

## # Background

iRobot has a series of wifi-connected robotic vacuum cleaners available for sale worldwide. These robots are capable of autonomously navigating a home to vacuum its floors. Upon mission completion, they send a summary report of the mission to cloud services, where it is processed and stored as a row in a **Postgres database**. However, any cleaning mission performed while the robot is not connected to wifi (either by user's choice or a faulty connection) will not be saved in the database. In addition, there are occasional periods where cloud services malfunction and no missions are reported, resulting in **discrete periods of data loss**.

These robots are programmed with an automatic recharge and resume function, which means that when the robot detects its **battery levels reaching critically low levels**, it will navigate back to the charging dock if available and **charge for up to 90 minutes before resuming the mission**. In addition, **if a robot becomes stuck on an obstacle in its environment or is manually paused by a button press, it will cease cleaning for up to 90 minutes before terminating the mission**. **If the user restarts the mission with a button press within 90 minutes of the pause, the robot will continue cleaning normally**. The number of minutes spent cleaning, charging, or paused are reported for each mission, as is the mission outcome (a field describing whether the mission was cancelled, the robot got stuck, the battery died, or the robot completed the job successfully).

## # Data

### 1. mission\_data.csv.bz2

This table contains details of cleaning missions for a sample of 10,000 wifi-connected robots

The columns are defined as follows:

\* \*\***robotid**\*\* : a unique robot identifier

\* \*\***datetime**\*\* : a date string that represents the **start time of a mission in GMT**

\* \*\***nmssn**\*\* : mission number. This information comes from an internal counter on the robot that increments +1 per mission. **Be aware that the complete mission history from mission 1 may not be included for each robot (due to missions being run before the robot was connected to wifi or data loss)**. The max mission number per robot should reflect its total number of missions to date reported to the database.

\* \*\***runm**\*\* : this is the **time in minutes that the robot spent actually cleaning**.

\* \*\***chrgm**\*\* : this is the **time in minutes that a robot spent charging**.

\* \*\***pausem**\*\* : this is the **time in minutes that a robot spent paused**.

\* \*\***outcome**\*\* : this is the end status of a mission. **"Cncl"** indicates that the **mission was cancelled** by the user. **"Stuck"** means the robot got **stuck on an obstacle, and was not rescued within 90 minutes**, so

could not return to the dock. "Bat" means the robot's battery grew too low for it to return to the dock. "Ok" means the robot successfully completed cleaning the space and returned to the dock.

## 2. geo\_data.csv.bz2

This table contains details of the robot's geographic location.

The columns are defined as follows:

- \* \*\*robotid\*\*\*: unique robot identifier
- \* \*\*country\_cd\*\*\*: 2-letter ISO country code
- \* \*\*timezone\*\*\*: robot's timezone (from IANA/Olson database)

## # Tasks

Perform data analysis exploring use patterns of the typical robot user per country. Include relevant visualizations where appropriate, and address any possible effects of data loss on your findings.

### 1. Are there geographic differences in robot usage?

- Consider all descriptive features of a mission, including when and how frequently it occurred.
- If applicable, comment on how trends in these features might impact design decisions for the hardware, battery, or navigation algorithms of robots sold in different regions.

### 2. Calculate the time between consecutive missions for each robot ("inter-mission interval" or "IMI"). Describe any interesting relationships between IMI and other features of robot behavior.

### 3. \*\*BONUS\*\*: We are aware that data loss exists among the mission records, but are unsure of the cause. Quantify the extent of the loss, differentiating between discrete catastrophic events and random mission loss for individual robots. Investigate whether this loss is uniform or whether it may be impacting other analyses.

## # Tools and Submission

Please perform all data analysis in Python (either v2 or v3 is acceptable). You are welcome to use any Python libraries you want. Your final submission should be a single document, either a Jupyter notebook with inline figures or a PDF that includes both write-ups and figures. Be sure to describe your hypotheses, methods, reasoning, and findings (including null findings) in addition to the code required to answer each question. You should expect this task to take 4-5 hours.