

(626) 782-2096
Costa Mesa, CA
tsoi.kev@gmail.com

Kevin Tsoi

katsoi.github.io

GitHub: katsoi
LinkedIn: tsoiso

SKILLS

Tools and Languages	Python, SQL (HiveQL, T-SQL), Tableau, Git, ArcGIS, Excel
Libraries	Pandas, PySpark, Numpy, Seaborn, Scikit-learn, Jupyter, ArcPy, Re
Technical Skills	ETL, Data Modeling (Regression, Classification, Clustering), Statistics, Experimental design, Hypothesis testing

EDUCATION

M.S. Mathematics	Aug 2020 — Dec 2022
<i>Cal Poly Pomona</i>	<i>Pomona, CA</i>
<i>Thesis: Clustering and Recommendation Systems for Instacart Delivery Orders</i>	

EXPERIENCE

Data Analyst - Underwriting	June 2023 — Present
<i>AAA Auto Club Enterprises</i>	<i>Costa Mesa, CA</i>

- Led the creation and deployment of automated python geocoding scripts, boosting data processing efficiency by geocoding millions of addresses and integrating them into a SQL Server database.
- Created detailed ArcGIS Pro maps to visualize fire and brush risk zones, supporting spatial data analysis and decision making.
- Engineered a Python-based automated mail-merge system to notify insured clients of renewal changes, ensuring compliance to Department of Insurance regulations.
- Act as a key consultant in automation projects, focusing on data engineering and data modeling initiatives.

ARCHES Staff / Analyst + Lead Tutor	Aug 2020 — July 2022
<i>Cal Poly Pomona Foundation</i>	<i>Pomona, CA</i>

- Analyzed the effectiveness of the program using survey data provided from participants and created performance reports using Excel and LaTeX.
- Mentored students with disabilities in the ARCHES program whose goal is to provide academic assistance to students with disabilities.
- Provided mathematics and statistics tutoring from various lower division to upper division courses: (basic algebra to calculus to probability and statistics).

Hackathon / DataFest 2019	May 2019
<i>Chapman University</i>	<i>Orange, CA</i>

- *Medalist for Best Use of Data Forensics* out of 30 teams.
- Handled and analyzed large data sets with millions of entries involving the Canadian National Women Rugby Sevens Team using SQL.
- Compared two week training schedules of players to tourney placements by creating visualizations using Tableau.

PROJECTS

Determining Humanitarian Aid to Underdeveloped Countries

- Used a Kaggle data set that contains information on socio-economic and health factors of 167 countries to determine which countries are in need of financial aid.
- Utilized K-means and hierarchical clustering methods to cluster countries into four levels of overall development.

Forecasting Air Quality in Naples, Italy

- Analyzed and assessed carbon monoxide levels of Naples, Italy from 2004 to 2005 provided by the UCI Learning Repository.
- Applied linear smoothing and Kalman filtering methods to approximate missing values in the data.
- Constructed forecast models using seasonal ARIMA models and cross-validated them on the basis of AIC and BIC.