设计考量

我们注意到批发数据集在5月或6月出现了不同程度的缺失，故目标预测7月1日-7日之前，进行对5月和6月的数据预测并填充，最后再进行预测

数据集划分

考虑到最后2个月的数据缺失，故增大训练集的长度。为了保证训练数据/测试数据比例接近0.2，我们将分配给训练数据时长4月的时间序列，将随后的时长1月的时间序列分配给验证数据，并通过键值对建立映射。训练数据进行模型训练，随后使用模型进行长达一个月的预测，并计算出随后1个月的预测结果，并使用均方根误差进行性能量化。

模型搭建

训练基于ARIMA模型的时间序列预测模型，同时对每个时间序列数据执行超参数搜索以找到最佳的ARIMA配置。函数接受训练数据和验证数据的字典，并返回一个包含模型和RMSE（均方根误差）的字典，其中每个键对应一个时间序列数据的模型和RMSE。

训练时执行多个训练周期（epochs），在每个周期内进行以下操作：

针对每个时间序列数据进行训练，遍历每个时间序列。

针对每个时间序列的每种可能的差分（d）值（d\_values\_to\_try），执行以下步骤：

训练自动ARIMA模型，通过自动选择ARIMA模型的超参数。

在验证集上进行预测，并计算RMSE。

如果当前的RMSE值比之前的最佳RMSE值更小，更新最佳模型、最佳d值和最佳RMSE值。

将每个时间序列的最佳模型存储在 best\_models 列表中。