 2020

Skills

Languages: Python (Numpy, Pandas, Scikit-learn), R (tidyverse, Quanteda), SQL
Analytics: A/B Testing, Statistical Modeling, Predictive Analysis, Clustering & Classification
Visualizations: Tableau, Python (Matplotlib, Seaborn, Plotly), R (ggplot2, ggmap, Plotly, DT, leaflet), Microsoft Excel

WORK EXPERIENCE

Meituan - The largest Chinese delivery App — **Beijing, China**
Business Analyst Intern — May 2021– Aug 2021

- Extracting and aggregating data of bio-pharmaceutical companies from multiple sources, and completed data validation and cleaning in preparation for future data analysis and reporting
- Communicated with stakeholders and business leaders on trends, barriers, insights, and data-driven recommendations on new opportunities of online Direct-to-Patient market
- Tracked and evaluated key metrics of Chinese pharmacy market, and analyzed pharmacy market segmentation and conducted competitive and industrial analysis for business improvement
- Provided data visualization support for business strategy decision making by creating dynamic and interactive maps for pharmacy market segmentation and disease distribution

KPMG — **Beijing, China**
Strategy Analyst Intern — Aug 2020 – Oct 2020

- Participated project of BeiGene medicine brand and customer positioning, and design of Eli Lilly's Patient Assistance Programs (PAP), and delivered two 60+ slides presentation for business strategy
- Translated analytic insights into tangible and actionable solutions for client to by performing deep dive analysis of market status quo, competitor's marketplace and SWOT analysis
- Unlocked market growth opportunity for new product and identified target customer profile by analyzing the survey data, customer behavior, retention rate, drop-off reason and barriers

Inspur — **Jinan, China**
Data Analyst Intern — Jul 2019 – Sept 2019

- Acquires and compiles unstructured and structured data of top e-commerce platforms from multiple sources and verifies its quality, accuracy and reasonableness for the future analysis
- Analyzed the performance of key metrics of Chinese e-commerce platforms in SQL, and generated data-driven industrial reports
- Predicted the e-commerce industry market size and future trend by leveraging Time Series Model ARIMA
- Collaborated with fellows in the development and implementation of index model to better evaluate e-commerce penetration, growth, and economic influence in different regions

PROJECTS

Bank Customer Churn Prediction Project | Columbia University — Feb 2022 – May 2022

- Performed Exploratory Data Analysis (EDA) to discover the patterns and gather insights from dataset; cleaned data and recoded categorical variables by OneHotEncoder
- Predicted the customer churn status by training 3 Supervised Learning Models including Logistic Regression, K-Nearest Neighbor, Random Forest with GridSearchCV
- Evaluated the model performance by Confusion Matrix, ROC curve and AUC, achieved 90% accuracy, 88% AUC and 85% precision, and ranked the most important features, which suggested that age, salary, credit and balance are tightly related to the response variable

Customer Review Natural Language Processing Project | Columbia University — Feb 2022 – May 2022

- Preprocessed the review text dataset by removing missing values, removing unicode characters, removing stop words, and stemming; and extracted features by Term Frequency – Inverse Document Frequency (TF-IDF) for future analysis
- Performed K-means clustering and Latent Dirichlet Analysis for review segmentation to identify the review topics and provided market strategy recommendation to different clusters of customers

Lovelytics Customer Segment Project | Columbia University — Sep 2021 – Dec 2021

- Completed 90k+ data cleaning work and utilized EDA to explore the data distribution and visualize correlation between consuming behavior and different characteristics of customers
- Applied K-means clustering algorithm and elbow curve to identify key customer segments and visualized the resulting clusters with the help of PCA that shrinks the dimensions in Python – (Matplotlib; Seaborn); 4 clusters showed differences between customers' figures that lead to different purchase behaviors
- Developed a comprehensive customer segmentation and identified the target market for Epsilon, which suggested that Epsilon should direct its resources to married customers with children, with high income and high net worth, to maximize the profit and loyalty to brand