

Kavita Shantiwal

Experienced Data Analyst skilled in Python, SQL, and machine learning, with proficiency in data visualization using Power BI, Seaborn, and Matplotlib. Expertise in predictive analytics and deep learning demonstrated through various projects. Open to Data Scientist roles, offering strong analytical abilities and a proven track record of transforming data into actionable insights.

Mobile: +918278669336 Email: kavita.shantiwal66@gmail.com Github: [Kavita-Shantiwal/data_science_work](https://github.com/Kavita-Shantiwal/data_science_work): my work on data science projects (github.com)
Linkedin: <https://www.linkedin.com/in/kavita-shantiwal/>

SKILLS

Programming & Databases <ul style="list-style-type: none">◦ Languages: Python, SQL◦ Databases: MySQL◦ Frameworks: Flask, Django Data Management, ETL & Wrangling <ul style="list-style-type: none">◦ ETL Processes (Extract, Transform, Load)◦ Data Pre-processing, Cleaning & Transformation◦ Pandas, NumPy Data Analysis & Visualization <ul style="list-style-type: none">◦ Libraries: Matplotlib, Seaborn, Plotly	<ul style="list-style-type: none">◦ Tools: Power BI, Excel Machine Learning & AI <ul style="list-style-type: none">◦ Techniques: Supervised & Unsupervised Learning, Prediction, Sequential Pattern Mining, Outlier Detection◦ Algorithms: Logistic Regression, Decision Trees, Random Forest, k-Nearest Neighbors (kNN), Clustering, Classification◦ Frameworks: Scikit-learn, TensorFlow, Keras, PyTorch, Flask, Django (for model deployment & web integration)	Statistical & Analytical Methods <ul style="list-style-type: none">◦ Regression Analysis◦ Statistical Modeling Specialized Areas <ul style="list-style-type: none">◦ Data Mining: Classification, Clustering, Regression, Artificial Neural Networks
--	---	--

EXPERIENCE

Zucol, Jaipur— Data Analyst March 2023 – June 2024

- Conducted exploratory data analysis (EDA) using **Pandas**, **NumPy**, and **Matplotlib**, identifying key trends and patterns.
- Automated data extraction from **databases** with **Python** scripts, streamlining data processing workflows.
- Cleaned datasets by handling outliers, missing values, and discrepancies to ensure data quality.
- Developed predictive models using **Scikit-Learn**, facilitating data-driven decision-making.
- Collaborated with cross-functional teams to translate data insights into actionable strategies, enhancing business outcomes.

Samyak IT Solutions Pvt Ltd — Data Science (Intern) April 2022 – August 2022

- Performed **data collection and cleaning** processes to ensure the accuracy and completeness of records using **Python** and **SQL**.
- Conducted extensive **data visualization** and **statistical analysis** with **Power BI** and **Seaborn**.
- Generated comprehensive **reports** and **dashboards** to provide insights.
- Enhanced operational efficiency, contributing to a 50% growth in performance through data-driven insights.

PROJECTS

Walmart Prediction

[data_science_work/4N-New_walmart_pred.ipynb](#) at main · Kavita-Shantiwal/data_science_work · GitHub

- Collected and preprocessed data using Python libraries like **Pandas** and **NumPy**.
- Applied **supervised learning** techniques with a **Random Forest** model for classification.
- Achieved 87% **accuracy** in the final predictions.
- Utilized **Scikit-learn** for model implementation and evaluation.

Dragon Real Estate Housing Price Prediction

[data_science_work/Dragon.ipynb](#) at main · Kavita-Shantiwal/data_science_work (github.com)

- Utilized **Pandas** and **Scikit-learn** to preprocess and clean housing data, handling missing values and feature scaling.
- Implemented a **machine learning** pipeline with RandomForestRegressor for predicting house prices.
- Conducted **train-test split** and stratified shuffle split for model validation.
- Achieved a cross-validated RMSE of 3.49 and saved the trained model using Joblib.
- **Deployed model using Flask API** to serve predictions as a web service.

Covid prediction

[data_science_work/Covid_prediction.ipynb](#) at main · Kavita-Shantiwal/data_science_work (github.com)

- Developed a predictive model for COVID-19 diagnosis using **Logistic Regression**, **K-Nearest Neighbors**, **XGBoost**, and **Decision Tree** algorithms.
- Achieved an accuracy of approximately 90% on the test dataset.
- Conducted exploratory data analysis (EDA), preprocessing, and model evaluation using **Python** libraries like **Pandas**, **Matplotlib**, **Seaborn**, and **Scikit-learn**.
- Presented model comparison visually using **Plotly**, demonstrating proficiency in data analysis and **visualization** techniques.
- **Developed a Django-based web dashboard** to allow users to input patient data and receive predictions

CERTIFICATE

Data Science & Machine Learning (Samyak Infotech)

EDUCATION

Kanoria PG College, Jaipur- B.Sc. (Maths)