# Bea Li

shell0081@gmail.com github.com/bli36 my portfolio

## Experience

**Data Engineer** **May 2018 – current**

*San Francisco International Airport - Information Technology Department* *South San Francisco, CA*

- Ingested and integrated petabyte-scale datasets from diverse internal and external APIs into **Snowflake** and Relational Database, **PostgresSQL**, supporting both machine learning models and operational reporting dashboards for airport-wide decision-making.
- Designed, implemented, and maintained scalable real-time and batch data pipelines using **AWS**, **DBT**, and **Snowflake**, ensuring consistent, high-quality data delivery to analytics and ML workflows.
- Created optimized data models leveraging star schemas, snowflake schemas, and normalization techniques, improving query performance and data accuracy for cross-department analytics by 25%.
- Developed interactive dashboards in JavaScript and ReactJS to visualize operational KPIs, reducing report turnaround time for executives and managers.
- Automated ingestion and processing workflows using **AWS EventBridge**, **Snowflake Snowpipe** and CI/CD pipelines (Jenkins/GitHub Actions), reducing manual intervention and deployment time by 40%.
- Partnered with data scientists, product managers, and business stakeholders to refine data requirements, troubleshoot data quality issues, and design scalable solutions.
- Mentored intern engineers, conducting code reviews, providing technical guidance, and leading design sessions to promote best practices in ETL design and data modeling.

**Data Engineer Intern** **Jan 2018 – May 2018**

*IPMD* *Berkeley, CA*

- ETL gigabytes of facial expression images, ingested images via Python scripts, stored in Google Cloud Storage, performing emotion recognition using **TensorFlow**.

## Selected Projects

**Airfield Incident Data Platform**

- Designed and implemented an end-to-end **ETL** pipeline to ingest, parse, and transform high-volume, deeply nested semi-structured airfield incident data from Veoci APIs. Leveraged **DBT** for complex relational modeling between entities such as involved persons, responsible companies, facilities, aircraft, and runways.
- Optimized data transformations and loaded cleaned, analysis-ready tables and views into Snowflake, enabling the Airfield Data Platform to support real-time analytics and cross-team reporting. Reduced downstream query times by 25% through schema optimization.
- Built a text-analytics proof of concept using **Snowflake Cortex Analyst** to automatically summarize first-responder notes, improving incident reporting efficiency for airport operations and reducing manual review time by 30%.

**Airport Operation Dashboard**

- Engineered an automated real-time traffic data ingestion pipeline to capture and process live Bay Area traffic and incident data from the INRIX API using AWS Lambda for serverless execution. Ensured low-latency processing for near-real-time operational monitoring.
- Designed and implemented robust ETL workflows in **Snowflake**, integrating multiple data sources into unified, analytics-ready datasets. Applied data quality checks and schema design principles to improve query performance by 20% for downstream analytics.
- Developed analytical views to monitor travel times, congestion trends, and incidents on key access routes to and from SFO, enabling airport operations to proactively manage traffic flow during peak travel hours.
- Created a segment range-matching algorithm to accurately align SFO campus geospatial segments with INRIX traffic segments, increasing data mapping accuracy from 85% to 98% and improving the reliability of location-based traffic insights.

**Virtual Taxi Queue**

- Conducted **Exploratory Data Analysis** in **Python** to identify gaps and anomalies in time series data, facilitating predictive modeling and feature selecting.
- Designed and implemented crucial machine learning data transformation steps: **one-hot encoding**, **k-fold cross-validation** and more, utilizing **Python** and the **TensorFlow** API. Achieved a notable improvement in overall test accuracy results, with an increase of 10% to 15% during model training.
- Optimized the model training process and documented each session using **MLflow**, enhancing performance monitoring efficiency. Additionally, developed a web app in **ReactJS** for streamlined access and management.

**TNC Dashboard and Monitoring System**

- Designed and implemented a **ETL** process for processing the driving records of billions of TNC vehicles, resulting in a 30% improvement in query efficiency using index keys for BI team dashboards.
- Built cost-efficient, high-performance **PostgreSQL** views to streamline reporting and reduce compute costs, enabling analysts to access key metrics without complex queries.
- Enhanced a **NodeJS** application utilizing TNC vehicle data on **AWS Lambda** and **API Gateway**, ensuring the sustained stability of the component responsible for estimating ground speed within the mobile app.
- Developed an automated daily data-quality alert system using Python and AWS Lambda to detect statistical anomalies in ingestion metrics, proactively notifying stakeholders and enabling same-day remediation of potential data integrity issues.

**Airport Concession Metrics**

- Conducted comprehensive sentimental analysis on reviews of airport concessions, incorporating aspect tagging to derive insights into customer opinions using Python **NLTK**.
- Developed dashboards using **d3** and web applications with **NodeJS** to visualize key findings from the reviews.
- Presented findings to management and contributed insights on concession performance.

## Technical Skills

**Data Engineering**: Python, SQL, AWS, Snowflake, Airflow, TensorFlow
**Deployment**: Git, Kubernetes, Docker, Jenkins

## Education

**University of California, Berkeley** **August 2015 – May 2018**
*Bachelor Degree in Applied Mathematics, minor in Computer Science* *Berkeley, CA*

**Mt. San Antonio College** **August 2012 – May 2015**
*General education* *Walnut, CA*

## Extracurricular

**Toastmaster International - Plane Talk** **August 2022 – Present**
*Webmaster, Active member* *SFO commission*

- Act as a webmaster for club members, resolving technical issues related to agenda generation, role assignments, and organizing members' pathways on the club's webpage.
- Attend Bi-weekly meeting as an active member and speaker, growing on presentation mastery pathway while sharing my life and passion.

**Cal Hacks** **October 2025**
*Volunteer* *UC Berkeley*

- Act as a volunteer to coordinate the event in stadium, assigning food and tickets to hackers..