

## 数据格式相关说明

### ts (时间戳)

**解释:** 这个字段表示记录的时间戳，记录了数据采集的具体时间，精确到毫秒。时间戳是时间序列数据的核心字段，帮助我们了解数据随时间的变化趋势。

### device\_sn (设备编号)

**解释:** 该字段表示设备的唯一标识符，用于区分不同的设备。每个设备都有一个唯一的设备编号 (device\_sn)。

### region (区域)

**解释:** 表示设备所在的地理区域或实验室。例如，501- 5G 通信联合创新实验室 是某个特定区域的设备所在地点。

### 电压 (Voltage)

**解释:** 电压是设备的电源电压，表示电流驱动设备工作的电势差。电压对设备的运行至关重要，过高或过低的电压都会影响设备的正常工作。

### 电流 (Current)

**解释:** 电流是设备通过电导体流动的电荷量，通常与设备的功耗相关。电流的变化可以反映设备负载的变化。

### 色温 (Color Temperature)

**解释:** 色温是光源发出的光的颜色属性，通常用于描述光的“暖”或“冷”。色温较高（如 5000K 以上）表示光源偏蓝色，而较低的色温（如 2700K）则表示偏黄色。

### 亮度 (Brightness)

**解释:** 亮度表示设备或环境的光照强度。亮度较高时，表示环境光较强，反之则表示光线较弱。

### 温度 (Temperature)

**解释:** 温度是设备或环境的热量状态。设备的温度过高可能表示过载或故障，因此监测温度是设备维护和管理的一部分。

### 湿度 (Humidity)

**解释:** 湿度表示环境中水蒸气的含量，较高的湿度可能影响设备的正常工作，甚至导致设备故障。湿度监测可以帮助判断设备是否处于合适的工作环境。

**人体红外感应;1-有人;0-无人 (Human Infrared Sensing; 1 - Presence; 0 - Absence)**

**解释:** 该字段记录了人体是否被红外传感器检测到。1 表示有人，0 表示无人。它用于监测房间或区域内是否有人活动。

**状态;1-开;0-关 (Status; 1 - On; 0 - Off)**

**解释:** 表示设备的工作状态。1 表示设备开启，0 表示设备关闭。这个字段可以帮助分析设备在不同时间的工作和休眠状态。

## 数据预处理

因为我们的数据量比较多，数据的维度也比较多，所以在对数据进行预处理之后我们增加了一步降采样。

**数据清洗**的目的是提高数据质量和一致性，包括处理缺失值、重复数据、异常值、格式问题等。

**降采样**的目的是减少数据量，常见的降采样方法包括随机降采样、分层抽样、聚类降采样、时间序列降采样和欠采样等。

**利用 Python 处理缺失值并删除含有缺失值的行**

缺失值是数据清洗中的一个重要部分。通过删除包含缺失值的行，可以确保后续分析的准确性。

**步骤：**

1. 使用 Pandas 读取数据，并查找包含缺失值的行。
2. 删除这些包含缺失值的行，确保数据完整性。

**按照时间进行数据排查，删除重复数据并进行异常值检测**

在数据清洗中，时间排序和去重是确保数据一致性的重要步骤。同时，异常值的检测帮助识别数据中的不合理值。

**步骤：**

1. 将数据按时间进行排序，确保时间顺序正确。
2. 删除重复的行，确保每条记录的唯一性。
3. 使用统计方法检测异常值并删除。

**进行时间序列降采样操作**

由于数据的时间间隔很短（每分钟一次采样），需要通过降采样来减少数据量，每隔四行进行一次采样，以便优化处理效率。

**步骤：**

1. 将数据按时间顺序排列（如果还没有）。
2. 使用 `.iloc` 索引每四行数据并选择一个值进行降采样。
3. 根据需求选择平均值、最大值、最小值等来进行采样。

## 可视化

**图 1：多指标时间变化**

图 1 的目的是展示该区域内几个关键设备参数（电压、电流、温度、湿度等）随时间变化的趋势，以及设备间的一些关联性。

- **左上图（电压变化）：**展示了电压随时间的变化情况。可以用来观察设备是否在稳定的电压下运行，或者电压波动是否与设备的行为相关。
- **右上图（电流变化）：**展示了电流随时间的变化。电流通常与设备的负载有关，因此该图有助于分析设备负载的变化。
- **左下图（温度变化）：**展示了温度随时间的变化。可以帮助识别设备是否过热，或者环境条件（如温度）的变化是否对设备运行产生影响。
- **右下图（湿度变化）：**展示了湿度随时间的变化。湿度的变化可能对设备的运行产生影响，尤其是在敏感的设备中。
- **底部左图（电压 vs 电流）：**这个散点图展示了电压与电流之间的关系。通过这个图，可以分析电压与电流之间的相关性。例如，是否电流增加时电压也随之增加。
- **底部右图（人体红外感应次数）：**使用 countplot 绘制人体红外感应的频次。这里显示了在不同时间段内（例如：有人的时间段 vs 无人的时间段）感应的次数，帮助分析该区域设备与人体活动的相关性。

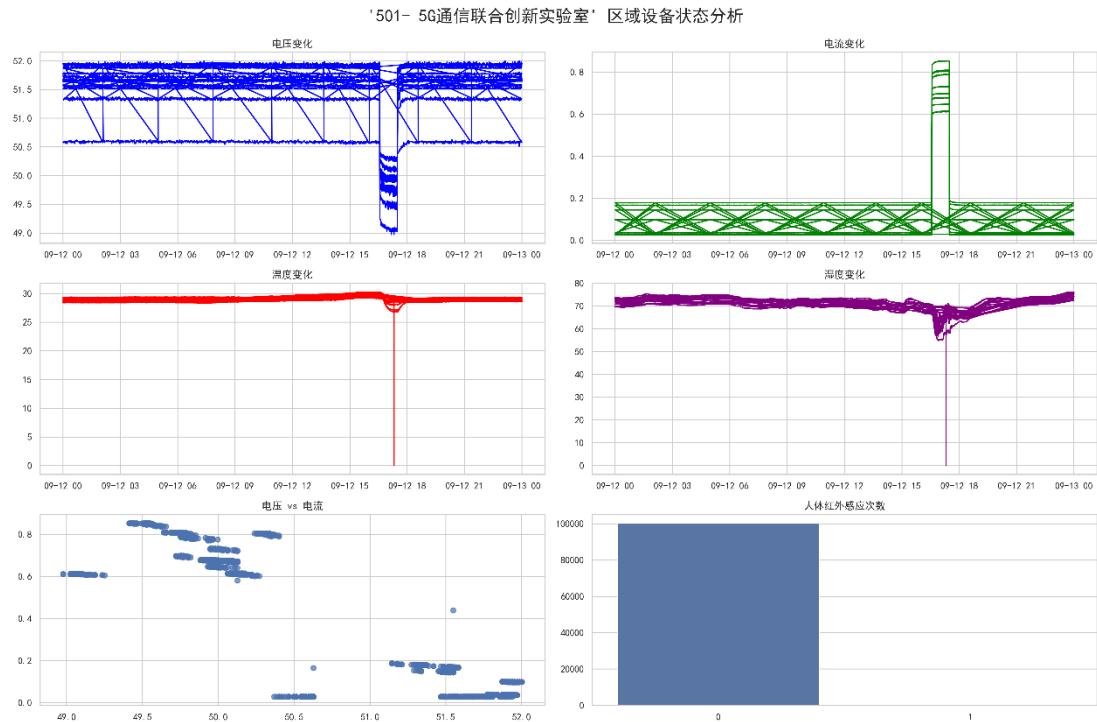


图 2：区域-日期感应热力图

图 2 的目的是展示该区域内设备在不同日期的“人体红外感应”情况，通过热力图直观地查看每个设备在各日期的活动情况。

通过数据透视表 (pivot\_table) 将每个设备在每个日期的“人体红外感应”值聚合起来, 并使用 heatmap 绘制热力图。颜色深浅代表了该设备在某一天感应的强度 (即人体感应的总次数)。深色表示感应强烈, 浅色表示感应较弱。

这个图有助于识别某些设备在特定日期的活动频率, 帮助管理人员了解设备是否在某些日期或时间内有过度的活动, 进而作出相应的调整。

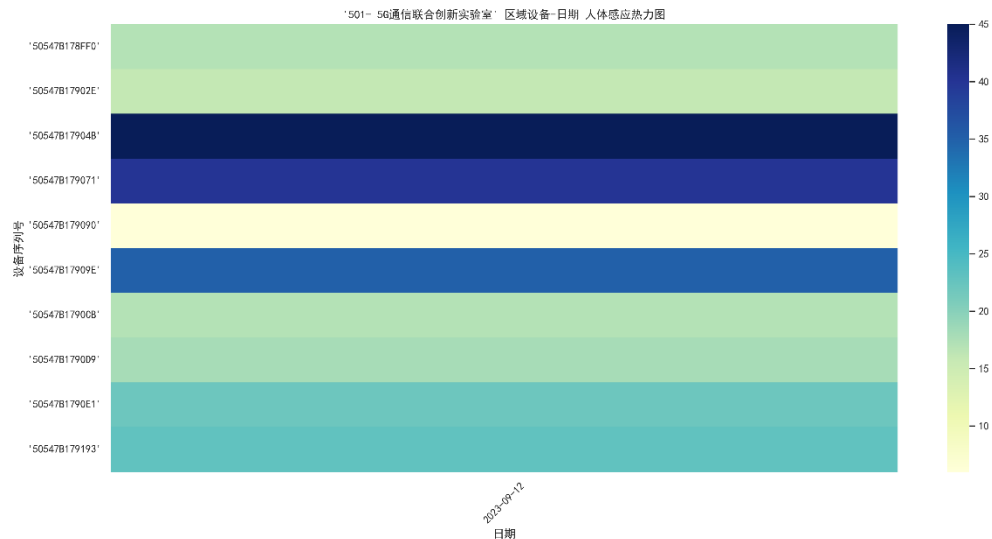


图 3: 区域内设备总感应次数

图 3 的目的是展示该区域内每个设备的总人体感应次数。

通过分组并计算每个设备的“人体红外感应”总和, 使用柱状图(barplot) 展示各设备的总感应次数。

图中的柱子代表每个设备在分析周期内的感应次数。有助于了解哪些设备频繁感应, 哪些设备则较少, 可能存在设备故障或环境条件的影响。

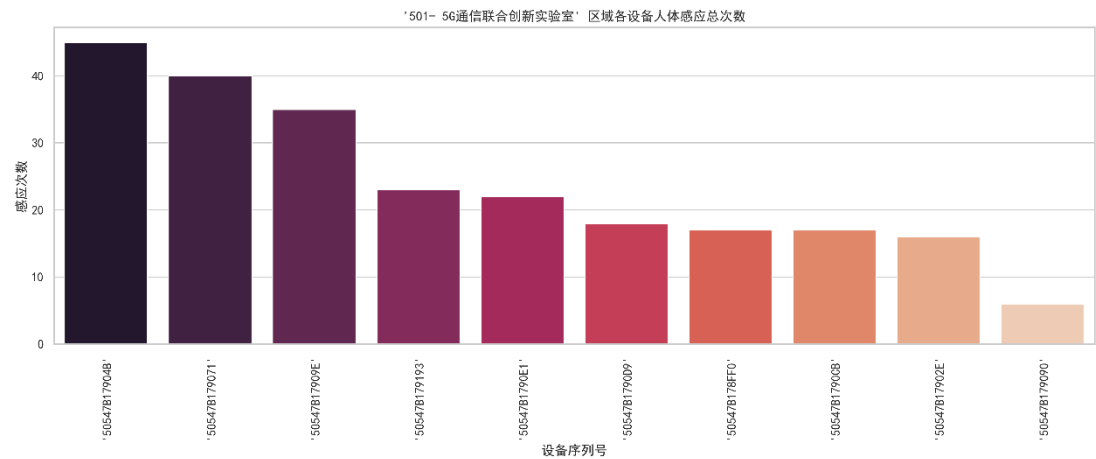
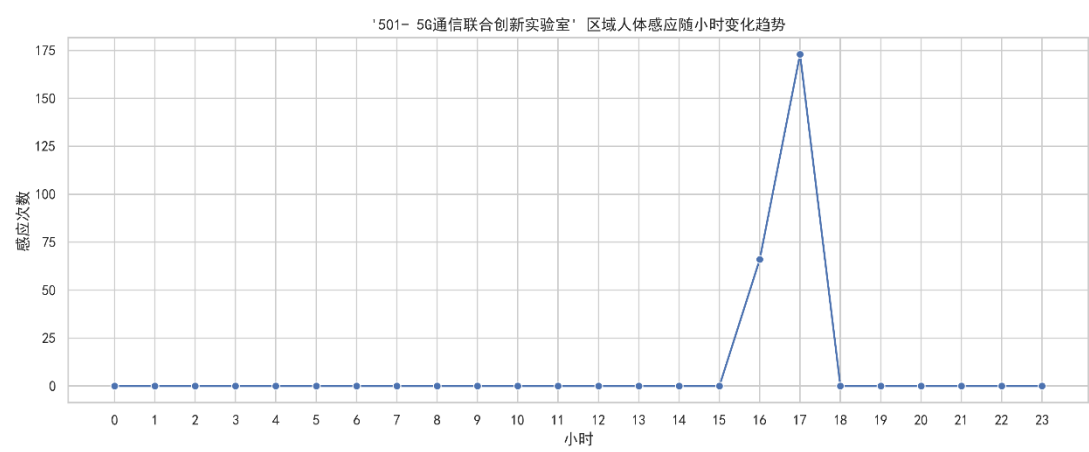


图 4：人体感应按小时变化趋势

图 4 的目的是展示该区域内人体红外感应的小时变化趋势。

该图通过按小时聚合“人体红外感应”数据，绘制了感应次数随小时变化的趋势线。每个小时的感应次数显示在纵轴，小时数（0-23 小时）显示在横轴。

该图可以帮助分析在一天的哪些时间段内设备感应的次数最多或最少。比如，某个区域可能在工作时间段内频繁感应，而在晚上则几乎没有人进入。



总结

图 1：展示了电压、电流、温度、湿度、设备间的关系以及人体红外感应的频率。

图 2：通过热力图展示了设备在不同日期的感应强度，帮助识别日期间的变化模式。

图 3：展示了每个设备的总感应次数，帮助了解设备的活动频率。

图 4：展示了人体感应随小时变化的趋势，揭示了设备的工作模式与时间的关系。

## 5. 交互式时间序列图 (Interactive Time Series Plot with Plotly)

**功能:** 这个图展示了每个区域中几个传感器数据随时间变化的趋势,包括 电压、电流、亮度、温度 和 湿度。

**作用:**

展示这些传感器的数据如何随着时间变化 (例如, 某个设备的电压或电流随时间的波动情况)。

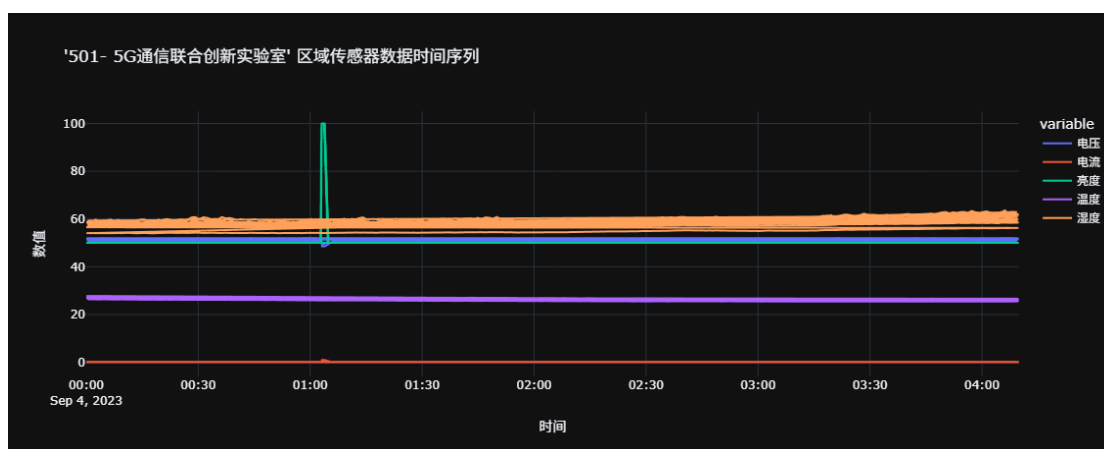
可以帮助分析设备是否在正常的工作范围内, 或者是否有任何异常波动。

用户可以交互式地查看每个时间点的数据变化, 帮助识别任何潜在的设备问题或环境变化。

**图表特性:**

时间 (ts) 在 X 轴上, 传感器数据 (电压、电流、亮度、温度、湿度) 在 Y 轴上。

由于使用了 Plotly, 这个图是交互式的, 可以放大、缩小、查看具体的点数据。



## 6. 交互式热力图 (Interactive Heatmap with Plotly)

**功能:** 这个图通过热力图展示了 电压 在不同时间点 (按日期) 的变化情况。

**作用:**

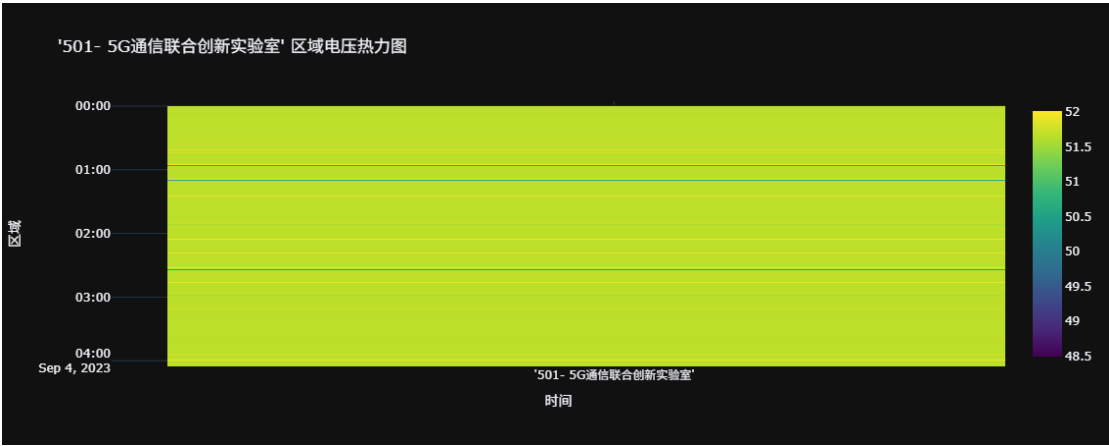
用来观察某个区域在不同日期的电压变化。通过热力图的颜色深浅来表达电压的高低。

热力图是通过数据透视表 (pivot\_table) 生成的, 其中 X 轴是日期, Y 轴是区域 (设备序列号), 颜色则代表电压值。

这个图有助于识别某些特定日期或时段内的异常电压波动，或者设备的异常工作模式。

**图表特性：**

热力图采用 Viridis 配色方案，颜色从浅到深表示电压值的变化，深色表示电压较高，浅色表示电压较低。



**7. 3D 散点图 (3D Scatter Plot with Plotly)**

**功能：**这个图通过 3D 散点图展示了 电压、电流 和 色温 之间的关系，并且 湿度 用颜色表示，亮度 用大小表示。

**作用：**

通过三维散点图展示电压、电流和色温之间的多维关系，颜色和大小分别代表湿度和亮度，帮助分析这些参数之间的潜在关系。

适用于识别电压与电流或其他传感器数据（如色温、湿度）之间的相关性，特别是在高维度下分析变量之间的交互作用。

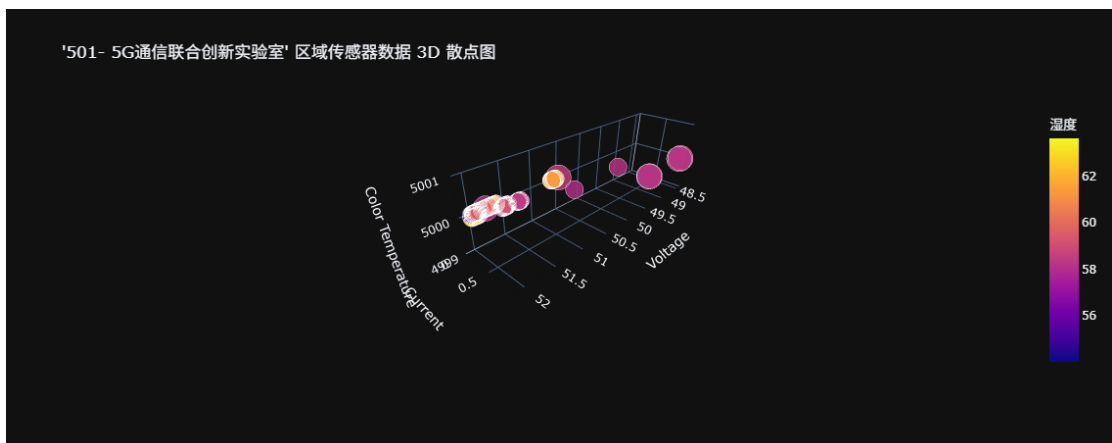
通过 3D 散点图，您可以在空间上更直观地理解这些传感器数据如何相互影响。

**图表特性：**

电压、电流 和 色温 是 X、Y 和 Z 轴，湿度 用颜色表示，亮度 用散点的大小表示。

该图是交互式的，可以通过旋转和缩放来探索数据的不同维度。





## 8. 桑基图 (Sankey Diagram)

**功能：**桑基图用于展示数据流动，特别是显示不同状态之间的转移。

**作用：**

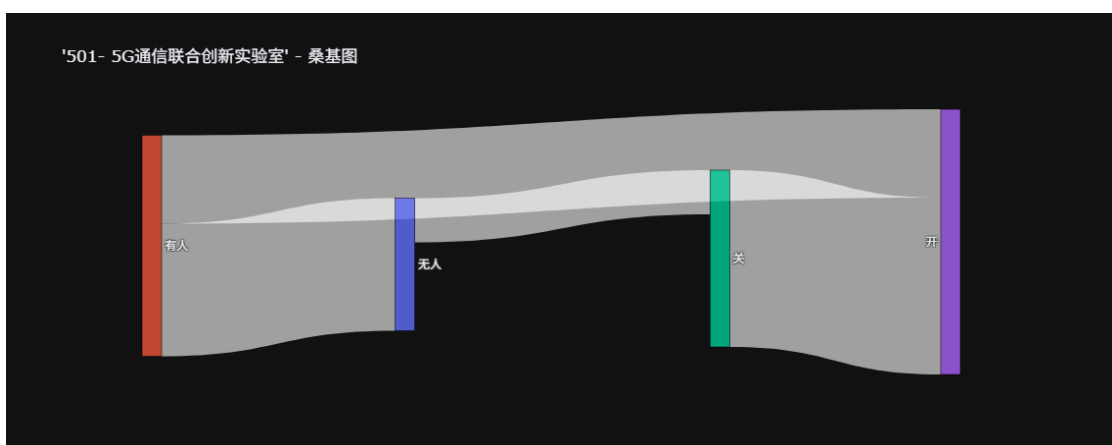
用于分析设备状态（如“开”与“关”，“有人”与“无人”）之间的转移情况。通过桑基图，可以清楚地看到数据是如何从一个状态流向另一个状态的。

在此图中，“无人”到“开”，“有人”到“关”的数据流量可以帮助了解设备的状态转变模式，进而优化设备管理和调度。

**图表特性：**

显示了数据流的源（source）、目标（target）和流量（value）。每个流动箭头的宽度代表数据流的大小，宽度越大，流量越大。

桑基图可以帮助揭示设备从一个状态到另一个状态的常见模式。



## 9. 密度图 (Density Plot with Seaborn)

**功能：**密度图展示了 电压、电流 和 色温 的概率分布。

## 作用：

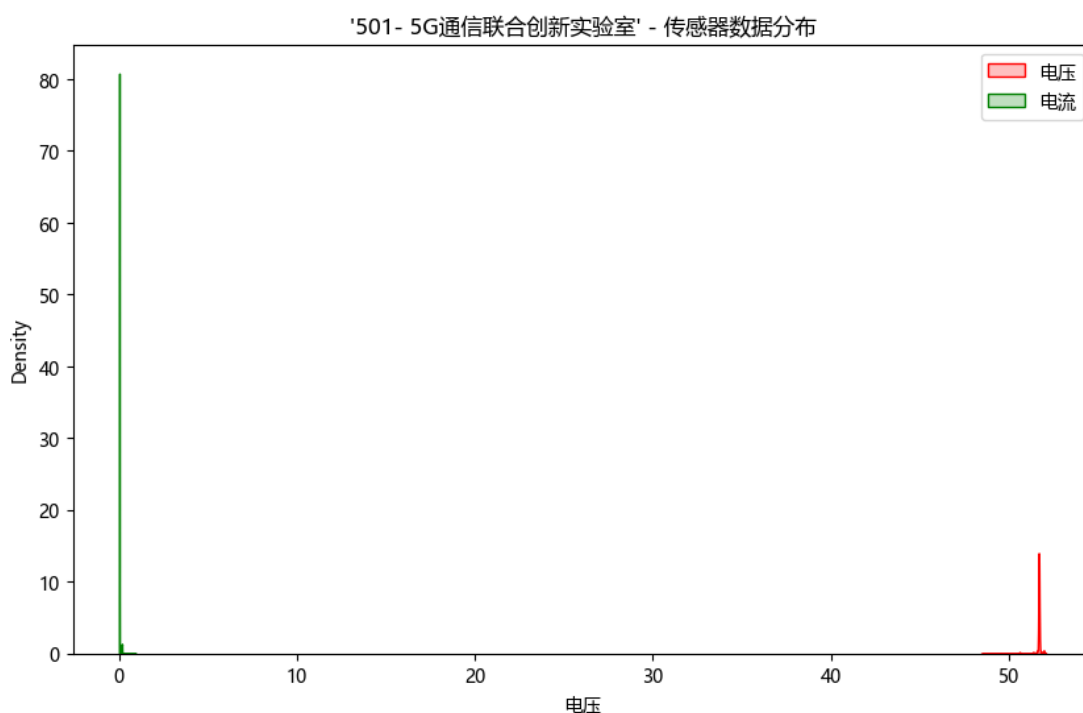
通过密度图，可以查看这些传感器数据的分布情况，识别数据的集中区域、偏态分布或异常值。

该图可以帮助了解设备的工作范围，例如电压通常在什么范围内变化，电流和色温的分布是否有明显的峰值。

## 图表特性：

每条曲线代表不同传感器数据的概率密度，密度较大的地方说明该变量值出现的频率较高。

通过 `seaborn` 绘制的密度图能够帮助您直观地分析数据分布情况，并识别任何异常分布。



## 总结

**交互式时间序列图：**展示不同传感器数据随时间的变化，帮助分析设备的运行状态。

**交互式热力图：**展示电压随时间的变化，帮助了解每个日期的电压波动。

**3D 散点图：**展示电压、电流、色温等传感器数据的多维关系，并通过颜色和大小表达湿度和亮度。

**桑基图：**展示不同设备状态之间的流动，帮助分析状态转变的模式。

**密度图：**展示不同传感器数据的分布情况，识别数据集中和分散的区域。