

Alwin Wilson

Thrissur, Kerala, 680711

☎ 8113051668 ✉ alwinw372@gmail.com <https://linkedin.com/in/alwinw/> <https://github.com/AlwinW64/>

Projects

Movie Hit/Flop Prediction & Revenue Analysis | *Scikit-learn, Streamlit* Nov 2025 - Dec 2025

- Built an end-to-end ML pipeline using Pipeline + ColumnTransformer, encoding numeric, categorical, and multi-label features via log transforms, imputation, scaling, one-hot encoding, and MultiLabelBinarizer
- Identified and removed 1,593 corrupted budget/revenue records (<\$5K), improving regression performance (R^2 : 0.40 \rightarrow 0.66, RMSE: 5.06 \rightarrow 1.07, log scale) and increasing F1-score (0.84 \rightarrow 0.88)
- Engineered features using log transformations, median imputation, scaling, one-hot encoding, and a custom MultiLabelBinarizer wrapper for genres and production entities
- Trained and compared regression models; Random Forest achieved best performance (\sim 0.88 F1)
- Deployed an interactive Streamlit application supporting real-time predictions and model comparison
- GitHub Repository Link: github.com/AlwinW64/CineAnalytics

Air Quality Monitoring & Trend Analysis | *Pandas, Matplotlib, Tkinter* Apr 2025 - May 2025

- Cleaned and preprocessed large Indian air-quality datasets; performed EDA to uncover AQI and pollutant trends
- Created visualizations including time-series plots, distributions, correlation heatmaps, and state/city comparisons
- Applied statistical techniques (z-scores, hypothesis testing) to identify statistically significant air-quality deviations
- Highlighted key pollutant patterns and regional differences, forming a base for future forecasting and dashboard features
- GitHub Repository Link: github.com/AlwinW64/Air_Quality_Monitoring_and_Trend_Analysis

Training

Data Science with Excel, Tableau and R June 2025 - July 2025

- Completed a 6-week training program focused on data analysis and visualization tools.
- Gained hands-on experience with multiple datasets using Excel, Tableau, and RStudio.
- Learned data cleaning, preprocessing, and basic statistical analysis using Excel functions and pivot tables.
- Designed interactive dashboards and visual analytics in Tableau.
- **Tech:** MS Excel, Tableau, R Programming

Certificates

NDG Linux Unhatched Jan 2025 - Apr 2025

Cisco Networking Academy — [Certificate Link](#)

Cloud Computing Jan 2025 - Apr 2025

NPTEL — [Certificate Link](#)

Technical Skills

Languages: Python, C++, Java, SQL

Libraries & Frameworks: Pandas, Scikit-learn

DevOps: Linux, Git, GitHub

Data Analytics: Excel, Tableau, Power BI

Education

Lovely Professional University, Punjab 2023 - Present
Computer Science and Engineering — CGPA: 8.15 *Jalandhar, Punjab*

SN Higher Secondary School 2020 - 2022
12th with Science — Percentage: 97 *Irinjalakuda, Kerala*

St. Dominic's Convent School 2011 - 2020
10th with Science — Percentage: 92.8 *Irinjalakuda, Kerala*

