**Internship Completion Report**  
**Name:** Daud Shah  
**Internship Program:** Data Science Internship  
**Company/Institution:** Cognifyz  
**Duration:** 1 month

**📌 Internship Overview**

This internship aimed to gain hands-on experience in **data analysis, visualization, and machine learning**. The internship focused on cleaning, analyzing, and extracting insights from restaurant data using various data science techniques.

**🔹 Level 1: Exploratory Data Analysis (EDA)**

**✔ Task 1: Data Cleaning & Preprocessing**

* Handled missing values and outliers in the dataset.
* Detected and corrected **corrupted text data** (e.g., encoding issues).
* Converted categorical variables into numerical format.

**✔ Task 2: Descriptive Analysis**

* Explored dataset statistics using **mean, median, mode, and standard deviation**.
* Identified trends in restaurant ratings, cost, and votes.

**✔ Task 3: Geospatial Analysis**

* Mapped restaurant locations using **latitude and longitude**.
* Analyzed city-wise distribution of restaurants and their ratings.

**🔹 Level 2: Feature Engineering & Price Analysis**

**✔ Task 1: Table Booking & Online Delivery**

* Calculated the percentage of restaurants offering **table booking and online delivery**.
* Compared average ratings of restaurants with and without these services.
* Analyzed how price range affects the availability of online delivery.

**✔ Task 2: Price Range Analysis**

* Identified the most common price range among restaurants.
* Calculated **average ratings for each price range**.
* Determined the **rating color associated with the highest average rating**.

**✔ Task 3: Feature Engineering**

* Extracted new features such as **restaurant name length and address length**.
* Encoded categorical variables like **'Has Table Booking' and 'Has Online Delivery'** for better model performance.

**🔹 Level 3: Machine Learning & Data Visualization**

**✔ Task 1: Predictive Modeling**

* Built a **Linear Regression model** to predict restaurant ratings.
* Tested **Decision Tree and Random Forest models**, where **Random Forest performed best (R² = 0.94)**.
* Identified **key factors influencing restaurant ratings** (price range, votes, online delivery).

**✔ Task 2: Customer Preference Analysis**

* Analyzed **which cuisines received the highest ratings**.
* Identified the **most popular cuisines based on votes**.
* Found that certain cuisines tend to receive higher ratings than others.

**✔ Task 3: Data Visualization**

* Created **histograms, bar plots, pie charts, and scatter plots** to visualize data.
* Compared **ratings across different cuisines and cities**.
* Analyzed relationships between **price, votes, and table booking with ratings**.

**🎯 Key Achievements & Skills Improved**

**✅ Technical Skills Gained:**

✔ **Data Cleaning & Preprocessing** (handling missing values, encoding issues)  
✔ **Exploratory Data Analysis (EDA)** using Pandas & Matplotlib  
✔ **Geospatial Analysis** (mapping locations with latitude/longitude)  
✔ **Feature Engineering** (extracting and encoding new features)  
✔ **Predictive Modeling** (Linear Regression, Decision Trees, Random Forest)  
✔ **Data Visualization** using Seaborn, Matplotlib, and Folium  
✔ **Outlier Detection & Handling**  
✔ **Statistical Analysis & Correlation Studies**

**✅ Soft Skills Gained:**

✔ **Problem-Solving & Critical Thinking**  
✔ **Data Interpretation & Decision-Making**  
✔ **Report Writing & Documentation**  
✔ **Time Management & Task Prioritization**

**📌 Final Summary & Conclusion**

This internship provided a strong foundation in **data science and analytics**. The tasks covered **real-world data analysis, machine learning, and data visualization techniques**, preparing me for professional roles in data science.

With a strong grasp of **data preprocessing, feature engineering, model building, and visualization**, I am now well-equipped to handle **complex datasets and extract meaningful insights**.

✅ **Internship Successfully Completed!** 🎉