企业财务数据分析与造假识别

摘要

随着市场经济的发展，企业之间的竞争日益加剧，为了追求企业价值最大化和经营管理的有效性，企业需要通过财务报表数据来正确评估自身的财务状况，以便判断其发展的前景。然而，伴随着这些机会和挑战的是财务报表造假的风险，一些企业可能会为了各种目的，如逃避税务、获取融资或提高市值，而对其财务数据进行造假。

为了对企业财务数据进行各项分析来预测企业营收等情况，并有效筛查财务造假的企业。本次作品使用 Python 进行编写代码，绘图使用 matplotlib 等工具。

针对任务一，主要进行财务数据处理，主要分为读取、提取、筛选、合并、删除等操作。在 Python 中，可以使用 pandas 库来提取数据，例如使用 read_csv( ) 从 CSV 文件中读取数据并使用 dropna( )函数删除包含空值的行；通过使用 columns 属性，可以获取对象的列名,如提取“Stk_ind.csv”中字段为“Indnme”和“Nindnme”的相应数据；使用 merge()函数将两个 DataFrame 对象按照指定的键进行合并；使用 datetime 模块来处理日期和时间，要将“Accper”字段的日期数据转换为“YYYY-mm-dd”格式，可以使用 strftime()方法。

针对任务二，进行财务数据指标分析及可视化，主要利用 Matplotlib 进行绘图，首先绘制出相关的“行业营业利润对比分析”图进行分析并比较可得出营业利润率均值排名第 1 的行业大类，进而逐步得出该行业排名最高的细类以及盈利最高的企业。在绘制的图形中亦可以观察指定时间的财务数据的盈利亏损等情况。

针对任务三，对企业利润预测及财务造假识别，利用 pandas 库中 df.corr()函数计算各个指标与利润总额的相关性，并利用热力图清楚的观察到相关性程度；建立的模型预测利润总额，最后进行筛选出涉嫌财务造假企业。

关键词：Anaconda；Jupyter；matplotlib

目录

1 财务数据处理

分别对企业营收利润数据“LR.csv”、企业资产负债数据“ZCFZ.csv”以及企业对应的行业“Stk_ind.csv“等进行数据处理。

任务 1.1 筛选数据

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），如图 1-1 所示。呈现筛选后的数据行数、列数，如图 1-2 所示。

表 1

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

```
import pandas as pd

# 读取CSV文件
data = pd.read_csv('LR.csv')

# 提取指定列的数据
selected_columns = ['Stkcd', 'Accper', 'Typrep', 'B001000000', 'B001100000', 'B001101000', 'B
data0 = data[selected_columns]
```

```
# 筛选出字段“Typrep”值为“A”的数据
filtered_data = data0.loc[data["Typrep"] == "A"]

# 将筛选出的数据另存为文件“LR_1.csv”，并设置文件编码为UTF-8
filtered_data.to_csv("LR_1.csv", index=False, encoding="utf-8")
```

图 1-1 提取指定列的数据，筛选数据并另存为“LR_1.csv”

```
filtered_data.shape
```

```
(33414, 15)
```

图 1-2 筛选后的数据行数、列数

"Typrep"是一个用于描述上市公司财务报表类型的变量，其取值可以是"A"或"B"。当其值为"A"时，表示报表为合并报表；当其值为"B"时，表示报表为母公司报表。由上述处理与分析得出，筛选出字段“Typrep”值为“A”的数据为 33414 行，15 列。

任务 1.2 提取数据

（1）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，如图 1-3 所示。

```
import pandas as pd

# 读取 LR_1.csv
lr_data = pd.read_csv('LR_1.csv')

# 读取 ZCFZ.csv
zcfz_data = pd.read_csv('ZCFZ.csv')

# 提取 ZCFZ.csv 中的列字段为 A002000000 和 A001000000
zcfz_selected_columns = ['Stkcd', 'Accper', 'Typrep', 'A002000000', 'A001000000']
zcfz_filtered_data = zcfz_data[zcfz_selected_columns]

# 合并到 LR_1.csv 中
merged_data = pd.merge(lr_data, zcfz_filtered_data, on=['Stkcd', 'Accper', 'Typrep'], how='left')
```

图 1-3

（2）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），如图 1-4 所示。

```
# 读取 Stk_ind.csv
stk_ind_data = pd.read_csv('Stk_ind.csv')

# 提取 Stk_ind.csv 中的字段为 Indnme 和 Nindnme
stk_ind_selected_columns = ['Stkcd', 'Indnme', 'Nindnme']
stk_ind_filtered_data = stk_ind_data[stk_ind_selected_columns]

# 合并到 LR_1.csv 中
merged_data = pd.merge(merged_data, stk_ind_filtered_data, on='Stkcd', how='left')

# 将合并后的数据另存为 LR_2.csv（编码为UTF-8）
merged_data.to_csv('LR_2.csv', index=False, encoding='utf-8')
```

图 1-4

（3）将合并后数据的行数、列数呈现出，如图 1-5 所示。

```
merged_data.shape
```

```
(33414, 19)
```

图 1-5 合并后数据的行数、列数

由上述处理与分析得出，合并后数据为 33414 行，19 列，相比于任务 1.1 列数增加。

（4）根据上述字段“Stkcd”，提取“Stk_ind.csv”中字段为“Indnme”和“Nindnme”的相应数据后，打开上述文件夹出现乱码情况，因原本为 GB312 的编码格式，需要在记事本中将编码格式转换为 UTF-8 编码，如图 1-6 所示。

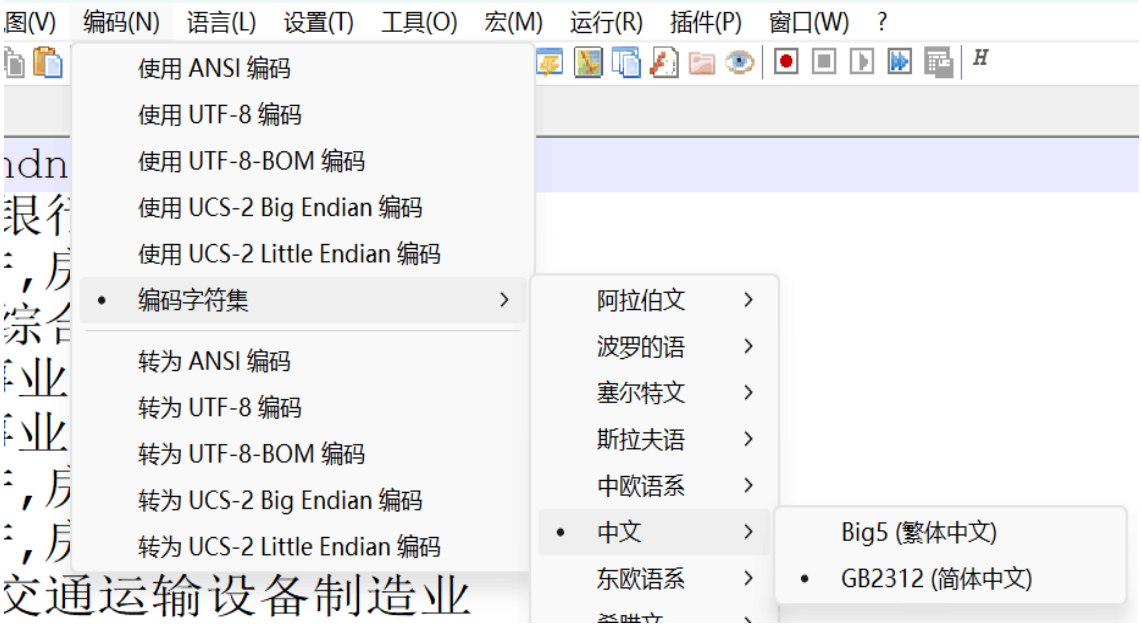


图 1-6 转换编码格式

任务 1.3 删除数据列

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），如图 1-7 所示。并呈现处理后数据的列数，如图 1-8 所示。

```
import pandas as pd

# 读取 LR_2.csv
lr2_data = pd.read_csv('LR_2.csv')

# 计算每列的空值占比
na_percentages = lr2_data.isna().mean()

# 筛选空值占比低于 70% 的列
filtered_columns = na_percentages[na_percentages < 0.7].index

# 保留筛选后的列数据
filtered_data = lr2_data[filtered_columns]

# 将处理后的数据另存为 LR_3.csv（编码为 UTF-8）
filtered_data.to_csv('LR_3.csv', index=False, encoding='utf-8')
```

图 1-7 删除空值占比达 70%及以上的数据列并另存为文件“LR_3.csv”

```
filtered_data.shape
```

```
(33414, 17)
```

图 1-8 处理后数据的列数

依据上述要求，对“LR_2.csv”空值占比达 70%及以上的数据列进行了删除处理，由分析得出空值占比达 70%及以上的数据列有两列，即处理后的数据为 17 列。

任务 1.4 删除包含空值的行

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）并呈现处理后数据的行数。如图 1-9 所示。

```
import pandas as pd

# 读取 LR_3.csv
lr3_data = pd.read_csv('LR_3.csv')

# 删除包含空值的行
filtered_data = lr3_data.dropna()

# 将处理后的数据另存为 LR_4.csv（编码为 UTF-8）
filtered_data.to_csv('LR_4.csv', index=False, encoding='utf-8')
```

图 1-9 删除包含空值的行并另存为文件“LR_4.csv”

```
filtered_data.shape
```

```
(30888, 17)
```

观察上述要求分析结果得知将删除“LR_3.csv”包含空值的行后，行数变为 30888 行。

任务 1.5 将字段“Accper”的日期数据转换格式

读取“LR_4.csv”，将字段“Accper”的日期数据转换为“YYYY-mm-dd”的格式，如图 1-8 所示。

```python
import pandas as pd

# 读取 LR_4.csv
lr4_data = pd.read_csv('LR_4.csv')

# 将字段 "Accper" 转换为日期时间对象
lr4_data['Accper'] = pd.to_datetime(lr4_data['Accper'])

# 将日期时间对象格式化为 "YYYY-mm-dd" 格式的字符串
lr4_data['Accper'] = lr4_data['Accper'].dt.strftime('%Y-%m-%d')

# 将处理后的数据另存为 LR_5.csv（编码为 UTF-8）
lr4_data.to_csv('LR_5.csv', index=False, encoding='utf-8')
```

```python
lr4_data.head()
```

	Stkcd	Accper	Typrep	B001000000	B001100000	B001101000	B001200000	B001201000	B001207000
0	600696	2018-03-31	A	8.669264e+06	18784589.8	18784589.8	1.068992e+07	7.810739e+06	204595.11
1	547	2018-03-31	A	9.675773e+07	517945246.5	517945246.5	4.266916e+08	2.729517e+08	789386.98
2	2772	2018-03-31	A	8.953362e+07	267161661.1	267161661.1	1.898233e+08	1.527985e+08	694050.24
3	818	2018-03-31	A	1.741463e+08	907539547.9	907539547.9	7.338691e+08	6.535983e+08	15336943.77
4	300568	2018-03-31	A	1.069983e+08	171703944.2	171703944.2	1.212891e+08	8.066557e+07	2874172.01

图 1-8 将字段“Accper”的日期数据转换为“YYYY-mm-dd”的格式

由上述将字段“Accper”的日期数据转换为“YYYY-mm-dd”的格式的处理操作，如将 Accper 中的“2018-1-31”转换为“2018-01-31”，其余数据同如上例所转换。

任务 1.6 利润率和资产负债率的计算

（1）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，如图 1-9 所示。

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）

```python
import pandas as pd

# 读取 LR_5.csv
lr5_data = pd.read_csv('LR_5.csv')

# 计算利润率
lr5_data['利润率'] = lr5_data['B001000000'] / lr5_data['B001100000']

# 计算资产负债率
lr5_data['资产负债率'] = lr5_data['A002000000'] / lr5_data['A001000000']
```

```python
lr5_data.head()
```

B001200000	B001201000	B001207000	B001209000	B001210000	B001211000	B001212000	A002000000	A001000000	Indnme	Nindnme	利润率	资产负债率
068992e+07	7.810739e+06	204595.11	281510.68	2.260219e+06	115431.05	17427.60	4.437834e+08	7.455484e+08	房地产	房地产业	0.461509	0.595244
266916e+08	2.729517e+08	789386.98	12657162.87	1.344988e+08	2518200.78	3276348.85	1.681114e+09	7.442314e+09	工业	信息技术业	0.186811	0.225886
898233e+08	1.527985e+08	694050.24	30133399.88	9.984417e+06	-3692930.13	-94158.98	1.422487e+09	4.176132e+09	综合	农业	0.335129	0.340623
338691e+08	6.535983e+08	15336943.77	24841913.58	3.475893e+07	3701613.61	1631376.36	1.322972e+09	4.061964e+09	工业	化学原料及化学制品制造业	0.191888	0.325698
212891e+08	8.066557e+07	2874172.01	3784963.29	2.709902e+07	6086240.16	779161.74	1.485785e+09	2.852590e+09	工业	化学原料及化学制品制造业	0.623156	0.520855

图 1-9 计算对应的利润率和资产负债率并追加对应字段

（2）分别删除表中利润率、资产负债率不在[-300%，300%]范围内的行，将处理后数据另存为文件“LR_new.csv”（文件编码设置为 UTF-8），并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，如图 1-10。

```python
# 删除利润率不在 [-300%, 300%] 范围内的行
lr5_data = lr5_data[(lr5_data['利润率'] >= -3) & (lr5_data['利润率'] <= 3)]

# 删除资产负债率不在 [-300%, 300%] 范围内的行
lr5_data = lr5_data[(lr5_data['资产负债率'] >= -3) & (lr5_data['资产负债率'] <= 3)]

# 将处理后的数据另存为 LR_new.csv（编码为 UTF-8）
lr5_data.to_csv('LR_new.csv', index=False, encoding='utf-8')

# 获取处理后的数据行数和列数
num_rows, num_columns = lr5_data.shape

# 输出处理后的数据行数和列数
print(f"处理后的数据行数：{num_rows}")
print(f"处理后的数据列数：{num_columns}")

# 输出前 5 个企业的利润率和资产负债率
print("前 5 个企业的利润率和资产负债率：")
print(lr5_data[['利润率', '资产负债率']].head(5))
```

```
处理后的数据行数：30690
处理后的数据列数：19
前 5 个企业的利润率和资产负债率：
      利润率      资产负债率
0  0.461509  0.595244
1  0.186811  0.225886
2  0.335129  0.340623
3  0.191888  0.325698
4  0.623156  0.520855
```

图 1-10

利润率是指公司的利润与其销售额之间的比率，而资产负债率是指公司的总负债与总资产之间的比率。由上述所出的前五个利润率与资产负债率观察到公司的盈利能力越强，所承担得负债率也高，从而得出企业得盈利能力越强，但同时也将会承担了更多的债务。

2 财务数据指标分析及可视化

使用可视化工具将计算得到的财务指标进行可视化展示，分析行业营业利润以及行业企业营收等关系，进而分析企业的财务状况与经营成果。

任务 2.1 绘制相关的“行业营业利润对比分析”图

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，并进行分析。

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

2.1.1 绘制“各行业大类的利润对比”柱状图

绘制“2019 年 9 月各行业大类的利润对比”的柱状图，如图 2-1 所示。

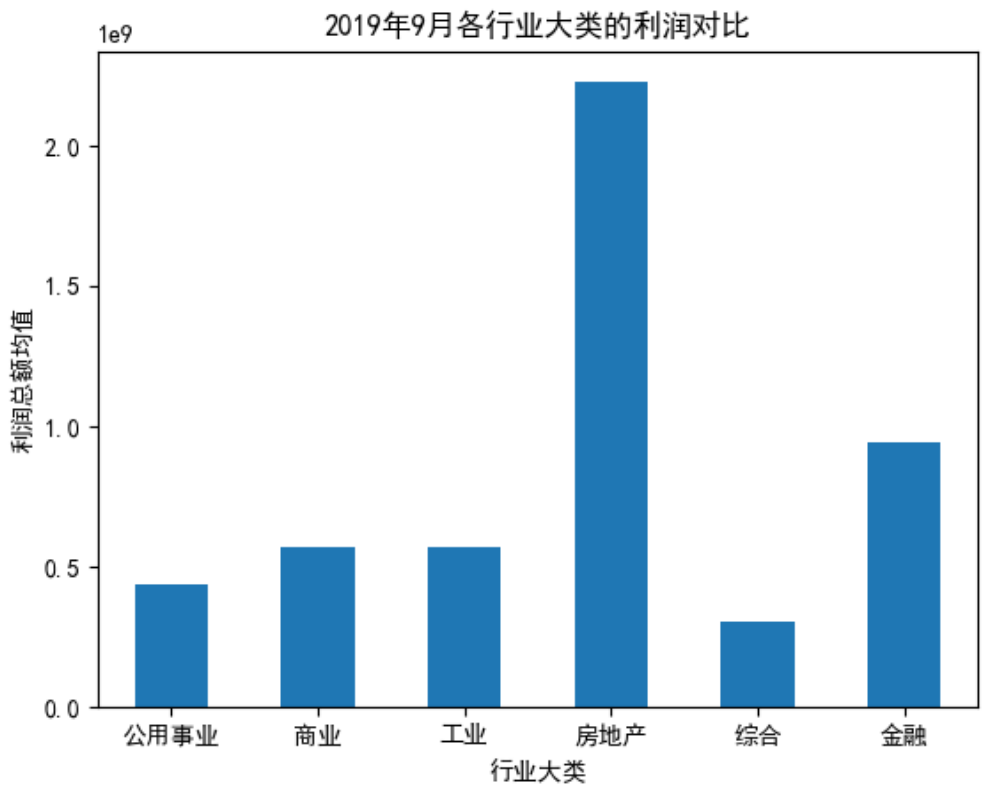


图 2-1“2019 年 9 月各业大类的利润对比”柱状图

该柱状图绘制的分别统计不同行业大类 2019 年 9 月利润总额的均值，根据图像所展示的，可以得出以下结论：

1.该图中展示了六个行业大类，分别是公用事业、商业、工业和、房地产、综合、金融。

2.在这六个行业大类中，利润总额均值最高的是房地产，高达 2.0 亿元；其次是金融，为 1 亿元；再次是工业以及商业较高于 0.5 亿元；公用事业与综合较低低于 0.5 亿元；房地产和金融两个行业的利润总额均值较高，说明这两个行业的盈利能力较强。公用事业与综合两个行业的利润总额均值较低，说明这两个行业的盈利能力较弱。

依据图像以及实际情况得出，因在 2019 年一些房企在行业调整中抓住了城镇化及城市发展的结构性等机会，使得总资产保持高增速，因此房地产行业的盈利能力最高，而公用事业与综合两个行业的盈利能力较差，还待提高。特别是石油和天然气开采业等都出现了利润下降的情况。这主要受到国际油价波动、国内经济增速放缓以及环保政策的影响。

2.1.2 绘制“各行业大类利润率变化”折线图

绘制“2018 年 1 月至 2019 年 9 月各行业大类利润率变化”柱状图，如图 2-2 所示。

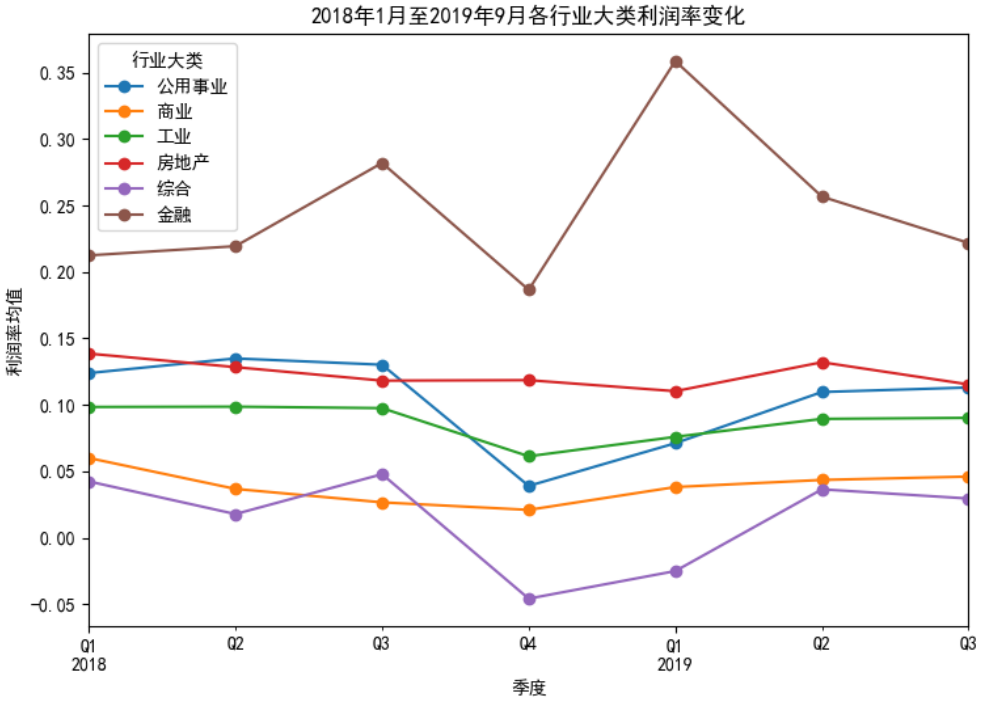


图 2-2“各行业大类利润率变化”柱状图

根据图中显示的 2018 年 1 月至 2019 年 9 月各行业大类利润率变化情况可得：

1.金融的利润率均值最高，综合的利润率均值最低。

2.在六个行业中，只有公用事业行业的利润率均值有所增长，金融行业利润率 2019 后一直处于下降趋势。

3.从整体上看，除开金融行业波动较大，其余五个行业在 2018 年 1 月至 2019 年 9 月期间的利润率均值变化都比较平稳，没有出现明显的波动。

根据以上数据，可以看出各个行业之间的竞争关系较为激烈，而且随着时间的推移，各个行业的利润率都呈现出下降趋势或是缓慢增长趋势。

任务 2.2 绘制相关的“行业企业营收分析”图

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，并进行分析。

表 4

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）

由任务 2.1 分析得出 2019 年 9 月营业利润率均值排名第 1 的行业大类为金融行业，因此需绘制金融行业相关的“行业企业营收分析”图。

2.2.1 绘制“该行业各细类利润率”对比柱状图

绘制“2019 年该行业各细类利润率对比”柱状图，如图 2-3 所示。

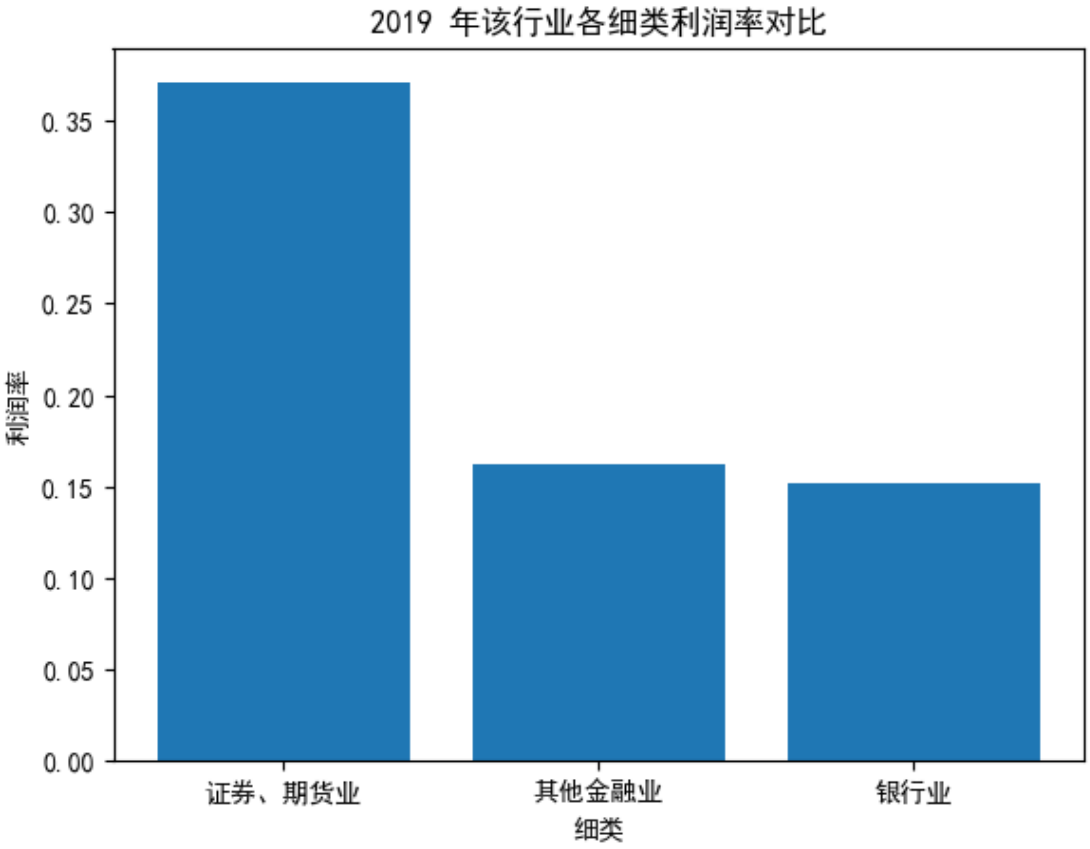


图 2-3“2019 年该行业各细类利润率对比”柱状图

该柱状图显示的是利润率排名前 3 细类的利润率，前三名为证券、期货业、其他金融业以及银行业。利润率最高的细类为证券、期货业，其中排名第三的银行业，该细类与证券、期货业的利润率有明显差距。其他金融业与银行业的利润率差异较小，初步判定两者盈利能力大相径庭。在 2019 年里证券、期货业的盈利遥遥领先。

2.2.2 绘制“排名第 1 细类的企业利润率对比”柱状图

绘制“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，如图 2-4 所示。

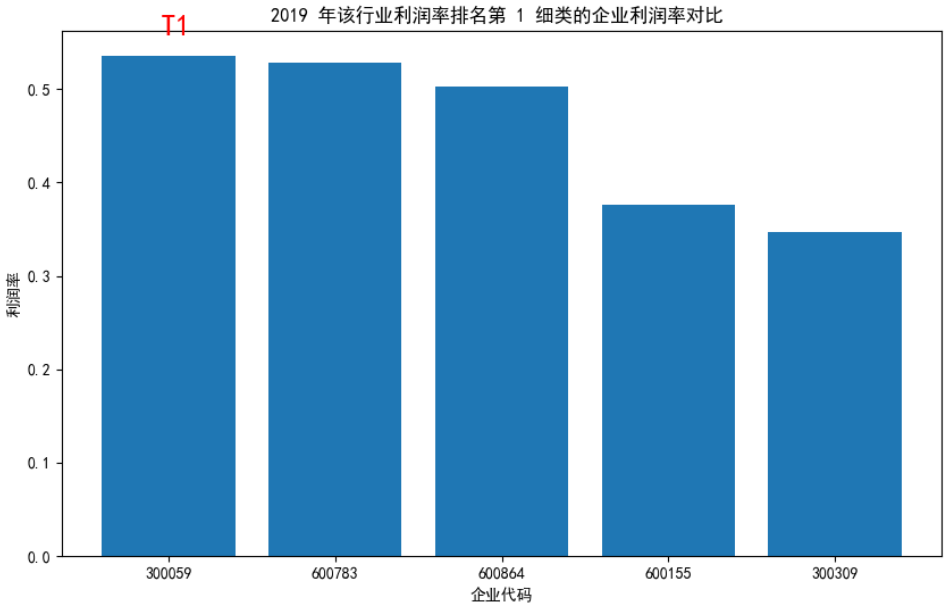


图 2-4“2019 年该行业利润率排名第 1 细类的企业利润率对比”柱状图

由任务 2.2.1 得知，排名第一的细类为证券、期货业，因此该柱状图显示的是 2019 年该行业利润率证券、期货业排名前 5 的企业利润率对比。

1.前五企业的企业代码分别是“300059“，“600783”，“600864”，“600155”，“300309”。排名前 2 的两个企业的利润率差异较小，说明企业代码为“30059”与“600783”的企业竞争较为激烈。

2.企业代码为“600864”的企业与排名前 2 的企业差距相比于企业代码为 600155“与”300309“的企业差距小一些，极有可能追赶上排名前 2 的企业。

3.企业代码为 600155“与”300309“的企业利润率差异较小，互相竞争较为激烈，但是较难追赶上前 3 个企业的利润率。

2.2.3 绘制“企业“T1”营业总成本分析”饼图

绘制“2019 年企业“T1”营业总成本分析“的饼图，如图 2-5 所示。

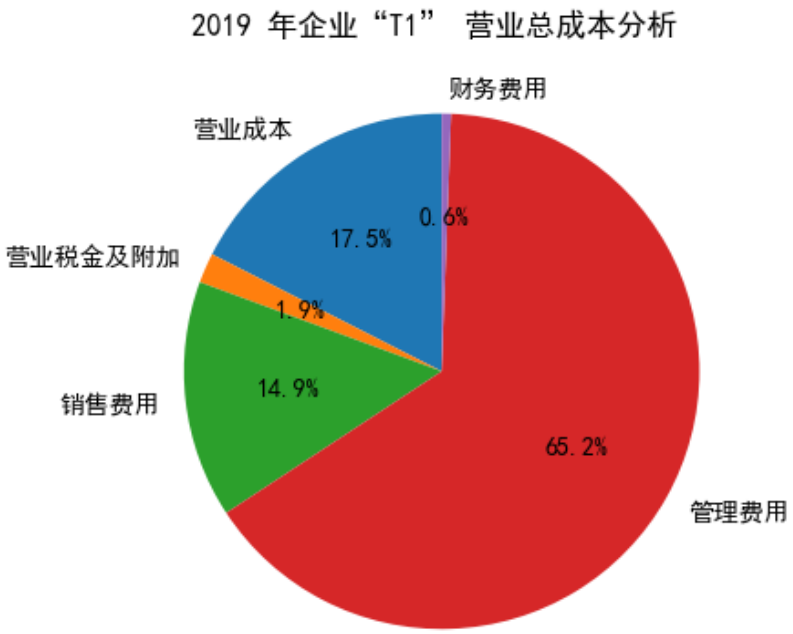


图 2-5“2019 年企业“T1”营业总成本分析“的饼图

根据绘制饼图显示，2019 年企业"T1"营业总成本分析中，营业成本占比为 17.5%,营业税金及附加占比为 14.9%,销售费用占比为 65.2%,管理费用占比为 1.9%。财务费用占比为 0.6%。明显得出管理费用占比最高，财务费用最低。

企业规模较大会导致管理费用增加，随着企业规模的扩大，管理层的人数也会相应增加，同时需要更大的办公场所和更多的办公设备，这些都会增加企业的管理费用。此外，大型企业还需要进行更多的培训和人力资源管理，这也会增加管理费用。因此可以得出企业“T1”的规模是相比较大的；企业财务费用最低可能原因是当公司的流动资金充裕，银行存款金额大大高于公司贷款融资额时，公司可能会把闲置的银行存款用于购买理财产品或者改存协议存款，尽管收益率远小于公司融资利率，但由于利息收入大于利息支出，这可能会导致公司的财务费用降低。

从整体观察，各类费用均在正常范围内，并没有造假等行为并且在行业内规模较好。

2.2.4 绘制“企业“T1”经营情况分析”柱状折线组合图

绘制“企业“T1”经营情况分析”柱状折线组合图，如图 2-6 所示

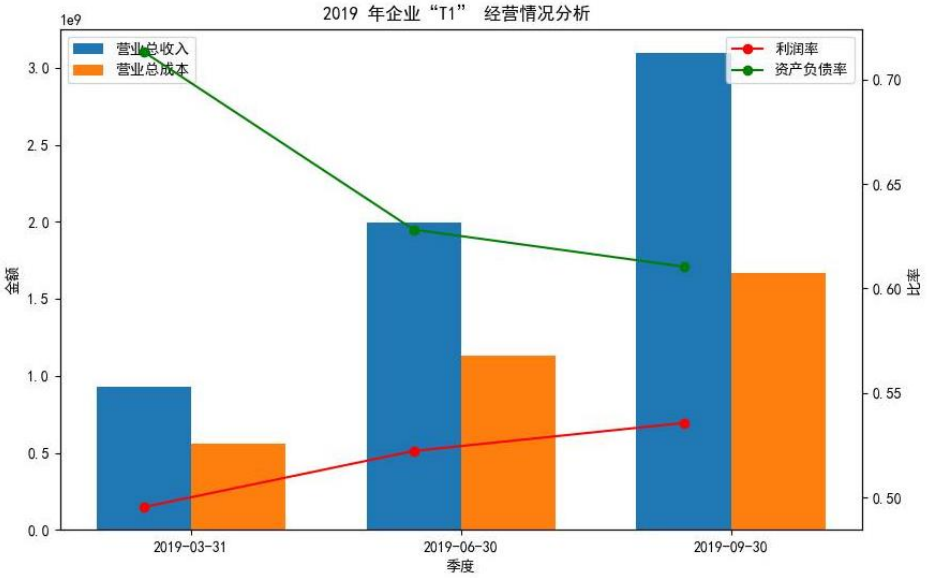


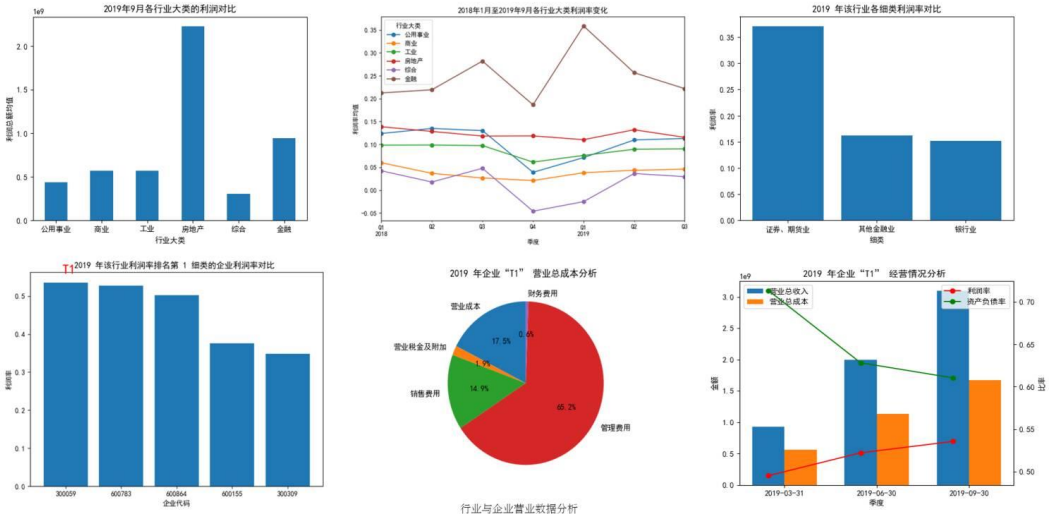
图 2-6 绘制“企业“T1”经营情况分析”柱状折线组合图

根据柱状图可以明显得到三个季度营业总收入大于营业总成本，说明这三个季度该企业是盈利的；三个季度的总收入逐季度在增加，并随着总收入的增加总成本也在增加，呈现相同的趋势；

根据利润率和资产负债率的变化折线图可以发现，企业“T1”的利润率与资产负债率呈现相反趋势，利润率在随着季度逐渐上升，资产负债率随着季度逐渐减少。总体可看出，“企业“T1”在该三个季度下的盈利可观。

任务 2.3 制作“行业与企业营业数据分析”大屏

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”，如图 2-7 所示。



3 企业利润预测及财务造假识别

任务 3.1 计算各个指标与利润总额的相关性

3.3.1 计算各个指标与利润总额的相关性

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，表 5 为最高的 5 个指标相关度汇总。

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# 读取CSV文件
df = pd.read_csv('financial_data.csv')

# 计算各个指标与利润总额的相关性
correlations = df.corr()['LRZE'].sort_values(ascending=False)
```

```
correlations.head(6)
```

```
LRZE        1.000000
YYSR        0.782726
YWFY        0.772832
YYCB        0.737736
YYSJJFJ     0.565440
ZCJZSS      0.238524
Name: LRZE, dtype: float64
```

图 3-1 相关度最高的 5 个指标

表 5

指标名称	相关度
YYSR(营业收入)	0.782726
YWFY(业务费用)	0.772832
YYCB(营业成本)	0.737736
YYSJJFJ(营业税金及附加)	0.5654040
ZCJZSS(资产减值损失)	0.238524

3.3.2 绘制热力图显示相关性

由于热力图可以更好的观察各个指标与利润总额的相关性，如图 3-2 所示。

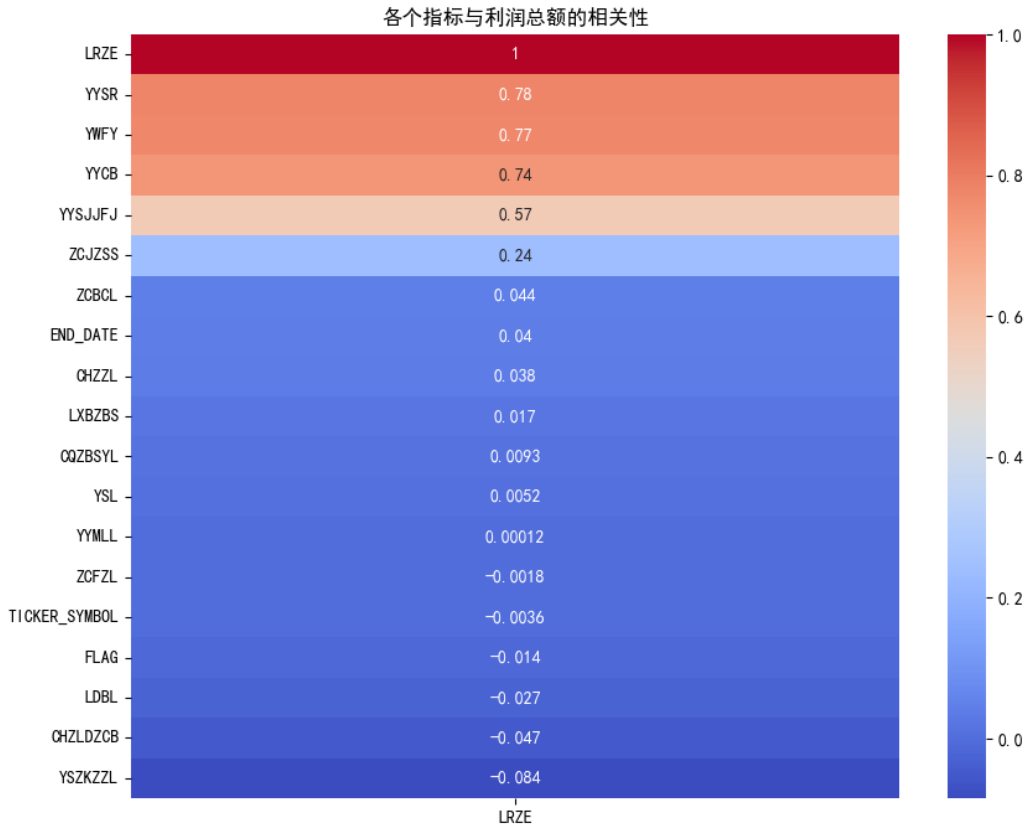


图 5-2 各个指标与利润总额的相关性

热力图的红色区域表示相关性较高，蓝色区域表示相关性较低，该热力图中指标从上往下相关性依次递减，因此可以更清楚的得出 5 个最高指标为 YYSR、YWFY、YYCB、YYSJJFJ、ZCJZSS。

任务 3.2 预测给定企业的利润总额

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呈现，如表6所示。

表 6 预测结果

TICKER_SYMBOL	LRZE
4953174	105929295.025
4961537	156625596.322

4962538	172955564.8
4968740	99710928.015
4973917	79885625.4267
4978589	605882605.95
4978721	96785058.64
4986535	146360630.15
4990739	217382535.475
4990942	-40173019.695

任务 3.3 识别涉嫌财务造假企业

“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）。请利用表中所列关键因子，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。

表7　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

识别涉嫌财务造假企业的分析方法如下：

1. 首先进行数据分析，区别出“FLAG”字段中造假与不造假所在行的数据，如图 5-3 所示。

	A	B	C	D	E	F
1	LDBL	ZCFZL	CHZZL	ZCBCL	YSZKZZL	FLAG
175	1.15064	0.36486	0.12133	-0.36636	1.43432	1
527	1.61333	0.37156	0.23607	0.07823	1.66536	1
703	1.55687	0.41672	0.39161	0.08502	0.93503	1
1388	1.02777	0.51225	0.13344	-0.21801	1.55261	1
1403	1.09905	0.9664	0.28064	0.0083	0.88926	1
1610	2.49918	0.44147	0.37916	0.09431	1.24374	1
1874	14.6546	0.06356	3.63049	0.05551	1.01201	1

图 5-3 区别出“FLAG”字段

2.对表 7 中的关键因子所对应的数据进行分析，对比造假与不造假得出的数据绘制出相应的箱线图。由箱线图可看出造假中的异常值与不造假的正常值进而识别出涉嫌财务造假的企业。

4 总结

在企业财务数据分析中主要根据分析行业营业利润以及行业企业收营状况去评判某时间段内盈利以及去评判靠前的行业或企业。财务造假的动机通常是为了做大利润和规模用于上市、发行债券等，或者是为了减少利润进行逃税，根据需要查看公司的财务报表，包括资产负债表、利润表等，以及根据利润总额与各项指标的相关性等综合去评判出是否财务造假。