

JOSEPH TRAN

✉ tranjoseph11@gmail.com ☎ (484) 538-5604 📍 Philadelphia, PA in joseph-l-tran 🌐 JLtran11

SUMMARY

Process oriented data analyst well-versed in cleaning, interpreting, and analyzing data to deliver valuable insights via data-driven methods. Experienced in using pandas, stats_models, glob, matplotlib, and seaborn libraries to produce summaries and reports that support public health policy recommendations.

SKILLS

PYTHON: Pandas, Scikit-learn, Numpy, Seaborn, Stats_models

DATA SCIENCE: Classification, regression, clustering, feature engineering/selection, data wrangling, data cleaning, time series analysis, Pyspark, Keras

DATA VISUALIZATION: Tableau, Matplotlib, Power BI, Plotly

SQL: Joining, funnel analysis, sub-queries, temporary tables

EXCEL: pivot tables, v-lookup, formulas/functions

STATISTICAL ANALYSIS: A/B testing, ANOVA, chi-squared, t-test, ARIMA, hypothesis testing

PROJECTS

Music Recommender

Mar. 2020 - Aug. 2020

- Sourced #nowplaying-rs dataset of 17m listening events to design a music recommender system
- Data was cleaned; three separate files representing listening event log, associated hashtags, and sentiment scores were combined into one data frame
- Matrix factorization algorithms from ALS and lightfm are used to find latent features of the user/item matrix to make user predictions
- Precision @ k scores are compared between models with lightfm and ALS model scoring 0.125 and 0.068, respectively

Loan Default Predictor

Sept. 2020 - Feb. 2021

- Utilized Imperial College of London's loan default dataset of 150k samples and 755 features to develop loan prediction model
- Filter and ensemble feature selection methods are used in addition to dimensionality reduction to find an optimal set of features for the model
- Various classification algorithms are tested to find the optimal model and parameters via random_searchCV.
- Best model was found to be XGBoost with SMOTE over-sampling yielding 0.677, 0.635 train and test ROC scores

EXPERIENCE

Springboard, Data Science Fellow, Philadelphia, PA

Feb. 2020 - Feb. 2021

- Leveraged bootstrap sampling to make inferences on populations
- Constructed SQL queries to track conversion rates across multiple marketing channels and websites
- Built ETL pipeline to clean and transform data from multiple sources
- Performed model selection with random search inside scikit-learn pipeline to find best model

KH Global Health Consulting, Data Science Fellow, Remote

Apr. 2020 - Oct. 2020

- Provided data cleaning, management, and reporting services for health systems geographic data from Sierra Leone
- Led development of automated data file ingestion with python and glob; reduced load time from 1 week to several minutes
- Created and maintained Sankey diagrams to track importation and distribution of pharmaceutical products
- Ad hoc statistical analysis on health system data

Freelance Consultant, Project Manager, Lancaster, PA

Jan. 2018 - Jan. 2019

- Coordinated with stakeholders to remodel and relaunch independent motel into a franchise destination
- Negotiated deadlines and budgets with local property developers and tradesmen
- Synchronized expectations of all involved parties with weekly budget updates and progress tracking with Gantt chart
- Project completed two weeks ahead of schedule, 6% under budget

Robson Forensic, Mechanical Engineer Intern, Lancaster, PA

May 2016 - Aug. 2016

- Created finite element model to study stress and deflection as a result of wind interactions; hand calculations within 7 % of computer model
- Responsible for design calculations and product development for intellectual property claims
- Produced mechanical drawings relating to intellectual property claims
- Assisted with evidence collection, documentation relating to forensic engineering field

EDUCATION

SUNY Maritime

Bachelor's of Engineering Mechanical Engineering

Lehigh Carbon Community College

Associate of Science Engineering