Sanjana Tiwari

tiwarisanjana0703@gmail.com — +91 8982760380 — LinkedIn/sanjana — GitHub/sanjana

Education

Vellore Institute of Technology November 2022 – Present
B.Tech, Electronics and Communication Engineering, (Specialization in AI and Cybernetics) **CGPA: 8.15 / 10**
Relevant Coursework: JAVA, Neural Networks, Machine Learning, Deep Learning

Skills

- **Programming Languages:** Python, Java, SQL
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn
- **Tools & Platforms:** MS Excel, PostgreSQL, Power BI, Jupyter Notebook, MS Office, GitHub
- **Other Skills:** Exploratory Data Analysis(EDA), ETL Processes, Content Writing, Effective Communication

Projects

Ink And Insight / Excel, SQL, Python, Machine Learning, Power BI, **GitHub**

- Managed the end-to-end data pipeline by ingesting and structuring raw sales data into a PostgreSQL database, utilizing and optimizing complex SQL queries to identify key business drivers, specifically noting the top 5 publishers contribute 60%+ of unit sales.
- Conducted extensive Exploratory Data Analysis (EDA) in Python (Pandas, Seaborn) to uncover critical sales drivers, proving that books rated 4.5+ have a 3x higher average sales volume, generating actionable insights for product strategy.
- Developed a predictive sales forecast model using Scikit-learn, which achieved a reliable prediction accuracy of 165-unit Mean Absolute Error (MAE), providing the basis for future inventory and demand planning.
- Translated complex analysis into easily digestible business insights by designing and deploying a professional Power BI dashboard, featuring 3 strategic KPIs and 5 interactive visuals to guide executive decision-making on marketing spend and inventory management.

The Rhythm of Rest / Excel, SQL, Python, Machine Learning, Power BI, **GitHub**

- Designed and implemented a robust, end-to-end ETL pipeline using Python (Pandas) to process, clean, and transform a sleep health dataset (374 records, 13 features), ensuring data integrity before loading into a PostgreSQL database.
- Executed advanced data retrieval and manipulation by constructing 7+ analytical SQL queries directly utilized for reporting, serving as the foundational metrics within the final business intelligence solution.
- Leveraged Scikit-learn (Random Forest) to construct a classification model, followed by a rigorous feature importance analysis that precisely identified 'Stress Level' as the most critical driver of sleep quality.
- Communicated complex analytical findings effectively by developing a dynamic Power BI dashboard, which combined 9 visuals (4 KPIs, 5 charts) and utilized custom DAX measures to provide stakeholders with data-driven insights on key health metrics.

Behind The Headline / Excel, SQL, Python, Machine Learning, Power BI, **GitHub**

- Designed and implemented a complete data analytics pipeline using Python (Pandas) for the ETL process, cleaning, categorizing, and structuring a database of over 44,000 news articles and storing the data in PostgreSQL.
- Developed a predictive system using Python/Scikit-learn for text classification, achieving a high-confidence rate of 99.4% accuracy in identifying false news content on a test set of 8,900+ articles.
- Constructed an interactive Power BI business intelligence dashboard linked directly to the PostgreSQL database, integrating insights derived from 10+ intricate SQL queries.
- Delivered actionable insights on misinformation patterns through the dashboard, featuring 5+ charts and 4 KPIs, and enabling stakeholders to perform dynamic filtering of 44,000 articles by subject and year.

Certifications

- Applied Machine Learning Using Python (Coursera)

Extracurriculars

Content Team (Health-o-Tech Club)

- Authored and optimized effective event scripts and anchoring speeches for signature, high-profile events, demonstrating the ability to translate complex information into clear, persuasive narratives for broad audiences.