

1.图表解读:

**标题:** 图表的标题是“Initial Model Performance Comparison (MSE)”，意为“初始模型性能比较 (MSE)”。

**X 轴:** 表示模型的名称，有两个模型被比较，分别是 Model1 和 Model2。

**Y 轴:** 表示均方误差 (MSE) 的值，MSE 越低表示模型的预测精度越高。

**条形图:**

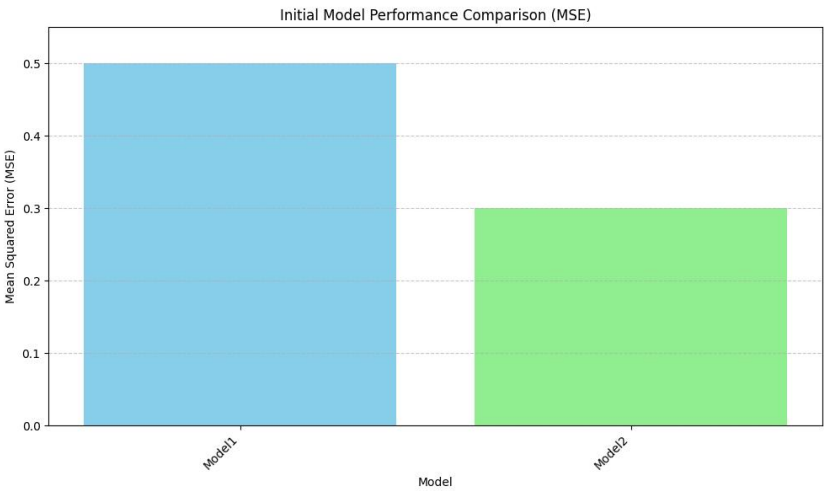
Model1 的 MSE 值较高，大约为 0.5。

Model2 的 MSE 值较低，大约为 0.3。

**颜色:** 两个模型的条形使用了不同的颜色，Model1 使用了蓝色，Model2 使用了绿色，以便于区分。

**网格线:** 图中有网格线，帮助更准确地读取 MSE 值。

**结论:** 从图中可以看出，Model2 的初始性能优于 Model1，因为它的 MSE 值更低。这意味着在没有经过优化的情况下，Model2 在测试数据集上的预测误差更小，预测结果更准确。这种比较可以帮助我们选择更好的初始模型，或者确定哪些模型需要进一步的优化。



2.图表解读:

**标题:** 图表的标题是“Model Performance Comparison: Initial vs. Optimized (MSE)”，意为“模型性能比较：初始 vs. 优化后 (MSE)”。

**X 轴:** 表示模型的名称，有两个模型被比较，分别是 Model1 和 Model2。

**Y 轴:** 表示均方误差 (MSE) 的值，MSE 越低表示模型的预测精度越高。

**条形图:**

每个模型都有两根条形，分别代表初始 MSE (蓝色) 和优化后的 MSE (绿色)。

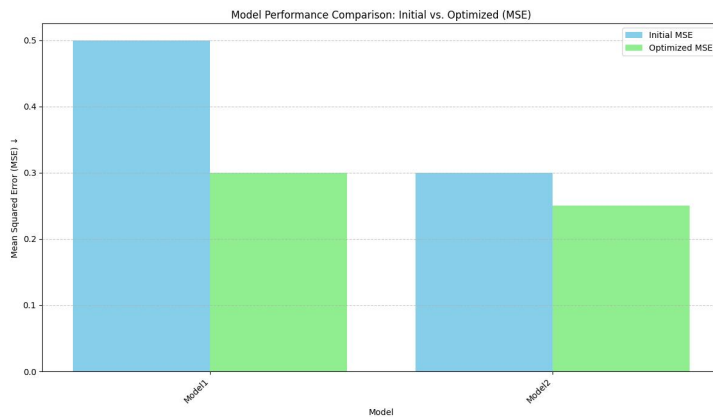
Model1 的初始 MSE 值较高，大约为 0.5，优化后降低到大约 0.3。

Model2 的初始 MSE 值较低，大约为 0.3，优化后进一步降低到大约 0.25。

**图例:** 图例说明了颜色的含义，蓝色代表“Initial MSE” (初始 MSE)，绿色代表“Optimized MSE” (优化后的 MSE)。

**网格线:** 图中有网格线，帮助更准确地读取 MSE 值。

**结论:** 从图中可以看出，两个模型在经过优化后，其 MSE 值都有所降低，这意味着模型的预测精度得到了提升。对于 Model1，优化显著降低了 MSE 值，表明优化对 Model1 的性能提升有较大影响。对于 Model2，虽然初始 MSE 值已经较低，但优化仍然进一步降低了 MSE 值，显示出优化过程对提升模型性能的积极作用。这种比较可以帮助我们理解优化过程对模型性能的具体影响，以及不同模型在优化前后的性能变化。



### 3.图表解读:

**标题:** 虽然图中没有明确的标题,但内容明显是关于特征重要性的分析。

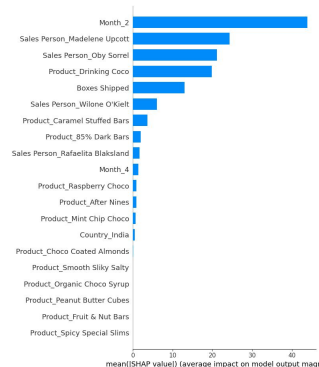
**X 轴:** 表示特征对模型输出的平均影响大小,即  $\text{mean}(|\text{SHAP value}|)$ ,也就是所有样本中该特征 SHAP 值绝对值的平均。数值越大,表示该特征对模型输出的影响越大。

**Y 轴:** 列出了各个特征的名称。

**条形图:**

每个条形代表一个特征,条形的长度表示该特征对模型输出的平均影响大小。从图中可以看出,Month\_2 对模型输出的影响最大,其次是 Sales Person\_Madeleine Upcott 和 Sales Person\_Oby Sorrel。其他特征如 Product\_Drinking Coco、Boxes Shipped 等也对模型有显著影响,但影响程度依次递减。一些特征如 Product\_Smooth Silky Salty、Product\_Organic Choco Syrup 等对模型的影响非常小。

**结论:** 此图有助于识别哪些特征是模型预测的关键驱动因素。通过分析这些特征,可以更好地理解模型的行为,以及哪些因素对模型预测结果影响最大。这种信息对于特征选择、模型优化和业务决策都是有价值的。使用场景:特征重要性分析常用于提高模型的可解释性,特别是在需要向非技术人员解释模型预测结果时。可以帮助数据科学家识别和专注于最重要的特征,从而提高模型的性能和效率。在业务环境中,了解哪些特征对预测结果影响最大,可以帮助制定更有针对性的策略和决策。



### 4.图表解读:

**X 轴:** 表示 SHAP 值,即特征对模型输出的影响大小。正值表示特征增加了模型的预测值,负值表示特征减少了模型的预测值。

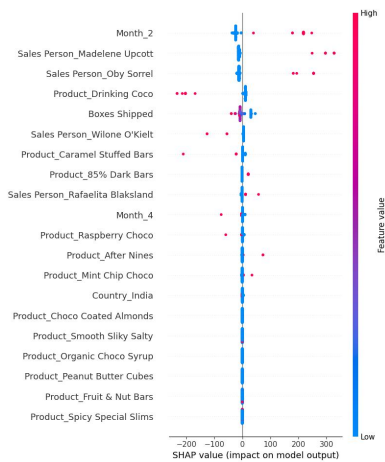
**Y 轴:** 列出了各个特征的名称。

**点的颜色:** 表示特征值的大小,颜色从蓝色(低值)到红色(高值)渐变。图例显示了颜色与特征值的对应关系。

**点的分布：**每个特征的点分布表示该特征对不同样本预测的影响。点越远离 0，表示该特征对模型输出的影响越大。点的颜色（特征值大小）可以帮助理解特征值的变化如何影响模型的预测。

**结论：**从图中可以看出，Month\_2、Sales Person\_Madeleine Upcott 和 Sales Person\_Oby Sorrel 等特征对模型输出有较大的影响。特征值的大小（颜色）与 SHAP 值的关系可以帮助理解特征值的变化如何影响模型的预测。例如，对于 Month\_2，高特征值（红色点）倾向于增加模型的预测值。这种图表有助于识别哪些特征是模型预测的关键驱动因素，以及这些特征如何影响模型的预测结果。

**使用场景：**特征重要性分析常用于提高模型的可解释性，特别是在需要向非技术人员解释模型预测结果时。可以帮助数据科学家识别和专注于最重要的特征，从而提高模型的性能和效率。在业务环境中，了解哪些特征对预测结果影响最大，可以帮助制定更有针对性的策略和决策。



## 5.图表解释

这张图片是一个 SHAP（SHapley Additive exPlanations）决策图（Decision Plot），用于展示模型对特定样本的预测是如何由各个特征的贡献累加而成的。

**X 轴：**表示模型的输出值，即预测值。

**Y 轴：**列出了各个特征的名称，这些特征按照对模型输出影响的大小排序。

**线条：**每条线代表一个样本的预测路径。线的起点是模型的基线预测值（即没有特征信息时的预测值），线的终点是考虑所有特征后的最终预测值。特征按照对预测值影响的大小排序，最上面的特征对预测值的影响最大。

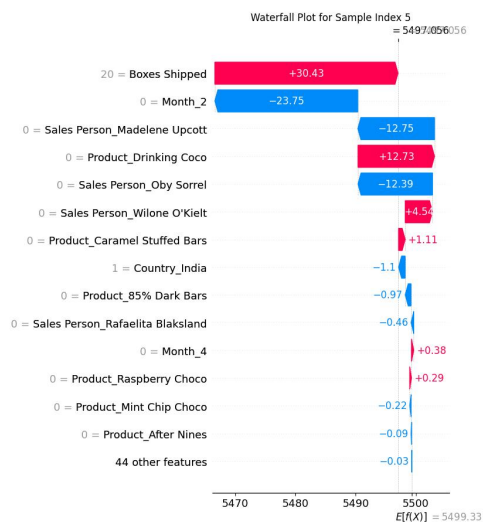
**颜色：**线条的颜色表示特征值的大小，颜色从蓝色（低值）到红色（高值）渐变。颜色的变化可以帮助理解特征值的变化如何影响模型的预测。

**结论：**从图中可以看出，Month\_2、Product\_Drinking Coco、Sales Person\_Madeleine Upcott 等特征对模型的预测有显著影响。特征值的大小（颜色）与对预测的影响（线条的方向和长度）相结合，可以帮助理解特征值的变化如何影响模型的预测。例如，对于 Month\_2，高特征值（红色）倾向于增加模型的预测值。这种图表有助于识别哪些特征是模型预测的关键驱动因素，以及这些特征如何影响模型的预测结果。

**使用场景：**决策图常用于解释单个样本的预测结果，帮助理解模型的预测逻辑。可以帮助数据科学家和业务分析师理解模型的行为，以及特征如何影响预测结果。在需要向非技术人员解释模型预测结果时，这种图表可以提供直观的解释。

**注意事项：**决策图通常用于展示单个样本的预测路径，但图中显示了多个样本的路径，这

可能会使得图表难以解读。在实际应用中，通常会选择一个或几个特定的样本来展示决策图，以便更清晰地解释模型的预测逻辑。



### 7.图表解读:

这张图片是一个 SHAP (SHapley Additive exPlanations) 瀑布图 (Waterfall Plot)，用于详细分析单个样本 (样本索引为 0) 的预测路径，展示了每个特征对该样本预测值的贡献。

顶部数值: = 5458.084084 表示模型对样本索引 0 的最终预测值。

底部数值:  $E[f(X)] = 5499.33$  表示模型的期望预测值 (即所有特征对预测值的平均贡献)。

特征贡献: 每个矩形条表示一个特征对预测值的贡献。蓝色条表示特征贡献为负, 即减少预测值; 红色条表示特征贡献为正, 即增加预测值。条的长度表示贡献的大小, 负值向左延伸, 正值向右延伸。

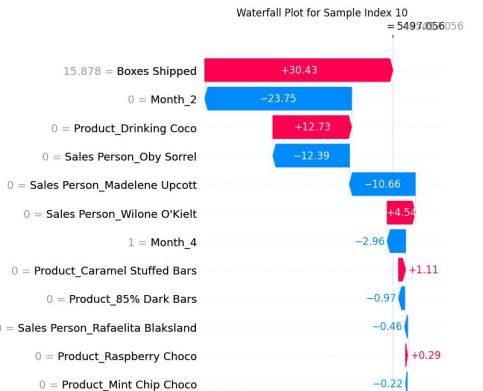
**特征名称:** 每个条上方列出了特征的名称, 例如 Month\_2、Sales Person\_Madeleine Upcott 等。特征名称旁边的数值 (如 0 =、60 =) 表示特征值, 其中 0 通常表示特征的基线值或缺失值。

**具体分析:** Month\_2 特征对预测值有最大的负向贡献 (-23.75), 意味着这个月份可能与较低的预测值相关。Sales Person\_Madeleine Upcott 和 Sales Person\_Oby Sorrel 也有显著的负向贡献, 表明这些销售人员可能与较低的销售额相关。

Product\_Drinking Coco 特征有正向贡献 (+11.71), 表明这种产品可能与较高的销售额相关。其他特征如 Boxes Shipped、Sales Person\_Wilone O'Kiel 等也有不同程度的贡献, 但影响较小。

**结论:** 瀑布图清晰地展示了每个特征如何推动或拉低模型的预测值, 从而帮助理解模型的预测逻辑。通过分析这些贡献, 可以识别出哪些特征对模型预测最为关键, 以及它们是如何影响预测结果的。这种图表对于提高模型的可解释性和透明度非常有用, 特别是在需要向非技术人员解释模型预测结果时。

**使用场景:** 瀑布图适用于详细分析单个样本的预测结果, 帮助理解模型的预测逻辑。可以帮助数据科学家和业务分析师理解模型的行为, 以及特征如何影响预测结果。在需要向非技术人员解释模型预测结果时, 这种图表可以提供直观的解释。



## 8.图表解读:

这张图片是一个 SHAP (SHapley Additive exPlanations) 瀑布图 (Waterfall Plot), 用于分析特定样本 (样本索引为 10) 的预测值是如何由各个特征的贡献累加而成的。

**顶部数值:** = 5497.056056 表示模型对样本索引 10 的最终预测值。底部数值:  $E[f(X)] = 5499.33$  表示模型的期望预测值 (即所有特征对预测值的平均贡献)。

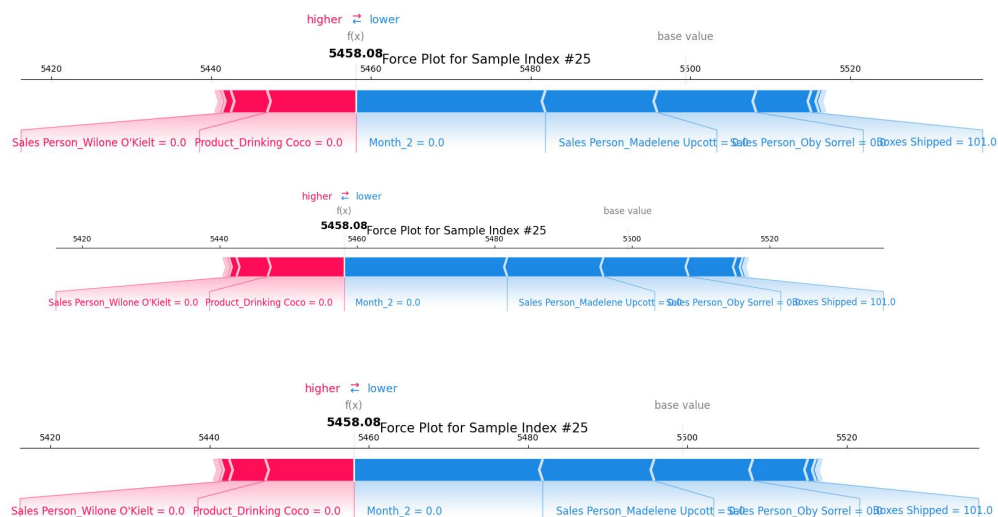
**特征贡献:** 每个矩形条表示一个特征对预测值的贡献。蓝色条表示特征贡献为负, 即减少预测值; 红色条表示特征贡献为正, 即增加预测值。条的长度表示贡献的大小, 负值向左延伸, 正值向右延伸。

**特征名称和值:** 每个条上方列出了特征的名称, 例如 Boxes Shipped、Month\_2 等。特征名称旁边的数值 (如 15.878 =) 表示特征的实际观测值。

**具体分析:** Boxes Shipped 特征对预测值有最大的正向贡献 (+30.43), 意味着这个特征的值显著增加了预测值。Month\_2 特征有负向贡献 (-23.75), 表明这个月份可能与较低的预测值相关。Product\_Drinking Coco 特征有正向贡献 (+12.73), 表明这种产品可能与较高的预测值相关。Sales Person\_Oby Sorrel 特征有负向贡献 (-12.39), 表明这位销售人员可能与较低的销售额相关。

**结论:** 瀑布图清晰地展示了每个特征如何推动或拉低模型的预测值, 从而帮助理解模型的预测逻辑。通过分析这些贡献, 可以识别出哪些特征对模型预测最为关键, 以及它们是如何影响预测结果的。这种图表对于提高模型的可解释性和透明度非常有用, 特别是在需要向非技术人员解释模型预测结果时。

**使用场景:** 瀑布图适用于详细分析单个样本的预测结果, 帮助理解模型的预测逻辑。可以帮助数据科学家和业务分析师理解模型的行为, 以及特征如何影响预测结果。需要向非技术人员解释模型预测结果时, 这种图表可以提供直观的解释。



## 9.图表解读:

这些图片是 SHAP 力图 (Force Plots), 用于解释三个不同样本 (索引分别为 7、15 和 25) 的模型预测。每个力图展示了各个特征如何影响模型的预测结果, 以及这些影响是如何累加到最终预测值的。

**顶部数值:** 每个图表顶部的数值 (如 5239.66、5458.08) 表示模型对相应样本的最终预测值。

**底部数值:**  $E[f(X)]$  表示模型的基线预测值, 即所有特征对预测值的平均贡献。



**特征贡献：**每个箭头代表一个特征对预测值的贡献。

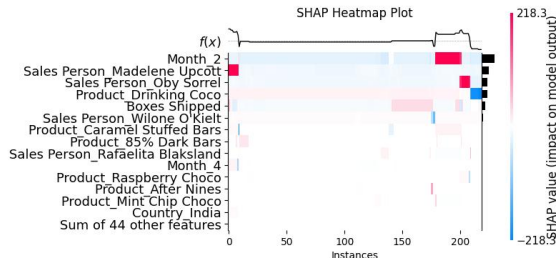
箭头的方向表示特征对预测值的影响方向：向右（红色）表示增加预测值，向左（蓝色）表示减少预测值。箭头的长度表示特征对预测值的影响大小。

特征名称和值：每个箭头旁边列出了特征的名称和值，例如 Sales Person\_Wilone O'Kiet = 0.0 表示特征 Sales Person\_Wilone O'Kiet 的值为 0。特征值旁边的数字（如 1.0、0.0）表示特征的实际观测值。

**具体分析：**样本索引 #7 Sales Person\_Wilone O'Kiet = 0.0 对预测值有显著的负向影响（降低预测值）。Product\_Drinking Coco = 1.0 对预测值有正向影响（增加预测值）。其他特征如 Month\_2、Sales Person\_Oby Sorrel 等也有不同程度的影响。样本索引 #15 Sales Person\_Wilone O'Kiet = 0.0 和 Product\_Drinking Coco = 0.0 对预测值有显著的负向影响。Month\_2 = 0.0 和 Sales Person\_Madelen Upcott = 5 对预测值有正向影响。Boxes Shipped = 249.0 进一步增加了预测值。样本索引 #25 Sales Person\_Wilone O'Kiet = 0.0 和 Product\_Drinking Coco = 0.0 对预测值有显著的负向影响。Month\_2 = 0.0 和 Sales Person\_Madelen Upcott = 5 对预测值有正向影响。Boxes Shipped = 101.0 也对预测值有正向影响。

**结论：**SHAP 力图清晰地展示了每个特征如何推动或拉低模型的预测值，从而帮助理解模型的预测逻辑。通过分析这些贡献，可以识别出哪些特征对模型预测最为关键，以及它们是如何影响预测结果的。这种图表对于提高模型的可解释性和透明度非常有用，特别是在需要向非技术人员解释模型预测结果时。

**使用场景：**力图适用于详细分析单个样本的预测结果，帮助理解模型的预测逻辑。可以帮助数据科学家和业务分析师理解模型的行为，以及特征如何影响预测结果。在需要向非技术人员解释模型预测结果时，这种图表可以提供直观的解释。



## 10.图表解读：

X 轴：表示样本的索引，即数据集中的每个实例。

Y 轴：列出了模型中使用的特征名称。

颜色：表示 SHAP 值的大小，颜色从蓝色（负向影响，即降低预测值）到红色（正向影响，即增加预测值）渐变。颜色条上的数值表示 SHAP 值的范围，从-218.3 到 218.3。

颜色强度：颜色越深，表示该特征对模型输出的影响越大。

分布：每个特征对应的颜色条的分布显示了该特征对不同样本预测结果的影响。

**具体分析：**Month\_2、Sales Person\_Madelen Upcott 和 Sales Person\_Oby Sorrel 等特征在图中显示为较大的红色或蓝色块，表明这些特征对模型预测结果有较大的影响。特征 Boxes Shipped 也显示为较大的红色块，表明它对预测结果有显著的正向影响。其他特征如 Product\_Drinking Coco、Sales Person\_Wilone O'Kiel 等对模型预测结果的影响较小，颜色块较小且分布较为分散。

**结论：**SHAP 热图有助于识别哪些特征对模型预测结果有较大的影响。通过分析这些特征的影响，可以更好地理解模型的行为和预测逻辑。这种图表对于提高模型的可解释性和透明度

非常有用，尤其是在需要向非技术人员解释模型预测结果时。

**使用场景：**SHAP 热图适用于全局特征重要性分析，帮助理解模型的整体行为。可以帮助数据科学家和业务分析师识别关键特征，从而进行特征选择或模型优化。在业务环境中，了解哪些特征对预测结果影响最大，可以帮助制定更有针对性的策略和决策。