DHRUV GAJJAR

(847) 890-1847 | DHRUV.GAJJAR.9@GMAIL.COM | DHRUVGAJJAR.COM | GITHUB.COM/DHRUVGAJJAR

EXPERIENCE

Amgen (Formerly Horizon Therapeutics) - Deerfield, IL

Data Engineer II, Data Science & Advanced Analytics   *Jun 2022 - Present*

- Piloted Neo4j and spearheaded an entity resolution project that matched over 11 million anonymized patient records across 4 different sources resulting in a 90% reuction of duplicate leads. Published finalized process in a PMSA white paper
- Administered working environments for data scientists, including Linux servers and Google Dataproc, ensuring seamless and efficient workflows. Made continual improvements to workflow, processes, utilities, and automated jobs
- Implemented and advocated for Git adoption among our team, streamlining collaboration and codebase management
- Partnered with data scientists on team to deploy models in production ML pipelines; ensured robust ML/Dev Ops practices with providing clean and reliable data, automating job scheduling, integrated Spark and other essential libraries, and effectively storing and retrieving model artifacts and metrics
- Developed and maintained automated ETL scripts to process the team's main data sources from various vendors totalling 200+ billion rows over each month. Published final datasets in Google Cloud, Databricks, and Azure SQL
- Designed and managed an internal logging and QC package to track and alert on ETL results and ML model performance. This ensured downstream processes were accurate and reliable reduced the frequency of pipeline reruns by over 90%
- Managed several integration initiatives post-acquisition: migrating >600 TB of data from GCP Storage to AWS, recreating working environments in Databricks, and refactoring exisitng pipelines and processes
- Assigned tasks and provided thorough feedback during code reviews to junior team members and consultants
- Independently managed relationships with multiple vendors, ensuring smooth collaboration and timely resolution of issues

Senior Analyst, Strategy & Commercial Analytics   *Nov 2020 - Jun 2022*

- Utilized Sklearn linear regression to predict newly prescribed patient adherence curves to aid in early outreach strategies, resulting in a 10% increase in patient adherence and a 5% increase in revenue over the following year
- Optimized existing Python sales reporting processes by implementing automated pipelines and establishing best practices for package management, documentation, and source control. Decreased runtime by 66% and error checking time by 50%
- Improved automated lead finding alert accuracy and timeliness by tying in external insurance claims with internal patient enrollment data, updating business rules, and adding a logging and QC process enabling the sales and patient access teams to react to and onboard potiential patients faster
- Developed KPIs, reports, and deep-dive projects with executive team as first data hire by gaining a keen understanding of business needs. Enhanced awareness of brand performance and patient trends across 5 previously siloed teams

Axtria - Thousand Oaks, CA

Decision Science Associate   *Apr 2020 - Oct 2020*

- Regularly presented Tableau dashboards developed through a Python backend which spanned every business unit and monitored sales effectiveness and account engagement empowering leadership to adapt quickly to COVID-19 disruptions
- Collaborated with an offshore team of 2-3 individuals by delegating tasks, leading working sessions, and providing constructive feedback, resulting in timely completion of tasks, positive feedback from stakeholders, and improved quality of work

Decision Science Analyst   *Mar 2019 - Apr 2020*

- Conducted a patient journey analysis using Python and SQL on a 10+ billion row insurance claims dataset which identfied key milestones slowing down patients' access to diagnosis and treatment, reducing early disease detection by 100+ days
- Designed multiple Tableau dashboards to replace the client's exisisting PowerPoint dashboards and various diconnected sales data sources, decreasing the 1-3 daily hours spent on manual calculation, analysis, and visualization

EDUCATION

Georgia Institute of Technology - MS in Computer Science (Specialization: Computing Systems)
University of Illinois Urbana-Champagin - BS in Engineering Physics (Specialization: Computational Physics)

SKILLS

Languages: Python, SQL, Scala, Java, BASH, Powershell, HTML/CSS, Javascript, R
Libraries & Frameworks: Pandas, Pyspark, GraphDataScience, React, JUnit
Developer Tools: Git, Neo4j, Google Cloud, Azure, AWS, Apache Spark, Snowflake, Databricks, Linux, Android Studio, Docker, Delta