

Joseph Lin

San Francisco, CA • thejosephlin93@gmail.com • (408) 646-3982 • [joseph-l.in](https://www.linkedin.com/in/josephlin11) • [in/josephlin11](https://www.linkedin.com/in/josephlin11)

Multi-faceted Data Scientist with over four years of hands-on experience that includes data analysis, engineering, visualization, automation, and tool development to scale data infrastructure and empower well-informed decision making within data-driven organizations.

EXPERIENCE

Data Scientist, Doximity - San Francisco, CA

Nov 2018 - present

Doximity is the largest network of medical professionals in the US.

- Core member of ads data science team responsible for all analytics, prioritization, and predictive modeling for Doximity ads, their leading revenue generator. Utilizing an Agent Based Model approach, I use simulated data and decision science to assist data-driven product decisions.
- Key contributor to consolidation of company dimensional model in Snowflake (dimensional, fact, report tables). Engineered ETL's (PySpark, Pandas, SQLAlchemy, SQL) deployed on DAGs in Airflow. Built dashboards, performed ad hoc ad data analysis (i.e. targeting, engagement, performance metrics), created supply and demand forecasts, and established Looker infrastructure.
- Created 12 health-checks for the ads ecosystem to alert stakeholders of abnormal system behavior. Developed framework and repository to centralize company data health-checks.

Marketing Data Analyst, MZ - Palo Alto, CA

Mar 2018 - Sept 2018

MZ is a global leader in mobile gaming with titles that include Game of War, Mobile Strike, and Final Fantasy XV: A New Empire.

- Overhauled data pipelines for Cognant (MZ's performance marketing agency) by automating data cleaning and analysis at multiple granularities using Python (Pandas, NumPy, Scikit-learn) and SQL.
- Supported research science, marketing analytics, data platform, and executive team initiatives ranging from fraud detection improvements, modelling partner KPI predictors, revamping data pipelines and dependency configurations to meet reporting SLA's, and executing ad hoc deep-dive requests for the CRO.
- Configured client budgets and contributed to Holt-Winter/ARIMA libraries which forecast KPI's for use on workbooks and dashboards across analytics teams.
- Deployed alerting tool which integrated Google, Slack, MySQL, Vertica API's to share projections with leadership and identify inconsistencies/gaps with data platform team.

Business Intelligence Analyst, Genentech - South San Francisco, CA

July 2015 - Feb 2018

Genentech is a leading biotechnology company that discovers, develops, manufactures, and commercializes medicines.

- Contributed to a cost reduction initiative and led development of data visualizations for 27 cost centers to monitor \$14M materials budget, leading to \$1.2M in first year savings. Implemented an outlier classification system to notify management of spending anomalies.
- Led development of proprietary scheduling software that coordinated 1900+ production runs among 50+ researchers and reduced scheduling logistics time 70%.

SKILLS

Databases: Snowflake, AWS S3, Vertica, Oracle 11g, MySQL

Data Analysis: Apache Spark, Python (NumPy, Pandas, Scikit-learn, SQL Alchemy), PostgreSQL, Excel, R (dplyr, tidyr)

Visualization: Looker, Tableau, Superset, D3.js, Seaborn, matplotlib, plot.ly, bokeh

Other Languages: JavaScript, Google Script, Shell, HTML

Concepts/Tools: Dimensional Modelling, Outlier Detection, Forecasting, Performance Marketing, AdNetworks (SSP, DSP, SANs), Bidding, Logistics and Supply Chain, Git, Airflow, Data Structures, Fluent in Chinese and English translation

EDUCATION & CERTIFICATIONS

B.S. Biochemistry/Cell Biology, University of California, San Diego - La Jolla, CA (2015)

Coursera - Johns Hopkins Data Science Specialization (Courses included R Programming, Getting and Cleaning Data, Regression Models, Reproducible Research, Practical Machine Learning, Exploratory Data Analysis, Statistical Inference, Developing Data Products)