工作内容

# 工作成果及验收形式

1. 提交数据分析与总结报告
2. ppt远程汇报及答辩

# 技能掌握能力与目标

Python基本使用、数据分析、机器学习、数据可视化及MongoDB数据库操作。

# 数据集说明及目标

xx省份的XX银行针对辖区内的4个区域的多家支行的300余个网点的客户所作的调查问卷。调查问卷旨在了解客户对线上线下渠道的行为偏好和交易习惯，内容涵盖人口特征、业务办理渠道偏好、产品持有量、线上线下渠道协同建议等4个部分，共设计了20个封闭式问题，1个开放性问题。问卷以excel格式，其中每个文件包含多个sheet页，每个sheet为一个调查表。

该问卷用于分析线上线下渠道的交易的总体趋势。对客户的线上线下渠道偏好和客户行为需求进行了数理分析，着力探究不同人口特征客户对渠道选择行为偏好的影响，研究分析行内线上交易渠道和线下交易渠道交易规模的变化趋势，为银行定位目标客户、开展线上渠道和线下渠道的精准营销提供参考，最终提出多渠道协同平台的处理流程、技术解决方案及理论建议。

# 项目具体要求

1. 实现文件夹下所有调查问卷的遍历读取；注意其中包含.xlse和.et两种格式；
2. 实现所有问卷结果保存在MongoDB数据库中；有的题目是单选、有的是多选；
3. 客户统计与分析，包括问卷总数、有效问卷数量统计；受访者性别、年龄、学历统计与分析，并以可视化图形展示；
4. 客户日常行为分析，包括每月使用营业网点、使用手机银行的主力军、客户日常消费的主流支付方式、办理银行基础业务的首选、主要功能和常用功能个数分析和办理银行业务的主要渠道的分析及可视化展示；（注意考虑年龄、性别及学历的条件）
5. 客户日常行为回归分析，客户的客观个体特征：性别、年龄、学历；行为特征：月均去网点频次、月均使用手机银行APP频次；日常的习惯：日常消费所用支付方式、网点办理业务常用方式；
6. 客户与银行黏连度特征分析及偏好分析，包括营业网点办理业务、网点到店频次与线上渠道使用频次、客户与银行黏连度的特征分析、平均到网点频次与银行产品持有数量、手机银行使用频次与持有银行产品数量、客户看重银行服务的渠道偏好等进行分析及可视化展示；
7. 聚类分析，针对多个指标，进行聚类可视化展示；
8. 用户画像构建，以年龄、性别和学历等为基准，建立不同类型的用户画像，并进行可视化展示；并进行4个区域下的同种类型的用户画像对比；
9. 针对测试数据集，利用测试表1已知的用户基本信息，使用多种算法判断其后续是否为可以挖掘的新客户，并说明使用的各种算法的优劣；
10. 针对新用户的挖掘，问卷表中哪个基础性问题可以删除？
11. 该银行将加大推进线上银行业务，以数据进行分析，如何开展精准推荐挖掘新客户以及提供哪些服务可以吸引新客户？