ANANYA GUPTA
Data Scientist/Analyst/Engineer

(+65) 85755248 | ananyagupta99200@gmail.com | linkedin.com/in/ananyagupta8576 | ananya8576.github.io

EDUCATION

National University of Singapore (NUS)   Aug 2020 - May 2024
Bachelor of Science (Honours)   Singapore
- Major in Data Science and Analytics with Second Major in Innovation and Design Programme
- **Academic Achievements**: Dean's List for 2023/24 Semester 1, Academic Bursary Award 2021
- **Leadership**: NUS Statistics and Data Science Society - Vice President, NUS GRC Group - Marketing Director
- **Relevant modules**: Statistical Modelling, Data Structures & Algorithms, Big Data Systems, Machine Learning, Artificial Intelligence, Data Visualisation, Data Modelling & Computation, Natural Language Processing

TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, SAS, Java, JavaScript, TypeScript
- **Data Science Tools:** Jupyter, MATLAB, Tableau, React.js, Git, Docker, Apache Hadoop, Apache Spark, Netlify
- **Database Platforms:** AWS, PostgreSQL, NoSQL, Firebase, Sisence, FiveTran
- **Additional data-driven skills**: Data Analysis, Data Visualisation, ML, AI, Data Wrangling, Web Development, NLP, CV

EXPERIENCE

Data/ML Intern | Warden   Jan 2024 - Mar 2024
- Built Large Language Model (LLM) for extracting and summarising important financial information with Typescript to make recommendations for document functionality with a front-end interface through node.js, yarn, Docker for the end-users.

Data Science Intern | Oversea-Chinese Banking Corporation, Limited (OCBC)   May 2023 - Dec 2023
- Researched and initiated statistical models and visualisation frameworks on historical data for performance forecasting and time-series correlation to align the current metrics with the targets, expediting performance review and decision making processes.
- Built ETL tools and automation algorithms using Python, SQL, Visual Basic to streamline the data extraction and transformation processes. Optimised the data pipelines with extensive data wrangling and manipulation in SAS programs, to improve efficiency by 200% and better data accuracy of the global metrics while querying and updating the data in a higher data warehouse.
- Established database schema with data dictionaries to develop reporting structure and align key metrics to fine-tune the data mart.
- Deployed Tableau dashboards end to end from data ingestion to user-centric, real-time visualisations for consumer segments.

Data Analysis and Operations Intern | Nugit   Dec 2021 - Jul 2022
- Analysed and visualized raw CSV files and APIs to create interactive reports for clients like J&J, Kellogg's using Tableau and R.
- Achieved 91% client renewal rate by data cleaning, conducting back-end, front-end checks for accuracy and seamless ingestion.
- Assessed product features with Product and Engineering team to troubleshoot bugs, deploy custom metrics using SQL and Java.

COURSE PROJECTS

Verifying Authenticity: Identifying Scams and Fakes in Online Dating (Python, CV, ML)   Feb 2024 - Apr 2024
- Developed a model to identify scam online profiles using image and textual data to mitigate the financial and emotional risks posed at victims. Achieved an accuracy of 0.964 and working further to enhance its practical use.

Enhancing Natural Language Inference with AI-Generated Explanations (Python, NLP, LLM)   Feb 2024 - Apr 2024
- Trained various versions of GPT-2 Models to compare the effectiveness of machine-generated explanations against human-annotated explanations in enhancing the NLI tasks using metrics like accuracy, precision, F1-score, and BLEU score.

Transaction Simulation Model for Construcshare (Python, API, Git, Docker)   Feb 2023 - Apr 2023
- Developed an agent-based simulation model (ABM) in Python from scratch, to mimic the e-commerce transactions at Construcshare, simulating how feature changes affect the website engagement and interaction for evaluating current strategies.

Post-discharge information management (Java, React.js, CSS, HTML, Firebase)   Jan 2022 - Dec 2022
- Built a web-application and creation tool, using React.js and Firebase database, that extracts data from the hospital database and organises it to make information accessible to patients, avoiding delayed, incomplete or complex information dispersal.

Text Summarisation and Simplification (Python, MLOps)   Feb 2022 - Apr 2022
- Trained various machine learning models like RNN, LSTM, Transformer, GRU in Python, to compare the output from different research papers using quantitative metrics like precision and recall, followed by qualitative metrics like brevity.

2-Factor Authentication App - Orbital Project 2021 (PostgreSQL, CSS, HTML, JavaScript)   May 2021 - Aug 2021
- Built a 2-FA Password Manager using SQL, CSS, HTML, JavaScript, to securely store & auto-fill passcodes with single click.