OSCIN JAYPAL

✉ jaypaloscin@gmail.com | 📞 +1 551 339 6397 | **in** **LinkedIn** | 📷 **Github** | **Portfolio**

Summary

- Passionate about applying expertise in quantitative analysis, data mining and econometric/ML model development to drive business solutions.
- Proficient in **Python** (pandas, matplotlib, seaborn), **SQL**, and **Scala** for **data analysis**, **visualization**, and **model building**, with a solid understanding of **statistical techniques** such as **regression analysis**, **clustering algorithms**, and **hypothesis testing**.
- Experienced in architecting and developing scalable data pipelines and workflows, supporting data integration, transformation, data manipulation, model building, integrating business applications and loading processes.
- Hands-on expertise in **Azure cloud services**, including **Azure Data Factory**, **Databricks**, **Synapse Analytics**, and **Apache Spark**, with a focus on **automation** and **CI/CD pipeline management** using **Azure DevOps** and **Git Repositories**.
- Proficient in building and maintaining **Azure DevOps (ADO)** pipelines and other Azure Infrastructure components, ensuring efficient and reliable delivery of data solutions through automated deployment and integration workflows.
- Committed to implementing best practices in software development, including code reviews, testing, and continuous integration/continuous deployment (CI/CD) processes, to maintain high standards of code quality and operational efficiency.
- Dedicated to ensuring data quality, integrity, and governance by implementing robust data validation, monitoring, and auditing processes.
- Strong analytical and problem-solving skills in designing and implementing scalable data pipelines, managing large datasets, and extracting meaningful insights from complex data. Adept at collaborating with cross-functional teams to understand business needs and leverage data for actionable insights.
- Committed to continuous learning and professional development, staying abreast of the latest trends and technologies in data science and cloud computing.

Education

New York Institute of Technology - 3.77 GPA | Master of Science, Computer Science | Sept 2022 – May 2024

Gujarat Technological University (India) - 3.80 GPA | Bachelor of Computer Engineering | July 2017 – June 2021

Experience

Jr. Data Engineer | MOOMOTH MEDIA PVT LTD (India) | July 2021 – June 2022 (1 year)

- Optimized ETL processes, reducing processing time by 30% and improving system performance. Developed a Power BI dashboard for data anomaly detection.
- Migrated legacy pipelines to the cloud, reducing maintenance and enhancing scalability; improved skills in Apache Spark, Airflow and Azure Data Lake. I also gained experience in DevOps, Git Repos, and CI/CD pipelines for code deployment.
- Collaborated cross-functionally on data processing and transformation; documented processes to ensure consistency and knowledge sharing.

Data Analyst | MEGH TECHNOLOGIES PVT LTD (India) | Jun 2020 - June 2021 (1 year)

- Proficient in Python libraries (pandas, matplotlib, seaborn) for data analysis and visualization.
- Experienced in SQL for querying and manipulating relational databases.
- Strong understanding of statistical concepts, including hypothesis testing, regression analysis, and clustering algorithms. Excellent communication skills for articulating complex analytical findings.

Skills

- **Programming Language:** Python, R, Java, Javascript and Scala.
- **Database management Systems:** Mysql, Microsoft SQL Server, MongoDB, PostgreSQL, Kusto(KQL).
- **Cloud Computing Services:** Microsoft Azure, Microsoft Fabric, Azure Databricks, Azure Data Factory, Azure SQL, Azure Data Lake, Azure HDInsight, Azure Synapse Analytics, Azure Cosmos DB, Logic Apps.
- **Machine Learning and Predictive Analysis:** Open CV, TensorFlow, Keras, Scikit-Learn, MATLAB, Pickle, Seaborn, Numpy, Classification, Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, SVM (Support Vector Machine), KNN, K-means clustering, Natural Language Processing (NLP), PyTorch, PySpark.
- **Big Data Technologies:** Apache Spark, Apache Kafka, HDFS, Hive, Apache Flink.
- **Data Visualization and Business Intelligence Tools:** Power BI, Tableau.

Academic Projects

Software Defect Prediction Analysis

- Led predictive analysis for software defects using data analytics and machine learning, significantly improving software reliability and quality.

Twitter Dataset Analysis

- Analyzed Twitter data with sentiment analysis and behavioral clustering, transforming insights into visual reports for strategic use.

Image Caption Generator

- Designed a system merging CNN and RNN to generate accurate image captions, demonstrating advanced proficiency in Python.

Temperature Screening System with Facial Recognition

- Created a Python-based temperature screening system with facial recognition and AI, featuring an automated email alert system for high temperatures.