TOM CHENG



DATA SCIENTIST

[TomHCheng@gmail.com](mailto:TomHCheng@gmail.com) 650.240.6147 3701 Alameda de las Pulgas, San Mateo, CA 94403

[LinkedIn](https://www.linkedin.com/in/tom-hunagyi-cheng/) [GitHub](https://github.com/tc18fwd/) [TableauPublic](https://public.tableau.com/profile/tom.cheng4159#!/)

SUMMARY



*I’m a data scientist specialized in improving business performances through data analysis and sales forecast to find optimal performance criteria for the business to be successful. I have extensive experience in design, implementation, and analysis of data-driven studies/decisions for businesses.*

EDUCATION



**Springboard** May 2020 - Current

*This online data science cohort consists of 500+ hours of intensive curriculums in data science, machine learning, Python, and SQL. Including one guided capstone and two selective capstone projects (see projects).*

**San Jose State University** Summer 2010 - Winter 2012

*B.A. Psychology 2012*

EXPERIENCE



**Super Cue Cafe**,*Co-owner*,SF and San Mateo Jan. 2013 - May 2018

*Collected data to forecast sales on a weekly basis to help inventory and staffing efficiently.*

*Analyze product sales seasonally to help decide on which products to keep, replace, or promote.*

*Stores staffed according to the forecasts had yearly net gain increasing from 5~20%.*

**U.S. Army**,*92A - Logistical Supply Specialist*,South Korea June 2003 - June 2006

*This is where I trained to be disciplined. I began as a follower, learned teamwork, and developed leadership.*

SKILLS



**SOFTWARE AND PROGRAMMING LANGUAGES:** *Python, Tableau, SQL, MS Excel, MS PowerPoint*

**DATA ANALYSIS:** *Pandas, NumPy, statsmodels, Scikit-learn, SciPy*

**DATA VISUALIZATION:** *Seaborn, Matplotlib*

**PREDICTIVE MODELS:** *ARIMA(X), SARIMA(X), BATS, T/BATS, HWE/SES, VARMA(X)*

PROJECTS



[SuperCue Time Series Forecast (click to view)](https://github.com/tc18fwd/SpringBoard/blob/master/Capstone%20Two/README.md) Sept. 2020 - Nov. 2020

*Using Python to clean and implement data to several time series models. I was able to make a model that's able to forecast sales on an hourly level, and capture three seasonality: daily, weekly, and yearly.*

[CA House values Exploratory Data Analysis (click to view)](https://public.tableau.com/profile/tom.cheng4159#!/vizhome/CS3CAhousevalues-EDA/California1to4BRhousesEDA) Dec. 2020

After downloading, cleaning, re-organizing data from zillow-research. Here is the exploratory data analysis of California house' estimated values from 1996 Jan to 2020 May.

[CA House values forecast with ARMA (click to view)](https://public.tableau.com/profile/tom.cheng4159#!/vizhome/CAhousevalueforecastusingzillowestandARMAmodel/CAhousevalueforecastusingzillowestimateandARMAmodel) or with [VARMA (click to view)](https://public.tableau.com/profile/tom.cheng4159#!/vizhome/CS3-forecastandinvestCAhouses2020-2025/CAHouseValues) Jan. 2020

After testing several different models: SES, ARIMAX, VARMAX, VARMA, TBATS. ARMA had the best overall score. The 1st slide shows the accuracy of the prediction, on the 2nd slide you get to choose the % gain you’re looking for in 5 yrs by BR/zipcode, 3rd slide is to check the specific feature’s past data and predictions.