# 需求分析报告

项目背景

本次项目展示旨在通过Spark技术对多个电子商务相关的数据集进行分析处理，并将分析结果通过积木报表进行可视化展示。团队成员将从电子商务综合性及欺诈分析、京东平台冰箱销售数据分析以及化妆品电商数据分析三个方面进行数据处理和展示。

项目目标

利用Spark进行实时数据处理，提升数据分析效率。

使用积木报表进行数据的可视化展示，提高数据结果的可读性。

通过对不同数据集的分析，提供有价值的商业洞察和策略建议。

需求分析

1. 电子商务综合性及欺诈分析

需求数据集：

数据来源：Kaggle

数据大小：320.8MB

数据内容：电子商务交易的全面信息，包含交易金额、支付方式、产品类别、客户信息、设备信息、地址信息以及欺诈标记等。

需求功能：

分析不同年龄段网购人群占比，发现主要的网购用户群体。

分析不同支付方式在不同金额区间的使用情况。

分析账户注册时长与欺诈率之间的关系，制定防范措施。

分析不同产品类别的订单金额和市场需求。

目标：

提供针对不同年龄段的市场策略。

制定账户安全措施，降低欺诈风险。

推广高价产品，吸引高消费用户。

2. 京东平台冰箱销售数据分析

需求数据集：

数据来源：公开网络数据

数据大小：约200M订单数据

数据内容：京东2020年5月25日的大件家电-冰箱的订单数据

需求功能：

统计各个品牌冰箱的销售总量及占比。

计算各个品牌产品优惠后售价的平均值，并排名前五。

统计各省份的冰箱销量，并进行地图展示。

分析山东省销量前五的品牌及其销量。

目标：

了解不同品牌冰箱的市场占有率。

提供各品牌产品定价策略参考。

分析不同区域的市场需求，制定区域营销策略。

3. 化妆品电商数据分析

需求数据集：

数据来源：Kaggle

数据大小：520MB，4635839行

数据内容：某化妆品电商十一月所有已注册用户的ID及其行为数据，包括浏览、加入购物车、移除购物车、购买等行为。

需求功能：

分析11月销量前五的品牌。

根据RFM模型划分用户价值，并统计各类用户的数量及占比。

分析11月当月的用户转换率。

分析双十一临周的ARPU变化。

分析双十一当天的分时销售量。

目标：

了解畅销品牌及其市场表现。

通过RFM模型划分用户价值，为市场活动提供参考。

提供用户转换率分析，提高平台的用户活跃度和转化率。

分析双十一促销活动对销售的影响，制定更有效的促销策略。

数据处理和可视化需求

技术工具：

Spark：用于大规模数据处理和分析。

SparkSQL：用于结构化数据的查询和分析。

Scala：用于编写高效的Spark应用程序。

积木报表：用于数据结果的可视化展示。

处理流程：

数据获取和预处理：从Kaggle和公开网络获取数据，对数据进行清洗和预处理。

数据分析：使用SparkSQL进行数据分析，提取所需的商业洞察。

数据可视化：利用积木报表将分析结果进行可视化展示，提高数据的可读性和决策支持能力。

结论

通过对不同电子商务数据集的深入分析，可以为平台和商家提供有价值的商业洞察和策略建议。利用Spark进行数据处理，积木报表进行可视化展示，能够高效地完成数据分析任务。项目的需求明确，目标清晰，具备较高的实施可行性和商业价值。