

# MANUEL RUEDA

[manuel.rueda@columbia.edu](mailto:manuel.rueda@columbia.edu)

New York, NY, USA 10033

(929) 235-3220

---

## **EDUCATION**

### **M.S. in Data Science** (Dec 2016, Columbia University, GPA 3.65)

- Algorithms, Machine Learning (ML), Bayesian ML, Natural Language Processing (NLP), Computational Models of Social Meaning, Data Visualization, Distributed Computing, GIS & Spatial Analysis.
- Capstone Project (Goldman-Sachs): Use text similarity measures and deep learning tools to programmatically quantify changes on thousands of company reports.

### **B.Sc. in Economics** (Dec 2009, Instituto Tecnológico de Monterrey - ITESM, GPA 3.50)

- Econometrics, Time Series, Multivariate Analysis, Statistics, Mathematics for Engineering, Linear Algebra.

## **PROFESSIONAL EXPERIENCE**

### **Data Science Consultant at Kora Management LP.** (June – Dec 2016, NY)

Kora is a global hedge fund that leverages on data science to make long term investment decisions.

- Scrapped unstructured transactional data from commerce websites, aggregated and analyzed it with *Spark* (*Scrapy*, *AWS EMR*).
- Applied feature engineering on this data to build predictive models for companies' performance (*python*, *Spark*).
- Implemented a text polarity tweet classifier for brands that collects the data, analyzes it and displays the results as time series on an interactive dashboard (*python*, *MongoDB*, *R*, *Shiny*).
- Performed statistical analysis and inference on survey data (*python*).

### **Data Science Intern at Trinnacle Capital Mgmt.** (Feb – May 2016, NY)

Trinnacle is a quantitative hedge fund that uses unconventional data sources to discover market opportunities.

- Created and administered HDFS and SQL databases with financial and market data (*Hadoop*, *MySQL*, *AWS*).
- Built a model to identify short-term investment opportunities. For this I used a large dataset of historical minute-frequency data covering the 10 years, and applied backtesting (i.e. time series cross-validation) to find the best model parameters (*R*).
- Created a model that uses mobile GPS data to estimate the number of customers visiting a target business, and then predict its revenue. The first test case predicted with accuracy superior to Bloomberg's estimates (*R*, *Shiny*).

### **Market Data Team Lead at MSCI** (Jan 2010 – Aug 2015, MX)

Asides from maintaining the MSCI global indices, the company provides models for risk and optimization analysis.

- Built and managed the Market Data Productions team, leading 3 analysts.
- Worked with the global Research and Project Management teams to improve our risk models and incorporate clients' suggestions.
- Developed interactive dashboards to analyze >10,000 time series efficiently (*SQL*, *R*, *SpotFire*).
- Developed scripts in R to backtest models and report the results in a markdown file(*SQL*, *R*).

### **TA for Columbia's edX Courses** (Dec 2015 – Feb 2016, NY)

- Statistical Thinking for Data Science and Analytics.
- Machine Learning for Data Science and Analytics.

## **PERSONAL PROJECTS** (<http://masta-g3.github.io>)

**Gentrification in NYC** (Dec 2015): Geo-spatial analysis on the distribution of socio-economical variables in NYC (*R*, *QGIS*, *CartoDB*).

**CULPA Sentiment** (Feb 2015): Sentiment analysis and visualization of Columbia's academic course ratings (*Python*, *HTML*, *D3*).

**MoMA Through Time** (Feb 2015): HTML visualization of exhibitions patterns on the museum; awarded best of the contest (*R*, *D3*).

**Summer Networks** (Summer 2016): Several attempts to build a neural network that can write short poems (*numpy*, *theano*).

**Linear Content Blog** (Ongoing): Personal blog where I post small experiments on Machine Learning and NLP.

## **GENERAL SKILLS**

**Visualization:** SpotFire, Shiny, D3 & P5.

**Programming:** R & Python.

**DB & Cloud Computing:** SQL, AWS & Spark.

**GIS Analysis:** QGIS, GeoDa & CartoDB.

**Web:** HTML, CSS & JS.