# EDUCATION AND HONORS

jacobnlinao@gmail.com

[linkedin.com/in/jacobnicholasl](http://linkedin.com/in/jacobnicholasl)

Portfolio: [jacoblinao.netlify.app](https://jacoblinao.netlify.app/)

714.353.6338

JACOB LINAO

**University of California, Irvine** Master of Science – Business Analytics (MSBA), Concentration in Data Science May 2022

* Merage Scholarship Recipient (Top 20%), GPA: 3.84/4.00
* Courses: Data Science Statistics, Adv. Machine Leaning, Adv. Predictive Analytics and Forecasting, Marketing Analytics, Big Data and Cloud Computing, Natural Language Processing, Customer and Social Analytics, Management Analytics

**California State Polytechnic University, Pomona** Bachelor of Science -Information Systems & Business Intelligence (B.S.) May 2021

* Dean’s List (x3), President’s List (x2); ACT Score: 32
* Director of Data for Digital Analytics Club, Head Teaching Assistant for College of Business, VP of Academics for Zeta Phi Rho

# SKILLS

**Programming Languages:** Python (Scikit-Learn, NumPy, Pandas, Matplotlib, TensorFlow), SQL, PL/SQL, R, Git, Linux, AMPL

**Data Tools:** Pyspark, AWS EMR + S3 + Athena, SSMS, Oracle DB, NoSQL, DynamoDB, Tableau, Office, Power BI, WEKA, Alteryx

# PROFESSIONAL EXPERIENCE

**Data Scientist Intern** Diamond Bar, CA

Niagara Bottling May 2020 - Aug 2021

* Developed and deployed an **NLP** Random Forrest model in **Python** as a user facing web application in **Streamlit** to resolve a department wide issue, reducing ticket assignment error by 18% saving est. $25,000/YR in resource allocation
* Led 8 **SQL** projects by designing functions, triggers, procedures, and advanced ad-hoc queries to achieve several department results such as a 91% cycle time decrease, fully eliminating data capture error for 2 product lines, and automating 6 daily reports
* Increased operational efficiency by 50% through **EDA** in Python and **PowerBI**, enabling team visibility of historical trends and suggesting actionable recommendations to mitigate system errors
* Drove bi-weekly team meetings for 10 analysts and management over KPI reports, ticket trends, and data-driven suggestions

**Data Analyst Intern** Burbank, CA

NBCUniversalSep 2020 - Dec 2020

* Selected as the sole intern to lead the data migration in a global ERP integration of SAP HANA, responsible for all data tasks from extraction, aggregation, analysis of data, and liaising insights to an international team of directors
* Cleaned over 10 million rows of SAP records using **Python**, distilling down to 4 **KPIs** and designed 20+ **Tableau** visualizations to assist several departments with their operational tracking needs and drill down analysis
* Designed an efficient **ETL** package to pull data from 3 disparate sources, staging into a consolidated table in Oracle, and connecting to web application servers -- providing a department dashboard of comparison trends and data discrepancies

# PROJECTS AND RESEARCH

**Hackathon Data Scientist**Gainesville, FL

IBM Aug 2021 - Present

* Utilized time-series algorithms including **ARIMA** and **Holt’s Method** to assess seasonality in household energy usage trends
* Built end-to-end data pipeline of data collection via web scraping, feature engineering, model building, and model ensembling

**Data Science Research Lead** ([*bit.ly/3CBlD0L*](https://bit.ly/3CBlD0L))Remote

Hawaii International Conference on System SciencesMay 2021 – Aug 2021

* Won Best Project - Created a **TFIDF** Logistic Regression model for segmenting customer sentiments and utilized **LSA** for topic modeling to find 3 positive/negative areas of business focus to combat the ill effects of Covid on LA restaurants

**Data Analytics Research Assistant + Hackathon Finalist** ([*bit.ly/3Bykl5f*](https://bit.ly/3Bykl5f))Pomona, CA

Center for Customer Insights and Digital MarketingJan 2020 - Mar 2020

* Led a team to place runners-up in GFK’s data science hackathon by building a product and presenting insights to VP of Marketing
* Conducted a successful **A/B test** to measure school website efficacy, resulting in the addition of a profitable UI feature

[Starbucks Survey KNN](https://github.com/jnlinao/Starbucks--Consumer_Clustering) – Silhouette method to find optimal k, **cluster** **analysis**, **lin-reg** on clusters to predict customer satisfaction

[AWS Forecasting](https://github.com/jnlinao/PriceDiamonds--Pyspark_Classification) – **Big data** wrangling on RDDs, **Gradient Boosted Trees**, grid-search parameter tuning to predict diamond prices

[Real Madrid Case Study](https://drive.google.com/file/d/1L9uUP_vTFSbXGMR5FXOIEKkrlBNeE0av/view) – Used data, financial, and marketing analysis and models to draft a data driven solution plan for soccer team