

ALICE LI

Mountain View, CA | 818-606-2071 | yuanxilialice@gmail.com | www.linkedin.com/in/alice-li-yuanxi

SKILLS

Programming Languages: SQL, Python (NumPy, Pandas, SciPy, Scikit-learn)

Software: Advanced Excel (Solver, Pivot Table, V-lookup), Tableau, Spark, Jupyter Notebook, MySQL, Google Analytics

Techniques: Regression, Logistic, A/B Testing, Cluster, Classification, PCA, KNN, NLP, Decision Trees, RNN, CNN, Time Series

EDUCATION

Pepperdine University, Malibu, CA

Aug. 2018 – Aug. 2019

M.S. in Business Analytics

Relevant Courses: SQL and Database Management, Simulation & Optimization, Visualization, Machine Learning, Prescriptive Models, Data Mining, Exploratory Data Analysis

Binghamton University, SUNY, Binghamton, NY

Aug. 2013 – Jan. 2018

Bachelor's in Mathematical Science specialized in Statistics.

Relevant Courses: Probability and Statistics, R Programming, Linear Modeling

WORK EXPERIENCE

Data Analyst - Phoenix Satellite TV (U.S.) Inc., Irwindale, CA

Sep. 2019 – Present

- Liaised with 3 business unit stakeholders to discover & define key metrics for data-driven measurement of business performance & decision-making & incorporated the metrics in existing data monitoring framework.
- Tracked & analyzed 100+ KPIs for website, mobile apps & social media pages in interactive **Tableau** dashboards.
- Preprocessed web traffic data in **SQL** & built a logistic regression in **Python** to predict campaign success rate with 74.9% accuracy.
- Discovered 5 factors that most impact campaign success, summarized model output & insights in a **Jupyter** notebook & shared the findings with the marketing team.
- Designed & conducted 15+ **A/B tests**, digitized & automated product research and supported Product & UX teams in discovery & analysis of features for a new digital product offering, reduced offering development time by ~30%.
- Conducted and presented deep dive analysis for all levels of the organization to deliver actionable recommendations resulting in a conversion rate increase of 16%

Marketing Analyst Intern - Cinema Libre Studio, Burbank, CA

May 2019 – Aug. 2019

- Designed and implemented 50+ marketing campaigns in **Google Analytics & Google Ads**, applied regression & adjusted the keywords, bidding, and segmentation strategy based on A/B tests and improved the overall conversion rate by 19%
- Collaborated with the design team & provided data-driven insights for re-design & overhaul of Ad texts and images based on CTR & CPI rates, with a potential boost in CTR of 20+%.
- Performed customer segmentation using **k-means clustering** in **R** based on 25 key metrics generated from the customer profile data available in company **SQL** database & showcased the findings with the team in **Google Data Studio**.
- Built automated email programs for 11 customer segments, incorporated **A/B testing** for subject line, content, CTAs, and email cadence within each group, implemented the findings & boosted email campaign conversion rate by 23%.

PROJECTS

Bank Customer Churn Prediction and Analysis

Aug. 2020

- Cleansed & processed 2M+ labeled records for bank customer communication & transactions using **Spark**, performed EDA to visualize impact of features, transformed 12 categorical features and standardized all the relevant features.
- Developed Random Forest, Logistic Regression & **SVM** models in **PySpark** and predicted customer churn probability, evaluated the models using a confusion matrix, selected **Logistic Regression** with an accuracy of 89.2%.
- Applied regularization & 5-fold cross validation and boosted model performance by 14%, determined feature importance and isolated top 3 factors that affected the customer churn & devised corresponding churn mitigation strategies.

Prime Video View Time Exploration and Prediction

Jan. 2020

- Analyzed Amazon prime data for 7K+ unique movies along with number of unique views, total view time, IMDb ratings, duration, language, maturity rating & a plot summary and explored video view time variation with each parameter.
- Preprocessed raw plot summary text data using **NLP** & generated relevant top 500 tokens, treated missing values & outliers, encoded & scaled categorical features in **Python**.
- Constructed Linear Regression & **Random Forest** models to forecast the total view time for each movie, optimized model parameters and selected Linear Regression with a lower RMSE of 0.096.
- Fine-tuned the **Linear Regression** model with stepwise regression & reduced the model parameters from 14 to 8 using AIC & quantified the impact of each factor on the view time performance for Amazon Prime.

House Prices - Advanced Regression Techniques

Dec. 2019

- Extracted 4598 housing records with 81 features, imputed Missing values, executed feature engineering, and explored the distribution of sale prices with the various features under a log transformation.
- Trained & tuned an **XGBoost** model with 7-fold cross validation & identified features that impact house prices with 91.77% accuracy.
- Conducted **PCA** and reduced model input parameters to 4 Principal components, refit the model & improved accuracy by ~3%.