

# Binh Hoang

## Projects

### Forecasting Monthly Superstore Sales with SARIMA Time Series Model

- Trained, validated, and tested a (0, 0, 0), (0, 1, 1, 12) SARIMA time series model, achieving a test RMSE of \$15k (equivalent to 0.6 standard deviations)

### A User-Empowered Anime Recommender System

- Utilized content-based and collaborative filtering techniques to make anime recommendations
- Productionized recommender system on Flask app, built with Python, HTML, and CSS, where user can control adventurous level of recommendations
- Built Docker container to deploy web scraping script across multiple Google Cloud instances

### From Despair to Hope: A Sentiment Analysis of Queer and Trans Asian Pacific Islander Blogs

- Performed NMF topic modeling and VADER sentiment analysis to uncover sentiment trends between beginning, middle, and end of blog posts
- Stored text and metadata scraped from tumblr on MongoDB database
- Visualized document-topic clusters using t-SNE and Plotly

### Optimizing Flight Choices for Punctuality

- Modeled flight delay probabilities using logistic regression, optimized for precision and recall
- Wrote tips to help travelers choose flights with highest probability of being on-time based on logistic regression model coefficients
- Created SQL database to store and query 5 million+ rows of 2015 U.S. flight data

### Predicting Global Box Office Revenue of Japanese v. American Animated Films

- Built separate linear regression models for Japanese and American animated films to predict global box office revenue
- Scraped imdb.com data on 3,000 Japanese and American animated films using BeautifulSoup

### Expanding NYC Subway Station Accessibility

- Combined NYC subway turnstile data with Google Maps API and census data to recommend top 5 Manhattan subway stations to prioritize elevator installation

## Experience

**Metis** June 2020 - Sept. 2020  
*Data Scientist* New York, NY

- Completed 12-week accredited data science bootcamp, demonstrating skills in Python, SQL, data wrangling, machine learning, and data visualizations

**D. E. Shaw & Co.** Aug. 2015 - Mar. 2020  
*Treasury Associate/Data Analyst* New York, NY

- Managed financing and securities lending for \$30 billion US, Canadian, and Brazil equity strategies, driving improvements in reporting, connectivity with trading desks, and relationships with equity finance and delta one brokers
- Modeled financing and margin costs on Excel when launching new trading strategies or expanding set of banking providers, helping senior managers make decisions on which banks to do business with
- Designed and productionized weekly FX PB allocation report, reducing daily margin posted by \$10-50 mn
- Developed borrow rates monitoring report, saving \$4 mn/year in borrow costs from 2018 to 2020
- Co-authored analysis based on lend income modeling on whether to start lending equities in D. E. Shaw's institutional asset management arm
- Oversaw dev projects to create financing metrics dashboard, build rate negotiation savings calculator, and update securities lending UI

## Contact

✉ binh.hoang7@gmail.com  
🌐 <https://binh748.github.io/my-blog/>  
☎ (860) 965-8121  
📍 New York, NY  
in binh748  
📄 binh748

## Education

Yale University  
B.A. Humanities (with distinction) 2015  
GPA: 3.86/4.00  
Cum Laude

## Skills

### Languages

Python  
SQL  
HTML  
CSS

### Experimentation

A/B Testing

### Machine Learning

Regression  
Classification  
Natural Language Processing

### Data Tools

Google Cloud  
Docker  
Pandas  
Numpy  
Ski-kit Learn  
Statsmodels  
Git  
Spacy  
NLTK  
Gensim  
VADER  
BeautifulSoup  
Selenium

### Data Visualizations

Tableau  
Flask  
Seaborn  
Matplotlib  
Plotly

### Databases

SQL  
MongoDB