Chatgpt Prompt for Dataset Generation

Please create a spreadsheet with **100,000 rows** and **20 columns** based on the specifications below. The data should represent one month in the year **2024** and include realistic variations and correlations among columns.

General Instructions:

1. **Booking ID**:
    - Ensure all Booking IDs are unique, except for 3 rows where the same Booking ID is repeated.
    - Format: 10-digit IDs starting with "CNR" followed by digits (e.g., CNR12345678).
2. **Dates and Time**:
    - **Date**: Use the format `YYYY-MM-DD`. Spread bookings unevenly across the month, with higher bookings at the start and end of the month.
    - **Time**: Use the format `hh:mm:ss`.
3. **Columns**:
    - **Booking Status**:
        - Categories: `Complete`, `Cancelled by customer`, `Cancelled by Driver`, `Driver not found`.
        - Ensure **62%** of bookings are successful (`Complete`).
        - **Customer cancellations** should not exceed **7%** and **Driver cancellations** should not exceed **18%**.
        - Keep **Incomplete Rides** under **6%**.
    - **Vehicle Type**:
        - Categories: `Auto`, `Prime Plus`, `Prime Sedan`, `Mini`, `Bike`, `eBike`, `Prime SUV`.
        - Keep the number of bookings unevenly distributed among vehicle types.
    - **Pickup Location** and **Drop Location**:
        - Use 50 areas in Pune (e.g., Kothrud, Shivajinagar).
    - **Average VTAT (Vehicle Time to Arrive)** and **Average CTAT (Customer Time to Arrive)**:
        - Only include values for successful bookings.
    - **Cancellation Reasons**:
        - **Customer Cancellation Reasons**:
            - Driver is not moving towards pickup location: **24%**.
            - Driver asked to cancel: **20%**.
            - AC is not working (only for 4-wheelers): **3%**.
            - Change of plans: **22%**.
            - Wrong Address: **31%**.
        - **Driver Cancellation Reasons**:
            - Personal & Car-related issues: **20%**.

        - Customer-related issue: **40%**.
        - Customer was coughing/sick: **14%**.
        - More than permitted people: **26%**.
    - **Incomplete Rides Reasons**:
        - Categories: `Customer Demand`, `Vehicle Breakdown`, `Other Issue`.
    - **Booking Value**:
        - Keep **70%** of bookings under ₹500.
        - Keep **28%** of bookings between ₹500 and ₹700.
        - Keep the remaining **2%** above ₹700.
        - Higher booking values should occur on weekends.
    - **Ride Distance**:
        - Use values up to 1 decimal point.
        - Assign higher distances to four-wheelers, with varying proportions for each vehicle type.
    - **Driver Ratings**:
        - Values: 1 to 5 (in uneven proportions).
            - **40%**: 3
            - **25%**: 4
            - **20%**: 1
            - **10%**: 5
            - **5%**: 2
        - Include ratings only for successful rides.
    - **Customer Ratings**:
        - Values: 1 to 5 (in uneven proportions).
            - **37%**: 5
            - **23%**: 4
            - **16%**: 3
            - **15%**: 1
            - **9%**: 2
        - Include ratings only for successful rides.
    - **Payment Method**:
        - Categories: `Cash`, `UPI`, `Card`.
        - Keep higher proportions for `Cash` and `UPI`.
4. **Other Requirements**:
    - Increase bookings on weekends and match days.
    - Correlate higher booking values with longer ride distances and four-wheelers.
    - Ensure variation in cancellation reasons and booking numbers by vehicle type.

Final Note:

Use the exact column names provided below:

1. Date
2. Time
3. Booking ID

4. Booking Status
5. Customer ID
6. Vehicle Type
7. Pickup Location
8. Drop Location
9. Avg VTAT
10. Avg CTAT
11. Cancelled Rides by Customer
12. Reason for Cancelling by Customer
13. Cancelled Rides by Driver
14. Reason for Cancelling by Driver
15. Incomplete Rides
16. Incomplete Rides Reason
17. Booking Value
18. Ride Distance
19. Driver Ratings
20. Customer Ratings
21. Payment Method