DATAQUEST: Rules & Guidelines

Thank you for registering for **DataQuest**, a competition designed to challenge your **data cleaning, SQL querying, and data visualization skills**! Please read the following rules and guidelines carefully to ensure a smooth experience.

Event Overview & Timeline

Round 1: Data Cleaning & Dashboard Creation (Online)

📅 **Date:** 20th March 2025
⏰ **Time:** 3:00 pm– 8:00 pm(5 hours, including submissions)
🖥️ **Mode:** Online
🔹 **Tasks:**

- **Data Cleaning** (25 Points)
- **SQL Query Challenges** (Bonus - 25 Points)
- **Data Visualization Dashboard** (50 Points)

Top-performing teams from Round 1 will qualify for the final round.

Round 2: Presentation (Offline)

📅 **Date:** 22nd March 2025
⏰ **Time:** 12:00 pm - 2:00 pm
📍 **Mode:** Offline (UVCE Campus)
🔹 **Finalists must present:**

- **Data Cleaning Approach**
- **Dashboard Insights and innovation**
- **Extra Features** (if any)

Finalists will be expected to explain their approach, insights, conclusions. They may also be questioned on potential improvements for the platform.

Rules & Guidelines

1. Submission Format

Each team must submit the following:

1. **Cleaned Data** – Upload the cleaned datasets or database file to Google Drive.
2. **Dashboard Submission** – Share **one** of the following:
   - A **hosted website link** (if using web technologies).
   - A **Power BI/Tableau link** or a **Google Drive file link**.
   - A **short recorded walkthrough** (if hosting is not possible).
3. **Source Code** – Provide a **GitHub repository** containing:
   - Data cleaning scripts (Python, SQL, Jupyter Notebooks, etc.).
   - SQL queries used for analysis (Optional as it is bonus)
   - Dashboard code **(if using web technologies)**

---

2. Data Files & Tools

- For **data cleaning and dashboard creation**, participants will receive **CSV files and an SQLite (.db) database**, allowing flexibility in data processing. They can use any tool of their choice.
- For the **SQL bonus challenge**, participants will be provided **a separate clean SQLite (.db) file** containing **a dataset with different values but the same database schema** for all teams. This ensures fairness while preventing direct copying of solutions.
- **Additional relevant features may bring extra points**, encouraging participants to go beyond the basic requirements.

---

3. Technology & Preparation

- Teams are advised to finalize their tech stack (**Python, SQL, JavaScript, Power BI, etc.**) before the event. Keep in mind you will be judged based on your technical choices as well.
- Ensure your system is set up with the required libraries, database connectors, and software beforehand.
- **For SQLite files:** Participants should be familiar with SQLite commands and tools to query `.db` files.
- **For hosted submissions:** Teams can use free hosting platforms like **GitHub Pages, Vercel, Streamlit, Flask on Render, etc**.

4. Scoring Criteria

Round 1: Online Challenge (100 Points)

- **Data Cleaning (25 Points)**
  - Handle inconsistencies like duplicate entries, NULL values, incorrect pricing, etc.
  - Maintain referential integrity.
  - Optimize the cleaning process for efficiency.
- **SQL Query Challenges (Bonus - 25 Points)**
  - Participants will solve **6 SQL queries** on a provided .db file.
  - Each correct query earns points based on **efficiency and accuracy**.
- **Dashboard creation (50 Points)**
  - Clarity & effectiveness of visuals
  - Creativity & uniqueness in dashboard design
  - Insightfulness of data analysis
  - Proper use of the cleaned dataset
  - Proper quantity and quality of visuals
  - Any extra relevant features

Round 2: Presentation (50 Points)

Finalists must present:

- Their **data cleaning approach** (How they handled inconsistencies and improved data quality).
- **Key insights from their dashboard** (Major trends and patterns observed).
- **Potential areas of improvement** identified through data (e.g., user engagement trends, subscription behavior, content performance).
- **Justification for their technical choices** (Why they used specific tools, techniques, and visualizations).
- **Q&A Session:** Judges may ask teams about potential platform improvements based on their findings. Teams may also choose to include these suggestions in their presentation.

5. Team Guidelines

- **Team Size:** 1-3 members.
- Each team must ensure smooth coordination and fair contribution from all members.
- One member per team must fill out the provided Google Form for submission.
- Teams are advised to allocate at least **10 minutes before the deadline** to complete the form and ensure all required files are correctly uploaded.

6. Fair Play & Integrity

- Any form of **plagiarism, using AI for querying and cleaning,** or using **pre-built dashboards** without actual work on data cleaning will result in **disqualification**.
- Unethical practices (e.g., copying queries, sharing solutions) will result in **immediate disqualification**.

AI Policy: Participants must perform all data cleaning, SQL queries, and dashboard creation manually. The use of AI tools **(e.g., ChatGPT, Copilot)** for query generation, data processing, or dashboard design is strictly prohibited. Judges may verify submissions for originality, and any AI-generated work will **result in disqualification**.

We encourage everyone to be honest in their work and embrace the true spirit of DataQuest. This event is not just about competition but also about learning, teamwork, and having fun. As the saying goes,

"The strength of the team is each individual member. The strength of each member is the team." **Give it your best, collaborate well, and enjoy the challenge!**

We hope you have a great time analyzing, visualizing, and uncovering insights in **DataQuest**! Give it your best shot, and most importantly, enjoy the process. **All the best!**

For any queries, contact:
Anagha S Shastri: +91 94803 35832
Sumanth G: +91 78924 98948
Varsha Shubhashri: +91 80505 02782