



Delivery Bot Socket Analysis Report

Huidi “Judy” Yang

June 13, 2025

Statement of Purpose

本报告旨在分析部署于服务机器人充电站和公共电脑终端的智能插座的活动日志，目标是刻画其运行行为并识别随时间变化的使用模式。通过时间序列处理、聚类技术和时间特征分析，我们试图揭示这两种设备状态转换和使用规律背后的结构。对于服务机器人，重点在于理解其充电与非充电状态之间的转换；而对于公共电脑，则需探查更为细致和重叠的能耗与系统活动模式。对比分析凸显了聚类可解释性、时间规律性和设备特定行为方面的差异。这些见解可用于制定调度计划、优化能耗、检测异常，以及在更广泛的智慧设施管理背景下实施预测性维护。

Executive Summary

本报告深入分析了来自两个不同环境的智能插座使用数据：一个是酒店配送机器人，另一个是公共电脑终端。研究的主要目标是揭示运行模式与能耗行为，并为设施层面的智慧能耗管理提供改进方向。

核心方法包括时间序列特征提取、基于 PCA 的降维、聚类（K-Means、DBSCAN 及人工分段）、以及时间使用模式的刻画。聚类方法能够有效识别公共电脑的不同活动状态，而配送机器人则表现出更为极端且规律的行为，难以通过算法聚类，需要人工分段处理。

时间分析显示，公共电脑在深夜和清晨处于稳定的低功耗状态；而机器人则在特定时间段内展现出高度规律化的充电模式。比较结果强调了不同设备角色对其与电力基础设施互动方式的影响。

本研究所获得的见解可用于制定实际的调度方案、优化能耗及异常检测。此外，本报告展示的分析框架——结合特征聚类与时间特征刻画——可扩展应用于其他智能设备，实现实时监测与预测性管理。

Contents

1	About the Data	6
1.1	equipmentName	6
1.2	functionType	6
1.3	time	7
1.4	value	7
2	Data Preprocessing	7
2.1	Timestamp Parsing	7
2.2	Data Reshaping	7
2.3	Filtering for Relevant Equipment	7
2.4	Data Cleaning	8
2.5	Derived Features	8
I	配送机器人分析	9
3	探索性数据分析 (EDA)	9
3.1	相关性分析	9
3.2	功能变量分布	10
3.2.1	Signal Strength (CSQ)	10
3.2.2	Current	11
3.2.3	Last Hour Power Factor	12
3.2.4	Leakage	12
3.2.5	Partial Power Factor (PartPF)	13
3.2.6	Phase Angle	14
3.2.7	Power	14
3.2.8	Power Factor	15
3.2.9	Relay Status	16
3.2.10	Temperature	16
3.2.11	Total Power Factor	17
3.2.12	Voltage	18
3.3	时间序列分析	18
4	基于 K-Means 与 PCA 的聚类分析	20
4.1	主成分载荷与解释	20
4.2	轮廓系数分析	21

4.3	K-Means 聚类结果	22
4.4	状态分布	23
4.5	基于马尔可夫链的状态转移分析	25
5	时间使用模式分析	26
5.1	按时段的使用模式	26
5.2	按两小时区间的使用模式	27
II	公共电脑插座分析	27
6	探索性数据分析 (EDA)	28
6.1	相关性分析	28
6.2	各功能变量的分布	29
6.2.1	Signal Strength (CSQ)	29
6.2.2	Current	30
6.2.3	Last Hour Power Factor	30
6.2.4	Leakage	31
6.2.5	Partial Power Factor	32
6.2.6	Phase Angle	32
6.2.7	Power	33
6.2.8	Power Factor	34
6.2.9	Relay Status	34
6.2.10	Temperature	35
6.2.11	Total Power Factor	36
6.2.12	Voltage	36
6.3	时间序列分析	37
7	聚类分析	39
7.1	通过 PCA 降维	39
7.2	通过轮廓分析选择最佳簇数	39
7.3	K-Means 聚类结果	40
7.4	基于密度的 DBSCAN 聚类	40
7.5	人工聚类	42
7.6	聚类结果解释	44
7.7	时间使用模式	45

III 对比分析：	
配送机器人 vs. 公共电脑	49
8 状态聚类对比	49
9 时间使用模式对比	49
10 可解释性与干预需求	50
IV 结论	51

本报告中引用的所有数据集、产品文档和分析脚本均可在以下 [GitHub 仓库](#) 中获取。本报告共分为五个部分。第一部分概述了整个项目并介绍所使用的数据集。第二部分深入分析了与配送机器人相关的智能插座。第三部分聚焦于连接酒店公共电脑的智能插座分析。第四部分则提供了对这两种设备的对比性评估，突出其使用模式中的异同。最后，第五部分总结了关键发现，并提出可供运维优化的可行建议。

1 About the Data

本数据集由住友酒店的尹必凡提供，应 Alex Hon 的请求整理。数据集包含两个智能插座在不同时间点记录的测量数据。每条记录包括四个变量：equipmentName、functionType、time 和 value。

1.1 equipmentName

该列包含两个唯一标识符：

- Zprime_Socket_01：对应于配送机器人所使用的智能插座。
- Zprime_Socket_02：对应于连接公共电脑的智能插座。

1.2 functionType

插座共记录了 12 种功能类型。每一种对应一个电气或环境参数，其定义可参考设备协议文档（安驿电管家系列产品协议文档-V1.07 1.pdf）：

- CSQ：信号强度百分比（0.00%-100.00%，值越大表示信号越强）
- Current：电流（单位：安培 A）
- LastHourPF：上一小时耗电量（单位：千瓦时 kWh）
- Leakage：漏电流（单位：毫安 mA）
- PartPF：当天当前时刻的累计耗电量（单位：千瓦时 kWh）
- PhaseAngle：相位角（单位：度）
- Power：有功功率（单位：瓦 W）
- PowerFactor：功率因数（无量纲）
- RelayStatus：继电器状态（0 = 断开，1 = 闭合）

- **Temperature**: 温度 (单位: 摄氏度 °C)
- **TotalPF**: 总累计耗电量 (单位: 千瓦时 kWh)
- **Voltage**: 电压 (单位: 伏 V)

1.3 time

记录测量值的时间戳, 采用带有 UTC 偏移的 ISO 8601 格式。

1.4 value

指定 `functionType` 对应的数值读数, 记录于特定时间点。

2 Data Preprocessing

在开展分析之前, 我对原始插座数据进行了预处理, 以便于时间序列分析和聚类建模。

2.1 Timestamp Parsing

原始数据集中时间戳采用 ISO 8601 格式并包含 UTC 偏移。我使用 `pandas.to_datetime()` 将其转换为 `datetime` 对象, 便于提取如小时、星期几、时间段等特征。此外, 为使结果更具本地化和实际意义, 我将所有时间转换为北京时间。

2.2 Data Reshaping

为了便于分析, 我将数据从长格式重塑为宽格式, 使每一行代表一个时间戳, 每一列代表一个功能类型。最终得到的分钟级时间序列包括以下特征: `CSQ`、`Current`、`LastHourPF`、`Leakage`、`PartPF`、`PhaseAngle`、`Power`、`PowerFactor`、`RelayStatus`、`Temperature`、`TotalPF` 和 `Voltage`。

2.3 Filtering for Relevant Equipment

数据集包含两个设备的记录。在后续分析中, 我根据 `equipmentName` (`Zprime_Socket_01` 和 `Zprime_Socket_02`) 对数据进行了筛选, 以分别对应后续章节的分析。

2.4 Data Cleaning

首先，我检查并删除了重复记录。同时确认数据中不存在缺失值，从而保证后续 PCA 和聚类分析在完整数据上进行，无需插补。

2.5 Derived Features

为了支持聚类与行为刻画，我衍生出以下几个变量：

- `duration_min`: 机器人处于某一推断状态的持续时间（单位：分钟），在完成状态标注后计算。
- `time_of_day`: 一天中的大致时间段分类变量（早晨、下午、傍晚、夜间）。
- `2hr_bin`: 以两小时为单位的更细粒度时间分组（如 08:00--10:00）。

以上预处理步骤为从插座数据中提取有意义的行为模式打下了坚实基础。

Part I

配送机器人分析

3 探索性数据分析 (EDA)

在完成数据预处理后，我进行了探索性数据分析，以初步了解数据的结构和分布情况。

3.1 相关性分析

为了识别智能插座各个功能变量之间的潜在冗余性，我计算了所有 `functionType` 变量之间的 Pearson 相关系数矩阵。生成的热力图如图 1 所示。可以观察到若干强正相关关系——例如，`Current` 与 `PhaseAngle`、`Power` 和 `PowerFactor` 高度相关。这三个变量彼此之间也存在强相关性，说明它们可能反映了电力消耗的重叠方面。由于存在多重共线性，后续分析中引入了降维技术。

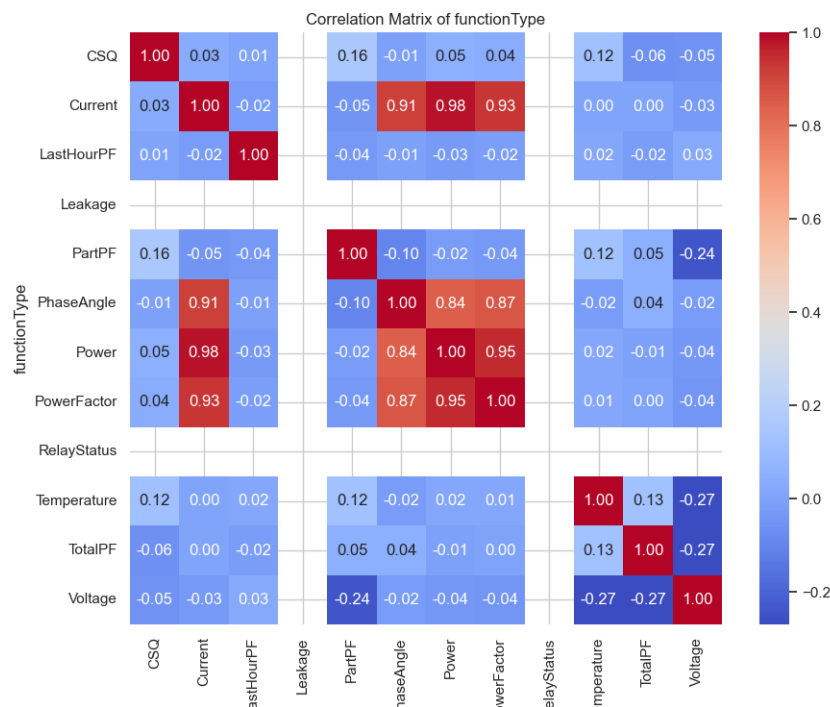


Figure 1: 各功能变量之间的相关性热力图。颜色越暖表示正相关性越强，颜色越冷则表示相关性较弱。

3.2 功能变量分布

为更好地理解配送机器人所使用的智能插座的行为，我对每个记录的电力功能指标进行了分布可视化。这些指标代表关键的电力与信号参数，其中许多可能揭示能耗模式、设备状态转换，或运行异常。

3.2.1 Signal Strength (CSQ)

CSQ 表示蜂窝信号质量，单位为百分比。该指标对于评估无线通信的可靠性尤为关键，尤其是对依赖移动网络进行遥测报告或接收指令的配送机器人而言。

图 2 显示了信号强度的分布。数值呈现多峰分布，主要集中在 65% 至 75% 区间，说明设备通常处于中等到较强信号质量的区域。该范围内的峰值可能对应于设备常驻点（如充电站），而极端值的缺失表明整体连接性较为稳定。

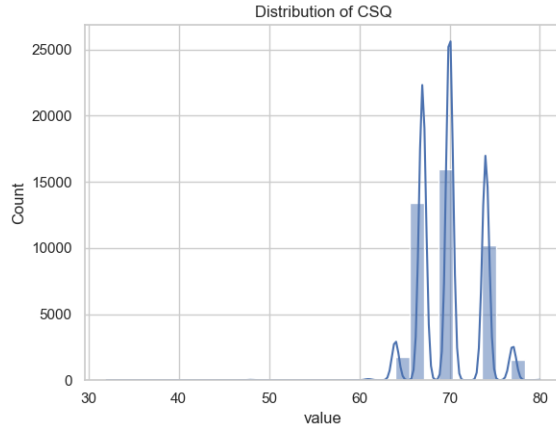


Figure 2: 信号强度（CSQ）分布。

3.2.2 Current

Current 表示插座所承载的电流（单位：安培）。该指标与机器人的能耗直接相关，可用于推断其活动周期，如充电或任务执行。

图 3 所示的分布高度偏斜，大多数值接近零，反映了设备处于长时间的空闲或低功耗状态。靠近 0.6 A 的小群体则可能代表周期性的充电事件，与机器人在空闲状态下停靠充电的行为相吻合。

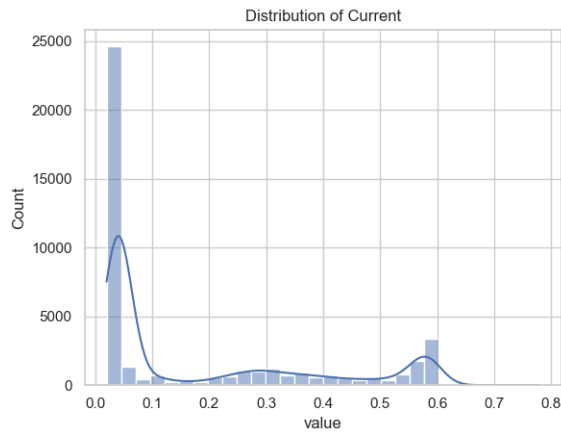


Figure 3: 插座所消耗电流（Current）的分布。

3.2.3 Last Hour Power Factor

LastHourPF 表示过去一小时内的平均功率因数，即能耗效率。值越低表示功率使用效率较差，接近 1 则说明用电较为高效。

从图 4 可见，大多数值集中在 0.02 到 0.06 之间，说明总体电力使用偏低，符合机器人以间歇性运行为主的使用模式。图中的节律性波峰暗示了设备运行的规律性，可能与每日配送任务周期一致。

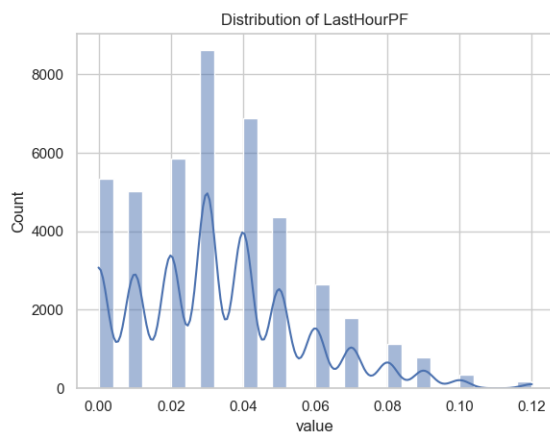


Figure 4: 上一小时能耗（LastHourPF）的分布。

3.2.4 Leakage

Leakage 反映电流是否在电路中异常泄漏，通常与设备安全相关。持续的漏电可能预示绝缘老化或设备故障。

如图 5 所示，所有 Leakage 的值均为 0，说明插座始终在安全电气范围内运行，或漏电监控在此设置中被禁用或不适用。

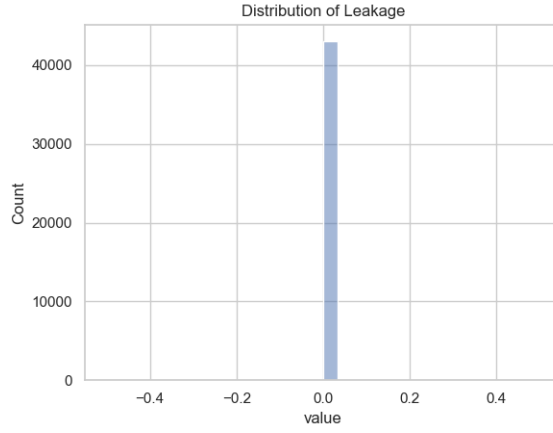


Figure 5: Leakage 分布：所有数值均为零。

3.2.5 Partial Power Factor (PartPF)

PartPF 表示某一短时间窗口下的功率因数估算值，相较 LastHourPF 更具时效性。

图 6 的分布相对平坦，范围为 0 到 0.8，在 0.2 附近略有峰值。这显示设备在日常使用中存在能耗效率的波动，可能与状态切换（如从空闲到充电）相关。

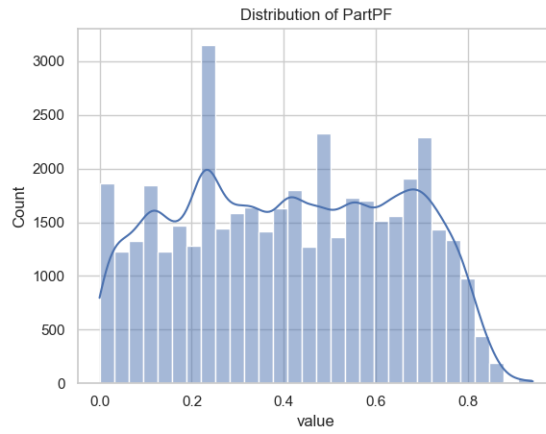


Figure 6: PartPF 的分布。

3.2.6 Phase Angle

PhaseAngle 表示电压与电流波形之间的相位差角度。它反映负载类型的变化，如从电阻性负载切换到感性负载。

图 7 展示了典型的双峰分布，峰值大致位于 115° 和 175° ，说明插座在两种主导状态之间切换，可能对应于空闲状态与充电状态的转换。该特征在后续状态分类中具有重要价值。

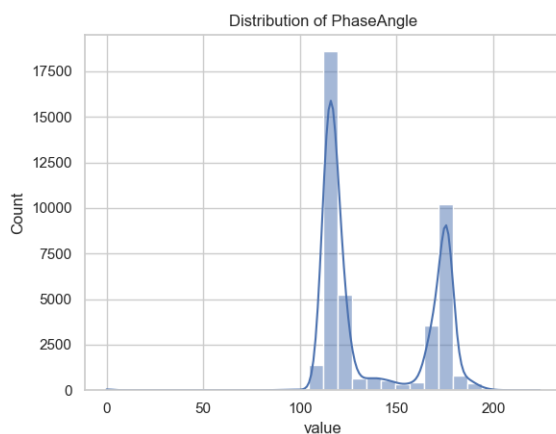


Figure 7: PhaseAngle 分布。

3.2.7 Power

图 8 显示了 Power（单位：瓦）的分布，反映真实功率消耗。大部分值接近 0，表明大多数时间处于空闲或低功耗状态。约 130 W 附近的次峰值可能代表充电周期。

该指标是判断机器人是否在充电的重要依据，并可支持“充电状态”与“非充电状态”的划分。

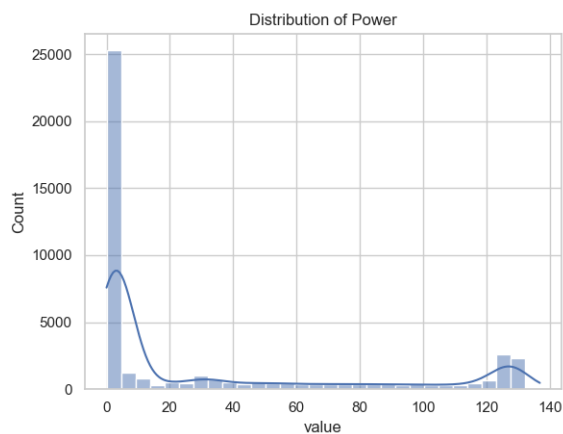


Figure 8: 实时功率消耗 (Power) 的分布。

3.2.8 Power Factor

图 9 展示了 PowerFactor 的分布，该指标表示有功功率与视在功率的比值，范围为 0 到 1。数据呈现明显的双峰形态，峰值分别出现在约 0.3 和 0.95。

这可能对应于两种工作模式：一种是低效用电（如待机状态），另一种是高效用电，可能为充电状态。

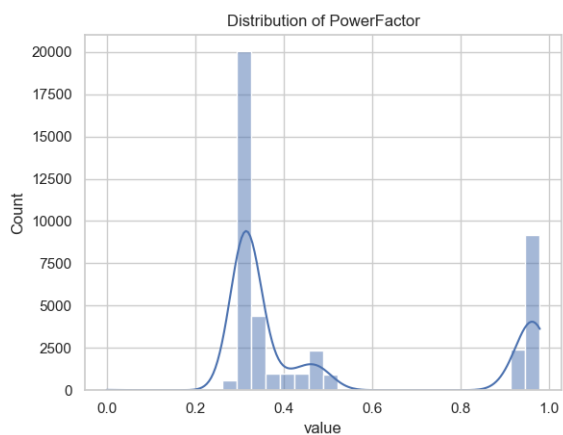


Figure 9: PowerFactor 的分布。

3.2.9 Relay Status

图 10 展示了 RelayStatus 的分布，该变量是一个二值指示器（1 为继电器闭合，0 为断开）。所有数值为 1，表明在整个观测期内继电器始终处于闭合状态。

继电器持续闭合意味着插座始终允许电流流通，进一步支持“机器人要么在充电，要么处于空闲但待命状态”的推论。

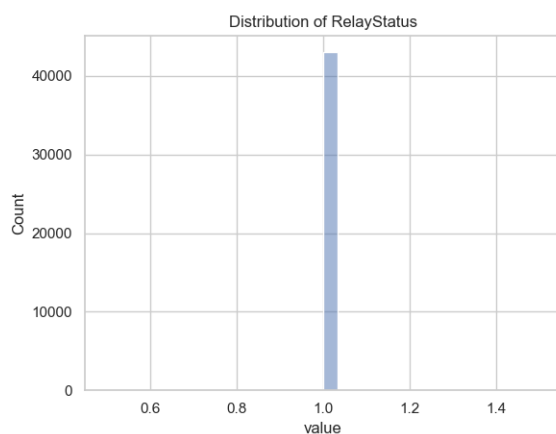


Figure 10: 继电器状态 (RelayStatus) 分布。

3.2.10 Temperature

图 11 展示了插座内部温度的分布。值集中于 35–38°C 的典型室温区间，出现多个尖锐峰值，可能源自测量误差或传感器精度限制。

Temperature 可用于监测过热风险或识别插座使用的异常情况。但本案例中数据较为稳定，说明所处环境在监测期内变化不大。

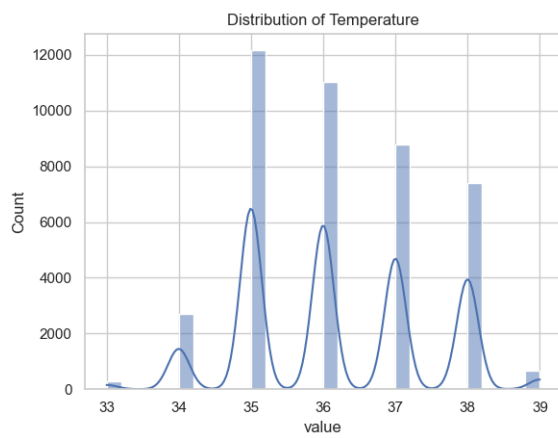


Figure 11: 内部温度 (Temperature) 分布。

3.2.11 Total Power Factor

图 12 显示了 TotalPF 的分布，可能表示功率因数的累计或平滑估算值。其值大致分布在 50 到 75 区间，呈均匀分布，未显示明显峰值。

该特征可能反映不同阶段的整体插座性能，但由于缺乏详细文档说明，其分析价值有限。

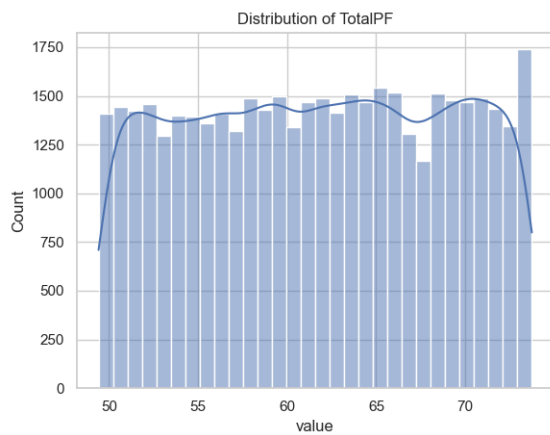


Figure 12: TotalPF 分布。

3.2.12 Voltage

图 13 展示了电压的分布，呈对称的钟形分布，峰值约在 225 V，符合商业或工业用电的标准。

Voltage 的稳定性表明供电质量稳定，说明电压波动在本项目中不是造成行为差异的主要因素。

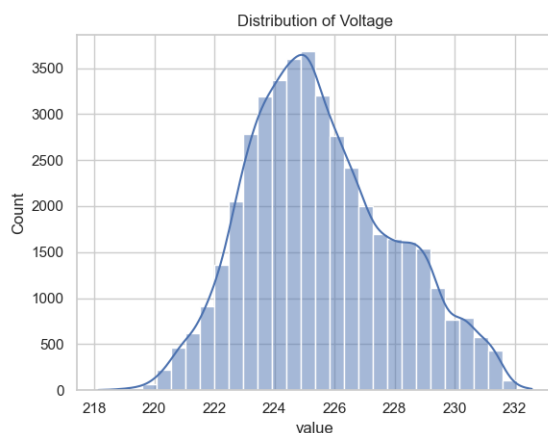


Figure 13: 线路电压 (Voltage) 分布。

3.3 时间序列分析

为了更好地理解智能插座记录的每个功能随时间的变化，我对观测期内的时间序列进行了分析。静态分布可揭示数值范围与特征模式，但无法捕捉随时间变化的周期性或趋势。由于智能插座与执行任务的移动机器人配合使用，了解各电力与环境参数的时间演变十分关键。

图 14 展示了所有功能值的时间序列图，主要观察结果如下：

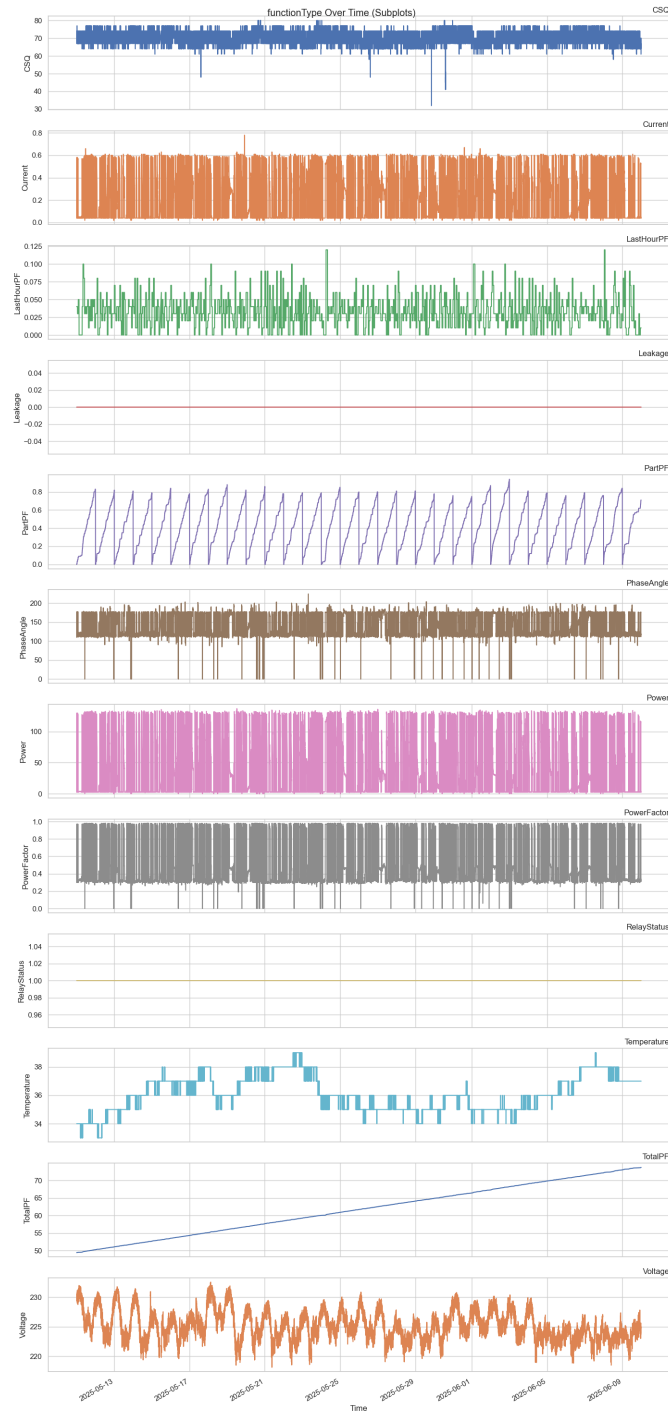


Figure 14: 数据收集期间各功能变量的时间序列图。

- **CSQ (信号强度)** 整体稳定，但有轻微波动，可能反映机器人在不同区域移动时的信号变化。
- **Current、Power、PowerFactor 与 PhaseAngle** 同步出现活动高峰，说明充电或高负载时段协调一致；而其他时段接近零，反映设备处于待机状态。
- **LastHourPF** 呈现噪声状态，且在后期呈下降趋势，可能表示充电次数减少或持续时间缩短。
- **Leakage 与 RelayStatus** 始终平稳，验证之前观察：无漏电、继电器始终闭合。
- **PartPF** 呈现清晰的锯齿图样，可能反映了定期归零或重置的机制。
- **Temperature** 随时间缓慢上升，可能受季节变换或储存环境温度积累的影响。
- **TotalPF** 稳步上升，或表示累计使用量增加，或反映能效随时间改善。
- **Voltage** 呈现日周期震荡，可能受电网波动或设施用电高峰影响。

该时间序列概览验证了存在日常周期性行为（如每日充电），同时表明后续建模或异常检测需引入动态特征或时间聚合策略。

4 基于 K-Means 与 PCA 的聚类分析

为了揭示插座使用行为中潜在的模式，我对预处理后的时间序列特征数据应用了 K-means 聚类方法。聚类有助于识别底层的使用状态或运行模式——例如空闲、充电、或过渡状态——这些都可通过多变量传感器读数来推断。由于特征之间存在不同的量纲与尺度，聚类能够提供一种基于数据的方式来对插座行为进行分类。

在聚类之前，我使用主成分分析（PCA）对特征空间进行了降维处理。该方法将原始数据投影到新的坐标系中，使得前几个主成分保留数据中大部分方差。通过将数据投影到二维空间中，可以更直观地可视化聚类效果，观察数据点的空间分布。

4.1 主成分载荷与解释

为更好理解各主成分的含义，我查看了载荷矩阵，它显示每个原始特征在主成分中的贡献程度。表 1 显示了 PC1 与 PC2 的载荷值。

Table 1: 前两个主成分的 PCA 载荷。

特征	PC1	PC2
Current	0.5101	0.0071
Power	0.5046	-0.0099
PowerFactor	0.5003	-0.0070
PhaseAngle	0.4824	0.0286
CSQ	0.0177	-0.2294
TotalPF	0.0085	-0.3852
Temperature	0.0044	-0.4884
RelayStatus	0.0000	0.0000
Leakage	0.0000	0.0000
LastHourPF	-0.0130	0.0486
Voltage	-0.0227	0.5975
PartPF	-0.0329	-0.4473

PC1 的解释：活动水平指标 第一主成分（PC1）解释了数据中最大部分的方差，并且与四个关键特征——Current、Power、PowerFactor 和 PhaseAngle——呈强正相关。这些变量均直接关联电力活动和能耗水平。

这些高正载荷集中表明，PC1 可被视为插座总体活动强度的代理指标：

- 较高的 PC1 值对应机器人处于充电或执行任务等高功耗状态；
- 较低的 PC1 值则反映设备处于空闲或待机等低功耗状态。

与活动无关的特征（如 Leakage、RelayStatus、CSQ）的载荷值接近 0，也进一步支持了该解释，即 PC1 主要受到电气活动维度的驱动。

因此，基于 PC1 的聚类主要将数据按运行强度进行区分，与前面函数类型的分布分析结果相吻合。

4.2 轮廓系数分析

为了确定聚类的最佳簇数，我计算了 $k = 2$ 到 $k = 10$ 各个簇数对应的 Silhouette Score。轮廓系数衡量每个样本与所属簇相比于其它簇的相似程度，值越高代表聚类效果越好、边界越清晰。

图 15 显示，当 $k = 2$ 时轮廓系数最高，之后随着 k 增大，提升幅度迅速减小。这表明 $k = 2$ 是最自然的数据划分方式，具有良好的紧凑性和分离性。这一趋势也与我们在分布分析中观察到的状态划分一致。尽管 $k = 3$ 的得分下降

幅度不大，提示三簇方案在解释上仍有一定价值，但出于清晰性和聚类质量的考虑，最终采用了两簇方案。

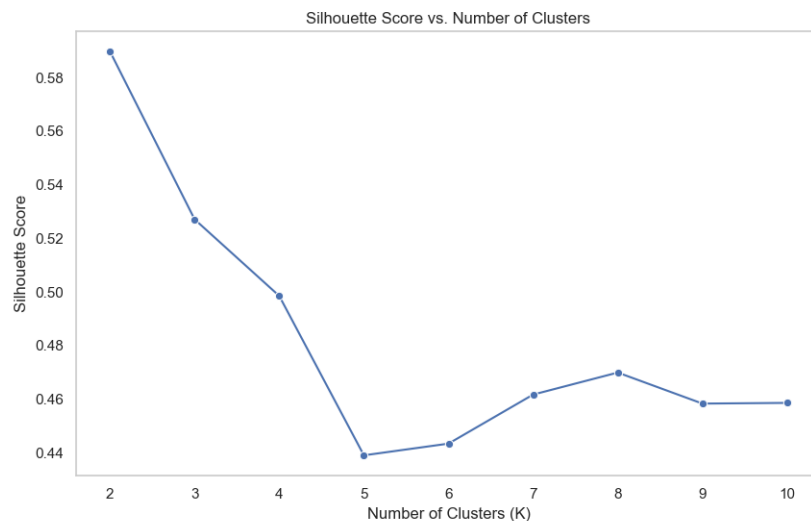


Figure 15: 不同聚类数 k 下的 Silhouette Score。

4.3 K-Means 聚类结果

图 16 展示了 $k = 2$ 时的 K-means 聚类结果。可以清晰地看到数据被划分为两个主要群组。这种二分可能对应“充电”与“非充电”状态，与 Current、Power 和 PowerFactor 的模式一致。

当将聚类数增加至 3（右图）时，出现了一个中间簇，可能代表过渡状态——例如短暂的唤醒期或处于待命与充电之间的中间阶段。这种细化对于行为识别或异常检测具有潜在意义。

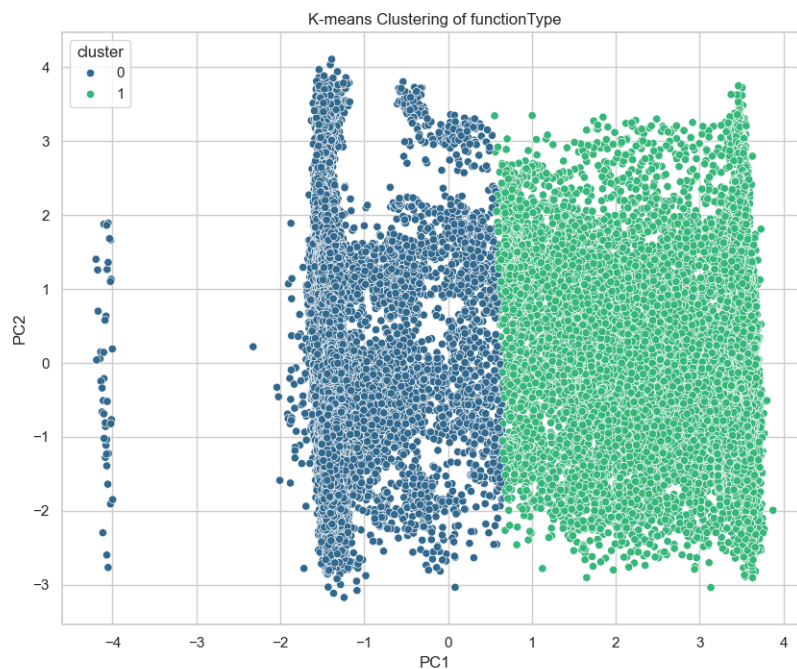


Figure 16: K-means 聚类在 PCA 投影下的结果 ($k = 2$)。

4.4 状态分布

完成聚类后，我们将每个时间点划分为两个运行状态：充电或活动。这些状态依据插座是否消耗显著电流进行定义，从而揭示配送机器人的行为模式。

图 17 显示约 66.2% 的时间处于充电状态，其余 33.8% 为活动状态，说明机器人在多数时间处于静止或停靠状态，与其周期性配送与返回的功能定位一致。

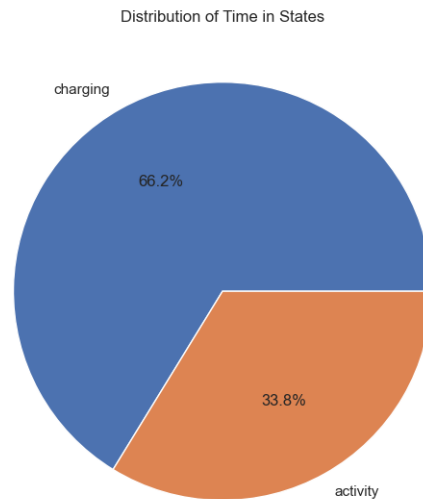


Figure 17: 在各运行状态下的时间分布。

我们进一步分析了状态随星期的变化（图 18）。尽管充电状态在每周各日均占主导地位，但周日与周一的活动比例略有上升，暗示配送任务在周初与周末期间更为频繁。周二至周五则较为稳定，可能反映出常规的每日运作模式。

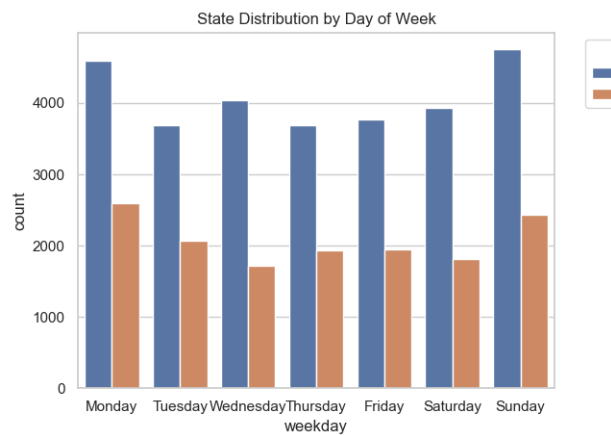


Figure 18: 按星期的状态分布情况。

另外还尝试绘制了如下探索性图表：

- 每小时的状态分布；
- 每日各状态的总时长；
- 状态持续时间的直方图；
- 状态持续时间随时间变化的折线图；

但由于数据噪声较大、模式不明显，上述图表未提供额外洞见，因此我们重点展示了更稳定、可解释性更强的按星期与总体状态分布。

4.5 基于马尔可夫链的状态转移分析

为了更深入理解“充电”与“活动”这两个运行状态之间的动态变化，我们构建了一个一阶马尔可夫链模型。该模型假设系统的未来状态仅由当前状态决定，可用于捕捉状态之间的转移概率。

图 19 展示了根据实际观测状态序列构建的状态转移矩阵。矩阵显示，两个状态均具有较高的自循环概率，尤其是充电状态，这表明系统在当前状态下会维持较长时间，才会发生切换。

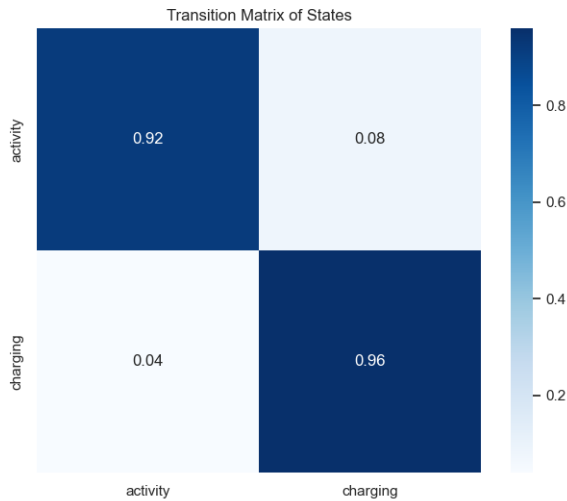


Figure 19: 马尔可夫链建模下的状态转移矩阵。

我们还计算了稳态分布——即在长期运行中系统在各状态下的平均占比。结果如下，与图 17 的可视化结果一致：

状态	稳态概率
Charging	0.662
Activity	0.338

这些结果进一步确认，配送机器人绝大多数时间处于低功耗或充电状态，仅在特定时间段短暂执行任务。

5 时间使用模式分析

为进一步探究机器人在不同时段的活动与充电行为，我们将状态时间线进行分段，并校正时区以反映本地使用习惯。分析采用了粗粒度与细粒度两种时间划分方式，以捕捉每日节律与运行窗口。

5.1 按时段的使用模式

图 20 展示了机器人在四个时间段（早晨、下午、晚上、夜间）中处于活动与充电状态的比例。机器人在所有时段均以充电为主，其中晚上（Evening）时段的活动比例最高（约 41%），表明机器人主要在晚上执行任务，可能与配送需求或排班安排有关。

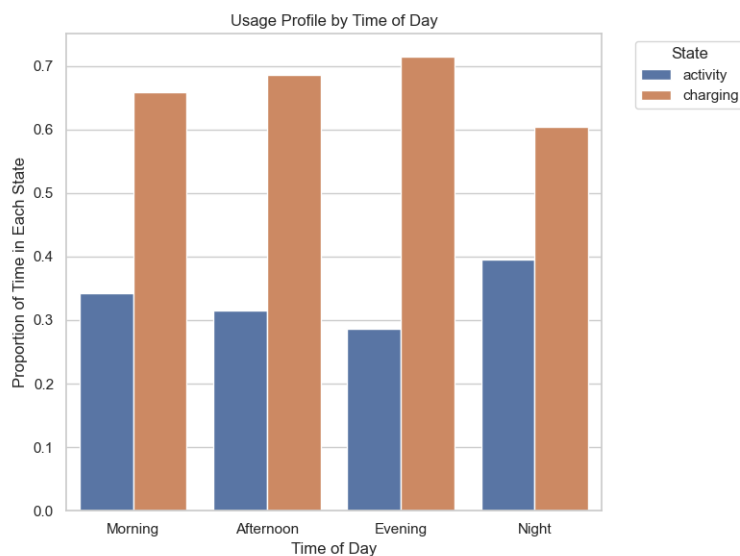


Figure 20: 按一天四个时段划分的状态占比。

5.2 按两小时区间的使用模式

为捕捉更细粒度的时间动态，图 21 展示了以两小时为单位的状态分布。可以看到，机器人在一天中状态有明显波动。活动最低时段出现在午后（如 14:00–16:00），而晚上（20:00–22:00）活动频率明显上升。凌晨（02:00–06:00）也有一定活动，可能用于非高峰配送或设备准备。

几乎所有时间区间中，充电状态占比均超过 60%，再次印证系统具有较高能耗密度，合理调度充电时间对系统稳定性尤为关键。

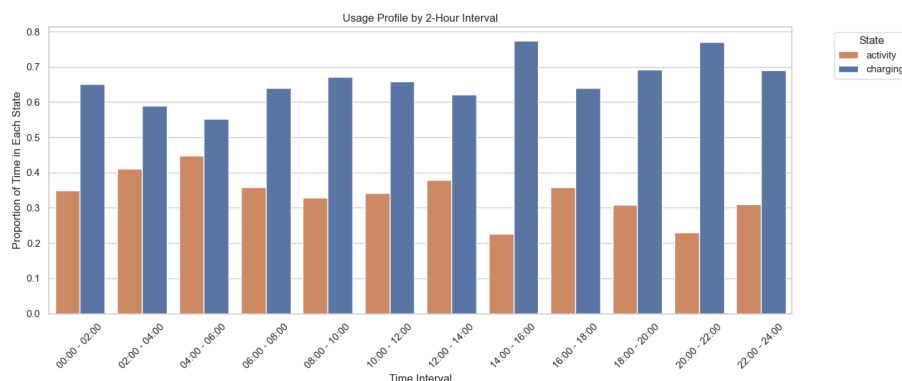


Figure 21: 以 2 小时为单位的使用状态分布。

Part II

公共电脑插座分析

6 探索性数据分析 (EDA)

6.1 相关性分析

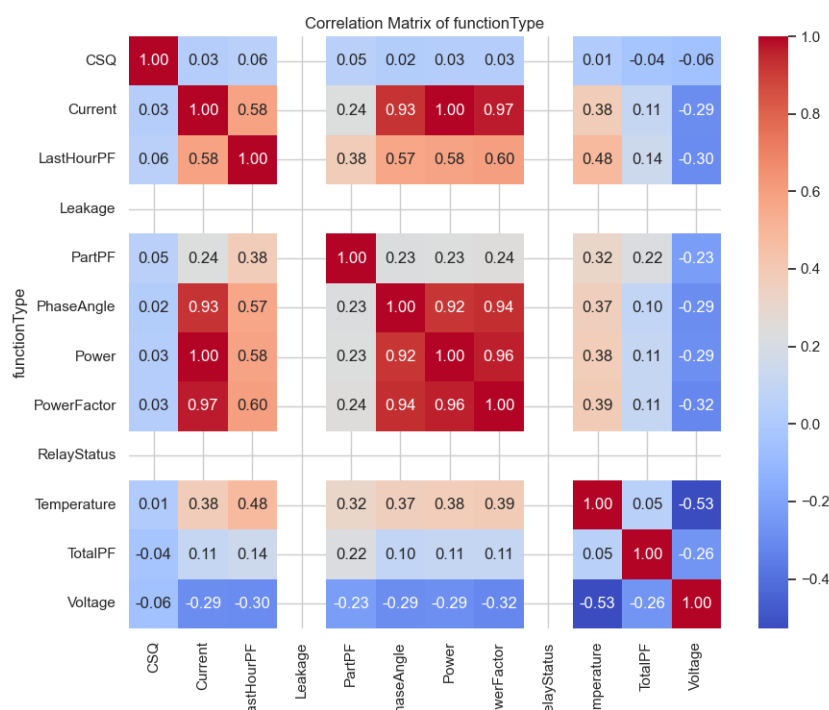


Figure 22: functionType 的相关性矩阵

图 22 展示了公共电脑插座中所有 functionType 变量之间的 Pearson 相关系数。主要观察如下：

- Current、Power 和 PowerFactor 之间存在显著正相关（均高于 0.9），表明这些变量在设备活跃时通常共同变化。

- PhaseAngle 与 Power 和 PowerFactor 同样高度相关，说明相位偏移与电能传输特性密切相关。
- LastHourPF 与 Current 及 PowerFactor 存在中等强度相关，提示功率因数随时间存在一定一致性。
- Temperature 与 Power、Current 和 PowerFactor 呈中等正相关，符合电流活动导致热量积累的预期。
- Voltage 与大多数其他变量之间相关性较弱或为负，可能反映出供电电压在不同使用强度下保持稳定。
- CSQ、Leakage 和 RelayStatus 与其他变量的相关性几乎为零，可能是因为在观测期间基本保持不变。

总体来看，相关矩阵揭示了围绕能耗与运行状态的信号集群，强化了电流、电力与热量输出之间的结构性关系。

6.2 各功能变量的分布

6.2.1 Signal Strength (CSQ)

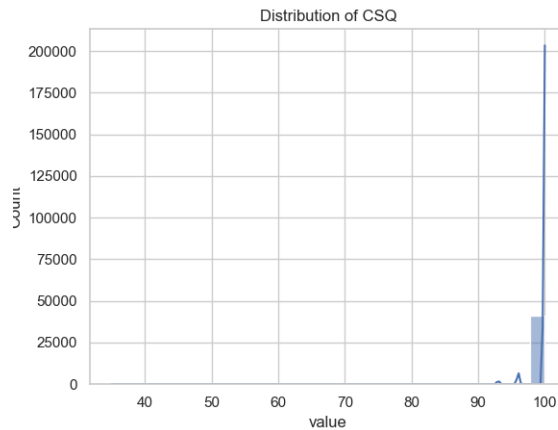


Figure 23: CSQ（信号强度）分布

图 23 显示了公共电脑的 CSQ 分布情况。该直方图在接近最大值 100 处呈现明显尖峰，几乎没有其他取值变化。这表明信号质量在整个观测期间始终表现优异，几乎没有波动，因此 CSQ 在该数据集中对区分设备状态并无显著作用。

6.2.2 Current

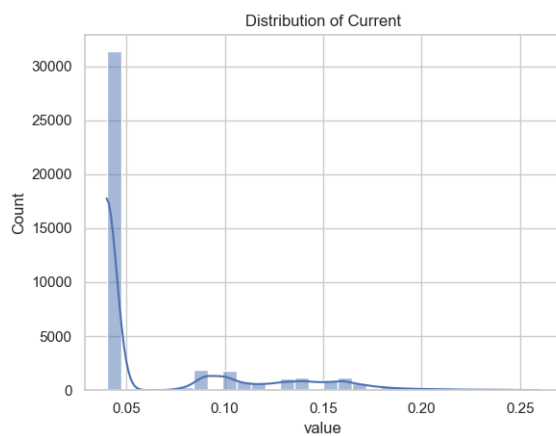


Figure 24: Current (电流) 分布

图 24 展示了电流值的分布。大部分数据集中在 0.05 附近，较小的峰值出现在 0.1、0.15 和 0.2，呈现多峰结构。这可能对应不同的负载状态或任务强度，长尾部分表示偶尔出现的高电流消耗，但此类情形较为罕见。

6.2.3 Last Hour Power Factor

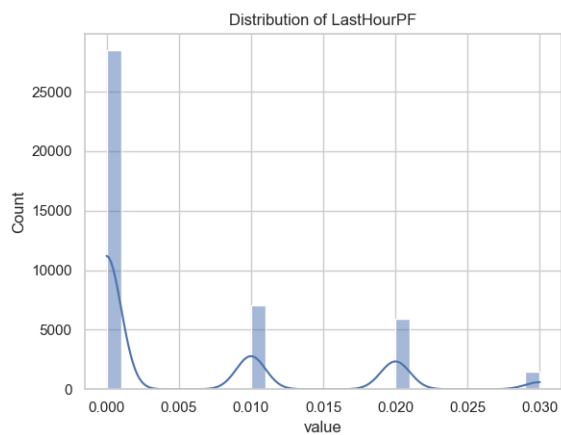


Figure 25: LastHourPF (上一小时功率因数) 分布

图 25 中的值紧密分布在 0 附近，出现了规则的尖峰（大约在 0.01、0.02 等处），提示功率因数读数可能存在离散化或取整现象。0 附近的强峰值可能对应长时间的空闲或轻度使用。

6.2.4 Leakage

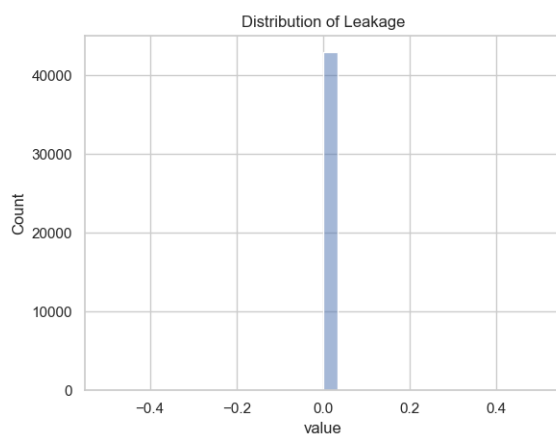


Figure 26: Leakage（漏电流）分布

图 26 显示所有记录值均为 0，说明监测期间未检测到漏电，或漏电电流低于传感器灵敏度，亦或该项未启用。

6.2.5 Partial Power Factor

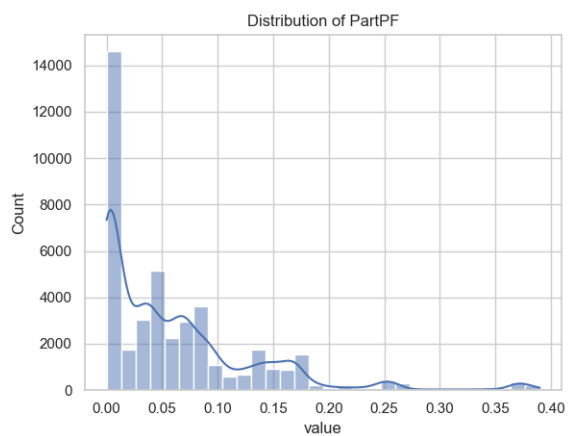


Figure 27: PartPF 分布

图 27 中，大多数值低于 0.1，主要集中于 0 附近，分布右偏，有长尾延伸至 0.4。这表明该设备经常处于低功率因数状态，反映出轻负载或空闲状态较常见。

6.2.6 Phase Angle

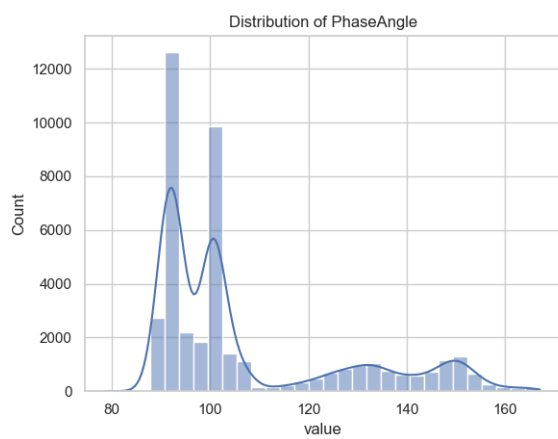


Figure 28: PhaseAngle (相位角) 分布

图 28 呈现多峰分布，主要集中在 90° 和 100° ，并在 140° 和 150° 附近出现次峰。这表明存在多个运行状态或连接设备具有不同相位特性。

6.2.7 Power

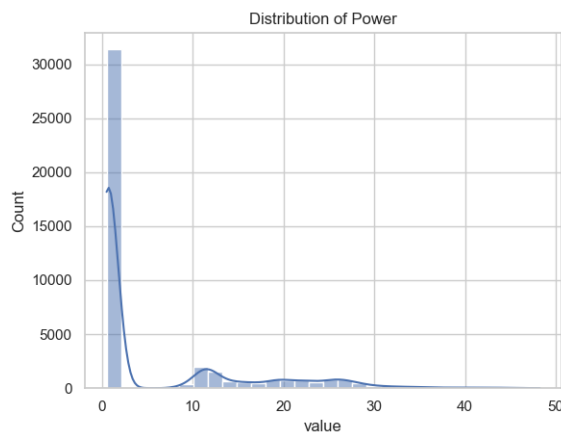


Figure 29: Power (功率) 分布

图 29 呈现明显的右偏分布，主峰在 0 附近，次峰出现在 10、20、30 等区间，反映出设备大部分时间能耗较低，但偶尔存在高功耗阶段，可能与特定设备或使用模式有关。

6.2.8 Power Factor

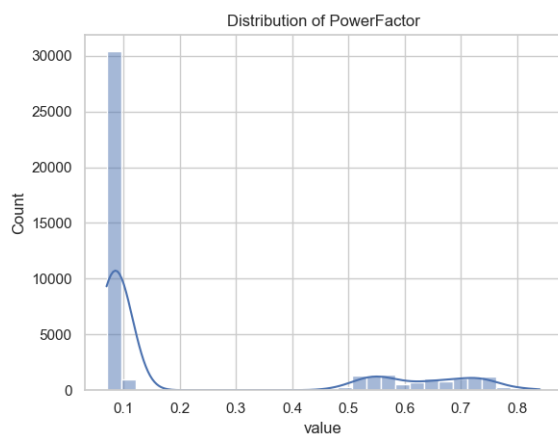


Figure 30: PowerFactor 分布

图 30 同样表现为右偏分布，在 0.1 附近有一个明显的主峰，并在 0.6 和 0.7 附近出现次峰。这可能说明该设备多处于功率因数较低的状态，仅在某些负载活跃阶段效率才有所提升。

6.2.9 Relay Status

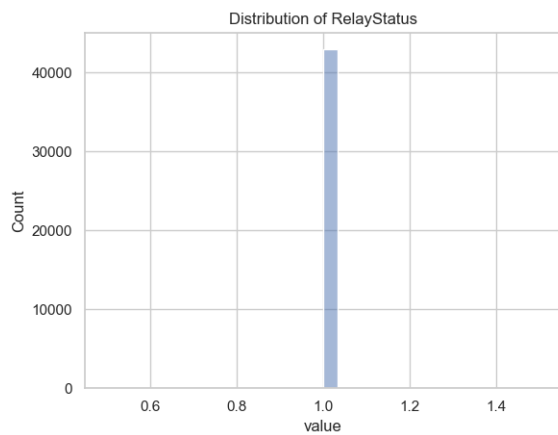


Figure 31: RelayStatus (继电器状态) 分布

图 31 显示所有取值集中在 1.0，表明继电器在整个观测期间几乎始终处于闭合（通电）状态，可能表示设备持续通电或很少被关闭。

6.2.10 Temperature

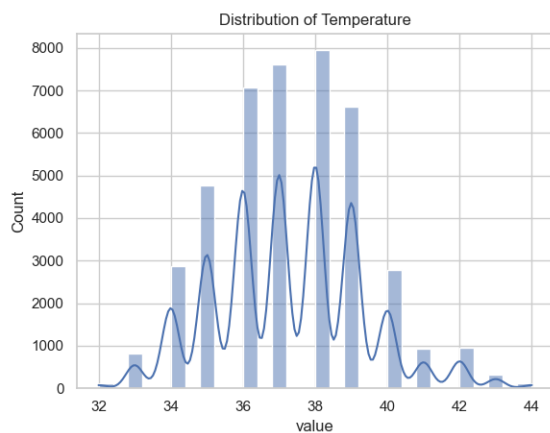


Figure 32: Temperature（温度）分布

图 32 显示温度分布大致符合正态分布，均值位于 37–38°C 之间，密度曲线中存在周期性波动，可能反映了热平衡或恒温控制机制。

6.2.11 Total Power Factor

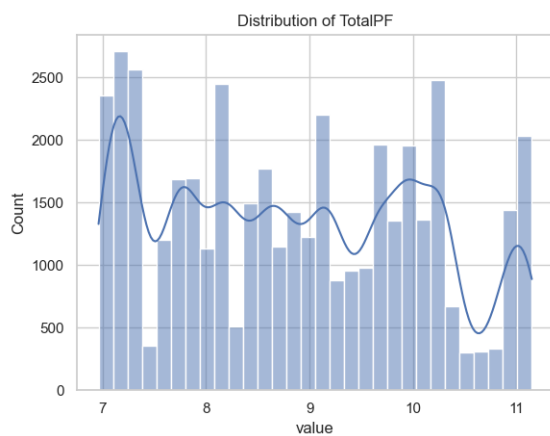


Figure 33: TotalPF 分布

图 33 的取值范围为 7 到 11，分布近似均匀或轻微多峰，说明累计功率因数在运行期间存在变化，可能受到负载类型或使用时间段影响。

6.2.12 Voltage

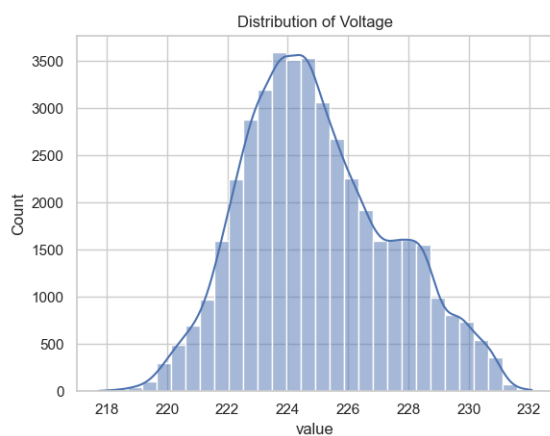


Figure 34: Voltage (电压) 分布

图 34 显示出接近正态分布，平均值约为 224 V，标准差较小，说明电压供给整体稳定，波动性低，符合商业用电标准。

6.3 时间序列分析

图 35 显示了所有 `functionType` 变量的时间序列变化，采用子图方式分别展示各变量行为，便于可视化比较：

- CSQ 大部分时间稳定维持高值，偶有下降反映短暂信号中断。
- Current、Power 与 PowerFactor 呈现高度间歇性的波动，表明设备存在突发性或周期性使用行为。
- Leakage 与 RelayStatus 几乎不变，表明系统配置静态或继电器开关极少变化。
- Temperature 呈现平缓波动，可能受到环境或内部温升影响。
- TotalPF 随时间稳步上升，可能代表累计功率因数的累积效应。
- Voltage 呈现强周期性波动，符合交流供电系统的日周期性特征。

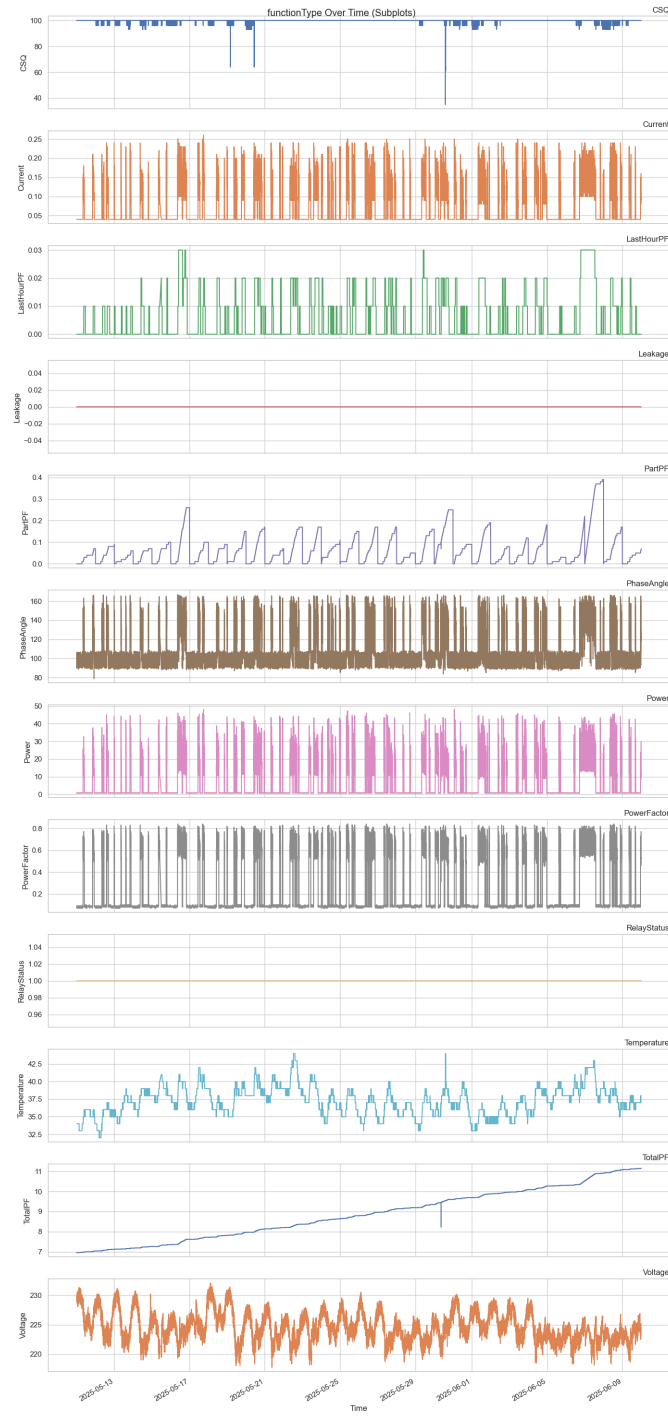


Figure 35: functionType 各变量的时间序列子图

7 聚类分析

为探索设备功能行为中的自然分组结构，我们对预处理后的数据集进行了聚类分析。

7.1 通过 PCA 降维

我们首先采用主成分分析 (PCA) 将高维特征空间压缩为两个主成分，以便于可视化解释。载荷矩阵显示，PC1 主要受到与有功功率相关的变量（如 Current、Power 和 PowerFactor）的影响，而 PC2 更多受到 Voltage、Temperature 和 PartPF 的驱动。

7.2 通过轮廓分析选择最佳簇数

为确定最优聚类数，我们计算了 $K = 2$ 至 $K = 10$ 的 Silhouette Score (图 36)。得分在 $k = 2$ 时达到最高，表明在该设定下簇间分离最明显。

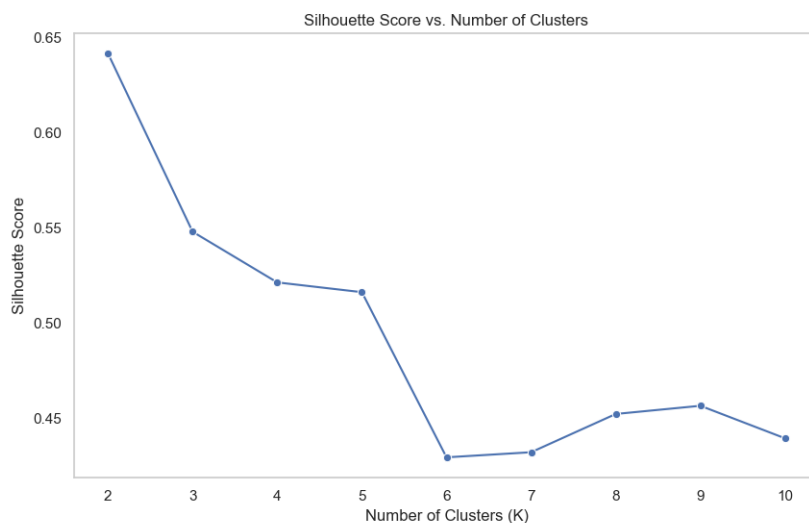


Figure 36: 不同聚类数下的 Silhouette Score

7.3 K-Means 聚类结果

在 $k = 2$ 的条件下，我们应用了 K-means 算法并在 PCA 投影空间中可视化结果（图 37）。然而，视觉检查表明该聚类结果并未与数据中的明显结构对齐，决策边界较为随意，K-means 未能捕捉到 PCA 投影所揭示的潜在结构。

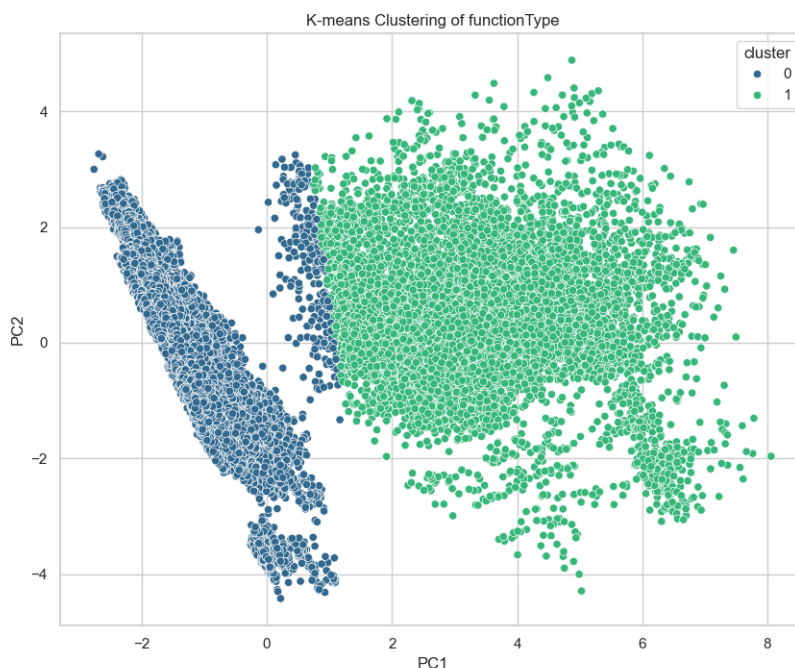


Figure 37: K-means 在 PCA 空间中的聚类结果 ($k = 2$)

7.4 基于密度的 DBSCAN 聚类

由于 K-means 聚类尽管具有较高的 Silhouette Score 却未能提供可解释的分组，我们尝试了密度聚类算法 DBSCAN。与 K-means 假设球状簇结构不同，DBSCAN 可识别任意形状的簇，并对噪声具有更强鲁棒性，适合处理 PCA 空间中不规则的自然分布。

我们采用 k 近邻距离图来选择 DBSCAN 的关键参数 ϵ （邻域半径）。该方法通过绘制每个点到其第 k 个最近邻的距离排序曲线（通常 $k = \text{min_samples}$ ），观察距离急剧上升的“肘部”位置作为 ϵ 的合理选择（图 38）。然而，即便经过

多次调参，DBSCAN 依然只识别出一个簇（图 39），表明数据中密度变化不足以支撑有效分组。

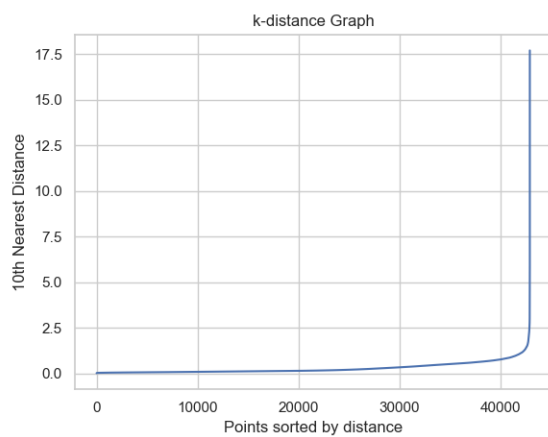


Figure 38: 用于选择 DBSCAN 参数的 k -近邻距离图

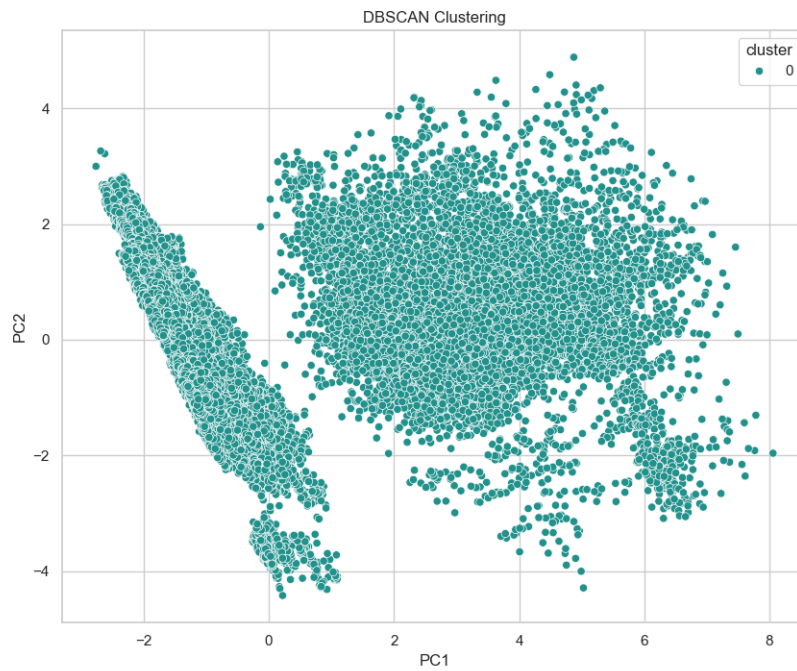


Figure 39: DBSCAN 聚类结果

7.5 人工聚类

考虑到 K-Means 和 DBSCAN 均未能实现直观有效的分组，我们采用人工聚类策略，在 PCA 空间中定义一条线性分界线。该决策边界形式为斜率 $m = -1.5$ 的直线：

$$y = -1.5x + b \quad (1)$$

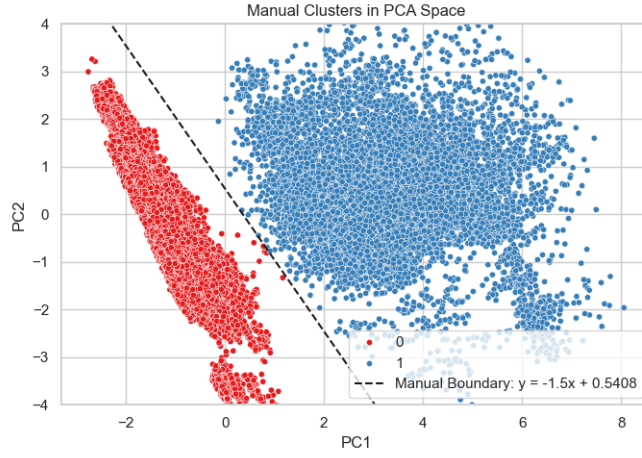


Figure 40: 基于人工决策边界 $y = -1.5x + 0.5408$ 的聚类结果。该方法通过在 PCA 空间中划分边界实现运行状态的可解释分组。

起初使用 $b = 0$ ，后通过观察点在边界法向量方向的投影进行调整，最终选定 $b = 0.3$ 作为边界，使两个点云的分离更清晰。

我们进一步将所有点投影至归一化法向量 $\vec{v} = [1, -1.5]^T / \|\cdot\|$ ，绘制投影值直方图（图 41）。在两个峰之间的谷值附近（约 -0.45 ）设置阈值，从而划定两簇：

$$y = -1.5x + 0.3 \quad (2)$$

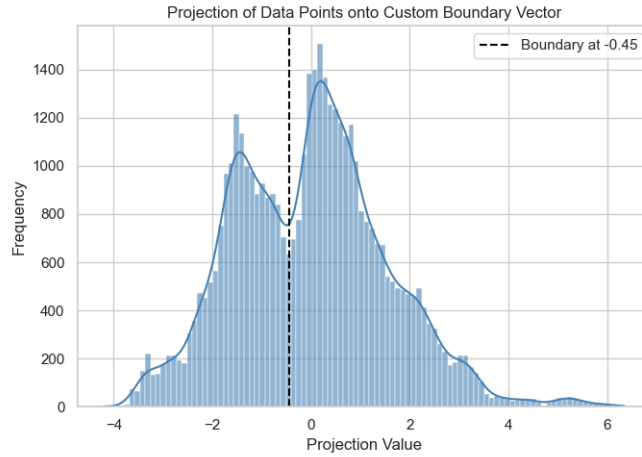


Figure 41: 点在边界法向方向的投影直方图。虚线位置表示所选阈值 -0.45 。清晰的双峰结构进一步验证了人工聚类的合理性。

边界下方的点被标记为 Cluster 0，上方的点为 Cluster 1。

7.6 聚类结果解释

通过 PCA 的载荷矩阵，我们可以对人工划分的两个聚类进行解释：

- PC1 主要受 Current、Power、PowerFactor 和 PhaseAngle 等功率相关变量影响，反映设备的整体能耗与运行强度；
- PC2 则更多受 Voltage、Temperature 和 PartPF 等环境或背景变量影响，可能反映外部条件变化。

functionType	PC1	PC2
Current	0.510	0.007
Power	0.505	-0.010
PowerFactor	0.500	-0.007
PhaseAngle	0.482	0.029
CSQ	0.018	-0.229
TotalPF	0.008	-0.385
Temperature	0.004	-0.488
RelayStatus	0.000	≈ 0.0
Leakage	≈ 0.0	≈ 0.0
LastHourPF	-0.013	0.049
Voltage	-0.023	0.598
PartPF	-0.033	-0.447

Table 2: PCA 载荷矩阵

由于决策边界既不垂直于 PC1，也不垂直于 PC2，聚类结果实际是能耗和环境因素的综合划分：

- **Cluster 0**: PC1 和 PC2 均较低，代表低能耗或待机状态；
- **Cluster 1**: PC1 较高、PC2 波动，代表活跃运行或高负载状态。

因此我们将 Cluster 0 标记为“idle”（待机），Cluster 1 标记为“operating”（运行）。换言之，Cluster 0 表示公共电脑处于待机状态，Cluster 1 则表示设备正在被主动使用。

7.7 时间使用模式

在识别出“待机”与“运行”两种状态后，我们进一步分析其随时间的分布，以理解设备在日/周周期下的使用行为。

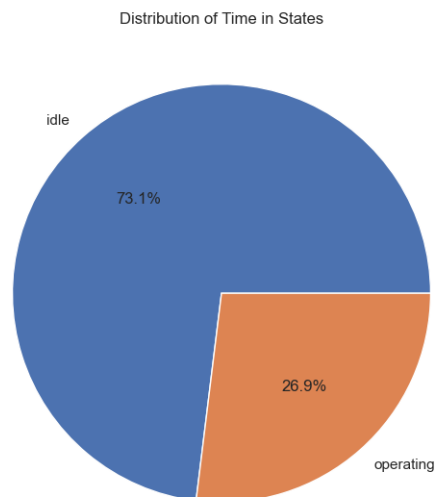


Figure 42: 总体状态分布。设备有 73.1% 的时间处于待机状态，仅 26.9% 时间处于运行状态。

总体使用比例。 图 42 显示设备绝大多数时间处于待机状态，说明活跃使用频率较低，可能与排程任务或外部触发事件有关。

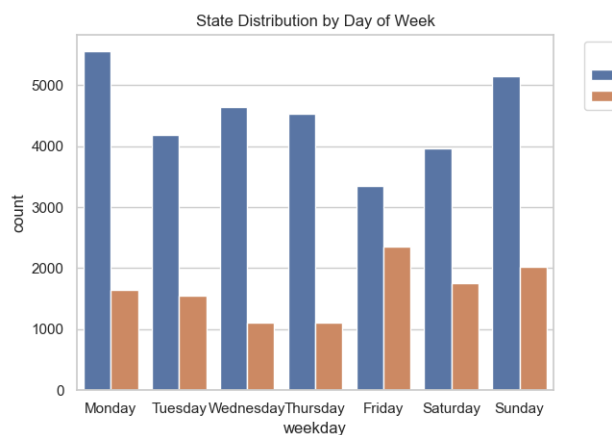


Figure 43: 按星期的状态分布。设备在工作日中段使用频率下降，周末略有回升。

星期趋势。 如图 43 所示，周二至周四设备使用减少，周五至周日略有上升。周一使用高峰可能反映系统在周初进行维护或初始化。

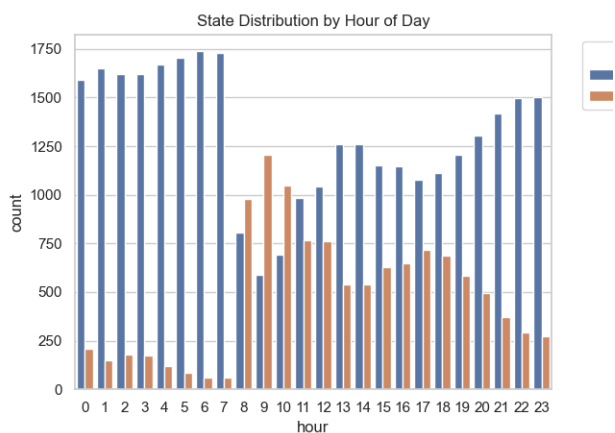


Figure 44: 按小时的状态分布。凌晨与深夜设备多为待机，08:00 至 18:00 期间活跃度明显上升。

小时行为模式。 图 44 显示出明显的日节律：08:00 后活动显著上升，午后逐渐回落，与一般办公时间吻合。

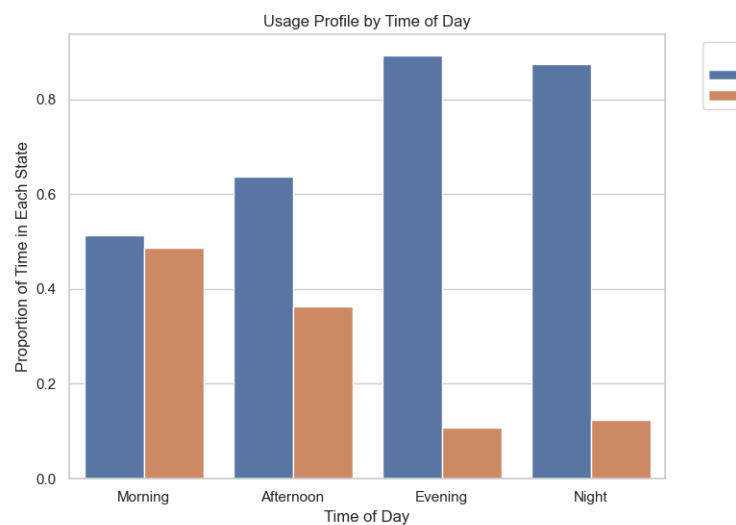


Figure 45: 按四大时段划分的使用状态。下午设备活动最为频繁，夜间与傍晚活动较少。

按时间段划分。 图 45 将小时趋势汇总为四个时段。下午活跃度最高，进一步确认设备以标准工作时段为主。

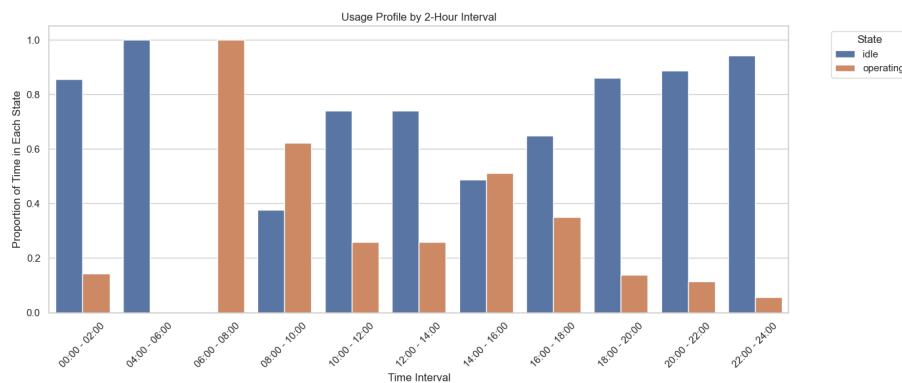


Figure 46: 每 2 小时的使用模式。早上 06:00-10:00 活跃度最高，夜间几乎为零。

更细粒度的使用模式。 图 46 显示，每日 06:00-10:00 是设备活跃的高峰期，午间略有下降，16:00-18:00 再次升高，可能表明存在分段式使用窗口。

总结。 总体而言，该设备展现出明显的日周期节律，主要在白天短时间内运行，其余时间处于待机状态。识别活跃阶段的触发因素，对于优化调度、维护及节能策略具有重要意义。

Part III

对比分析： 配送机器人 vs. 公共电脑

为整合不同智能插座部署场景中的洞见，我们对配送机器人和公共电脑的使用模式与聚类表现进行了对比分析。

8 状态聚类对比

两个设备均基于功能型变量（如功率、电压、电流）进行了无监督聚类分析。配送机器人的聚类结果与预期的运行模式（活动与 充电）高度一致，在 K-Means 与 PCA 投影中呈现出清晰的周期性使用特征。

相比之下，公共电脑的使用模式更连续、不易分割。K-Means 聚类无法识别出直观的行为分组，DBSCAN 在各种参数下也仅返回一个密集簇。这种模糊性可能源于电脑负载的变动性及其更为持续的使用特性。

因此，我们在 PCA 空间中引入人工划分边界，形成两个可解释的簇。主成分载荷分析显示，PC1 捕捉了总体功耗和运行强度，PC2 则反映了电压和温度等环境或系统状态变化。通过斜率为 -1.5 的边界将状态划分为待机（低功耗）与运行（高负载），这种人为干预对解析公共电脑更复杂的行为模式是必要的。

9 时间使用模式对比

时间序列分析进一步揭示了两者在行为模式上的差异。配送机器人表现出强烈的周期性，其状态在不同时间段和工作日之间的分布清晰可辨。其活跃高峰集中在晚上，深夜和清晨处于低使用率，这与任务调度或充电窗口高度一致。

而公共电脑虽使用模式更为连续，但依然展现出一定规律。低使用时段（如凌晨）依然存在，但状态转换频繁且分布更平均。按星期分析显示，设备在一周内均有使用，周末使用率下降不明显，表明其使用场景可能包括自动化任务或非人为驱动的行为。

10 可解释性与干预需求

总体而言，配送机器人的使用行为可通过少量分析自动推断；而公共电脑则需结合降维、投影分析与领域知识指导的阈值调整，才能获得有意义的操作状态。

这一对比凸显了在智能基础设施中面对异构设备时，采用灵活分析策略的重要性。

关键发现：

- 配送机器人具备高度规律性和聚类清晰度，受限于可预测的任务周期；
- 公共电脑行为状态重叠明显，需依赖更多解释性和人工参与分析；
- 时间使用轮廓反映设备用途差异：配送设备受限于时间窗口，公共终端则维持持续可用性。

Part IV

结论

本分析揭示了配送机器人与公共电脑在运行行为上的显著差异，为设施管理提供了可执行的见解。配送机器人在傍晚的集中活动与公共电脑的多状态能耗模式，为智能排程与节能策略提供了切入点。理解设备“何时”和“如何”使用，可进一步实现负载均衡、削峰填谷，以及在非高峰时段的自动关机或充电协议部署。

此外，通过本次分析构建的结构化数据处理流程为未来拓展提供了可扩展的基础。通过高精度识别设备状态与建模状态转换，可以开发实时监测面板、基于事件的告警系统、以及用于异常检测或预防性维护的预测模型。这些工具可推广至类似 IoT 场景，为智慧设施运行提供数据驱动的支持，助力实现节能与高效运维目标。