CURRICULUM VITAE

SHAZIA SULTANA, **Email Id:**shaz.wasif@gmail.com
B.E.- Information Technology (O.U.),

Address:
H.No:SRT 246,
Chandulal Baradari Colony
Hyderabad-500064**.**

Career Objective:

Aspiring Data Scientist/ Data Analyst with practical experience and strong foundation in Python, Numpy, Pandas, Matplotlib, Scikit learn, Seaborn, power Bi, SQL . Seeking a Junior Data Scientist/ Data Analyst position in a challenging environment that encourages the continuous learning and exposure to new ideas in Software and Information Technology where I can contribute towards organizational goals while achieving personal and professional growth within a dynamic data team.

Educational Background:

Degree	Subjects	College	Duration	University/ Board	Percentage %
Bachelor of Engineering	Information Technology	Deccan College Of Engineering & Technology	2002-2006	Osmania University	**83**
Intermediate	Maths, Physics, Chemistry.	Islamia Girls Junior College	2000-2002	Board of Intermediate Education, A.P.	**91**
S.S.C.	All	Quaint High School	2000	Board of Secondary Education	**80**

Skills

Programming Languages:

Python, C, C++, Java

Libraries & Tools:

NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, Power BI, Spreadsheets, Data Analysis, Data Modeling, Missingno, WordCloud.

Platforms & Environments:

Jupyter Notebook, Google Colab,Visual Studio, Kaggle, SSMS

Domain Knowledge:

Machine Learning, Deep Learning, Artificial Intelligence, Data Science, Statistics, DBMS (SQL)

Experience:

Data Science, Machine Learning, and AI Trainee & Intern

Centre for Training and Employment (CTE), Hyderabad

May 2025 – Present

Projects(Data Science):

Project-1 : Netflix Recommendation Engine:

Mentor/Guide : Anilkumar Manchikatla, IIT Alumnus & Industry Expert, Technical Director – Centre for Training and Employment**(CTE)**

Objective: To develop a machine learning-based recommendation engine that recommends movies to users on a platform like Netflix by analyzing user behavior, preferences, and viewing history, using collaborative and content-based filtering.

Technologies: Python, Scikit-learn, Pandas, NumPy, SQL, Collaborative Filtering, Content-Based Filtering.

Key Highlights:

- Imported and processed **user-movie interaction data** from CSV files using **Pandas.**
- Performed **Exploratory Data Analysis** to understand rating distribution, top-rated movies, and user preferences.
- Visualized rating patterns with **bar plots** and **scatter plots** to identify viewing behavior.

- Developed a movie rating prediction model using **Linear Regression**, **dimensionality reduction via PCA**, and other data standardization techniques.

- Evaluated the model performance through key evaluation metrics, including **MAE** (Mean Absolute Error), **MSE** (Mean Squared Error), and **RMSE** (Root Mean Squared Error).
- Designed a **Collaborative Filtering system** to recommend personalized movie suggestions based on user's similarity.
- Plotted Actual vs Predicted Ratings using **Seborn library** to visually interpret model performance.

Project-2 : Covid Analysis:

Mentor/Guide : Anilkumar Manchikatla, IIT Alumnus & Industry Expert, Technical Director – Centre for Training and Employment**(CTE)**

Objective: To extract actionable insights from **COVID-19** case data by applying data science techniques, forecasting future trends, visualizing key metrics and enabling insights into infection trends, growth rates, and recovery patterns using real-world data.

Technologies: Python, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, Time Series Analysis.

Key Highlights:

- Imported and processed **COVID-19** data from CSV files using **Pandas.**
- Performed **Weekly Summary Analysis** to track infection trends and to understand the pandemic's progression.
- Calculated **Daily Growth Rates(in %)** to measure the spread and slowdown of the virus.
- Identified the **peak day** of reported cases by processing daily case data and **applying time-series analysis.**
- Computed **Case Doubling Time** to determine the virus spread.
- **Forecasted future trends** using **Regression**.
- Performed **Correlation Analysis** between No. of Cases and other factors like mobility, weather, etc.
- Distinguished between **Active** and **Resolved Cases** to monitor recovery patterns.
- Calculated **Mortality** and **Recovery Rates** to identify the disease's severity.
- Visualized key metrics using **Matplotlib** to present clear, informative plots or graphs

Project-3 : Disney+ Hotstar Content Analysis and Recommendation System using Data Science:

Mentor/Guide : Anilkumar Manchikatla, IIT Alumnus & Industry Expert, Technical Director – Centre for Training and Employment**(CTE)**

Objective: To explore, clean, and analyze the Disney+ Hotstar dataset to Understand Content Trends by type, genre, and country, Uncover Content Addition Patterns over time, Visualize insights for better decision-making, Build a basic content-based recommendation system, and Demonstrate the real-world application of data science in the entertainment industry.

Technologies: Python, Pandas, NumPy, Matplotlib, Seaborn, WordCloud, Missingno, Content Based Filtering.

Key Highlights:

- Installed **Kaggle package**, Uploaded Kaggle.json File, Downloaded the dataset.
- Loaded the **csv file** into the **dataframe**.
- Performed **Data Overview** and **Data Cleaning** to identify problems like Missing Values,Duplicate Records, Outliers/Noise, Invalid values, etc
- Conducted **Exploratory Data Analysis(EDA)** on viewing patterns across genres, hours, devices, and days of the week.
- Visualized type distribution and peak watching Time using **Bar chart**, **Pie chart,** and **Heatmap**
- Generated weekly reports summarizing active users and top genres.
- Generated **Word Cloud** by preprocessing the descriptions.
- Used **missingno** library for visualizing Missing Data
- Performed **Content Duration Analysis** and displaying the result through **Histogram** and **Boxplot**
- Designed a **Content Based Recommendation system** using **cosine similarity.**
- Developed **Linear Regression Model** to predict duration based on genre, device, etc
- Predicted **subscriber churn Prediction** by using mock or activity data of the subscriber.
- Applied **Logistic Regression** for binary classification
- Evaluated the model using various evaluation metrics such as **accuracy**, **recall, confusion matrix**
- Built a tool to monitor daily, weekly, and monthly subscription revenue
- Generated **Word Cloud** by tokenizing titles and counting words frequency
- Analyzed user feedback text for sentiment **(Sentiment Analysis)**

Project-4 : YES Bank Monthly Closing Price Detection

Mentor/Guide :Vipul Sonawane, Data Scientist , Research Intern in IIT Madras
- Labmentix Solutions

Objective: To predict the monthly closing price of Yes Bank's stock using past data and machine learning models.

Technologies: Python, Pandas, NumPy, Matplotlib, Seaborn,sklearn

Key Highlights:

- Imported and processed **Yes Bank** data from CSV files using **Pandas.**
- Performed **Exploratory Data Analysis** on financial time-series data, focusing on key stock indicators such as high, low, open, and close values to identify market trends, volatility, and trading patterns.
- Visualized the **missing values** Using **Heat Map**
- Performed **Data Wrangling** to obtain cleaned and transformed code suitable for Data Analysis and Data Modeling.
- Analyzed and Visualized Rising and Falling trends in stock closing prices using **Line Plot** and **Heat Map**

- Detected **Volatility** and **Stable periods** useful in Risk Assessment by keeping track of Daily Price range.
- Identified Peak Time for entering and exiting the stock market by using **Scatter Plot**
- Performed **Hypothesis Testing** using **Welch's t-test** to obtain **P-Value**
- Conducted **Statistical Testing** using **Shapiro-Wilk Test** to obtain **P-Value**
- Implemented **Feature Engineering**, Handled **Outliers**, **Categorical Encoding**, **Rolling Statistics** to capture short-term trends
- Developed a stock close price prediction model using **Linear Regression**, **dimensionality reduction via PCA**,
- Evaluated the model using various evaluation metrics like **MAE** (Mean Absolute Error), **MSE** (Mean Squared Error), and **RMSE** (Root Mean Squared Error), **R2 Score**. **.**

Project-5 :Air Pollution Analysis:

Mentor/Guide : Anilkumar Manchikatla, IIT Alumnus & Industry Expert, Technical Director – Centre for Training and Employment**(CTE)**

Objective: To analyze Air Pollution data from different cities to identify trends, seasonal variations and the impact of pollutants on public health

Technologies: Python, Pandas, NumPy, Matplotlib,Seaborn

Key Highlights:

- Imported and processed **Air Quality** data from **kaggle** using **Pandas**
- Performed **Data Preprocessing Steps** to check missing values and handled them appropriately
- Identified the city with the highest average **PM2.5 Levels** and the month with the worst air pollution based on PM 2.5 levels using the **aggregate function**
- Computed the correlation between **PM2.5,PM10,NO2 and CO**.
- Visualized the data ,correlation between the pollutants and Trend Analysis **using Line Plot, Bar Chart, and Heat Map**
- Analyzed the distribution of NO2 levels across different months using **Box Plot**
- Performed **Statistical and Numpy-based Analysis** by computing mean, median, standard deviation, 90^{th} percentile, moving average, Log **transformation** using **Histograms**

Project-6 : Iris Flower Classification

Mentor/Guide : Anilkumar Manchikatla, IIT Alumnus & Industry Expert, Technical Director – Centre for Training and Employment**(CTE)**

Objective: To develop a supervised machine learning model that classifies Iris flowers into three species—Setosa, Versicolor, and Virginica—based on sepal and petal measurements, using data analysis and predictive modeling techniques.

Technologies:
Python, scikit-learn, Pandas, NumPy, Matplotlib, Seaborn

Key Highlights:

- Imported and cleaned the **Iris dataset** using Pandas for structured analysis.
- Performed **Exploratory Data Analysis (EDA**) to visualize **feature distributions** and identify species-specific patterns.
- Applied classification algorithms such as **Logistic Regression** to build predictive model.
- Evaluated model performance using **accuracy, confusion matrix.**
- Identified **petal length** and **petal width** as the most influential features for species classification.
- Visualized **Feature Coefficients** and **Feature Importance** for each Class using Seaborn and Matplotlib for intuitive interpretation.

Project-7 : Wine Classification

Mentor/Guide : Anilkumar Manchikatla, IIT Alumnus & Industry Expert, Technical Director – Centre for Training and Employment**(CTE)**

Objective : To build a machine learning model that accurately classifies wines based on their chemical attributes. The project aims to explore how features like alcohol content, flavonoids, and color intensity,etc contribute to varietal differentiation, and to evaluate the performance of various classification algorithms in accurately identifying wine types.

Technologies:
Python, scikit-learn, Pandas, NumPy, Matplotlib, Seaborn

Key Highlights:

- Built a **multi-class classification model** to identify wine cultivars using 13 chemical features from the Wine Recognition Dataset.
- Performed **data preprocessing, feature scaling,** and **exploratory data analysis** to uncover relationships between alcohol, flavanoids, and wine type.
- Applied **Logistic Regression (multi_class='ovr')** and achieved an **accuracy** of 97.2% on the test set.
- Visualized **feature importance** using **coefficient plots** and **bar charts** to interpret model behavior across classes.
- Plotted **decision boundaries** using 2D feature combinations (alcohol vs. total phenols) to illustrate model separability.
- Evaluated model performance using **confusion matrix, classification report,** and **heatmaps** for interpretability.
- Used **Seaborn pairplots** and **correlation analysis** to explore feature interactions and class distributions.

Project-8 : Loan Approval

Mentor/Guide : Anilkumar Manchikatla, IIT Alumnus & Industry Expert, Technical Director – Centre for Training and Employment**(CTE)**

Objective:
To build a **Decision Tree Classifier** that predicts loan approval status based on applicants' income and credit score, enabling banks to automate and streamline the loan evaluation process.

Technologies:
Python, scikit-learn, Pandas, NumPy, Matplotlib, Seaborn

Key Highlights:

- Developed a **decision tree model** to automate loan approval decisions based on **applicants' income and credit score.**
- Created a **synthetic dataset** simulating real-world banking scenarios and labeled outcomes as **Approved** or **Rejected**.
- Trained and visualized the **Decision Tree** using **Scikit-learn's tree.plot_tree** to interpret decision paths and feature splits.
- Evaluated model predictions on test data and validated outcomes against expected approval decisions.
- Demonstrated model interpretability by mapping income and credit score thresholds to approval outcomes.

Professional Experience:

- **Data Science and ML Intern** at Centre for Training And Employment**(CTE)**
- Recently joined **PlanetSpark** as an online **Mathematics Tutor.**
- Worked as an online **mathematics and computers tutor** at **Urban Pro.**
- Team Leader**(TL)**, **IT dept** at **SURE PUBLICATIONS(SIA GROUP) (6+ years)**

Role and Responsibilities:

- To **Prepare** copyright-free study material**(ALL IN ONES)** for **B.E., B.Tech, MCA, and MBA (computer -related subjects).**
- To **Review** content prepared by team members/content writers.
- To **Guide** proofreaders.
- To **Compile** comprehensive assessment paper for job seeking candidates that outlined key performance indicators, resulting in a selection process that improved candidate quality and reduced hiring time by 25% across these roles..
- To **Conduct Technical Interviews** for the candidates.

- Graduate Engineer Trainee at **VIJAI ELECTRICALS LIMITED**(**6** months).
- **Programming Instructor(C & C++) :** 1 year of experience teaching engineering students.

Special Achievements:

- Ranked **Tenth(X)** in the entire university **(O.U.)**
- Authored a book for **MCA** first year**: Computer Programming Using C++ (SPECTRUM SERIES, SURE PUBLICATIONS).**
- Got **distinction** in **B.E.(topper)**
- **Class topper** from I to IX
- **School topper** in X.
- **College topper** in Intermediate.

- Received **Excellence Award** (College topper).
- Received **Top Scorer Award** in **Programming in C & C++.**
- Received **Best Teacher Award** in school on Teacher's Day.

Strengths

- Good Communication Skills.
- Fast Learner, Creative, Multitasking, and Accountable.
- Ability to cultivate a resilient team environment by encouraging open communication and collaboration.
- Leadership and Teamwork abilities.
- Commitment and performance-oriented.

Personal Profile

Name	**:**	**SHAZIA SULTANA**
Father's Name	**:**	Syed ValiUllah Ahmed
Husband's name	**:**	M.B.A.Wasif
Date of Birth	**:**	17-03-1984
Gender	**:**	Female
Status	**:**	Married
Nationality	**:**	Indian
Languages known	**:**	English, Hindi, Urdu,Telugu(read and write only).
Religion	**:**	Islam
Hobbies	**:**	Reading books, playing games, listening to music, Browsing the net.
Address	**:**	H.No SRT246, Chandulal Baradari Colony, Hyderabad- 500064
Phone	**:**	8374349435, 8374539435
E-mail Address	**:**	**shaz.wasif@gmail.com**

Declaration

I hereby declare that all the information provided is true and correct to the best of my knowledge and belief.

Place: Hyderabad
Date: 07/17/2025

SHAZIA SULTANA
