Assumptions for Dataset Generation

1. Each episode starts with an authorized access card swipe and ends with the door closing event, if applicable.
2. This dataset simulates 1000 episodes, each representing a unique door access event.
3. **Tailgating** is defined as an event in which an authorized user is followed by multiple users without additional swipes.
4. Tailgating detection by the system is a rule-based classifier, resulting in false positives.
5. The human feedback column serves as the ground truth label, manually indicating if tailgating has occurred.
6. Multiple scenarios were included:
    a. **Normal entry (true negative)**: One person swipes and enters, closing the door behind themselves. **(~50%)**
    b. **Tailgating (true positive)**: Multiple entries without swiping after authorized access.**(~ 6%)**
    c. **Guest Entry (false positive)**: Employee invites a guest; system flags it. **(~17.5%)**
    d. **Janitor + Cart (false positive)**: Cart misclassified as human; system flags it. **(~10%)**
    e. **Door held open (false positive)**: Authorised user holds the door; multiple entries. **(~12.5%)**
    f. **Door left open (outlier)**: Janitor forgets to shut the door; episode ends without the door closed. **(~4%)**
7. Episode duration ranges vary by scenario:
    a. **Normal:** 3-10 seconds
    b. **Tailgating**: 5-30 seconds
    c. **Guest/Janitor:** 15-20 seconds
    d. **Door held open:** 15-60 seconds
    e. **Door left open:** 180-200 seconds
8. All the activity is simulated throughout the week. Peak of the activity is between 9-6 on weekdays, with sparse activity occurring round the clock.
9. Data is structured in a multi-row/episode format.
10. The final generated data can be visualized as follows: