Amanda Tran

atran163@calpoly.edu | (408) 335-9997
linkedin.com/in/amanda-tran | amanda-tran.com | github.com/atran163

EDUCATION

Bachelor of Science Statistics, Data Science Minor June 2021
California Polytechnic State University, San Luis Obispo

TECHNICAL SKILLS

- **Big Data Tools:** Pandas, NumPy, Scikit-Learn, TensorFlow, Airflow, Spark ML
- **Languages:** R, Python, Java, Javascript, SQL, PostgreSQL, SAS, HTML
- **Visualization:** Tableau, Matplotlib, Seaborn, Plotly, PowerBI, Looker, MS Excel
- **Analysis:** Regression, A/B Testing, ANOVA, Cluster Analysis, Classification, EDA

PROFESSIONAL EXPERIENCE

Data Analyst Researcher – *Cal Poly Corporation* Present

- Collaborated with Engineering Team to implement A/B experiments on different freeway designs
- Automated data processes with advanced R scripts to increase efficiency by 90% for dashboards
- Translated statistical analysis to inform team on the effectiveness of their freeway implementations
- Communicated weekly transportation research results involving bike lane and sidewalk accessibility

Cloud Computing Apprentice – *Under Data Scientist: Sara Kimmich* June 2020 – Sept 2020

- Developed data analysis presentations showcasing the basics of Airflow, Spark ML, PostgreSQL
- Managed an average of 4 concurrent data science projects to teach students of a UK-based nonprofit
- Implemented project management strategies for clients, reducing fulfillment hours by 75%
- Designed data visualizations and dashboards on open-source data using Seaborn and Tableau

Corporate Development Intern – *SMART Modular Technologies* June 2019 – Sept 2019

- Analyzed competitor's KPIs in the Specialty Memory industry for financial analysis reports
- Conducted ad hoc analyses of competitor strategies, business models, and marketing techniques
- Communicated post analysis results to VP's and Leadership to ensure alignment in business initiatives
- Evaluated potential impact of marketing strategies through A/B Testing

PROJECTS

Using NLP and Linear Classification to Classify Movie Reviews – *Data Classification* November 2020

- Automated classifications of 50,000 movie reviews as positive or negative using panda's NLTK library
- Built efficient logistic regression models and tuned model to achieve 82% overall accuracy
- Enhanced developer efficiency 25% by preprocessing dataset to deal with big data ramifications

Visualizing Music Qualities vis Spotify API – *Data Visualization* October 2020

- Fetched audio features for 400 most played songs over 4 decades to compare to Billboard top songs
- Created seamless visualizations in Tableau and Plotly showcasing the distinction between the two
- Performed PCA to reduce dimensionality of songs to cluster songs into 5 different emotions

Predicting Airbnb Prices and Neighborhood Boundaries – *Machine Learning* March 2020

- Utilized Airbnb's API to extract 50,000 Airbnb homes in New York City to create interactive visualizations
- Minimized model validation error by testing various models to predict the price of an Airbnb
- Used Random Forest Classifier to predict a listing's neighborhood with 92% accuracy

ACHIEVEMENTS AND LEADERSHIP

SLO Hacks – *Sponsorship and Lead Director* April 2019 – Present

- Fostered a collaborative team environment to secure monetary and in-kind sponsorship for main event
- Raised over $50,000 in swag, prizes, and food to give away to our hackathon winners and attendees