

# AASHISH NAIR

Hyattsville, MD 20781 | (814) 880-1838 | aashishnair8925@gmail.com

<https://www.linkedin.com/in/aashish-nair/> | <https://github.com/anair123> | <https://medium.com/@aashishnair>

## WORK EXPERIENCE

### Data Scientist

August 2021 - Current

*Optimal Solutions Group, LLC | College Park, MD*

- Reduced market research time for businesses by 80% by leveraging machine learning and natural language processing to identify and recommend potential contractors
- Utilized Python to collect ~2 TB of structured and unstructured data from federal agencies using APIs and web scraping tools
- Designed and built ETL pipelines that extracted and processed data from 20+ sources and then stored them in databases in the cloud
- Built chatbots powered by large language models using the Langchain framework to reduce the time needed for the internal research team to conduct information retrieval by 70%
- Enabled clients to achieve their business targets by generating analytic reports and deploying Power BI dashboards to evaluate their performance and identify pain points

### Data Science Intern

November 2020 – May 2021

*DailyBot | Washington, DC*

- Utilized the TensorFlow framework and natural language processing libraries to build a chatbot that automated manual tasks for the internal teams
- Facilitated the integration of the chatbot microservice into the company's existing backend

### Data Visualization Intern

May 2020 – August 2020

*BroadStreet | Washington, DC*

- Led a team of up to 5 people that created visualizations for Covid-19 data from the Midwest
- Utilized Tableau to create dashboards and capture Covid-19 spread for different demographics
- Independently collected and cleaned data on COVID-19 cases and deaths in the Midwest with Python

## EDUCATION

### Master of Science, Data Science

May 2021

*George Washington University | Washington DC*

### Bachelor of Science, Industrial Engineering

May 2019

*Pennsylvania State University | University Park*

## TECHNICAL SKILLS

- **Languages:** Python, R, SQL, Powershell, Bash
- **Database Management:** PostgreSQL, MySQL, SQL Server, Amazon Redshift, Snowflake
- **Machine Learning:** Classification, Regression, Clustering, Natural Language Processing
- **Data Visualization:** Power BI, Tableau, Plotly, Matplotlib, Seaborn
- **Data Engineering:** Apache Spark, Apache Airflow, Mage
- **Other:** ETL, A/B testing, Web Scraping, Time series analysis, Hypothesis testing, API

## PROJECTS

### Building a Custom Q&A Chatbot on a CPU | [GitHub](#) | [Medium Article](#)

October 2023

Trained a chatbot powered by the Llama2 model with research papers that can run offline on a CPU

### Detecting Faulty Water Pumps in Tanzania | [Github](#) | [Medium Article](#)

June 2023

Developed and deployed a web application powered by an XGBoost model that predicts the functionality of water pumps in Tanzania

### Analyzing Flight in U.S. Airports | [Github](#) | [Medium Article](#)

April 2023

Built a data pipeline using AWS and a dashboard with Power BI to examine U.S. flights and determine how the pandemic affected flight traffic in various regions, airports, and airlines.

## CERTIFICATIONS

### AWS Certified Cloud Practitioner

December 2022

- Offered by Amazon, which proves proficiency in cloud principles and AWS technologies