

# WARREN LEE

## CONTACT

✉ wal82012@gmail.com

☎ 415-860-3907

📍 Based in SF, CA 94030

in <https://www.linkedin.com/in/warren-a-lee/>

👤 warren-andrew

## SKILLS

**PROGRAMMING LANGUAGES & LIBRARIES:** Python, SQL, STATA, SAS, Pandas, NumPy, Scikit-learn, Statsmodel, Matplotlib, Seaborn, BeautifulSoup, Selenium, NLTK, GeoPandas, Gensim

**STATISTICAL METHODS & MACHINE LEARNING:** Regression Analysis, Classification, Unsupervised Learning, Natural Language Processing

**ENTERPRISE TOOLS:** Postgres, Tableau, Jupyter Notebook

## PROJECTS

### Data Driven Analysis of Neighborhoods in Vancouver, B.C. (In Progress)

This project seeks to show visible and important facets of the neighborhoods of Vancouver.

- Collected local business data from Yelp and neighborhood shapefiles from Government of Vancouver and calculated an incidence ratio to identify businesses over-indexed in each neighborhood.
- Tools used: Pandas, GeoPandas, Scikit-learn, Matplotlib, Seaborn, Tableau.

### Modeling Topics in Song Lyrics

Are there particular themes that songwriters tend to express, and can song topics be generalized across musical genres?

- Collected bag-of-words lyrics data from Million Song Dataset and modeled topics in song lyrics using the NMF algorithm from the Scikit-learn Python library to explore whether song topics can be generalized across musical genre.
- Tools used: SQLAlchemy, Pandas, Scikit-learn, Matplotlib, Seaborn.

### Predicting Bias against Women in Movies

Can characteristics about a movie's production team determine whether the movie passes the Bechdel Test, indicating bias against female roles?

- Collected data on movie production staff from IMDb, applied a Gender-Guesser Python package on the first name of staff, and created a logistic regression to predict whether a movie fails the Bechdel Test.
- Controlled for features affecting content, such as dummy variables for international production companies and top production companies.
- Tools used: SQLAlchemy, Pandas, Gender-Guesser, Scikit-learn, Matplotlib, Seaborn.

### Predicting Trading Card Price

Can observable characteristics about trading cards predict the value of cards?

- Scraped characteristics of Final Fantasy TCG trading cards from TCGplayer.com using Selenium and BeautifulSoup Python libraries and created a regression model to predict price.
- Cross-validated models based on mean average error to select Ridge regression as the final predictive model.
- Tools used: Selenium, BeautifulSoup, Pandas, Scikit-learn, Matplotlib, Seaborn.

## EDUCATION

Pomona College · Aug. 2012 to May 2016

B.A. Economics 2016

GPA: 3.86; Cum Laude

Recipient of the Margery Smith Briggs Prize, awarded to a non-music major student who made significant contributions to Pomona College Music Department.

Recipient of the Brystine Prize, awarded to the top student enrolled in courses in finance.

## EXPERIENCE

### Metis

Remote

Data Scientist · Jan. 2021 to Mar. 2021

- Metis is an immersive 12 week immersive data science bootcamp focused on project oriented learning.
- The core curriculum is centered around Python, statistics, supervised and unsupervised machine learning, exploratory data analysis, databases, and visualization techniques.
- Completed five self-designed data science projects from conception to presentation; including data collection, data management, exploratory data analysis, modeling, and visualizations.

### Charles River Associates International, Inc.

Oakland, CA

Consulting Associate/Analyst ·

Dec. 2019 to Dec. 2020, Apr. 2021 to Current

- Transitioned to the equivalent position from the SF office acquired from the economic consultancy Edgeworth Economics, LLC.
- Examined economic issues to quantify business damage exposure between \$2m - \$75m for 10+ intellectual property and antitrust lawsuits.

### San Francisco, CA

Managing Consultant · Apr. 2019 to Nov. 2019

- Crafted analyses with internal economic experts.
- Collaborated with end-clients to specify data required for objectives.
- Managed project workflow among consultant team.
- Served as a liaison with legal clients to explain complex economic analyses.
- Created PowerPoint presentations of findings for audiences of various technical backgrounds.

#### Selected Casework #1:

- Produced key analyses presented in patent infringement trial of a top global cardiac medical devices manufacturer that were used to determine monetary damages in jury ruling.
- Calculated reasonable royalty rates using Excel charged for infringing sales under multiple business scenarios using a discounted cash flow model, adjusting forecasts of projected infringing sales based on scenarios.
- Using STATA, processed market share data and historical company sales to identify lost profit damages due to infringing sales.
- Updated damage calculations in response to last-minute client requests based on rulings made during trial.

#### Selected Casework #2:

- Prepared valuation of standard essential patent portfolio asserted against a Fortune 100 telecom client to assist with patent license negotiations using value share analysis based on number of patent citations.
- Systematically cleaned strings of asserted patent numbers using STATA to match against database of US patents declared essential to technological standard.
- Scraped expiration data for all patents in database using Selenium and BeautifulSoup Python libraries from Google Patents webpage.
- Adjusted number of citations for patents based on age of patent, self-citations, and expiration over the time of negotiated license.
- Mapped patent value based on number of citations to a value distribution generated using Monte Carlo simulations.
- Calculated an upper-bound royalty rate for asserted portfolio to prepare counter-offer made by client.

### San Francisco, CA

Senior Consultant · July 2017 to Mar. 2019

- Led training sessions for new consultants in San Francisco office.
- Participated in office social committee, organizing quarterly social events under budgets between \$3k-\$4k.

### San Francisco, CA

Consultant · July 2016 to June 2017

- Analyzed 10 million+ observation datasets using SQL, Excel, Python and Stata to prepare economic and statistical models under tight deadlines used in expert reports and presentations submitted in patent and antitrust litigation cases.
- Authored market research reports on industry trends and key competitors for inclusion in expert reports and trial testimony.