

ANQI XU

510-424-3509 | anqi_xu@berkeley.edu | [Linkedin](#) | [Github](#) | [Website](#)

Education

University of California, Berkeley | Master of Engineering | GPA: 3.8 /4.0 08/2022 – 05/2023

- ◆ Industrial Engineering & Operations Research
- ◆ Coursework: Principles and Techniques of Data Science, Optimization Analytics, Applications in Data Analysis

Dalian University of Technology | Bachelor of Science | GPA: 3.9 /4.0 09/2018 – 07/2022

- ◆ Information Management and Information Systems

Technical Skills

Core skills: Machine Learning, Data Mining, Statistical Analysis, Data Interpretation, A/B testing, Data Visualization.

Programming Languages: R(ggplot2), SQL, Python (Pandas, NumPy, Scikit-learn), C, Java

Software: MySQL, Power BI, Tableau, MS Office (advanced Excel, Word, PowerPoint, Outlook)

Project Experience

Analysis of TMDB Movie Dataset and Revenue Prediction | Big data analysis, Predictive modeling | [Github](#) 10/2022 – 12/2022

- ◆ Crawled **57k+** movie data through TMDB Open API, performed exploratory data analysis using **SQL queries** in **Python** in terms of the release year, genre, ratings, runtime, etc., and built comparative visualizations in a **Jupyter Notebook**.
- ◆ Conducted **feature engineering** and preprocessed text data to build **Bag-of-Words** using NLTK in **Python** for modeling.
- ◆ Applied multiple **machine learning models** to predict movie revenue and validated the result through **the bootstrap method**.

Research on User Download Behavior based on Google Play App Data | ML Classification, Text Mining 11/2020 – 12/2020

- ◆ Created a data preprocessing pipeline on Google app data, utilized multiple **machine learning classifiers** (Logistic Regression, SVM, Random Forest) in **Python** to predict application ratings, and **achieved an accuracy of 89.3%** with SVM.
- ◆ Performed polarity analysis on product review in **Python** and recommended targeted suggestions to app developers.

Internship Experience

Monetization Product intern | Hypothesis testing, Statistical analysis, SQL | Bytedance, Beijing, China 02/2022 – 06/2022

- ◆ Designed **5 dashboards** to track ad metrics and analyzed sales data of **300+ advertisers** to boost revenue.
- ◆ Conducted **A/B tests** using **SQL** to test the hypothesis that extending the duration of opening screen ads from 3s to 5s will improve the performance of ads metrics on Toutiao, resulting in a **28.8% increase** in click-through rate.
- ◆ Collaborated with the product team and design team to promote the **go-to-market process** of 3D Layer News Feed Ads.
- ◆ Collected feedback from salespeople and advertisers to provide actionable insights for feature development.

Research Intern | Regression Modeling, Timeseries Forecasting, Python | University of Victoria, Remote | [Github](#) 07/2021 – 10/2021

- ◆ Predicted COVID-19 cases, deaths, and recoveries in **Python** using **Linear Regression, Auto-regressive**, etc, compared forecast accuracy using RMSE, and visualized confidence intervals of results with **Python Plotly**.
- ◆ Generated interactive plots predicting Canada's cumulative cases with ipywidgets in a **Jupyter Notebook**.

Commercial Product intern | Data Analysis, Product Management, Tableau | Meituan, Shanghai, China 06/2021 – 09/2021

- ◆ Monitored daily sales of ads and produced **20+ dashboards** in **Tableau** for weekly team meetings.
- ◆ Performed abnormal data analysis using **Excel Pivot table** and **identify root causes** with the advertising platform department.
- ◆ Led the launch of the new advertising product, including pricing, market research, ad testing, and promotion plans.

User Growth intern, Lead | Data Preprocessing, Data Visualization, R | Tencent, Shenzhen, China 08/2020 – 12/2020

- ◆ Cleaned the data of newly joined content creators, tracked user growth metrics, and applied **R** to build time series plots.
- ◆ Optimized the registration process based on user data and **boosted the conversion rate by 4%** in a month.