Data Scientist

Professional Profile

Passionate about data analysis and experiments, mainly focused on user behavior, experience, and engagement, with a solid background in data science and statistics, and extensive experience using data insights to drive business growth.

Education

2016 **University of California, Berkeley** Master of Science, Data Science

Relevant courses:

- Data Visualization and Communication
- Field Experiments
- Exploring and Analyzing Data
- Machine Learning at Scale
- Storing and Retrieving Data
- Machine Learning
- Research Design and Applications for Data Analysis
- Applied Regression and Time Series Analysis

2014 ***Universidad Politécnica de Madrid*** M.S. in Statistical and Computational Information Processing

Relevant courses:

- Data Mining
- Multivariate Analysis
- Time Series
- Neural Networks and Statistical Learning
- Regression and Prediction Methods
- Optimization Techniques
- Monte Carlo Techniques
- Numerical Methods in Finance
- Stochastic Models in Finance
- Bayesian Networks

2005 ***Universidad Politécnica de Madrid*** M.S. in Telecommunication Engineering

Focus Area: Radio communication systems (radar and mobile).

Fellowship: First year at University, due to Honors obtained last year at high school.

Skills

Programming / Statistics Big Data Visualization Others

Proficient: *R*, *Python*, *SQL Hadoop*, *Hive*, *MrJob Tableau Git*, *AWS*

Intermediate: *SPSS*, *SAS*, *Matlab Spark*, *Storm Bash*

Basic: *EViews*, *Demetra+ D3.js Gephi*, *Neo4j, QGIS*

Experience

DATA SCIENCE

Jan. 2016 – Mar. 2016 **Data Scientist**

- Designed and implemented the ETL pipeline for a predictive model of traffic on the main roads in eastern Spain (a project for the Spanish government).
- Automated scripts in *R* to extract, transform, clean (incl. anomaly detection), and load into *MySQL* data from multiple data sources: road traffic sensors, accidents, road works, weather.

Jun. 2014 – Sep. 2014 **Data Scientist**

- Designed an experiment for Google Spain (conducted in October 2014) to measure the impact of YouTube ads on the sales of a car manufacturer's dealer network.
- A matched-pair, cluster-randomized design, which involved selecting the test and control groups from a sample of 50+ cities in Spain (where geo-targeted ads were possible) based on their saleswise similarity over time, using wavelets (and *R*).

MANAGEMENT – SALES (Electrical Eng.)

Feb. 2009 – Aug. 2013 **Head of Sales, Spain & Portugal** – Test &Measurement dept.

- Applied analysis of sales and market trends to decide the direction of the department.
- Led a team of 7 people.

• Increased revenue by 6.3%, gross profit by 4.2%, and operating income by 146%, and achieved a 30% ratio of new customers (3x growth), by entering new markets and improving customer service and training.

SALES (Electrical Eng. & Telecom.)

Apr. 2008 – Jan. 2009 **Sales Engineer** – Test & Measurement dept.

• Promoted to head of sales after 5 months leading the sales team.

Sep. 2004 – Mar. 2008 **Sales & Application Engineer**

• Exceeded sales target every year from 2005 to 2007 (achieved 60% of the target in the first 3 months of 2008).

EDUCATION

Jul. 2002 – Jun. 2004 **Tutor of Differential & Integral Calculus, Physics, and Digital Electronic Circuits**

• Highest-rated professor in student surveys, in 4 of the 6 terms.

• Increased ratio of students passing the course by 25%.

Projects

2016 **SmartCam**

Capstone *Python*, *OpenCV*, *TensorFlow*, *AWS* (*EC2*, *S3*, *DynamoDB*)

A scalable cloud-based video monitoring system that features motion detection, face counting, and image recognition.

2015 **Implementation of the Shortest Path and PageRank algorithms with the Wikipedia graph dataset**

Machine Learning at Scale *Hadoop MrJob, Python*, *AWS EC2*, *AWS S3*

Using a graph dataset of almost half a million nodes.

2015 **Forest cover type prediction**

Machine Learning *Python*, *Scikit-Learn*, *Matplotlib*

A Kaggle competition: predictions of the predominant kind of tree cover, from strictly cartographic variables such as elevation

and soil type, using random forests, SVMs, kNNs, Naive Bayes, Gradient Descent, GMMs, …

2015 **Redefining the job search process**

Storing and Retrieving Data *Hadoop HDFS*, *Hive*, *Spark*, *Python*, *AWS EC2*, *Tableau*

A pipeline that combines data from Indeed API and the U.S. Census Bureau to select the best locations for data scientists

based on the number of job postings, housing cost, etc.

2015 **A fresh perspective on Citi Bike**

Data Visualization and Communication *Tableau*, *SQLite*

An interactive website to visualize NYC Citi Bike bicycle sharing service.

2015 **Investigating the effect of competition on the ability to solve arithmetic problems**

Field Experiments *R*

A randomized controlled trial in which 300+ participants were assigned to a control group or one of two test groups to

evaluate the effect of competition (being compared to no one or someone better or worse).

2014 **Prediction of customer churn for a mobile network carrier**

Data Mining *SAS*

Predictions from a sample of 45,000+ customers, using tree decisions, logistic regression, and neural networks.

2014 **Different models of Harmonized Index of Consumer Prices (HICP) in Spain**

Time Series *SPSS*, *Demetra+*

Forecasts based on exponential smoothing, ARIMA, and transfer function (using petrol price as independent variable) models.