

# JASON WU, MS

815.582.0411 • jasonwu1211@gmail.com • <https://www.linkedin.com/in/jasonwu1211/> • <https://github.com/JasonWu1211>

## DATA SCIENTIST | PREDICTIVE MODELING & MACHINE LEARNING

*Applying predictive modeling techniques to derive strategic data insights that identify and address business challenges, enhance decision-making processes, and facilitate the building of applications that secure a competitive edge*

Self-directed **Data science professional**, offering **2+ years** of professional experience and knowledge using the data science and data analytics function to uncover, interpret, and translate meaningful information into company roadmaps that improve business outcomes. **Inquisitive and technically-adept**, underpinned by substantial project experience using a variety of data mining and data analysis tools, creating machine learning algorithms, and building, testing, and implementing predictive models that solve challenging problems across a diverse set of business needs.

- Highly effective at planning and time management; capable of working independently, leading and contributing to teams, and prioritizing workloads to manage and meet project deadlines and **deliver practical solutions**.
- **Excellent written and verbal communication** skills to coordinate across teams and present data analyses in a straightforward manner for non-technical audiences. **Trilingual fluency in English, Mandarin, and Taiwanese**.
- **Thoughtful and organized problem solver** who applies a holistic approach to understand open-ended problems from multiple perspectives. Driven to learn and master new technologies and techniques.

### TECHNOLOGY SUMMARY

|                                 |   |
|---------------------------------|---|
| PROGRAMMING & DATA MANAGEMENT:  | Python, R, SQL, PL/SQL, Jupyter, Oracle   |
| MACHINE LEARNING & DATA MINING: | Clustering, decision tree learning, random forest, regression, classification, support vector machine, pattern recognition, anomaly detection, time series analysis |
| DATA VISUALIZATION & ANALYSIS:  | Tableau, ggplot, matplotlib, ggmap, plotly, Seaborn, Pandas, NumPy, Scikit-Learn  |
| BIG DATA & CLOUD SERVICES:      | AWS, Hadoop, Hive, Pig, MapReduce   |

### EDUCATION

**Master of Science (MS)**, *Predictive Analytics (Data Science)*, **GPA: 3.98** • DePaul University—Chicago, IL (2018)

**Bachelor of Business Administration (BBA)**, *International Business*, **GPA: 3.58** • Tamkang University—New Taipei, Taiwan (2013)

DataCamp (June 2018) | Data Analyst in Python (Cert)—License #40512 • Python Programmer (Cert)—License #40513

### CORE SKILLS & ATTRIBUTES HONED THROUGH EDUCATION & EXPERIENCE

Data Science • Teamwork Facilitation • Data Analysis & Data Mining • Project Management

Analytical & Quantitative Problem Solving • Recommender Systems • Business Intelligence (BI) Solutions • Data Visualization

Experimental Design (DoE) • Machine Learning Models • Statistics & Statistical Inference • Research & A/B Testing

Public Speaking & Presentation • Data Warehouse (ETL Processes) • Continuous Improvement • Practical Decision-Making

### PROFESSIONAL EXPERIENCE

**DATA SCIENTIST** | DataCo LLC

August 2018 to Present

*Brought on board to provide marketing solutions thru the use of predictive analytics. Working with almost all industries using all forms of data to **provide insight and actionable recommendations to marketers**. Developing analytical tools, machine learning models and reporting dashboards from scratch using all kinds of statistical and programming software including **SQL, python, R, Git** and more.*

- Leveraged **machine learning** to build **automated lead scoring systems** for clients and maintained the script daily
- Applied **statistical & hypothesis testing** for marketing segmentation and penetration research
- Analyzed **customer value** using marketing analysis technique including but not limited to Customer lifetime value, cohort and churn analysis

**SURVEY ASSISTANT | NORC at the University of Chicago**

June 2017 to November 2017

Brought on board to provide **strategic data analysis and support** for a National Science Foundation-sponsored project. Fielded a large volume of daily phone calls to gather up-to-date contact information, verify respondents' identity, and apply advanced Excel and database management skills to clean, query, and derive data insights to increase international survey response rates. Gained experience learning the importance of data integrity and the differences between good and bad data collection practices.

- Earned recognition from the manager and a **positive performance review** for producing high-quality work while managing the largest number of cases, between 30 to 40 per week, and effectively outperforming colleagues.
- Achieved data collection and cleaning for **87%** of cases and detected missing values for 14% of the cases.
- Applied **balanced interpersonal, technical, and communication skills** to persuade respondents to participate in the survey while ensuring the protection of each respondent's personal information.
- Undertook a side project focused on updating key contact information, based on the business and survey application needs, within the organization's database to **improve the survey mailing quality**.

**MARKETING DATA ANALYST | GET Educational**

January 2013 to December 2013

Hired in a part-time capacity while earning BAA in international business to manage and analyze student data using SQL. Contributed to the development of engaging marketing and promotional materials, based on data analysis and interpretation, to grow student enrollment levels. **Rapidly overcame the learning curve** using SQL and data visualization/histogram tools.

- Achieved a **20% increase** in response rates within a 2-month period by thinking 'outside the box' and promoting the program in a university setting, resulting in a **10% growth** in student annual enrollment rates.

---

**SELECT DATA SCIENCE PROJECTS**

---

**PATIENT BEHAVIOR STUDY, HEALTHCARE | PYTHON**

Investigated basic and advanced machine learning algorithms (Naïve Bayes, Ensemble, Support Vector Machine [SVM]) and created predictive models to analyze medical appointment datasets from Kaggle to determine why patients were missing appointments. Planned and executed experiments to optimize models' performance by selecting the most relevant key features and identify the best algorithm that would help hospitals predict show or no-show cases.

- Learned how to **build better results** by combining the advantages of multiple machine learning algorithms, including SVM, gradient boosting, and random forest to identify no-show cases based on more than **110,000 data entries**.
- Identified linear SVM as the optimal performer in terms of recall for no-show cases, **capturing 77%** of no-show cases, after analyzing various performance metrics.
- **Minimized model bias and inaccuracy** by applying an oversampling technique, SMOTE, which rebalanced the no-show cases in the dataset and produced more meaningful results.

**ANIME RECOMMENDER SYSTEM, ENTERTAINMENT | PYTHON**

**Single-handedly** developed a collaborative-based anime recommender system capable of generating a personalized list of anime recommendations based on database information comprising total user history and rating ID user feedback. Challenged to establish and train the appropriate evaluation matrix and key metrics to gauge the model's performance and accuracy in providing practical information to improve the user experience.

- Boosted the system's accuracy from **30% to 45%** after implementing a nDCG matrix that ranked the most relevant items at the top of the list, based on user input. **Recognized as the only person in the class to apply this approach**.
- **Streamlined the analysis** of large datasets by integrating a data pre-processing system that enhanced the rating data quality and increased the rating density **from 0.02% to 6.22%** to yield a more accurate prediction.
- **Maximized the results** by incorporating and fine-tuning key performance metrics including a nDCG matrix, coverage, and average list popularity, which **diversified** the items in the list.

**CHICAGO CRIME DATA VISUALIZATION, CRIME | R**

Used **Tableau, R statistical analysis, geographical heat maps**, and 2012 – 2016 crime data from Kaggle to investigate the different crimes, time, location, and the frequency in which they were committed throughout Chicago's various neighborhoods.

- Concluded that overall crime in Chicago has declined and that the highest crime rates specifically occur downtown, on the south side, and the area between Oak Park and downtown during the summer period from June to August