

# SIDDHARTH SURESH

New York, NY | +1 (202) 378-3623 | [siddharthsuresh3092@gmail.com](mailto:siddharthsuresh3092@gmail.com) | [linkedin.com/in/siddharthsuresh92](https://www.linkedin.com/in/siddharthsuresh92)

Process oriented data professional with 5+ years of experience in Data Science & Data Analytics, spanning across Consulting, Banking and FinTech. Proficient in working with large unstructured data and extracting insights worth up to \$3.5 million. Relevant skills include Programming, Machine Learning, Statistics, Problem Solving and Critical Thinking. Seeking to tackle complex real-world challenges and drive impactful solutions by leveraging data, statistics, and machine learning techniques.

## Skills

**Proficient:** Python, Pandas, Numpy, SQL, PySpark, R, Databricks, Scikit-Learn, TensorFlow, Apache Airflow, Web Scraping, Statistical Analysis, Quantitative Analysis, Forecasting, Bayesian Modelling, Data Mining, Data Transformations, Data Quality Assurance, Data Visualization, RNNs, LSTM, Git, Excel

**Familiar with:** AWS, GCP, A/B testing, Attribution Models, PyTorch, Agile Methodologies

## Relevant Professional Experience

### Data Product Associate, Yipitdata | New York, NY

August 2020 – Present

*Yipitdata provides alternative-data driven research to help institutional investors answer their key questions*

- Analyze large unstructured alternative data from 5+ sources using SQL, Python and PySpark to produce research reports for institutional investors, directly handling products contributing ~\$3.5 million in Annually Recurring Revenue
- Develop and maintain large scale web scraping systems written in Python using Object Oriented Programming to automate daily collection of 10 million+ rows of data
- Create and deploy statistical models using DAGs written in Airflow and Databricks to estimate company comparable KPIs consistently within a 2% margin of error
- Design comprehensive data visualization dashboards and jobs using PySpark and Databricks to run daily quality assurance checks and monitor data sanity across alternative data sources used in the products
- Lead the company wide PySpark training and built analysis frameworks to enable analysts to easily migrate SQL code into PySpark code thereby improving the return on analyst time by ~33%

### Data Science Intern, Capital One | McLean, VA

September 2019 – May 2020

*Capital One is a leading financial institution specializing in credit cards, auto loans, banking, and savings accounts with a focus on technology and data-driven decision-making*

- Analyzed complex retail banking data for several million customers using SQL and Python to identify the key drivers of attrition geared towards minimizing potential loss of capital for the bank
- Engineered 5 new features that helped better identify churn behavior by minimizing type I and type II errors by ~16%
- Developed a Machine Learning model using Scikit-learn along with a proprietary soft attrition metric to detect churn behavior amongst retail banking customers up to 3 months early, helping provide timely inputs to the teams handling pricing and retention strategy
- Created scripts in Python to process new data and finetune and/or retrain the model efficiently for upto 10 million rows
- Published a paper in Systems and Information Engineering Design Symposium 2020 titled "[Evaluating and Improving Attrition Models in the Retail Banking Industry](#)"

### Senior Assistant – Deal Advisory Services, BDO LLP | India

April 2017 – February 2019

*BDO is an international network of public accounting, tax, consulting and business advisory firms and is the fifth largest network in the world.*

- Advised 25+ Private Equity and Venture Capital clients on financial risks within their Target Investment's business operations by identifying key patterns and drivers of KPIs of the Target Company
- Analyzed large amounts of financial and commercial data to identify its potential impact on each line item of the financial statements
- Increased analyst efficiency by 90% by designing frameworks in Excel to automate the financial analysis processes
- Optimized the detection of key financial risks by using statistical analysis, increasing analyst productivity by ~25%

## Education & Certifications

- **M.S. in Data Science**, School of Data Science, University of Virginia | **GPA: 3.97/4** July 2019 – May 2020
- Passed Level 1 of the **CFA® program** June 2018
- Certification course on **Financial Markets & Investment Strategy** September 2016 - January 2017
- **B. Technology in Mechanical Engineering**, Shiv Nadar University | **GPA: 8.54/10** August 2011 – May 2015