Chhayansh Purohit

Dallas, TX | (945) 217-7878 | chhayanshpurohit@gmail.com | https://www.linkedin.com/in/Chhayanshp11/

EDUCATION

The University of Texas at Dallas **Aug 2022-May 2024**
Master of Science, Information Technology and Management GPA: 3.81/4.0
Graduate Teaching Assistant
Relevant Coursework: Big Data, Data Visualization, Advanced Statistics for Data Science, Business Analytics with R

Jabalpur Engineering College, India **Aug 2017-May 2021**
Bachelor of Technology, Computer Science and Engineering GPA: 3.7/4.0
Relevant Coursework: Data Structure and Algorithms, Database Management System, Software Engineering, Data Mining

SKILLS

Certifications	AWS Cloud Practitioner, Alteryx Certified Core Designer
Programming	Python (NumPy, Pandas), SQL, C, C++, R, Unix
Database	MySQL, PostgreSQL, SQL Server
Cloud	AWS (S3, EC2, RDS, Redshift, VPC, Glue, Athena, Lambda, Quick Sight), GCP, Azure
Data engineering	ETL Pipeline, ELT, Kafka, Airflow, Spark, Hadoop (HDFS, Sqoop, Hive), Snowflake, Docker, Tableau

WORK EXPERIENCE

Data Analytics Intern | LPL Financial | Austin, TX **May 2023–Present**

- Developed and executed a comprehensive data cleansing project using **SQL**, **NumPy** and **Pandas**, resulting in a 30% reduction in data errors and improving the accuracy of forecasting models.
- Enhanced data management and reporting capability by automating routine **ETL** tasks by designing **Alteryx Workflows** and **Batch Macros**, reducing manual operations by up to 35%.
- Collaborated with cross functional teams to implement a virtualization strategy using **Tableau**, reducing server costs by 20% while maintaining high availability of data for analysis purposes.

Data Engineer | Infosys Limited | India **Sep 2021–Jun 2022**

- Developed and implemented a Rules Engine for a leading healthcare company, resulting in 60% increase in efficiency for their Registry Quality Control Management System by capturing changed data and conducting logic rule checks using SQL.
- Designed and executed workflows and **SQL** queries to clean and transform raw clinical data from global sources, ensuring accurate and reliable information for analysis.
- Proactively monitored **data pipelines** and promptly resolved data quality issues, resulting in a 25% decrease in data-related incidents and ensuring data integrity and consistency throughout the pipeline.
- Automated job scheduling and reporting emails through Alteryx workflows, streamlining effective communication.

ACADEMIC PROJECT EXPERIENCE

Truck Fleet Risk Analysis using Sensor Data – [Hadoop, HDFS, Hive, Tableau]

- Ingested 10000 records to HDFS with Sqoop and analyzed using Hive and Pig to find top 5 risky drivers and over-used trucks.
- Integrated HDFS with Tableau for accidental trends, forecasting vehicle maintenance and driver safety with 92% accuracy.

Trending YouTube Video Analysis – [AWS Data Engineering, S3, Glue, Athena, Lambda]

- Designed an AWS pipeline to analyze YouTube videos trending data by category and trending matrices.
- Implemented Data Catalog & ETL transformations using AWS S3, Glue and Lambda, automating the process to 90%.
- Visualized various aspects of trending videos and gain insights using interactive Tableau Dashboards.

Healthcare Services in New York – [Tableau, Microsoft Excel, Python]

- Analyzed hospitalization data from multiple counties in New York to assess the quality of healthcare services provided by different hospitals and investigated how Covid-19 has impacted these services and people of various age groups.
- Performed data cleaning by leveraging Python libraries such as NumPy and Pandas for better understanding of dataset and used our findings to formulate hypothesis.
- Utilized Tableau to create innovative dashboards to validate the hypothesis and effectively visualize the data.

Terrorist Network Identification – [Data and Node Analysis]

- Examined 26/11 Mumbai Attack terrorist network dataset and designed a framework with a fitness function employing Prim's Algorithm for enrichment of key node evaluation on the terrorist network.

PUBLICATIONS

Trusted Infrastructure Design for Secure Virtualization in Cloud Computing Dec 2021–Sep 2022
An Informative Analysis of Encryption Algorithms Using Quantitative Fitness Function | Springer Jul 2019–Dec 2019