

# Muhammed Murshid

London, United Kingdom • +44 7587839147 • [muhammedmurshidpp0@gmail.com](mailto:muhammedmurshidpp0@gmail.com) • [linkedin.com/in/murshidpp](https://www.linkedin.com/in/murshidpp) • <https://murshi-xd.github.io/Portfolio/>

## Data Scientist

Detail-oriented and proactive Data Analyst with an engineering and data science background, skilled in SQL, Python, and Power BI. Improved forecasting accuracy by 10% and decision efficiency by 15%. Developed 85% accurate models, optimized inventory, and enhanced promotions by 12%, leading store layout changes that boosted sales by 7%.

## SKILLS

Python, MySQL, Tableau, Machine Learning, Data Science, Data Analysis, Image Processing, AWS, HTML, CSS, R, GitHub, Py torch, Numpy, Pandas, Seaborn, Matplotlib, Word, Outlook

## EDUCATION

### Master of Science in Applied Data Science

University Of Central Lancashire • Preston, United Kingdom • 09/2023 – 09/2024

### Bachelor Of Engineering in Computer Science

P.A College of Engineering (VTU) • India • GPA: 3.3/4 • 10/2019 – 05/2023

## WORK EXPERIENCE

### Codelab Systems

#### Data Analyst Intern

- Analyzed three years of hypermarket sales data using SQL and Python, identifying key trends and insights that led to a 10% improvement in sales forecasting accuracy.
- Developed regression models to forecast sales, achieving a prediction accuracy of over 85% for future sales trends, which facilitated better inventory and resource planning.
- Designed a Power BI dashboard to visualize sales data, helping stakeholders identify top products and customer patterns, resulting in a 15% boost in data-driven decision-making.
- Recommended targeted discount strategies and product offers based on customer buying behavior analysis, contributing to a 12% increase in promotional campaign effectiveness.
- Suggested strategic product placement changes within the store layout by analyzing customer traffic data, leading to a 7% increase in product visibility and sales.

## PROJECTS

### Image Classification System Development

- Developed and implemented an image classification system using both conventional

machine learning techniques (e.g., Random Forest, GradientBoosting) and deep learning models, achieving up to 98% accuracy with the ResNet-18 architecture.

- Optimized the performance of machine learning models through data preprocessing, feature extraction using Histogram of Oriented Gradients (HOG), and dimensionality reduction with Principal Component Analysis (PCA).
- Applied transfer learning by fine-tuning a pre-trained ResNet-18 model, which significantly improved classification accuracy and reduced the training time for a complex dataset.
- Utilized advanced data augmentation techniques and cross-validation to enhance the model's generalization capabilities, leading to superior performance in image classification tasks.

### **YouTube Video Analysis and Dashboard Development**

- Developed a Python-based solution using the YouTube Data API to automate the extraction of channel analytics data, including video metrics like views, likes, comments, and subscriber growth. This automated pipeline improved data retrieval efficiency, reducing manual extraction time by 80%.
- Implemented an AWS Lambda function to trigger periodic data extraction and storage, ensuring that channel data is continuously updated. The data was securely stored in Amazon S3, facilitating scalable storage for large datasets and enabling seamless data retrieval for further analysis.
- Deployed an AWS Glue Crawler to catalog the data stored in S3, followed by using AWS Athena to perform SQL queries on the structured data. This process provided real-time insights into video performance and audience engagement, improving query processing speed by 60% compared to traditional methods.

### **Weather Data Simulation for Photovoltaic Systems Performance Analysis**

University Of Central Lancashire

- Collected and preprocessed weather data from multiple sources (PVGIS, Sarah 2, Era 5) and University of Central Lancashire's PV system data to model photovoltaic system performance, achieving a discrepancy of only 2.10% to 5.31% compared to real output.
- Utilized Python's pvlib library in Jupyter Notebook to simulate DC and AC power outputs for the University of Central Lancashire PV system, improving annual solar energy production forecasts by 3.79% in accuracy.
- Conducted comparative analysis of real versus simulated data from the University of Central Lancashire's PV system, identifying Era 5 data as the most reliable with only a 2.10% variance from actual results.
- Optimized data preparation and cleaning methods using Pandas, NumPy, and Seaborn for University of Central Lancashire PV system data, reducing simulation setup time by 15% and enhancing data accuracy.
- Recommended the use of Era 5 meteorological datasets and pvlib simulation models for accurate photovoltaic system performance predictions, increasing reliability by 3.79% annually for similar university setups.

## **NYC Taxi Trip Duration Prediction Using Machine Learning Models**

- Analyzed a dataset of over 1.5 million NYC Yellow Taxi rides from January 2016, applying winsorization to handle outliers and dummy encoding to process categorical variables, leading to improved data quality and analysis accuracy.
- Implemented machine learning models, including XGBoost Regressor, Random Forest Regressor, and Multiple Linear Regression, to predict taxi trip durations, achieving a prediction accuracy improvement of 15% with XGBoost over other models.
- Developed and optimized data preprocessing pipeline, reducing processing time by 20% by efficiently handling duplicates, null values, and feature extraction for temporal data.
- Evaluated model performance using metrics such as Root Mean Squared Error (RMSE) and Variance, demonstrating XGBoost's superior accuracy with the lowest RMSE among tested models.
- Identified key factors influencing trip durations, such as pickup time, trip distance, and day of the week, enabling actionable insights for optimizing taxi operations and resource allocation.

## **CERTIFICATIONS**

### **Getting Started with Enterprise Data Science**

IBM

### **Ask Questions to Make Data-Driven Decisions**

Grow with Google on Coursera

### **Prepare Data for Exploration**

Grow with Google on Coursera

### **Machine Learning with Python**

Cognitive Class

### **Python 101 for Data Science**

Cognitive Class