Sanjana Tiwari

Shahdol, Madhya Pradesh
| Gmail | 8982760380 | Linkedin | GitHub |

EDUCATION

Vellore Institute of Technology (Bhopal) **2022 - 2026**
B.Tech, Electronics and Communication Engineering, (Specialization in AI and Cybernetics) **CGPA: 8.15**

TECHNICAL 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

PROJECT

Ink And Insight **GitHub**

- Executed a complete data lifecycle by ingesting and structuring raw sales data into a PostgreSQL database; wrote and optimised complex SQL queries for initial analysis, uncovering that the top 5 publishers generate over 60% of all units sold.
- Conducted in-depth exploratory data analysis (EDA) in Python (Pandas, Seaborn), confirming that books rated 4.5+ sell 3x more units on average. Subsequently, built a Scikit-learn machine learning model that predicts future sales with a 165-unit Mean Absolute Error.
- Designed and deployed a comprehensive business intelligence dashboard in Power BI featuring 3 key KPIs and 5 interactive visuals, effectively communicating data-driven insights from the analysis and model to inform strategic decisions on inventory and marketing.

The Rhythm of Rest **GitHub**

- Designed an end-to-end data pipeline with Python (Pandas) for the ETL process on a sleep health dataset containing 374 records and 13 original features; imported the cleaned and transformed data into a PostgreSQL database for it to act as the single source of truth for analysis.
- Constructed a classification model with Scikit-learn (Random Forest) to examine the drivers of sleep quality; conducted feature importance analysis to determine and order the top 3 predictors, with 'Stress Level' being the most correlated in effect on the target variable.
- Translated sophisticated data sets and model results into meaningful insights through creation of an interactive Power BI dashboard; the report combined results from 7+ analytical SQL queries and included 9 visuals (4 KPIs, 5 charts), including tailored DAX measures to monitor key metrics.

Behind The Headline **GitHub**

- Designed a complete data analytics pipeline to clean, categorize, and model a database of more than 44,000 news articles. Employed Python (Pandas) for the ETL operation, persisted and stored the structured data in PostgreSQL, and created an interactive dashboard in Power BI to display the final insights.
- Designed a text classification system with Python and Scikit-learn to attain 99.4% accuracy for identifying false news. The system was trained using TF-IDF features and tested on a dataset of over 8,900 articles, enabling it to identify linguistic patterns that distinguish real from constructed content.
- Developed and constructed an interactive Power BI dashboard linked with a PostgreSQL database, including 5+ charts and 4 KPIs to display insights obtained through more than 10 intricate SQL queries. The dashboard enables dynamic filtering of over 44,000 articles based on subject and year to analyze misinformation patterns.

LEADERSHIP/CO-CURRICULAR

- As a Content Team Member in Club Health-O-Tech, I wrote effective event scripts and anchoring speeches for signature events, including the marathon and the flagship event of the year.
- Worked together with the organizing committee to facilitate event planning and delivery, ensuring the smooth running of club activities and promoting a good team culture.

CERTIFICATION

- Applied Machine Learning Using Python (Coursera)