SANTHOSH KUMAR SURESH KUMAR

santhoshkumarsures@lewisu.edu | +1 (779) 390-8922 | LinkedIn | GitHub | Website

EXPERIENCE

Data Scientist Intern   June 2025 - Sept 2025
Novitium Pharma, USA

- Built a Selenium automation to pull pharmaceutic sales/marketing data from Clarivate and the Orange Book, then cleaned absent values and one hot encoded categories, cutting manual preprocessing time 50%. Also implemented real time scraping via SOAP, Requests, and Selenium, improving data quality, and delivered a multi-year growth and sales dashboard for insights.

Data Analyst   Feb 2024 - April 2024
Office of Commissioner of Police Chennai, India

- Built a Twitter sentiment analysis pipeline to identify and categorize customer complaints, reducing manual testing by 80%. Implemented real time data extraction using SOAP, Selenium, and Requests; translated non-English tweets with MBART and ran sentiment classification with RoBERTa base. Segregated complaints using sentiment scores and added plotting to track complaint metrics and trends.

Data Analyst   June 2023 - Dec 2023
Tamil Nadu Civil Supply and Corporation, India

- Analyzed sales and procurement data to uncover trends and deliver actionable insights, using statistical methods and machine learning to identify purchase behavior patterns and market dynamics. Built predictive models to forecast sales and optimize inventory planning (e .g ., buffer stock levels), improving functional efficiency. Produced careful reports and visualizations to communicate findings and support strategical decision making.

SKILLS

Programming: Python, R, SQL, HTML5, CSS, JS
Data Engineering: MySQL, PostgreSQL, MongoDB, HDFS, ETL, AWS CLI, PySpark
Analytics: Power BI, Tableau, Gephi, Looker, Microsoft Office, Machine Learning, Azure, SAS, Apache Spark

EDUCATION

Master of Science in Data Science   May 2026, GPA: 3.96

Bachelor of Technology in Artificial Intelligence and Data Science   May 2024, GPA: 9.16/10

PROJECTS

- **Economical Impact of COVID 19 (GDP, Cases, Unemployment)**: Built an interactive dashboard in Tableau to analyze relationships between U.S. GDP (1961–2021), COVID 19 case trends (2020), and unemployment metrics (1979–2022), cleaned data, engineered a common "Year" key, blended datasets, and practical EDA + correlation analysis to superficial economic and labor market patterns with time-based filtering and trend visualizations.

- **COVID 19 Mobility Patterns & Tourism**: Developed a Tableau dashboard to visualize pandemic driven mobility shifts across retail/recreation, grocery/pharmacy, parks, transit, workplaces, and residential categories using Google Community Mobility data (2020–2022) and Kaggle tourism/mobility data (1995–2021), performed EDA, category wise and comparative analysis, and delivered multi-year KPI/trend views to highlight sector level recovery patterns.

- **Character Network Analysis (A Song of Ice and Fire)**: Performed interpersonal network analysis on character interactions across five book level datasets (nodes/edges with weighted relationships), including data cleaning/preprocessing, and created advanced network visualizations in Gephi using multiple layout algorithms (e.g ., ForceAtlas2, Fruchterman Reingold, Yifan Hu, OpenOrd). Computed degree/betweenness centrality, eigenvector centrality, and community detection/modularity to identify cogent characters and faction level clusters, surfacing narrative structure and key "bridge" nodes.

- **Uber Data Analytics (NYC 2023 | GCP)**: Analyzed the "NYC 2023" dataset to drive insights for improving urban open transportation accessibility, reducing traffic congestion, and increasing ridership and rider satisfaction. Built an end to end, scalable analytics pipeline on Google Cloud Platform: data ingestion and orchestration with Mage ETL, storage in Google Cloud Storage, transformation and querying in BigQuery, and executive ready Looker dashboards for KPI tracking and stakeholder reporting. Designed with data security, governance, and cloud scalability in mind to model a quotable GCP based transit analytics solution.

HONORS AND CERTIFICATIONS

Certifications: Microsoft Power Platform fundamentals, Microsoft Azure AI fundamentals, Microsoft Azure fundamentals, Preparing Data for Analysis with Microsoft Excel, Harnessing the Power of Data with Power BI, Data Analysis and Visualization with Power BI

Publications:

- Optimizing Healthcare Delivery with Seamless Integration of Clinical Decision Support EHR Systems (doi.org/10.1063/5.0255988)
- Revolutionizing Transportation Insights using GCP Services in Uber's Data Analytics (click).