# Nachiket Patel

Jersey City, NJ  -  Email me on Indeed:

An adroit Data Scientist who is better at statistics than any other software engineer and better at software engineering than any other statistician. Deep knowledge of statistical theory and state-of-the-art data mining techniques. Hands-on experience with machine learning algorithm in R programming and Python. Technical skills on big data processing tools knowledge of SQL, NoSQL and fluency with Python with big data intensive computer science background.

Sponsorship required to work in the US

WORK EXPERIENCE

## Data Scientist

New York Data Science Academy   -   New York, NY  -

June 2017 to September 2017

New York, NY Jun 2017 - Sep 2017
Data Science fellowship involving hands on R, Python, SQL, Spark, Hadoop and Hive
development as well as machine learning, big data, visualization, web scraping, statistics and analytics
• Collaborated with an ad-tech firm to analyze over ~350GB of household-level data; leveraged Principal component analysis to deal with high-dimensional data, K-means clustering to find the relationships in data, location analyses and supervised learning methods to develop propensity recommendations (e.g., users with specific household behaviors more likely to buy through direct mail channels)
• Applied random forest, gradient boosting and linear regression machine learning algorithms combined with feature engineering to minimize the mean absolute error of Zillow's home value in a kaggle competition team
• Leveraged scrapy tool to web scrape Hackmageddon.com and conducted text mining analyses using Natural Language Processing and K-means clustering methods to identify and categorize prevalence, target areas, and what attack they used to drive up overall cost of cyber-attacks worldwide
• Analyzed NYC Vehicle Motor Collision data and conducted exploratory data analyses to understand patterns in frequency and location for NYC collisions and pedestrians getting injured due to collisions -- built an interactive
Shiny application to visualize vehicle collisions and most pedestrians injured in real time

## Software Engineer Data

Narola Infotech   -   Gujarat, IN  -

February 2013 to April 2015

•Migrated the client facing database from Postgres to MongoDB leading to a 90% decrease in query response times
•Developed SQL code to identify, analyse, and interpret trends in large datasets; retrieve data from SQL to clean up data systems and wrote data migration script to transform existing database
•Loaded the aggregate data into a relational database for reporting, dash boarding and ad-hoc analyses, which revealed ways to lower operating costs and offset the rising cost of programming
•Collaborated with project managers, engineering teams, and client representatives to coordinate design and implementation of application; led end-to-end management of project deliverables to ensure on-time delivery

•Implemented the database infrastructure, database enhancements, database performance tuning and Provide remote support for maintenance of databases and servers
•Database design, data modelling, data warehousing using SQL, Teradata SQL assistant and ETL tools

EDUCATION

## Master of Science in Information

New York Institute of Technology   -   New York, NY

January 2016 to Present

## Bachelor of Science in Computer Science

Rajasthan Vidyapeeth

August 2008 to April 2012

SKILLS

SQL (2 years), APACHE (Less than 1 year), APACHE HADOOP HDFS (Less than 1 year), APACHE HADOOP MAPREDUCE (Less than 1 year), AWS (Less than 1 year)

LINKS

http://www.linkedin.com/in/0xiNach

https://blog.nycdatascience.com/author/0xinach

https://github.com/0xiNach

CERTIFICATIONS/LICENSES

## Python Bootcamp

May 2016 to Present

## Data Science Bootcamp

June 2017 to September 2017

Relevant Coursework: Data Science Toolkit, including Github, SQL, Unix; Data Analysis and Visualization with R(e.g dplyr, ggplot2, knitr, R Shiny); Data Analysis and Visualization with Python(e.g Numpy, Scipy, Matplotlib, Pandas); Natural Language Processing(NLP); Statistics; Algorithms in Computer Science with Python; Machine Learning with R; Machine Learning with Python; Parallel Computing, Hadoop & Spark.

ADDITIONAL INFORMATION

SKILLS
Data Science: R (dplyr, ggplot2, shiny, caret) Python (scipy, numpy, pandas, seaborn, matplotlib, scikit-learn, scrapy, selenium) SQL (MySQL) D3.JS NoSQL(MongoDB) Apache

Spark(RDD, MLib) Apache Hadoop(MapReduce, HDFS) Apache Hive AWS

Machine Learning: Linear Regression Logistic Regression Ridge Regression Lasso Regression Cross Validation Decision Trees Random Forest Support Vector Machine Gradient Boosting K-Means Clustering Natural Language Processing PCA Tensorflow

Tools: RStudio Jupyter Sublime text Git Scrapy Selenium Xcode