Ameya Damle

+44 7767956960 | ameyadamleuk@gmail | linkedin.com/in/ameya-damle | github.com/daameya | Reading, United Kingdom

EDUCATION

University of Reading, Reading UK **September 22 – September 23**
Master of Data Science and Advanced Computing Merit (2:1)
Coursework: Data Analysis, Machine Learning, Big Data, Cloud computing, Artificial Intelligence

University of Mumbai, India **June 16 – November 20**
Bachelor of Science in Statistics CGPA: 6.13/10
Coursework: Hypothesis Testing, Probability Theory, Bayesian Theory, Estimation Theory, ANOVA

TECHNICAL SKILLS

Programming / Scripting Languages: Python, R, SQL, UNIX Scripting, HTML, CSS, JavaScript
Database and Tools: PostgreSQL, MySQL, Oracle SQL Developer, MongoDB, Apache Hive
Data Analysis and Software: GCP, AWS, Snowflake, Databricks, SAS, Airflow, Tableau, Power BI, DBT
Framework: NLTK, TensorFlow, Keras, Scikit-Learn, PySpark, Flask, Hadoop

PROFESSIONAL EXPERIENCE

Supervisor, University of Reading **November 22 – January 24**
Technologies: Descriptive Statistics, Casual Inference, Significance Tests, EDA, SAS, MS Excel, Power BI

- Analysed disparities among diverse data sets including Beverages, water, food stock, and financial metrics and assisted in finance operations in quarterly utilities restocking using python programming.
- Used statistical analysis techniques pertaining descriptive statistics, casual inference, and significance tests within **SAS**, this would aid the analytics by determining which factors provide better conservation.
- Maintaining Data through databases (MySQL), Azure cloud (Data Factory- Data Ingestion pipelines), Excel for further data manipulation.

Data Analyst, HDFC ERGO General Insurance **February 22 – July 22**
Technologies: SQL Server, Power BI, Data Factory, Databricks

- Responsible for delivering, maintaining, and enhancing the NLP Engine service for some of the top-tier insurers across the world.
- Built and optimized features using machine learning techniques such as SVM, logistic regressions, conditional random field model, negation handling, etc.
- Worked with senior data analysts and data engineers to find the trends in the data and support in developing Visual reports and dashboards through visual tools like Power BI and Qlik.

INTERNSHIP EXPERIENCE

Data Science Intern, Data Glacier **June 23 – September 23**
Technologies: Python, Git, Docker, Flask, Airflow, Docker

- Developed and maintained data ingestion pipelines using Apache Airflow, Docker, and Databricks, ensuring efficient and reliable data processing.
- Implemented MLops practices like DVC (Data Version Control) and GitHub Actions, to machine learning models.
- Transformed nested JSON data to flat data structure, created data model and migrated to cloud databases.

British Airways Data Science Intern, Forage **September 23 - November 23**

- Scraped and analysed customer review data to uncover findings.
- Built a predictive model to understand factors that influence buying behaviour.

Cognizant Artificial Intelligence Intern, Forage **November 23 – January 24**

- Conducted exploratory data analysis using python and Google Collab for one of Cognizant's technology-led clients, Gala Groceries.
- Prepare a Python module that contains code to train a model and output the performance metrics for the Machine Learning engineering team.

ACADEMIC PROJECTS

Predictive Modeling for forecasting Flight Delay | PySpark, Jupyter, Databricks, Data Factory, Power BI, Azure Cloud

- **Designed** and **deployed** PySpark and machine learning pipeline on **cloud-based infrastructure**. Created and implemented a solution to predict Flight Delay with **200K+** data using Azure ML studio and Synapse.
- **Employed** feature engineering and used **Random Forest** for predictive modeling with an 85% accuracy rate.

Real-Time Music Analytics Pipeline | Kafka, Spark Streaming, Airflow, DBT, Python, Docker, GCP, Data Studio

- This project sets up a live data pipeline to capture user interactions on a mock music streaming platform (like Spotify), processing this data in real-time and storing it in a data lake every two minutes.
- Then, hourly **batch processes** transform this accumulated data to derive key insights like trending songs, user engagement levels, and demographic patterns, feeding into an analytics dashboard.

E-commerce Sales Analysis | Snowflake, Snowpark, Python, Tableau, Snowpark File API, Snowsight Dashboards

- Created virtual warehouse, databases, schema objects and internal stage to host all data in local machine and loaded Parquet, CSV and JSON into internal stage using **Snowpark File API**.
- **Transformed** the workflow of curating **snowsight** dashboards based on **KPIs** using snowflake's **tasks, streams and pipes**, to aid visualizations and reduce the time by 40%.

Data Engineering on YouTube Videos using AWS | S3, Glue, Athena, Lambda, Python, Tableau

- **Implemented** AWS Lambda for serverless data processing, automating the **conversion** of raw JSON data to Parquet format, resulting in enhanced data efficiency across the data pipeline by **20%**.
- **Built** scalable pipelines to extract data from S3 buckets, performed join transformations using Glue and Athena, loaded cleaned data, leading to a streamlined and **automated** data processing workflow, providing analysis-ready datasets.