# 一、业务分析

①本题分值：25 分

②考核时间：30min

③考核形式：技能操作

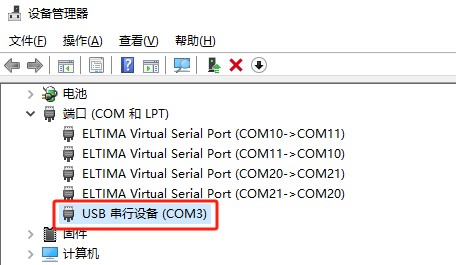
通过数据采集系统获取业务数据，并根据题目要求对数据进行预处理操作。

**1.1** 数据采集

根据题目要求进行业务分析，绘制数据采集程序流程图，编写相应程序实现从数据采集模块中采集数据 A.csv 和 B.csv 到编程计算机。

要求**:**

1. 将数据采集模块通过 USB 连接 PC，通过串口进行通信。设备管理器页面“端口”节点中可以查看到用于通信的串口 COM 号，通信波特率 19200，无校验。
2. 数据采集规则为：程序发送“startA”字符串，模块开始上传文件 A 数据，数据上传完成后返回“end”代表文件上传结束；程序发送“startB”字符串，模块开始上传文件 B 数据，数据上传完成后返回“end”代表文件上传结束。

  数据采集模块示例 USB 设备串口示例

1. 新建一个名为 1.docx 的 word 文件绘制采集流程图并编写程序 1.py 完成数据采集，完成后将流程图和程序代码及两份采集文件 A.csv 和 B.csv 复制到提交文件夹。
2. 模块通过状态指示灯显示工作状态，0.5Hz 闪烁代表等待命令，2Hz 闪烁代表数据上传中。

**1.2** 数据分析

编写程序对提供的 C.csv 和 D.csv 进行以下操作，必须采用 NoteBook(.ipynb) 格式提交，文件名为 1.ipynb，并将文件复制到提交文件夹。每道小题按照顺序在不同单元格中进行回答，如需跳过某道小题，请添加空白代码单元格。本题目共 6 题，6 个代码单元格。

要求**:**

1. 读取 C 数据文件并预览前 3 行。
2. 计算 C 数据文件内一共多少列。
3. 获取 C 数据文件内所有文本列。
4. 计算 C 数据文件内所有数字列的平均值。
5. 按照 C 数据文件内 Status 列分组后汇总所有数字列的总和。
6. 分析 C.csv 和 D.csv 的数据，根据分析结果合并两个文件数据（请用文字简要说明分析结果，文字以注释的形式编写）。

# 二、智能训练

①本题分值：35 分

②考核时间：42min

③考核形式：技能操作

根据提供的训练集，采用人工智能算法（算法不限）进行图片分类，保存训练后的模型，并调用模型对测试集数据进行评估。将训练程序代码、模型文件及测试代码复制到提交文件夹。

要求**:**

1、图片数据集规范如下：

1. 训练集为 2train.csv，测试集为 2test.csv，其中测试集数据不公开。
2. 第 0 列为标签，用数字 0，1，2，3 对每个类别进行编号。
3. 第 1~784 列为图片数据，图片大小为 28\*28。
4. 训练代码命名为 2Train.py，模型文件命名为 2Model，格式不限。
5. 测试代码命名为 2Test.py，运行后加载测试集数据，把测试分类结果绘制在测试集的前 10 张图片中，打印输出正确率等指标，创建 result 文件夹并把图片按 0.png、1.png、2.png…9.png 命名保存在该文件夹中。

# 图片保存结果示例三、智能系统设计（**35** 分）

①本题分值：35 分

②考核时间：42min

③考核形式：技能操作

根据提供的数据文件训练一个模型，用于预测未来的结果，保存训练后的模型，并调用模型对测试集数据进行评估。将训练程序代码、模型文件及测试代码复制到提交文件夹。

要求**:**

1、数据集规范如下：

1. 训练集为 3train.xlsx，测试集为 3test.xlsx，其中测试集数据不公开。
2. 内容为时间序列数据，例如：不同时间不同商店某种商品的价格。

2、训练代码命名为 3Train.py，模型文件命名为 3Model，格式不限。 3、测试代码命名为 3Test.py，加载测试集，运行后打印出最后 10 个时间周期的 MSE 值和预测结果。

# 四、培训与指导（**5** 分）

①本题分值：5 分

②考核时间：6min

③考核形式：文字作答

请说明基于人工智能的时间序列预测系统的数据采集流程及数据处理流程，并简要阐述潜在机会点和隐藏价值。新建一个名为 4.docx 的 word 文件作答，并将该文件复制到提交文件夹。