## 项目三：数据挖掘模型实践

**1.学时**

**课内学时：4学时**

**课外学时：2学时**

**2.项目目的与要求**

通过实践让学生深入理解数据挖掘模型--关联分析的应用及操作过程。

**3.项目环境**

**SQL Server BI**

**4.项目内容**

(1) 在依赖关系网络图，移动滑块至最强链接，有关联的商品有哪些？

(2) 在依赖关系网络图，移动滑块至所有链接，与ML Road Tire相关联的商品都有什么？分别是什么关系？

(3) 在“项集”选项卡，“筛选项集”框中输入“Touring Tire”并按回车，显示结果是什么？

(4) 任选个项集，右键选择“钻取”-“模型和结构列”，你选择的是什么样的项集？有多少订单包含该项集？

(5) 清空筛选项集，最低支持改为1500，有多少频繁项集？

(6) 选择“规则”选项卡，在“筛选规则”框中输入“Mountain-200”,选择”仅显示属性名称”，有多少规则包含“Mountain-200”？

**5.项目步骤**

(1)新建数据源视图，选择表vAssocSeqLineItems和表vAssocSeqOrders。

(2)选择表vAssocSeqLineItems的OderNumber列，向表vAssocSeqOrders的OrderNumber列拖动，使其建立关联关系。

(3)新建挖掘结构。“创建数据挖掘结构” 页面选择关联规则。数据源视图选择刚刚创建的数据源视图。

(4)在“指定表类型”页面，设定vAssocSeqLineItems为嵌套表，设定vAssocSeqOrders为事例表。

(5)在“指定定型数据”页面，选中OrderNumber为“键”，选中Model为“键”、“输入”、“可预测”，选中IncomeGroup和Region左侧的复选框。

(6)在“创建测试集”页面，“测试数据百分比”设为0。

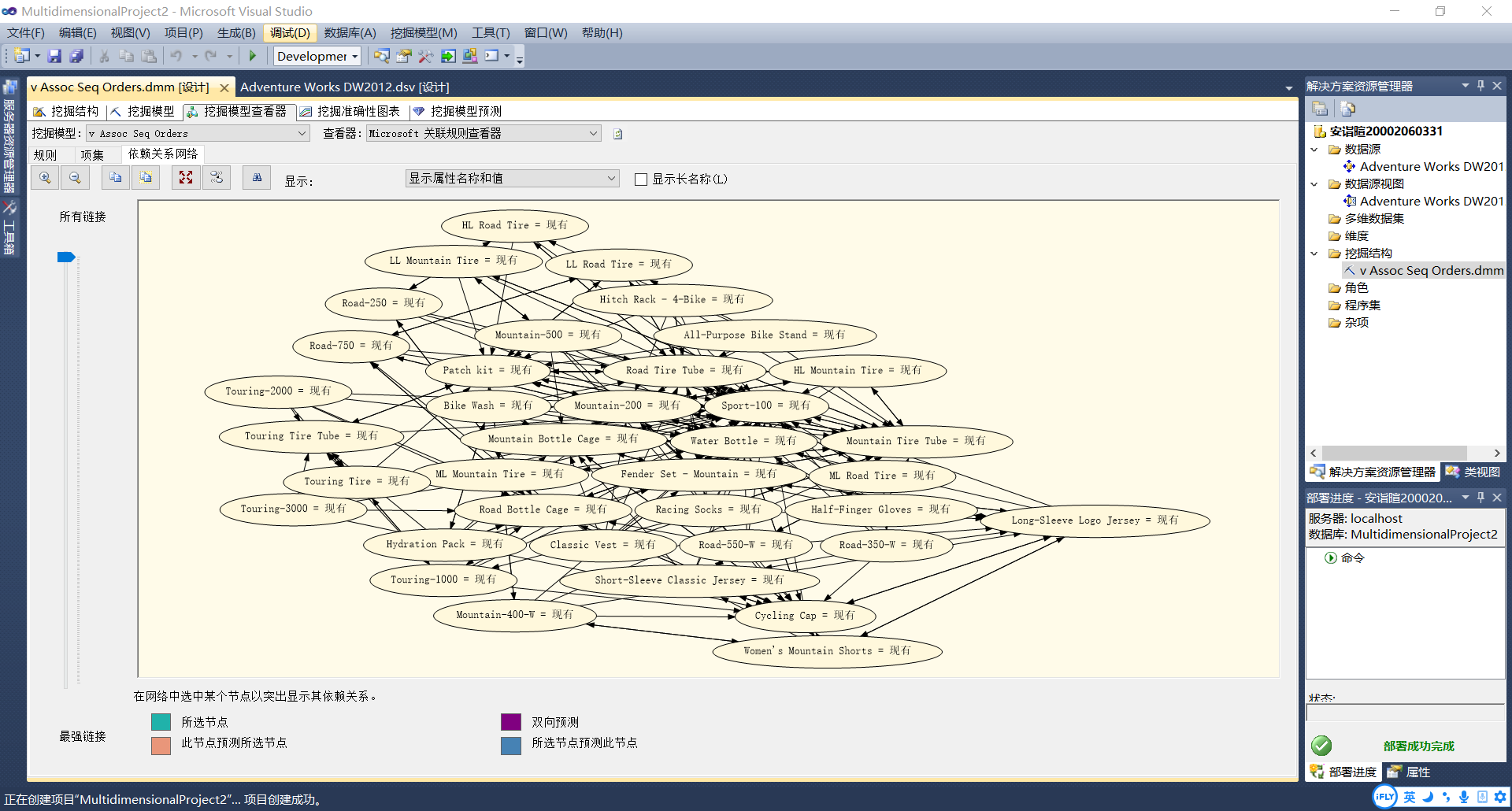
(7)在“完成向导”页“挖掘结构名称”及“挖掘模型名称”均可设为“Association”,选择“允许钻取”选项。完成模型设定。

(8)在“挖掘模型”选项卡,右键点击“Association”,选中“设置算法参数”，设置最小可信度为0.1, 设置最小支持度为0.001。

菜单“挖掘模型”下选择“处理挖掘结构和所有模型” 并运行。

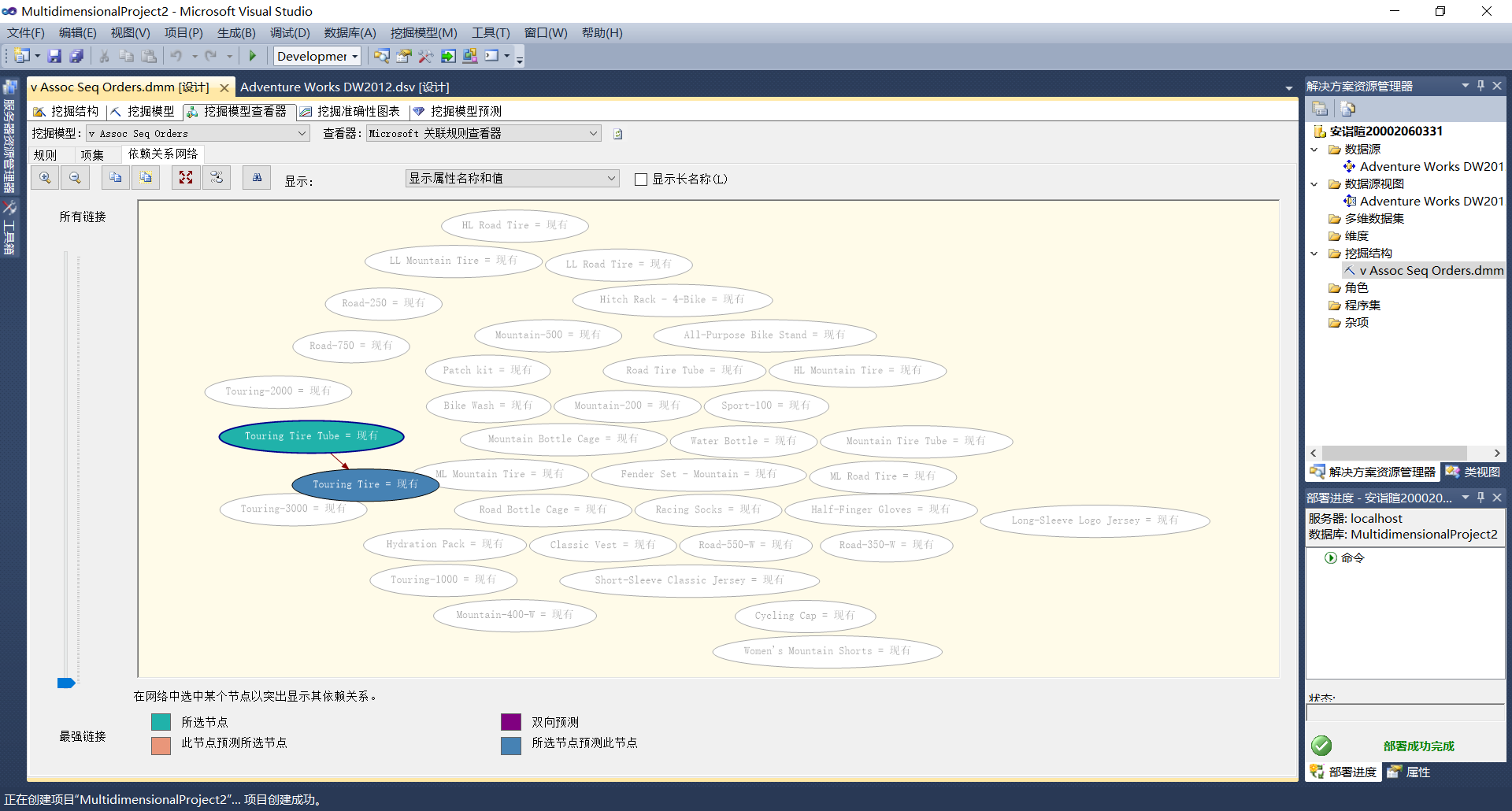
**6.成果物**

实验成果展示



(1) 在依赖关系网络图，移动滑块至最强链接，有关联的商品有哪些？

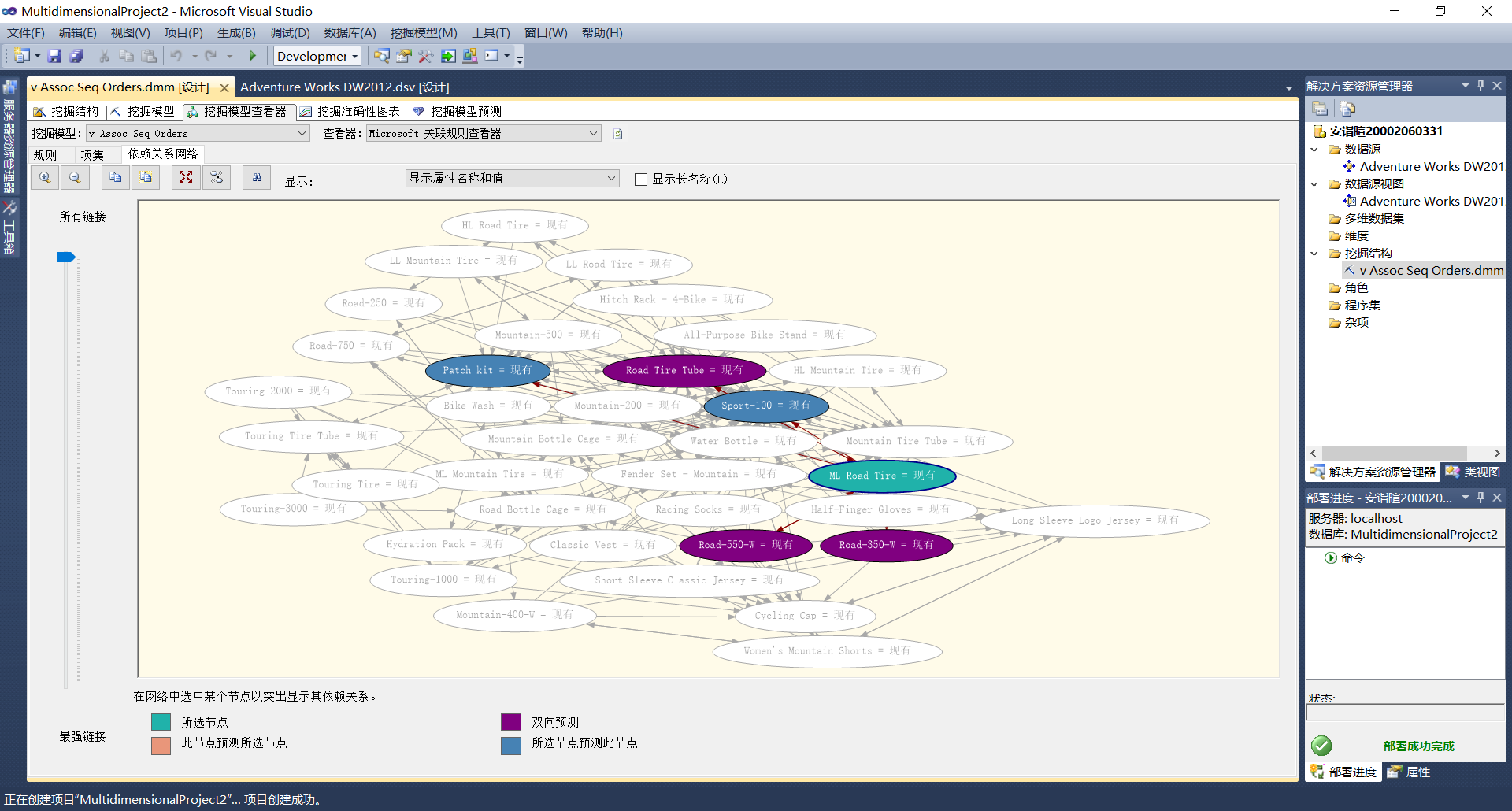
有关联的商品有Touring Tire Tube和Touring Tire



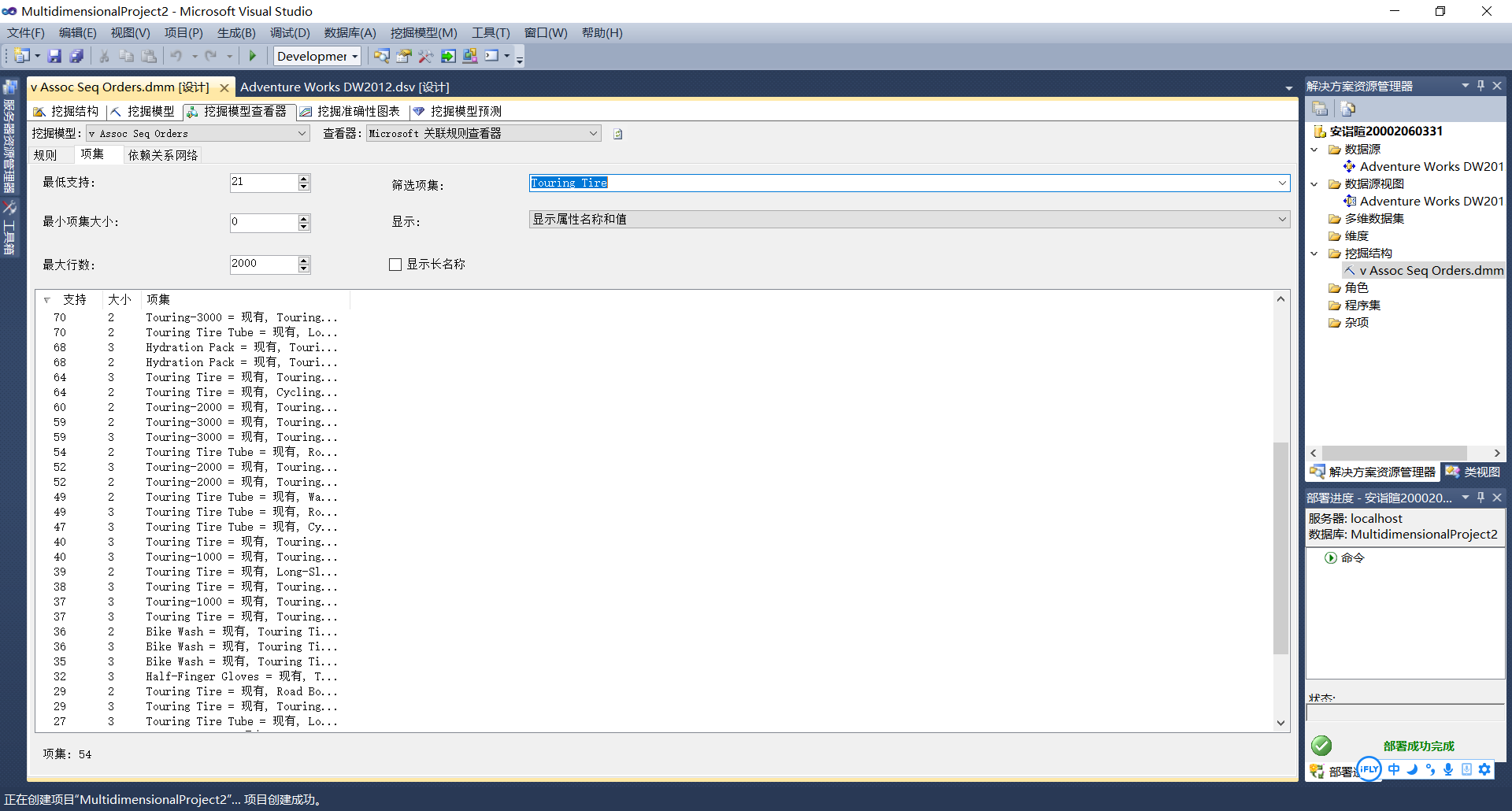
(2) 在依赖关系网络图，移动滑块至所有链接，与ML Road Tire相关联的商品都有什么？分别是什么关系？

与ML Road Tire相关联的商品有Patch kit, Sport-100, Road Tire Tube, Road-550-W, Road-350-W。

ML Road Tire与Road Tire Tube, Road-550-W, Road-350-W为双向预测关系；ML Road Tire单向预测Patch kit, Sport-100这两个节点。

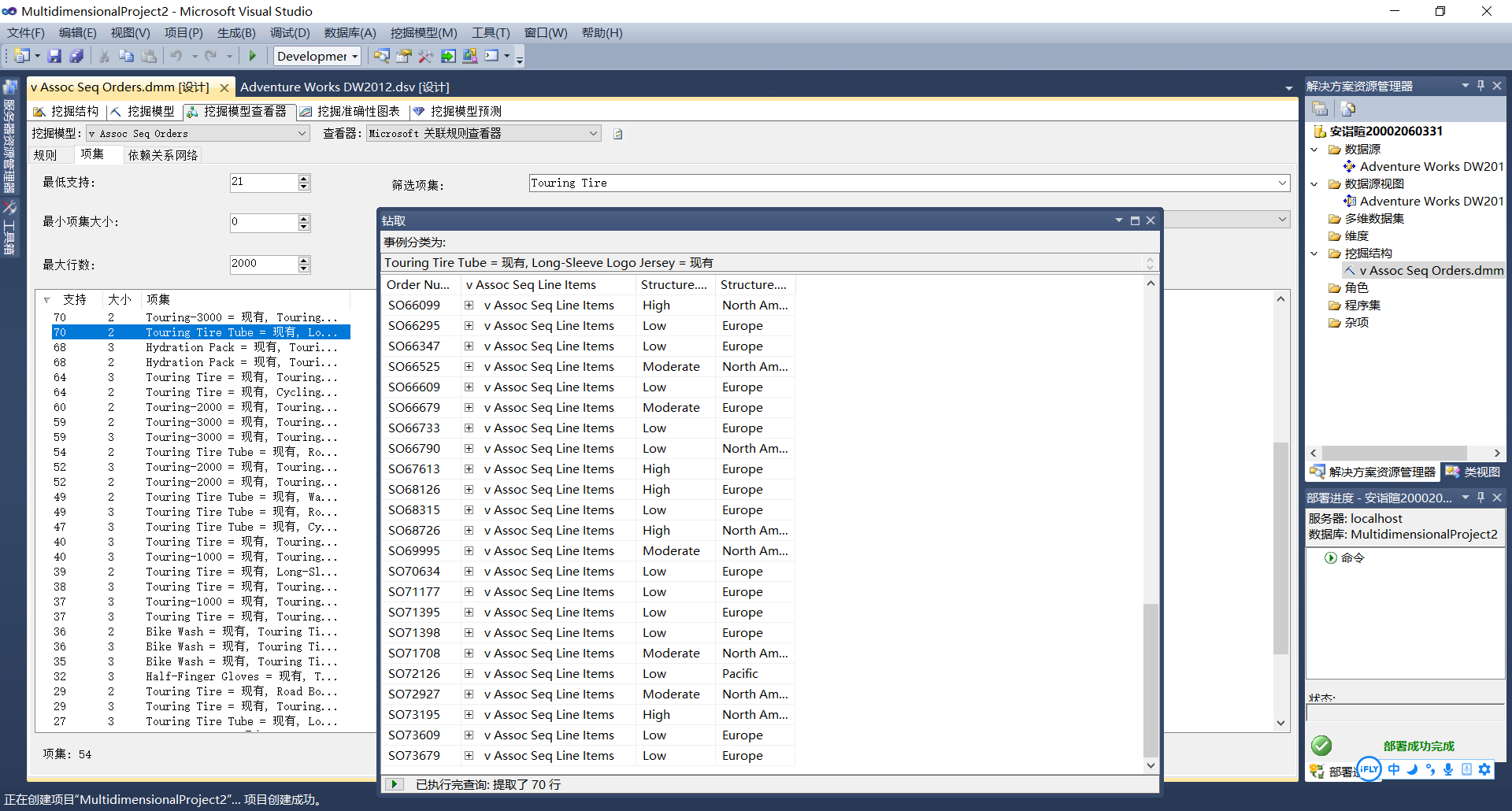


(3) 在“项集”选项卡，“筛选项集”框中输入“Touring Tire”并按回车，显示结果是什么？



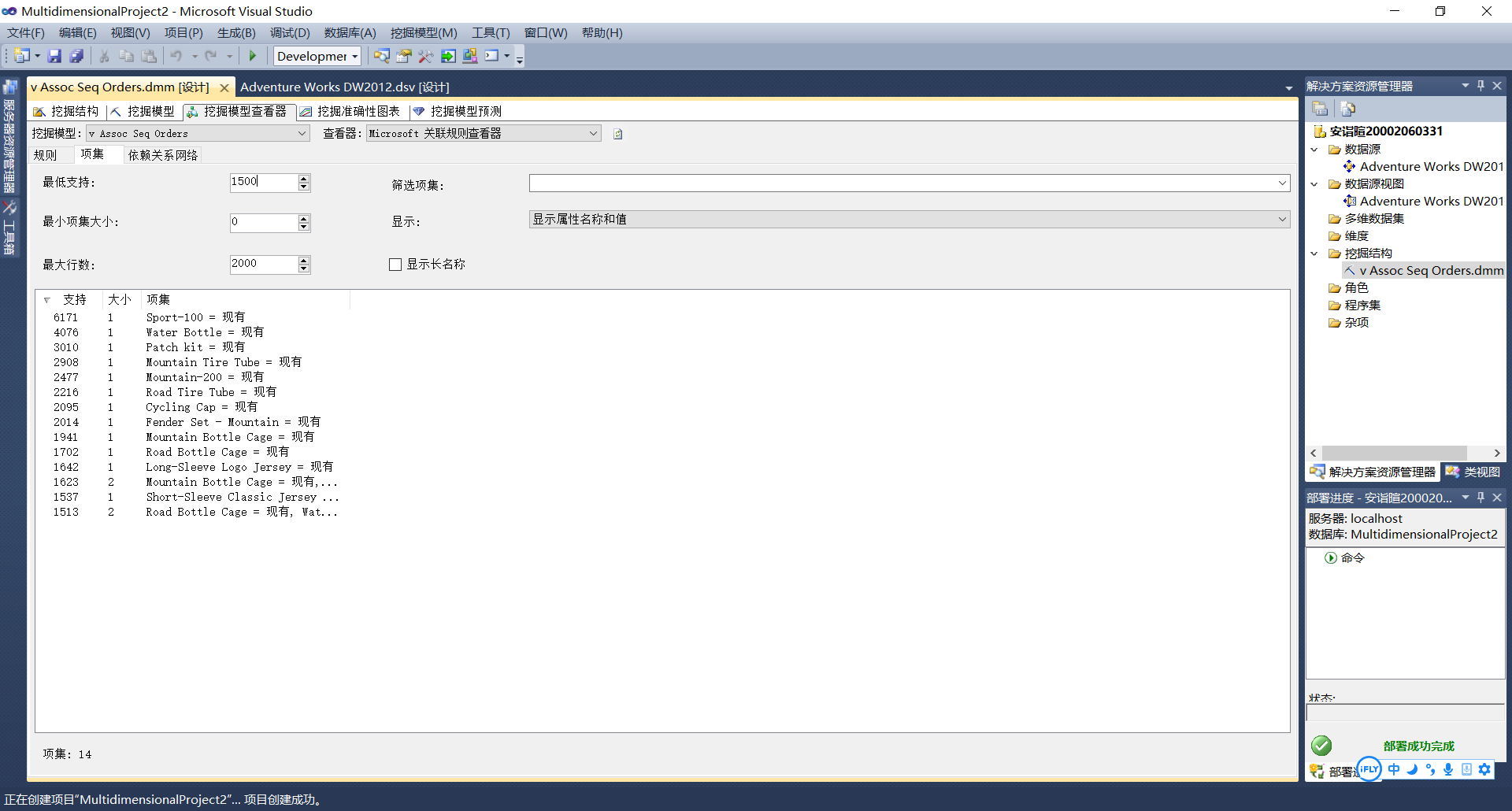
(4) 任选个项集，右键选择“钻取”-“模型和结构列”，你选择的是什么样的项集？有多少订单包含该项集？

我选择的项集是Long-Sleeve Logo Jersey,该项集包含70个订单。



(5) 清空筛选项集，最低支持改为1500，有多少频繁项集？

有14个频繁项集



(6) 选择“规则”选项卡，在“筛选规则”框中输入“Mountain-200”,选择”仅显示属性名称”，有多少规则包含“Mountain-200”？

有119个规则包含“Mountain-200”

