

“Build with Us | Deep Dive: Data Analysis in Contour”是发布在 learn.palantir.com 上的一个实战教程,由 Ontologize 团队(前 Palantir 工程师)制作 1。该课程专门针对 **Palantir Contour** 这一应用进行深度教学,指导用户如何从原始数据出发,通过一系列转换和可视化,最终构建出一个交互式的分析看板 1。

以下是根据来源对该课程内容的详细拆解:

1. 核心背景与业务场景

- 业务角色: 用户扮演 **Titanium Works Manufacturing** 公司的数据分析师 1。
- 分析目标: 利用设备数据(Equipment data)和零件数据(Parts data)进行分析,识别高风险机械,从而优先安排设备检查并确保生产效率 1。
- 核心工具: 主要使用 **Contour**(一个低代码分析环境),其特点是分析过程呈线性流向,并利用各种“板(Boards)”执行操作 2。

2. 分析全流程: 从预处理到深度分析

A. 基础操作与数据清洗

- 创建分析: 从零件数据集(Parts dataset)启动 Contour, 并进入**编辑模式(Editing mode)**进行工作 2。
- 摘要统计(**Summary Statistics**): 使用 **Calculation board**(计算板) 计算零件 ID 和设备 ID 的唯一计数(Unique count) 3, 4。
- 列管理: 通过 **Multi-column editor**(多列编辑器) 重命名列(如将 EQ ID 改为 Equipment ID)并删除不必要的列(如颜色、重量等) 4, 5。
- 数据修正:
- 使用 **Find and Replace**(查找与替换) 板清除字符串中的杂质(例如去除前缀 p-) 5。
- 使用 **Convert types**(类型转换) 将字符串字段(如 Purity)转换为双精度浮点数(Double),以便进行数值计算 6。
- 使用 **Filter**(过滤) 板按日期范围或特定条件筛选行 6。

B. 路径(Paths)与数据集成

- 路径的概念: Contour 的分析可以像浏览器标签页一样创建多个路径(**Paths**) 7。新路径可以从全新的数据集开始,也可以从现有分析的中间结果开始 7, 8。
- 数据连接(**Join**): 在名为“Parts and Equipment join”的路径中,将零件数据与设备数据集连接起来 8。连接过程中可以添加前缀以避免列名冲突,并设置匹配条件(如设备 ID) 8, 9。

C. 高级逻辑处理(Expression Board)

- **SQL** 式逻辑: 使用 **Expression board**(表达式板) 编写类似 SQL 的逻辑 9, 10。
- 派生新指标:
- 计算平均纯度(**Average Purity**): 利用开窗函数(Partitioning)按设备 ID 分组计算平均值 10。
- 生成检查警告(**Inspection Alert**): 使用 CASE WHEN 条件逻辑,根据设备检查时长和平均纯度自动生成风险等级(如高、中、低警告) 10, 11。

3. 可视化与参数化 (Visualizations & Parameters)

- **参数化交互**: 创建一个名为“Plant”的参数(**Parameter**), 并链接到设备工厂列 12。这允许最终用户通过多选下拉菜单过滤整个看板的数据, 而无需手动编写过滤条件 12, 13。
- **核心图表类型**:
- **直方图(Histogram)**: 展示不同等级警告的频率 14。Contour 的直方图支持交叉过滤, 点击某个条柱即可过滤下游所有数据 14。
- **透视表(Pivot Table)**: 按工厂和 ID 进行分组统计, 并可格式化单元格数据(如百分比或货币) 15。
- **散点图与条形图**: 展示平均纯度与零件计数的分布, 并支持**堆叠(Stacked)或分组(Grouped)**显示 16, 17。

4. 看板发布与成果分享

- **Dashboard(看板)**: 看板是 Contour 分析的一个精简、用户友好的视图 18。用户可以将分析过程中生成的直方图、透视表和图表“添加到看板(Add to dashboard)”中 18。
- **最终交付**:
- 在看板视图中, 之前设置的**参数(Parameters)****会出现在侧边栏, 供非技术用户进行筛选交互 18。
- 结果可以导出为 **PDF**, 或将数据板内容复制到 **Notepad** 应用中进行更正式的汇报 19。

5. 总结

该课程不仅教会用户如何使用单一的 Board, 更强调了如何通过 **Path** 的组织逻辑和 **Parameter** 的交互设计, 构建出一个具备工业级水平的设备风险监测系统 19。通过此流程, 原本杂乱的原始数据变成了能够指导生产决策的高风险机械检查清单 1, 19。