# Xinsheng(Sean) Zhang

Data Scientist | Data Analyst | Quantitative Analyst

55 River Drive S
Jersey City, NJ 07310

xz1757@nyu.edu
(218)393-0258

## EDUCATION

*New York University (NYU)*, New York, NY — *May 2018*
**M.S. in Data Science** — **GPA: 3.7/4.0**

*University of Minnesota Duluth (UMD)*, Duluth, MN — *May 2015*
**B.S. in Statistics, B.A. in Physics, Minor in Mathematics** — **GPA: 4.0/4.0**

## SKILLS SUMMARY

**Analytics:** SAS, Enterprise Miner, Mathematica, JMP, SQL/MySQL, H2O, Microsoft Excel
**Languages:** R, Python, PySpark, Scala

## CERTIFICATION

SAS Certified Statistical Business Analyst Using SAS 9 — *Jan. 2016*
SAS Certified Advanced Programmer for SAS 9 — *Dec. 2015*
SAS Certified Base Programmer for SAS 9 — *May 2015*

## WORK EXPERIENCE

**Research Assistant [Python]**, *Baysquare*, New York, NY — *Jan. 2017 - Present*
- Built KNN forecasting system to predict expected return.
- Implement k-mean clustering methods.

**Data Scientist Intern [Python, Spark, SQL]**, *NBC Universal*, New York, NY *Sep. 2016 - Dec. 2016*
- Enhanced both the quality and accuracy of forecasting results used for decision making.
- Integrated linear forecasting result from Nielsen raw data with different data wrangling tools.
- Improved the production algorithms to generate forecasting files more efficient and automated.

**Statistical Modeler Intern [R, SQL, H2O]**, *LexisNexis*, Orlando, FL — *June 2016 - Aug. 2016*
- Performed data manipulation, analysis, and visualization using R and Python.
- Designed and improved clinical risk stratification models and data mining procedures.
- Developed expertise in machine learning including CV, PCA, Gradient Boosting, K-Means Clustering.
- Tuned model parameters and tested model performance with different algorithms and various constrains.

**Data Analyst [R, SAS, SQL]**, *QuantGroup*, Beijing, China — *June 2015 - Aug. 2015*
- Cleaned large dataset from detailed cell line list using R and Python.
- Analyzed social network using R package igraph and interactive visualization software Gephi.
- Used Python to scrape financial related APPs information from baidu.com.
- Entered and updated data into database system (MySQL).

**Disease Model Research Assistant [Matlab]**, *UMD*, Duluth, MN — *Sept. 2014 - May 2015*
- Analyzed and built mathematical model to predict the outbreak of Ebola disease using Mathematica.
- Showed the relation between hospitalization rate with the transmission of the disease.
- Provided the alternative suggestion for controlling the disease through the mathematical computation.

## RECENT PROJECT

**Identify Terrorist Attack** , *NYU course project* — *Dec. 2016*
- Proposed an empirical model which learns from Global Terrorist Database and identifies the responsible terrorist group given a future attack.
- Performed feature engineering on text columns and used a one-hot encoding schema to transform all the categorical features.
- Examined models including Logistic Regression, Decision Tree, Random Forest, SVM, and Naive Bayes.