

Donald Pierce

Software Engineer

Physicist turned software engineer, passionate about alternative data and experienced in agile working.



don.j.pierc@gmail.com



(248) 977 0841



New York, NY



linkedin.com/in/donjpierce



github.com/donjpierce

SKILLS

Python3

pandas

Dask & Coiled

SQL

scikit-learn

Tableau

AWS Redshift & Athena

AWS S3

AWS Glue

AWS SageMaker

Flask/Dash

Node & JS

BigQuery

GCP BigQuery & Storage

Gremlin

Jinja

HTML

Git

INTERESTS

Designing the cities of the future

Playing jazz sax

Algo trading

WORK EXPERIENCE

Software Engineer Technology Crossover Ventures

07/2021 - Present

New York, NY

Venture Capital & Private Equity - Working within the Digital Intelligence Group (DIG)

Achievements/Tasks

- Directly contributed to a deal sourcing and CRM application for private technology markets, used by 100s of investors across the globe, with offices in New York, Menlo Park/SF, and London.
- Built "alt-data" pipelines to power ML models and a data-intensive application, processing TBs of data on public and private companies.
- Performed ingestion/ELT/ETL on new datasets ranging from 10GB-1TB+ in diverse environments: Snowflake, batch + async APIs, external/internal S3 buckets, distributed pre-processing pipelines in Dask and Hive/SQL.
- Built an automated QA system to analyze features of database tables and raise alerts to developers, aborting Airflow pipelines if data anomalies were found. The system compares key metrics and indicators from past versions of tables to the candidate data table before it is promoted.
- Automated QA system successfully prevented dozens of real data anomalies from corrupting production and user-facing systems, providing a test net to ~80% of core Airflow DAGs. System became core team infra.
- Initiated relationships with third-party data vendors, performed competitor analysis, and evaluated new data opportunities, resulting in valuable partnerships. Drove an initiative to scale data pipelines from Dask to Coiled increasing RAM and CPU headroom in distributed compute pipelines by multiples.
- Maintained and scaled data fusion pipelines, as well as information retrieval and clustering systems, allowing an increase in the number of matching entities by over 100%.
- Built pipelines that transform data across Sagemaker Notebooks, Heroku dynos, S3, and Athena tables to achieve a fully-integrated product. Became proficient in team data ops with 100+ pipelines in this stack.

Senior Analyst, Analytics Carat

02/2020 - 06/2021

New York, NY

Proctor & Gamble Line of Business

Achievements/Tasks

- Promoted to Manager, Analytics in May, 2021.
- Built an automated weekly reporting system in Python, SQL, and Tableau by joining data sources to build rich reports of ad metrics and business KPIs. Reduced regular report turnaround time from 1 month to 1 week.
- Lead two junior analysts in developing ETL pipelines, automating away 100s of hours of work per month.

Contact: Matthew Purdie-Smith - Matt.Purdie-Smith@carat.com

PERSONAL PROJECTS

Traffic (11/2018 - Present)

- Traffic simulation in Python. Reinforcement learning algorithm in TensorFlow. Open source: github.com/donjpierce/traffic

EDUCATION

Artificial Intelligence for Trading Nanodegree Udacity

07/2021 - 03/2022

Graduated
graduation.udacity.com/confirm/5GQDYVTD

Projects

- Smart Beta and Portfolio Optimization
- Sentiment Analysis with Neural Networks
- NLP on Financial Statements
- Backtesting with Multi-factor Models

B.A. Physics and Mathematics New York University

09/2014 - 05/2019

New York, NY