

KHANG LE (KAYLA) TAN

khangletan@hotmail.com | 515-868-1829 | [linkedin.com/in/khangletan](https://www.linkedin.com/in/khangletan) | khangletan.github.io

EXPERIENCE

REVOLOGY

Data Scientist

Iowa City, IA

04/2023 – Present

- Led end-to-end development and deployment of a machine learning model to predict healthcare claims recovery propensity using Python to optimize the work strategy
- Developed an anomaly detection system using Z-score and Isolation Forest algorithms using Python to proactively surface irregularities to business leaders
- Implemented a HIPAA-compliant data de-identification pipeline to anonymize PHI/PII, enabling safe sharing of real-world datasets for internal testing and client demonstrations
- Engineered and maintained multiple AWS Glue ETL pipelines using PySpark and SQL to consolidate client data from distributed S3 buckets across AWS accounts, supporting centralized analytics and efficient reporting
- Designed and implemented interactive Amazon QuickSight dashboards for multiple clients to support revenue cycle management, financial reporting, and operational efficiency
- Delivered monthly report with actionable insights to multiple clients, supporting data-driven decisions on financial performance

THE UNIVERSITY OF IOWA, TIPPIE COLLEGE OF BUSINESS

Iowa City, IA

Tippie Analytics Cooperative Intern - Team Leader

05/2022 – 07/2022

- Led a team of 11 and supervised them on projects while collaborating with 3 non-profit business partners
- Established ETL process by utilizing scalable SQL and R to ensure data quality and accuracy
- Unified CRM in Google Sheets approximately 6,000 customers and tracked ticket sales and donation activities
- Developed financial analysis using Tableau dashboards to evaluate YoY growth rate to support global overview and business strategies
- Designed cost-benefit analysis visualization to determine optimal solution, projected cost savings to reduce time consumption and improve future process efficiency
- Created R scripts and deployed interactive web application using Shiny App to promote automation for updated procedures for business partners to focus on reducing manual intervention
- Structured, engineered and launched a new inventory database in REDCap and automated workflow processes for product distribution
- Provided inventory analysis to reduce obsolete inventory, optimize inventory turnover rate, and forecast demand

HUAWEI TECHNOLOGIES

Petaling Jaya, Selangor

Department Secretary

01/2019 – 07/2021

- Led and collaborated global cross-functional projects with stakeholders across the company to improve operations and processes
- Processed and maintained massive data involving more than 140 countries using Advanced Excel (Index Match, VLOOKUP/HLOOKUP, Pivot Table, VBA)
- Conducted statistical analysis and established dashboards for internal stakeholders to track project progression
- Performed routine inspections and quality analysis to ensure business standards and safety regulations
- Created training materials and operating manuals with recommendations for future improvement

Payroll Analyst

12/2017 – 12/2018

- Compiled and processed payroll for employees in various countries (Taiwan, Korea, Zimbabwe, Kuwait) using Oracle, Excel and multiple internal software systems
 - Performed data analysis on payroll transactions to identify errors and drive compliance towards payroll guidelines
 - Participated in automation project and performed UAT testing to improve internal process quality and reduce manual intervention
-

PROJECTS

The Power of International Education - IENetwork Membership Price Analysis

- Conducted trend analysis on approximately 2,000 members, researched competitors and identified target market
- Provided insights and recommendations on potential pricing models to help optimize membership and leveraged members retention using RStudio and Shiny App

Itaewon Stampede Incident Analysis

- Collected over 72,000 tweets using Twitter API v2 in Jupyter Notebook utilizing Python programming language and performed text analysis with NLP and machine learning to extract valuable insights from unstructured data
- Implemented K-means & hierarchical clustering algorithms and LDA topic modeling to study topics of discussion surrounding the incident

Product Insights Analysis

- Performed large-scale web scraping in Jupyter Notebook using Python programming language to extract approximately 7,000 web articles and developed sentiment analysis to predict the sentiment of new sentences using classification algorithms
- Conducted K-means and hierarchical clustering algorithms to identify potential terms in a similar group and developed LDA topic modeling to map out wording that mostly captured in each topic

Bank Marketing Analysis

- Conducted ETL process on over 42,000 customers in Databricks server using HiveQL and Spark SQL
- Produced data analysis, data modeling, and data visualization to study client behavior and predict client subscription churn rate

All-NBA Team Analysis

- Built a Gradient Boosting Machine model with over 18,000 players in RStudio using R programming language to calculate the rating for NBA players within the last 40 years
- Utilized nbastaR package to directly scrape data from basketball-reference to predict which players will be selected to the All-NBA teams in 2022

Diabetes Analysis

- Utilized open-source machine learning and data visualization platform – Orange Data Mining Software
- Processed over 254,000 patients with CRISP-DM methodology to identify risk factors and provided recommendations on reductions of diabetes

Annual Compensation Analysis

- Extracted, transformed, and load over 63,000 compensation data using Microsoft Excel and performed statistical data analysis in JMP Statistical Software
- Successfully developed a STEM salary predictor model that determined the years of experience highly influence the annual compensation along with other variables like city, company, education, and ethnicity

SKILLS & CERTIFICATES

- **Certification:** Microsoft Certified Azure Fundamentals | Credential ID: 994321713
 - **Programming Languages:** Python (Matplotlib, NLTK, NumPy, Pandas, scikit-learn), R (dplyr, ggplot2, readr, shiny, tidyverse), SAS, SQL
 - **Application:** Alteryx, Excel, GitHub, JMP Statistical Software, Jupyter Notebook, Oracle, Orange Data Mining Software, Power BI, Tableau, Visual Studio Code
 - **Big Data:** API, BigQuery, Databricks, Hadoop, Hive, Microsoft SQL Server, Spark
 - **Cloud Computing:** AWS (S3, Glue, Lambda, SageMaker, Athena, RDS, Step Functions, CloudShell, Secrets Manager, QuickSight), Google Cloud Platform, Linux, Microsoft Azure, Windows
 - **Agile:** Confluence, Jira, Monday.com, SCRUM, Trello
 - **Methodologies:** Big Data Analytics, Cloud Computing, Cluster Analysis, Data Analysis, Data Integration, Data Visualization, ETL Process, Forecasting, Machine Learning, Natural Language Processing, Optimization Analysis, Predictive Modeling, Project Management, Simulation Analysis, Statistical Analysis
 - **Language Proficiencies:** English, Mandarin, Cantonese, Hokkien, Malay
-

EDUCATION

THE UNIVERSITY OF IOWA	Iowa City, IA
<u>Master of Science, Business Analytics STEM Designation</u>	12/2022
GPA: 3.9/4.0	
UPPER IOWA UNIVERSITY	Fayette, IA
<u>Bachelor of Science, Psychology and Communication</u>	10/2016

ADDITIONAL INFORMATION

<u>Technology Association of Iowa</u>	Des Moines, IA
• Iowa Tech Connect Class of 2023	12/2023
<u>The University of Iowa</u>	Iowa City, IA
• Business Analytics Department, Graduate Teaching Assistant	08/2022 – 12/2022
• Warren and Maryellen Boe Summer Research Fellowship, Graduate Research Assistant	05/2022 – 07/2022
• Iowa Business Analytics Case Competition, Student Chair	01/2022 – 04/2022