

SHUBHAM PATHANIA

Data Analyst | Data Scientist | Machine Learning Engineer

📞 +91 8219833828

@ shubhpathania619@gmail.com

🌐 <https://www.linkedin.com/in/shubham-pathania-0b4bb8147> 🌐 <https://github.com/ShubhPathania>

📍 Chandigarh

SUMMARY

Data-driven analyst with a proven ability to transform raw data into actionable insights.

Skilled in Python, SQL and machine learning. Expert in visualizing complex data using Excel, Tableau and Power BI to drive strategic decision-making. Seeking a challenging role to contribute to innovative data solutions.

PROJECTS

Zomato Restaurant Clustering and Sentiment Analysis

🌐 <https://github.com/ShubhPathania/Zomato-Restaurant-Clustering-and-Sentiment-Analysis>

Skills: Data Wrangling, EDA, Data Visualization, Text Preprocessing, Sentiment Analysis, Unsupervised Machine Learning, Text Classification, Topic Modeling

- Rigorously cleaned and preprocessed data for accuracy and integrity.
- Deeply explored user reviews and identified key topics using **topic modeling** to understand customer preferences and sentiment.
- Developed four **sentiment analysis** models, with **XGBoost** and **SVC** achieving over **94% accuracy** to extract actionable insights for restaurant improvement.
- Clustered restaurants using **K-means** and **Hierarchical Clustering** for personalized recommendations.

Bike Sharing Demand Prediction

🌐 <https://github.com/ShubhPathania/Bike-Sharing-Demand-Prediction>

Skills: Data Wrangling, EDA, Data Visualization, Hypothesis testing, Statistical Analysis, Supervised Machine Learning

- Conducted comprehensive data quality assurance and EDA, identifying key demand patterns and critical factors influencing bike rentals.
- Validated **hypotheses** on temperature, humidity and holiday influences, optimizing bike-sharing operations.
- Implemented **feature engineering** and **categorical encoding** to enhance model performance.
- Developed five regression models, with **Random Forest** achieving over **89% predictive accuracy** in predicting Seoul's bike rental demand based on weather and seasonal factors.

Product Dissection for Netflix

🌐 <https://github.com/ShubhPathania/Product-Dissection-for-Netflix>

Skills: SQL, ER Diagram, Schema Design, Case Study

- Analyzed Netflix's business model, outlining key components and their interactions with users, content and technology.
- Explored Netflix's recommendation system, identifying key algorithms and data points.
- Developed a data **schema** and **ER diagram** for efficient user, content and engagement management.

Play Store App Review Analysis

🌐 <https://github.com/ShubhPathania/Play-Store-App-Review-Analysis>

Skills: Data Wrangling, Exploratory Data Analysis, Data Visualization, Pandas, Numpy, Matplotlib, Seaborn, Data Analysis

- Cleaned and analyzed Google Play Store app data, uncovering insights into competitive categories, app size impact and user sentiment.
- Unveiled hidden patterns through advanced data visualization, enabling data-driven decision-making.
- Utilized **Tableau** to create interactive, **user-friendly visualizations** and **Dashboards** to explore and understand key trends and insights.
- Delivered actionable insights and recommendations through compelling data storytelling to drive apps success.

EDUCATION

Master of Computer Science in Data Science and Artificial Intelligence

Woolf University

📅 2023 - Present 📍 CA, USA

Master of Science in Biotechnology

Shoolini University

📅 2017 - 2019 📍 Solan, Himachal Pradesh

Bachelor of Science in Biotechnology

Chandigarh University

📅 2014 - 2017 📍 Mohali, Punjab

EXPERIENCE

Quality Control Officer

Glenmark Pharmaceuticals Limited

📅 11/2020 - 05/2023 📍 Nalagarh, Himachal Pradesh

CERTIFICATION

Advanced Certification in Full Stack Data Science and Artificial Intelligence

Indian Institute of Technology, Guwahati

One Month Skill Development Course on Machine Learning

Indian Institute of Technology, Mandi

TECH STACK

Expertise in Languages & Tools

Python, SQL, Tableau, Power BI, Advance Excel

Platform

Jupyter Notebook, Google Colab, VS Code, pgAdmin, DbVisualizer, Google Sheets

Framework

Numpy, Pandas, Matplotlib, Seaborn, SciPy, Plotly, Scikit-Learn, NLTK

KEY ACHIEVEMENTS



HackerRank

Earned 4 stars in Python & SQL