

## WENYUAN YANG

Address: 1200 W Huntington Dr, 91007, Arcadia, CA, U.S. Tel: 213-348-8786 Email: [wenyuany@usc.edu](mailto:wenyuany@usc.edu)

### EDUCATION EXPERIENCE

**University of Southern California, M.S., Econometrics and Quantitative Economics (Big Data Track)** Jan 2021 – Dec 2022

**GPA:** 3.88/4.0; **Core Courses:** Economic Applications of Machine Learning, Big Data Econometrics, Time Series Analysis

**Fudan University, B.S., Economics, Shanghai, China**

Sep 2016 - Jun 2020

**Core Courses:** Econometrics, Corporate Finance, Business Administration, Statistics, Game Theory, Python Programming

**Awards:** Outstanding Officer of the League Committee (Top 5%) ASEM-DUO Full Scholarship (Top 5%)

### WORKING EXPERIENCE

**KUAISHOU** (Short Video Platforms, 300 million DAU, SEHK:1024)

*Data Analysis Intern in Overseas Division*

Jun 2021 – Aug 2021

- Initiated data extraction by using **Advanced SQL Query** from 10+ datasets with massive daily operating data of short video platform and improved working efficiency by optimizing SQL commands
- Performed analysis to verify whether the promotion of video production tools has a positive impact on video creation by using key performance indicators, such as number of DAU, video production and video plays, etc., and built a **PowerBI dashboard** to present the findings

**Baidu** (580+million MAU, NASDAQ: BIDU)

*Strategy and Operation Intern, Intelligent Driving Department*

Dec 2020 – Apr 2021

- Analyzed development and market insights about autonomous driving and the smart transportation industry. Tracked the dynamics of key companies, including financing, strategic direction, operational data, etc.
- Assisted in carrying out industry chain research and scanning of investment targets, and conducted preliminary investigations and analysis on investment opportunities by evaluating the market size, and researched on the development of the business model by using **PEST analysis** and **market entry**

**Ping An Technology** | *Digital Platform Strategy Intern*

May 2020 – Nov 2020

- Evaluated customers satisfaction level with the insurance product and services through calculating their **Net Promoter Score** (NPS), and figured out main factors which significantly influenced their decisions in the customer journey
- Proposed several solutions such as launching new features and sending real-time notifications of the claim process. These product promotion measures increased the Net Promoter Score (NPS) from 77.9% to 83.5% for segment customer groups

### RESEARCH EXPERIENCE

**Seoul Bike Sharing Demand Prediction with Linear Regression and Ensemble methods**

Dec 2021 - Dec 2021

- Conducted research on predicting bike sharing demand in Seoul with weather conditions, bike count, and other attributes
- Cleaned raw data and visualized by histogram and heatmap to observe their distribution and correlation
- Processed and analyzed data with Python via **linear regression**, **Lasso**, **Decision Tree**, **Random Forest**, and **XGboost** models to predict demand for shared bicycles, where **XGboost** has the best prediction accuracy with the lowest MSE

### SELECTED PROJECTS

**Email delivery experiment on funding rate improvement**

Jun 2021 - Jul 2021

*Proposed an A/B test on 480000 users divided into 24 treatments groups to improve the funding rate*

- Calculated email open rate in treatment groups, visualized the results by heatmap with **Seaborn** to identify the most effective email and the best user segment group
- Designed an **A/B Test** on treatment groups and control groups via one-sample proportion **Z-test**, verified the statistical correlation between funding rates and email campaign strategy
- Performed **Time Series Analysis**, and offered the strategy to increase sending times to achieve a higher open rate

**Making predictions over amazon recommendation data set**

Jul 2021 - Aug 2021

*Predicted the recommendation's sentiment on Amazon's recommendation dataset via text analysis method*

- Processed text data using **Pandas**, **Numpy**, and **NLTK** (Natural Language ToolKit) to remove punctuations and affixes
- Converted text data into vectors with **n-gram**, **TF-IDF** method from scikit-learn packages
- Applied **Logistic Regression Learning** method and **Random Forest** method, both models achieved over 90% accuracy of measured by AUC on determining reviewer's satisfaction based on customers' text comments

### SKILLS

- Data Analysis Tools:** Python, SQL, R, Stata, Tableau, LaTeX
- Statistical Modeling:** Linear/Logistic Model, Model Selection & Regularization (LASSO, Ridge), Time Series Analysis
- Machine Learning:** Dimensionality Reduction & Feature Discovery (PCA); Classification (Logistic Regression, Decision Tree, Random Forest, K-Nearest-Neighbors, Support Vector Machine, Neural Network), Clustering