

# 2023 年“泰迪杯”数据分析技能赛

## B 题

### 企业财务数据分析与造假识别

#### 一、背景

财务数据是指企业经营活动和财务结果的数据记录，反映了企业的财务状况与经营成果。对行业、企业的财务数据进行分析，就是要评价其过去的经营业绩、衡量现在的财务状况、预测未来的发展趋势。财务数据分析不仅对企业生产经营管理有着重要的作用，而且对企业外部投资、贷款、赊销等决策有重要作用；经济管理部门也可以通过财务数据分析了解行业经济的运行状况，合理配置资源，科学调控经济运行，促进经济高质量发展。而财务造假则会对企业、投资者和整个市场带来极大的危害，准确甄别财务造假企业，帮助投资者识别风险，促进企业诚信经营，对营造良性的市场竞争环境有非常重要的现实意义。

#### 二、目标

1. 财务数据预处理。
2. 财务数据指标分析及可视化。
3. 建立企业利润预测模型，预测企业营收情况；并根据财务报表数据的特点，筛查财务造假的企业。

#### 三、任务

请根据提供的数据，选择分析工具完成以下任务，并撰写报告。

##### 任务 1 财务数据处理

任务 1.1 读取“LR.csv”，提取表 1 中所列字段的数据，筛选出字段“Typrep”值为“A”的数据，将筛选出的数据另存为文件“LR\_1.csv”（文件编码设置为 UTF-8），并在报告中呈现筛选后的数据行数、列数。

**表 1 数据指标筛选**

序号	指标名称	指标编号
1	证券代码	Stkcd
2	会计期间	Accper
3	报表类型	Typrep
4	利润总额	B001000000
5	营业总收入	B001100000
6	营业收入	B001101000
7	营业总成本	B001200000
8	营业成本	B001201000
9	营业税金及附加	B001207000
10	销售费用	B001209000
11	管理费用	B001210000
12	财务费用	B001211000
13	资产减值损失	B001212000
14	汇兑收益	B001303000
15	影响净利润的其他项目	B002300000

任务 1.2 读取“LR\_1.csv”、“ZCFZ.csv”、“Stk\_ind.csv”三个数据文件。根据“Stkcd”、“Accper”和“Typrep”三个字段，提取“ZCFZ.csv”中字段为“A002000000”和“A001000000”的相应数据，合并到“LR\_1.csv”中。根据字段“Stkcd”，提取“Stk\_ind.csv”中字段为“Indnme”和“Nindnme”的相应数据，合并到“LR\_1.csv”中。将完成合并的数据另存为文件“LR\_2.csv”（文件编码设置为 UTF-8），并在报告中呈现合并后数据的行数、列数。

任务 1.3 读取“LR\_2.csv”，删除空值占比达 70%及以上的数据列，将处理后的数据另存为文件“LR\_3.csv”（文件编码设置为 UTF-8），并在报告中呈现处理后数据的列数。

任务 1.4 读取“LR\_3.csv”，删除包含空值的行，将处理后的数据另存为文件“LR\_4.csv”（文件编码设置为 UTF-8），并在报告中呈现处理后数据的行数。

任务 1.5 读取“LR\_4.csv”，将字段“Accper”的日期数据转换为“YYYY-mm-dd”的格式，例如：“2018-1-31”转换为“2018-01-31”，将处理后的数据另存为文件“LR\_5.csv”（文件编码设置为 UTF-8）。

任务 1.6 读取“LR\_5.csv”，插入“利润率”和“资产负债率”两列。根据下表公式，计算对应的利润率和资产负债率，追加到“LR\_5.csv”对应字段。分别删除表中利润率、资产负债率不在[-300%，300%]范围内的行，将处理后的数

据另存为文件“LR\_new.csv”（文件编码设置为 UTF-8），并在报告中呈现处理后的数据行数、列数，及前 5 个企业的利润率、资产负债率。

**表 2 指标计算公式**

指标名称	计算方法
利润率	利润总额（B001000000）/营业总收入（B001100000）
资产负债率	负债合计（A002000000）/资产总计（A001000000）

## 任务 2 财务数据指标分析及可视化

任务 2.1 读取“LR\_new.csv”，根据表 3 要求统计数据，绘制相关的“行业营业利润对比分析”图，每张图表需在报告中进行呈现及分析。

**表 3 行业营业利润对比分析**

序号	图表标题	图表类型	图表内容
1	2019 年 9 月各行业大类的利润对比	柱状图	分别统计不同行业大类 2019 年 9 月利润总额的均值，并绘图
2	2018 年 1 月至 2019 年 9 月各行业大类利润率变化	折线图	分别统计不同行业大类 2018 年 1 月至 2019 年 9 月各季度利润率均值，在同一张图表中绘制各行业大类利润率均值变化折线图

任务 2.2 读取“LR\_new.csv”，根据任务 2.1 结果，确定 2019 年 9 月营业利润率均值排名第 1 的行业大类，并按表 4 要求绘制该行业大类相关的“行业企业营收分析”图，每张图表需在报告中进行呈现及分析。

**表 4 行业与企业营收分析**

序号	图表标题	图表类型	图表内容
1	2019 年该行业各细类利润率对比	柱状图	分别统计该行业大类不同细类 2019 年 9 月的利润率，绘制利润率排名前 3 细类的利润率柱状图
2	2019 年该行业利润率排名第 1 细类的企业利润率对比	柱状图	对该行业大类中利润率排名第 1 细类各企业 2019 年 9 月利润率进行排序，绘制该细类利润率排名前 5 企业的利润率柱状图（注：将该细类利润率排名第 1 的企业记为“T1”）
3	2019 年企业“T1”营业总成本分析	饼图	绘制企业“T1”2019 年 9 月财务报表的营业成本、营业税金及附加、销售费用、管理费用、财务费用的饼图
4	2019 年企业“T1”经营情况分析	柱状折线组合图	在同一张图表中，绘制企业“T1”2019 年 3 月、6 月、9 月三个季度营业总收入、营业总成本的柱状图，绘制利润率、资产负债率变化的折线图（可根据实际情况设置

			副坐标轴)
--	--	--	-------

任务 2.3 利用可视化大屏制作工具，将任务 2.1 和任务 2.2 所列的 6 张图制作成一个大屏，大屏命名为“行业与企业营业数据分析”，并在报告中呈现。要求大屏整体设计美观、布局清晰直观。

### 任务 3 企业利润预测及财务造假识别

任务 3.1 读取“financial\_data.csv”，计算各个指标与利润总额的相关性，挑选相关度最高的 5 个指标。

任务 3.2 利用挑选的 5 个指标建立企业利润预测模型，运用建立的模型预测“test.csv”表中给定企业的利润总额，并将预测结果以表格的形式在报告中呈现。

**表 5 结果展示样式**

TICKER_SYMBOL	LRZE
4953174	.....
4961537	.....
4962538	.....
4968740	.....
4973917	.....
4978589	.....
4978721	.....
4986535	.....
4990739	.....
4990942	.....

任务 3.3 “financial\_data.csv”中包含一个“FLAG”字段用于标识财务数据造假（“1”表示财务造假）。请利用表 6 所列关键因子，对样本数据“financial\_data.csv”进行分析，挖掘财务造假的识别特征。根据你们的分析，对“financial\_data\_new.csv”所列 5 个企业的财务数据进行筛查，识别其中唯一的 1 个涉嫌财务造假企业，并在报告中描述分析方法与结果。

表 6 关键因子

指标名称	指标编号
流动比率	LDBL
资产负债率	ZCFZL
存货周转率	CHZZL
资产报酬率	ZCBCL
应收账款周转率	YSZKZZL

#### 四、数据说明

赛题数据文件夹具体内容如下所示。

表 7 数据文件说明

文件名	内容说明
target_data.xlsx	赛题数据字段说明，包括：Stk_ind、ZCFZ、LR、financial_data 等
Stk_ind.csv	企业对应的行业
ZCFZ.csv	企业资产负债数据
LR.csv	企业营收利润数据
financial_data.csv	企业财务数据样本集
test.csv	企业财务数据预测集
financial_data_new.csv	包含 1 个数据造假的企业财务数据