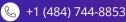
# Wei-Hua (Wafer) Hsu





github.com/wafer110



#### **SKILLS**

- Coding Languages: Python, R, and SQL.
- Data Science: EDA, Visualization, Interpretation, Tidyverse, Applied Statistics, and Regression.
- Machine Learning: SVM, PCA, kNN, Logistic Reg., LDA, QDA, Decision Trees, and Random Forests.

### **EDUCATION**

### Master of Science in Data Science with a concentration in Computer Science

American University, Washington, DC, USA **Graduated:** December 2020 **GPA:** 4.0/4.0

- Technical Leader & Table Co-host in AU Catholic Student Organization (2019-2020).
- Spring Cultural Festival Volunteer in International Student & Scholar Services (2019).

- Adobe Apps: Premiere, Photoshop, and After Effect.
- Web Developer: Shinyapps and HTML.
- NLP: Feature Engineering, Word Embedding, Text Mining, Sentiment Analysis, and Applied Machine Learning.
- Languages: English (fluent) and Mandarin (native).

## Bachelor of Business Administration in Communications Management

Shih Hsin University, Taipei, Taiwan

**Graduated:** July 2016 **GPA:** 3.9/4.0

- Winner (2015) & Student Leader Coordinator (2016) in Asia Innovation, Creativity, and Entrepreneurial Charrette.
- Excellent Work Award in National TBSA Innovative Planning and Marketing Competition (2015).

#### **WORK EXPERIENCE**

#### **ASAC Quantitative Supports & Statistics Tutor**

#### American University, Washington, DC

March 2020 -December 2020

• Helped more than 30 students with solving statistical, data analytics, and basic computer science problems.

#### **Data Engineer Intern (Natural Language Processing)**

#### AU Optronics Corporation, Taichung City, Taiwan

*June 2020 – August 2020* 

- Applied NLP using Feature Engineering (Clustering and Topic Modeling), Keyword Analysis, AUO self-building dictionary, Word Embedding (TF-IDF, Bag of Words, and Bert), and Word Cloud on shift notes.
- Produced daily visualization reports alerting management to potential manufacturing line problems.
- Acted as Intern Leader, liaised between management and interns, and managed entertaining residential life.

#### TV Production Assistant & Visual Content Specialist

#### Sanlih E-Television Co. Ltd, Taipei City, Taiwan

October 2018 – January 2019

- Broadcasted live streaming data, sports data, song lyrics, and graphics in a real-time production truck.
- Assisted the Production Director behind the screen to coordinate with live TV production flows.

#### Software Trainer & Freelanced Graphic Design Producer

#### Voxel Vision Technology Co. Ltd, Taipei City, Taiwan

December 2016 - December 2018

- Gathered, cleaned, and organized data through Excel and/or FTP Systems for multi-media productions.
- Produced visual content on CG Graphic Systems, streaming out on TV channels and live stream platforms.

#### **RELEVANT PROJECTS**

#### Utilized ML to Conduct Triple-Negative Breast Cancer Prognosis (Python, R, & Weka)

- Applied Feature Selections and Association Rules to define features that predict two-year survivability.
- Used Clustering and Classification methods to perform accurate TNBC prognosis.

#### Applied Natural Language Processing on Reddit Comments About Politics and Sports (Python)

- Identified trending issues in US politics, using Topic Modeling and Clustering.
- Trained Classification Model, resulting in remarkable accuracy in classifying similar themed subreddits.

#### How to Succeed in E-Sports: A Data Analysis on PlayerUnknown's Battlegrounds (R)

- Applied data preprocessing and data analytics using Tidyverse package to provide data analytics results.
- Visualized data correlations between variables including kill rate, damage rate, weapon/vehicle usage, survival time, moving distance, and standing positions on the game map utilizing ggplot2 package.

#### U.S. High School Track & Field and Cross-Country Data Analysis (R)

- Cross-analyzed track and field datasets with US statewide income and climate data to identify lower income athletes with better performance to recruit with Federal Work Study funds.
- Applied web scraping, data cleaning, and data analytics and visualized the statewide results on Shinyapps.

#### RECOMMENDATIONS

• Dr. Richard Ressler rressler@american.edu

• Prof. Scott Talan

talan@american.edu