ALYCE FAN

(213)-234-8789 | alycefyq@gmail.com | https://www.alycefan.com

EDUCATION

University of Southern California   **Los Angeles, CA**
Master of Science in Analytics, GPA: 3.84/4.00   August 2019 – May 2021
Coursework: Predictive Analytics, Data Visualization, Data Analytics Consulting, Database Management, Applied Statistics, Optimization

SKILLS AND CERTIFICATIONS

- **Programming Languages:** *Python, PySpark, R, SQL (SQL Server Management Studio), JavaScript.*
- **Tool Kits:** *GitLab/GitHub.*
- **Other:** *Google Analytics, Google BigQuery, Power BI, Tableau, Figma, Microsoft Excel.*
- **Certifications:** IBM Applied Data Science Specialization (Issued June 2020)

PROFESSIONAL EXPERIENCE

KRE Consulting (Client: Southern California Gas Company)   **Los Angeles, CA**
Data Scientist   July 2021 – Present
- Executed efficient queries, leveraged techniques such as partitioning and non-cluster indexing to optimize data storage and improve query runtime by **60%**.
- Constructed a *Grafana* dashboard to monitor database space usage, developed Key Performance Indicators to assure data quality, set up automatic email notification alerts. Reduced **2 hours** manual work daily.
- Created an interactive data-oriented web application using *Node.js* and *React.js*, assisted business team to detect leaks.

KRE Consulting (Client: Southern California Gas Company)   **Los Angeles, CA**
Analytics Associate Intern   September 2020 – June 2021
- Developed streamlined ETL pipelines via *PySpark* for large-scale data manipulation tasks, such as processing data across multiple databases and writing records to one table in million scale, reduced runtime from **2 hours** to **8 minutes**.
- Implemented K-Means and KNN to classify 6 million customers into 20 defined clusters based on consumption pattern.
- Built a *Splunk* dashboard to monitor the health of Spark jobs, visualizing execution time and data throughput for each job with key features such as time window pick and developing environments change.
- Worked with business team, implemented a dashboard by utilizing Power BI to visualize gas consumption anomalies.
- Managed company data by designing schema and migrating across SQL server databases.
- Performed code review through Git and implemented automatic code quality check through gitlab-ci.yml.

University of Southern California Integrated Media Systems Center   **Los Angeles, CA**
Research Assistant Directed by Dr. Luciano Nocera   January 2021 – May 2021
- Developed a dashboard in *Vue.js* for freight modeling research that displays sensors, truck states, and traffic.
- Visualized modeling results with heatmap, *Mapbox*, bootstrap table to improve client understanding.

Jaguar Land Rover   **Shanghai, China**
Digital Marketing Analyst Intern   March 2019 – June 2019
- Drafted Key Performance Indicators for conversion funnel and consumer profiles to track customer behavior.
- Designed the Brand App, modified user story descriptions based on data analytics, across multiple departments.
- Reported selling results monthly via Google Analytics and assisted the senior manager in adjusting market strategies.

GRADUATE STUDENT PROJECTS

Efficiency Problem Solution for RIOgaleao Airport | *Python*   February 2021 – May 2021
- Separated passengers and staff with statistical analysis based on Mac addresses data counts of connection times.
- Visualized passenger and staff hourly distribution patterns separately via interactive heat map and gridding technology.
- Performed time series analysis using ARIMA model to predict human traffic, helped staff be prepared in advance.
- Designed an application prototype using *Figma* for the airport to monitor people distribution in each grid.

Storytelling for Netflix | *JavaScript*   August 2020 – December 2020
- Performed exploratory data analysis with a *Kaggle* dataset of Netflix content and subscriptions using *Python*.
- Designed a website using Vue.js framework, containing responsive and interactive D3 Charts, a D3 map, and a *Mapbox* map to visualize Netflix producers and audiences, giving Netflix actionable insights for the future development.

Chinese Restaurant Open in Toronto | *Python*   May 2020 – June 2020
- Scraped demographical data from Wikipedia and geographical data from Foursquare API.
- Provided insights through interactive data visualization with Folium, Seaborn and Matplotlib.
- Implemented K-means clustering to classify restaurants into 6 clusters based on demographical and geographical patterns.