

# EDWARD LEEN

Mob: +44 7553139933 | [edwardleenuk@gmail.com](mailto:edwardleenuk@gmail.com) | Hertfordshire

## Personal Statement

Currently pursuing a MSc in Data Science to deepen expertise in machine learning and advanced analytics. Data Science & Analytics professional with 5+ years' experience in data engineering, automation, and analytics, including ML-based content validation for YouTube Music. Skilled in Python, data pipelines, and cross-functional collaboration to deliver intelligent solutions that enhance data quality and reduce manual effort. Seeking a growth oriented role in the field of Data Science to deliver measurable business impact through innovative analytics solutions while continuously expanding my technical capabilities.

---

## Education

### **University of Hertfordshire, MSc Data Science**

**Sept 2024 – Sept 2026**

Relevant modules include:

- Completed Machine Learning and Neural Networks course work, focusing on neural network theory and diverse applications like data mining and bio-computation.
- Delved into Data Mining and Discovery, honing skills in SQL database manipulation, data cleaning, and visualisation. Applied these techniques in practical workshops to analyse real-world datasets, uncovering trends and associations within complex data landscapes.
- Mastered essential data tools-Apache Spark, TensorFlow, Keras-in Applied Data Science, hands-on labs. Developed data processing pipelines using distributed computing and GPUs.

### **St. Xavier's College Mumbai, Bachelor in Software Development**

**Sept 2014 – June 2018**

Relevant Modules include:

- Python Programming: Developed robust data processing pipelines using Python, implementing efficient algorithms and leveraging libraries, such as Pandas, NumPy, and Matplotlib to automate analysis workflows and extract actionable insights from complex datasets.
- Data Warehousing and Mining: Implemented efficient data warehouse architectures while applying mining techniques to extract meaningful patterns from complex datasets, using classification and clustering algorithms to generate actionable business insights.

## Work experience

**Role: Senior Associate**

**May 2021 – Aug 2023**

**Client: Google**

**Project: YouTube Music – Metadata Quality Automation**

**Responsibilities:**

- Collaborated with a cross-functional team on a machine learning-supported content validation system to improve metadata accuracy for YouTube Music (e.g., song titles, artist names, album classification).
- Conducted data cleansing and preprocessing on large-scale music metadata and media files; identified inconsistencies and cleaned language tags, album/single classifications, and release orders.
- Supported the ML team with transfer learning experiments aimed at automating classification of songs, albums, and media content.
- Worked under the guidance of technical leads, contributing to scripts and QA processes that improved classification accuracy and reduced manual verification time.
- This hands-on exposure inspired me to pursue a Master's in Data Science to advance my skills in ML and data modelling.

**Tools: Python, Pandas, NumPy, Scikit-learn**

**Role: Trust and Safety Engineer**

**Feb 2020 – May 2021**

**Client: Google**

**Project: Google Account Approvals – Site Validation Automation**

**Responsibilities:**

- Designed and implemented an automated system to assess and approve websites for Google Ads publishing based on trust and safety guidelines.
- Built a scalable Python-based pipeline that evaluated websites for content quality, copyright compliance, presence of illegal downloads, and overall ethical standards.
- Developed a rules engine with a SQLite backend allowing dynamic updates to website evaluation criteria without code changes.
- Automated report generation by storing site evaluation results (site URL, approval status, and category) in structured tables for easy review and audit.
- Reduced manual review time by 80%, leading to an estimated \$50,000 annual cost saving by decreasing labour hours and increasing throughput.

**Tools: Python, SQLite**

**Role: Analyst – Customer Segmentation & Account Management**

**Aug 2018 – Feb 2020**

**Client: Google**

**Project: Google Cloud Customer Support Automation**

**Responsibilities:**

- Co-developed and maintained an automation module to manage customer account changes and segmentation in Google Cloud using Python and NoSQL.
- Automated client account upgrades (e.g., SMB to Enterprise) based on predefined valuation thresholds, triggered by Salesforce ticket requests.
- Implemented logic to detect and merge duplicate accounts periodically, maintaining data integrity.
- Enabled automated linking of child accounts to parent accounts based on client requests and approval workflows.
- Collaborated on drafting technical specifications and took ownership of module maintenance, including debugging and enhancement.
- Reduced manual effort by 70%, improving ticket turnaround time and supporting faster client onboarding and account corrections.

**Tools: Python, NoSQL, Salesforce**

## Technical Skills

<b>Data Engineering Tools</b>	<b>: Apache Spark, ETL Pipelines, Data Cleaning, Data Mining</b>
<b>Cloud Technologies</b>	<b>: Google Cloud Platform (GCP)</b>
<b>Databases</b>	<b>: NoSQL (Firestore), SQLite</b>
<b>Programming Languages</b>	<b>: Python, Shell Script</b>
<b>Operating Systems</b>	<b>: Chrome, Windows</b>
<b>Other Tools &amp; Platforms</b>	<b>: Salesforce, Scikit-learn, Pandas, NumPy, Power BI</b>

## Additional training

- Microsoft Azure AI Certificate (May 2025)
- PowerBI Certificate (Nov 2024)
- Google Data Analytics Certificate (Oct 2023)
- Lean Six Sigma -Green Belt (Feb 2023)