

1. 数据读取

```
In [1]: # -*- coding: utf-8 -*-
        """
        @author: EasternDay
        """

        import pandas as pd
        # 导入扩展库
        import re # 正则表达式库
        import collections # 词频统计库
        import numpy as np # numpy数据处理库
        import jieba # 结巴分词
        import wordcloud # 词云展示库
        from PIL import Image # 图像处理库
        import matplotlib.pyplot as plt # 图像展示库


        # 读取csv文件
        Company_info = pd.read_csv(
            "../数据/附件二_企业信息.csv", usecols=['企业代号', '企业名称'])
        Company_in = pd.read_csv(
            "../数据/附件二_进项发票信息.csv", usecols=['企业代号', '发票号码', '开票日期', '销方单位代号', '金额', '税额', '价税合计', '发票状态'])
        Company_out = pd.read_csv(
            "../数据/附件二_销项发票信息.csv", usecols=['企业代号', '发票号码', '开票日期', '购方单位代号', '金额', '税额', '价税合计', '发票状态'])
```

```
In [2]: Company_info.head()
```

Out[2]:

	企业代号	企业名称
0	E124	个体经营E124
1	E125	个体经营E125
2	E126	个体经营E126
3	E127	个体经营E127
4	E128	个体经营E128

```
In [3]: def getIn(id):
        data_in = Company_in[(Company_in["企业代号"] == id) &
                               (Company_in["发票状态"] == "有效发票")]
        return data_in["金额"].sum()


        def getInVar(id):
            data_in = Company_in[(Company_in["企业代号"] == id) &
                                   (Company_in["发票状态"] == "有效发票")]
            return data_in["金额"].var()


        def getOut(id):
            data_out = Company_out[(Company_out["企业代号"] == id)
                                    & (Company_out["发票状态"] == "有效发票")]
            return data_out["金额"].sum()


        def getOutVar(id):
            data_out = Company_out[(Company_out["企业代号"] == id)
                                    & (Company_out["发票状态"] == "有效发票")]
            return data_out["金额"].var()


        def getProfit(id):
            data_in = Company_in[(Company_in["企业代号"] == id) &
                                   (Company_in["发票状态"] == "有效发票")]
            data_out = Company_out[(Company_out["企业代号"] == id)
                                    & (Company_out["发票状态"] == "有效发票")]
            return data_out["税额"].sum() - data_in["税额"].sum()


        def getInNumber(id):
            data_in = Company_in[(Company_in["企业代号"] == id) &
                                   (Company_in["发票状态"] == "有效发票")]
            return data_in.shape[0]


        def getOutNumber(id):
            data_out = Company_out[(Company_out["企业代号"] == id)
                                    & (Company_out["发票状态"] == "有效发票")]
            return data_out.shape[0]


        def getUselessNumber(id):
            data_in = Company_in[(Company_in["企业代号"] == id) &
                                   (Company_in["发票状态"] == "作废发票")]
            data_out = Company_out[(Company_out["企业代号"] == id)
```

```

        & (Company_out["发票状态"] == "作废发票"])
    return data_out, shape[0]+data_in, shape[0]

# 使用apply函数, 如果city字段包含'ing'关键词, 则'判断'这一列赋值为1, 否则为0
Company_info['进项总金额'] = Company_info["企业代号"].apply(lambda x: getIn(x))
#Company_info['进项金额方差'] = Company_info["企业代号"].apply(lambda x: getInVar(x))
Company_info['销项总金额'] = Company_info["企业代号"].apply(lambda x: getOut(x))
#Company_info['销项金额方差'] = Company_info["企业代号"].apply(lambda x: getOutVar(x))
Company_info['总税额'] = Company_info["企业代号"].apply(lambda x: getProfit(x))
Company_info['进项数'] = Company_info["企业代号"].apply(lambda x: getInNumber(x))
Company_info['销项数'] = Company_info["企业代号"].apply(lambda x: getOutNumber(x))
Company_info['作废数'] = Company_info["企业代号"].apply(lambda x: getUselessNumber(x))
Company_info['进项数'].astype("float")
Company_info['销项数'].astype("float")
Company_info['作废数'].astype("float")
Company_info['作废比例'] = Company_info.apply(
    lambda x: x['作废数']/(x['进项数']+x['销项数']), axis=1)

def getMonthIn(id):
    data_in = Company_in[(Company_in["企业代号"] == id) &
                          (Company_in["发票状态"] == "作废发票")]
    data_in["开票日期"] = pd.to_datetime(data_in["开票日期"])
    return len(np.unique(data_in["开票日期"].dt.year.values))
def getMonthOut(id):
    data_out = Company_out[(Company_out["企业代号"] == id) &
                           (Company_out["发票状态"] == "作废发票")]
    data_out["开票日期"] = pd.to_datetime(data_out["开票日期"])
    return len(np.unique(data_out["开票日期"].dt.year.values))
Company_info['进项年数'] = Company_info["企业代号"].apply(lambda x: getMonthIn(x))
Company_info['销项年数'] = Company_info["企业代号"].apply(lambda x: getMonthOut(x))

def getGoodMonthIn(id):
    data_in = Company_in[(Company_in["企业代号"] == id) &
                          (Company_in["发票状态"] == "作废发票")]
    data_in["开票日期"] = pd.to_datetime(data_in["开票日期"])
    data_in["年"] = data_in["开票日期"].dt.year.values
    data_in["月"] = data_in["开票日期"].dt.month.values
    return len(np.unique(data_in["月"].values))
def getGoodMonthOut(id):
    data_out = Company_out[(Company_out["企业代号"] == id) &
                           (Company_out["发票状态"] == "作废发票")]
    data_out["开票日期"] = pd.to_datetime(data_out["开票日期"])
    data_out["年"] = data_out["开票日期"].dt.year.values
    data_out["月"] = data_out["开票日期"].dt.month.values
    return len(np.unique(data_out["月"].values))
Company_info['进项月数'] = Company_info["企业代号"].apply(lambda x: getGoodMonthIn(x))
Company_info['销项月数'] = Company_info["企业代号"].apply(lambda x: getGoodMonthOut(x))

def getCompanyType(id):
    a = re.findall(r'商贸|劳务|机械设备|贸易|管理|咨询|药|广告', id)
    b = re.findall(r'装饰|工程|建筑|建设|建筑|电子|电器|控制|机械|机电', id)
    c = re.findall(r'科技|信息|电脑|通讯', id)
    if(len(a) > 0):
        return 0
    if(len(b) > 0):
        return 1
    if(len(c) > 0):
        return 2
    return 3

"""
def getCompanyScale(x):
    type = x["公司类别"]
    money = max([x['进项总金额'], x['销项总金额']])
    if(type == 0):
        if(money >8e7):
            return 3
        elif(money>1e6):
            return 2
        else:
            return 1
    elif(type == 1):
        if(money >2e7):
            return 3
        elif(money>3e6):
            return 2
        else:
            return 1
    elif(type == 2):
        if(money >1e7):
            return 3
        elif(money>5e5):
            return 2
        else:
            return 1

```

```

else:
    if(money >2e7):
        return 3
    elif(money>1e6):
        return 2
    else:
        return 1
    """

def getCompanyScale(x):
    type = x["公司类别"]
    money = max([x['进项总金额'], x['销项总金额']])
    if(money >1e7):
        return 3
    elif(money>5e5):
        return 2
    else:
        return 1

Company_info["公司类别"] = Company_info["企业名称"].apply(lambda x: getCompanyType(x))
Company_info["公司规模"] = Company_info.apply(lambda x: getCompanyScale(x),axis=1)

def getCompanyIn(id):
    data_in = Company_in[(Company_in["企业代号"] == id) &
                          (Company_in["发票状态"] == "有效发票")]
    return list(set(data_in["销方单位代号"].tolist()))

def getCompanyOut(id):
    data_out = Company_out[(Company_out["企业代号"] == id) &
                           (Company_out["发票状态"] == "有效发票")]
    return list(set(data_out["购方单位代号"].tolist()))

Company_info['供应公司比重'] = Company_info.apply(
    lambda x: len(getCompanyIn(x["企业代号"]))/x["进项数"], axis=1)
Company_info['销售公司比重'] = Company_info.apply(
    lambda x: len(getCompanyOut(x["企业代号"]))/x["销项数"], axis=1)

Company_info

```

```

<ipython-input-3-d7a32b0e11db>:73: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```

data_in["开票日期"] = pd.to_datetime(data_in["开票日期"])

```

```

<ipython-input-3-d7a32b0e11db>:78: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```

data_out["开票日期"] = pd.to_datetime(data_out["开票日期"])

```

```

<ipython-input-3-d7a32b0e11db>:88: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```

data_in["开票日期"] = pd.to_datetime(data_in["开票日期"])

```

```

<ipython-input-3-d7a32b0e11db>:89: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```

data_in["年"] = data_in["开票日期"].dt.year.values

```

```

<ipython-input-3-d7a32b0e11db>:90: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```

data_in["月"] = data_in["开票日期"].dt.month.values

```

```

<ipython-input-3-d7a32b0e11db>:95: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```

data_out["开票日期"] = pd.to_datetime(data_out["开票日期"])

```

```

<ipython-input-3-d7a32b0e11db>:96: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```

data_out["年"] = data_out["开票日期"].dt.year.values

```

```

<ipython-input-3-d7a32b0e11db>:97: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
opy
data_out["月"] = data_out["开票日期"].dt.month.values
```

Out[3]:

	企业 代号	企业名称	进项总金额	销项总金额	总税额	进项 数	销项 数	作废 数	作废比例	进 项 年 数	销 项 年 数	进 项 月 数	销 项 月 数	公司 类别	公司 规模	供应公司 比重	销售公司 比重
0	E124	个体经营E124	7.855255e+08	7.417804e+08	7980118.45	15264	1099	2341	0.143067	3	3	12	12	3	3	0.091916	0.047316
1	E125	个体经营E125	9.323832e+08	9.412114e+08	14811182.88	17776	1380	2727	0.142357	3	3	12	12	3	3	0.100641	0.045652
2	E126	个体经营E126	1.127735e+08	5.206911e+08	1271020.12	515	1264	270	0.151771	3	3	8	12	3	3	0.139806	0.113133
3	E127	个体经营E127	1.646702e+06	6.519380e+08	19451364.52	1424	3904	148	0.027778	3	3	8	12	3	3	0.208567	0.006404
4	E128	个体经营E128	8.928033e+06	2.423902e+08	7866599.98	3134	1161	182	0.042375	3	3	12	10	3	3	0.169751	0.006029
...
297	E421	***保温材料有限公司	3.039381e+04	2.457008e+05	5594.88	19	28	1	0.021277	0	1	0	1	3	1	0.263158	0.178571
298	E422	***童装店	8.893900e+02	2.701831e+04	767.69	3	27	3	0.100000	0	2	0	3	3	1	0.666667	0.740741
299	E423	***通风设备有限公司	6.104614e+04	7.278640e+04	-5821.96	21	6	1	0.037037	0	1	0	1	3	1	0.380952	0.500000
300	E424	***贸易有限公司	2.543299e+05	7.689642e+04	-15362.61	25	37	6	0.096774	0	2	0	4	0	1	0.520000	0.567568
301	E425	***商贸有限公司	5.719625e+04	8.512621e+04	-2331.36	116	10	10	0.079365	1	2	2	2	0	1	0.232759	0.200000

302 rows × 17 columns

In [4]:

```
Company_info.to_excel("题目二_信贷数据分析.xlsx") #不保存行索引
Company_info
```

Out[4]:

	企业 代号	企业名称	进项总金额	销项总金额	总税额	进项 数	销项 数	作废 数	作废比例	进 项 年 数	销 项 年 数	进 项 月 数	销 项 月 数	公司 类别	公司 规模	供应公司 比重	销售公司 比重
0	E124	个体经营E124	7.855255e+08	7.417804e+08	7980118.45	15264	1099	2341	0.143067	3	3	12	12	3	3	0.091916	0.047316
1	E125	个体经营E125	9.323832e+08	9.412114e+08	14811182.88	17776	1380	2727	0.142357	3	3	12	12	3	3	0.100641	0.045652
2	E126	个体经营E126	1.127735e+08	5.206911e+08	1271020.12	515	1264	270	0.151771	3	3	8	12	3	3	0.139806	0.113133
3	E127	个体经营E127	1.646702e+06	6.519380e+08	19451364.52	1424	3904	148	0.027778	3	3	8	12	3	3	0.208567	0.006404
4	E128	个体经营E128	8.928033e+06	2.423902e+08	7866599.98	3134	1161	182	0.042375	3	3	12	10	3	3	0.169751	0.006029
...
297	E421	***保温材料有限公司	3.039381e+04	2.457008e+05	5594.88	19	28	1	0.021277	0	1	0	1	3	1	0.263158	0.178571
298	E422	***童装店	8.893900e+02	2.701831e+04	767.69	3	27	3	0.100000	0	2	0	3	3	1	0.666667	0.740741
299	E423	***通风设备有限公司	6.104614e+04	7.278640e+04	-5821.96	21	6	1	0.037037	0	1	0	1	3	1	0.380952	0.500000
300	E424	***贸易有限公司	2.543299e+05	7.689642e+04	-15362.61	25	37	6	0.096774	0	2	0	4	0	1	0.520000	0.567568
301	E425	***商贸有限公司	5.719625e+04	8.512621e+04	-2331.36	116	10	10	0.079365	1	2	2	2	0	1	0.232759	0.200000

302 rows × 17 columns

In [5]:

```
import joblib
# 下载本地模型
WY = joblib.load(filename="违约分析.model")
# 下面将数据集分为训练集和测试集
x = Company_info.iloc[:, 2:].values
Company_info['信贷风险'] = WY.predict(x)
Company_info
```

Out[5]:

	企业 代号	企业名称	进项总金额	销项总金额	总税额	进项 数	销项 数	作废 数	作废比例	进 项 年 数	销 项 年 数	进 项 月 数	销 项 月 数	公司 类别	公司 规模	供应公司 比重	销售公司 比重	信 贷 风 险
0	E124	个体经营E124	7.855255e+08	7.417804e+08	7980118.45	15264	1099	2341	0.143067	3	3	12	12	3	3	0.091916	0.047316	0
1	E125	个体经营E125	9.323832e+08	9.412114e+08	14811182.88	17776	1380	2727	0.142357	3	3	12	12	3	3	0.100641	0.045652	0
2	E126	个体经营E126	1.127735e+08	5.206911e+08	1271020.12	515	1264	270	0.151771	3	3	8	12	3	3	0.139806	0.113133	0
3	E127	个体经营E127	1.646702e+06	6.519380e+08	19451364.52	1424	3904	148	0.027778	3	3	8	12	3	3	0.208567	0.006404	0
4	E128	个体经营E128	8.928033e+06	2.423902e+08	7866599.98	3134	1161	182	0.042375	3	3	12	10	3	3	0.169751	0.006029	0

	企业 代号	企业名称	进项总金额	销项总金额	总税额	进项 数	销项 数	作废 数	作废比例	进 项 年 数	销 项 年 数	进 项 月 数	销 项 月 数	公 司 类 别	公 司 规 模	供应公司 比重	销售公司 比重	信 贷 风 险
...
297	E421	***保温材料 有限公司	3.039381e+04	2.457008e+05	5594.88	19	28	1	0.021277	0	1	0	1	3	1	0.263158	0.178571	1
298	E422	***童装店	8.893900e+02	2.701831e+04	767.69	3	27	3	0.100000	0	2	0	3	3	1	0.666667	0.740741	1
299	E423	***通风设备 有限公司	6.104614e+04	7.278640e+04	-5821.96	21	6	1	0.037037	0	1	0	1	3	1	0.380952	0.500000	1
300	E424	***贸易有限 公司	2.543299e+05	7.689642e+04	-15362.61	25	37	6	0.096774	0	2	0	4	0	1	0.520000	0.567568	1
301	E425	***商贸有限 公司	5.719625e+04	8.512621e+04	-2331.36	116	10	10	0.079365	1	2	2	2	0	1	0.232759	0.200000	1

302 rows × 18 columns

In [6]:

```
import joblib
# 下载本地模型
WY = joblib.load(filename="信誉分析.model")
# 下面将数据集分为训练集和测试集
x = Company_info.iloc[:, 2:-1].values
Company_info['信誉评级'] = WY.predict(x)
Company_info
```

Out[6]:

	企业 代号	企业名称	进项总金额	销项总金额	总税额	进项 数	销项 数	作废 数	作废比例	进 项 年 数	销 项 年 数	进 项 月 数	销 项 月 数	公 司 类 别	公 司 规 模	供应公司 比重	销售公司 比重	信 贷 风 险	信 誉 评 级
0	E124	个体经营 E124	7.855255e+08	7.417804e+08	7980118.45	15264	1099	2341	0.143067	3	3	12	12	3	3	0.091916	0.047316	0	80
1	E125	个体经营 E125	9.323832e+08	9.412114e+08	14811182.88	17776	1380	2727	0.142357	3	3	12	12	3	3	0.100641	0.045652	0	80
2	E126	个体经营 E126	1.127735e+08	5.206911e+08	1271020.12	515	1264	270	0.151771	3	3	8	12	3	3	0.139806	0.113133	0	80
3	E127	个体经营 E127	1.646702e+06	6.519380e+08	19451364.52	1424	3904	148	0.027778	3	3	8	12	3	3	0.208567	0.006404	0	80
4	E128	个体经营 E128	8.928033e+06	2.423902e+08	7866599.98	3134	1161	182	0.042375	3	3	12	10	3	3	0.169751	0.006029	0	60
...
297	E421	***保温材 料有限公司	3.039381e+04	2.457008e+05	5594.88	19	28	1	0.021277	0	1	0	1	3	1	0.263158	0.178571	1	20
298	E422	***童装店	8.893900e+02	2.701831e+04	767.69	3	27	3	0.100000	0	2	0	3	3	1	0.666667	0.740741	1	20
299	E423	***通风设 备有限公司	6.104614e+04	7.278640e+04	-5821.96	21	6	1	0.037037	0	1	0	1	3	1	0.380952	0.500000	1	20
300	E424	***贸易有 限公司	2.543299e+05	7.689642e+04	-15362.61	25	37	6	0.096774	0	2	0	4	0	1	0.520000	0.567568	1	20
301	E425	***商贸有 限公司	5.719625e+04	8.512621e+04	-2331.36	116	10	10	0.079365	1	2	2	2	0	1	0.232759	0.200000	1	20

302 rows × 19 columns

In [7]:

```
Company_info.to_excel("题目二_信贷数据分析.xlsx") #不保存行索引
Company_info
```

Out[7]:

	企业 代号	企业名称	进项总金额	销项总金额	总税额	进项 数	销项 数	作废 数	作废比例	进 项 年 数	销 项 年 数	进 项 月 数	销 项 月 数	公 司 类 别	公 司 规 模	供应公司 比重	销售公司 比重	信 贷 风 险	信 誉 评 级
0	E124	个体经营 E124	7.855255e+08	7.417804e+08	7980118.45	15264	1099	2341	0.143067	3	3	12	12	3	3	0.091916	0.047316	0	80
1	E125	个体经营 E125	9.323832e+08	9.412114e+08	14811182.88	17776	1380	2727	0.142357	3	3	12	12	3	3	0.100641	0.045652	0	80
2	E126	个体经营 E126	1.127735e+08	5.206911e+08	1271020.12	515	1264	270	0.151771	3	3	8	12	3	3	0.139806	0.113133	0	80
3	E127	个体经营 E127	1.646702e+06	6.519380e+08	19451364.52	1424	3904	148	0.027778	3	3	8	12	3	3	0.208567	0.006404	0	80
4	E128	个体经营 E128	8.928033e+06	2.423902e+08	7866599.98	3134	1161	182	0.042375	3	3	12	10	3	3	0.169751	0.006029	0	60
...
297	E421	***保温材 料有限公司	3.039381e+04	2.457008e+05	5594.88	19	28	1	0.021277	0	1	0	1	3	1	0.263158	0.178571	1	20

	企业 代号	企业名称	进项总金额	销项总金额	总税额	进项 数	销项 数	作废 数	作废比例	进项 年数	销项 年数	进项 月数	销项 月数	公 司 类 别	公 司 规 模	供应公司 比重	销售公司 比重	信 贷 风 险	信 誉 评 级
298	E422	***童装店	8.893900e+02	2.701831e+04	767.69	3	27	3	0.100000	0	2	0	3	3	1	0.666667	0.740741	1	20
299	E423	***通风设备有限公司	6.104614e+04	7.278640e+04	-5821.96	21	6	1	0.037037	0	1	0	1	3	1	0.380952	0.500000	1	20
300	E424	***贸易有限公司	2.543299e+05	7.689642e+04	-15362.61	25	37	6	0.096774	0	2	0	4	0	1	0.520000	0.567568	1	20
301	E425	***商贸有限公司	5.719625e+04	8.512621e+04	-2331.36	116	10	10	0.079365	1	2	2	2	0	1	0.232759	0.200000	1	20

302 rows × 19 columns

2. 出结果

```
In [8]: Company_mess = pd.read_excel(
        ".\\题目二_信贷数据分析.xlsx", sheet_name="Sheet1", usecols=["企业代号", "企业名称", "信誉评级", "信贷风险", "总税额", "进项总金额", "销项总金额", "作废比例", "公司规模", "供应公司比重", "销售公司比重", "进项年数", "销项年数", "进项月数", "销项月数", "公司类别"]
        #Company_mess = Company_info[["企业名称", "信誉评级", "是否违约", "信贷风险", "总税额", "进项总金额", "销项总金额", "作废比例", "公司规模", "是否违约", "信誉评级", "是否违约", "信贷风险", "进项总金额", "销项总金额", "总税额", "进项数", "销项数", "作废数", "作废比例", "公司类别"]]
        Company_mess
```

Out[8]:

	企业代 号	企业名称	进项总金额	销项总金额	总税额	作废比例	进项 年数	销项 年数	进项 月数	销项 月数	公 司 规 模	供应公司 比重	销售公司 比重	信 贷 风 险	信 誉 评 级
0	E124	个体经营E124	7.855255e+08	7.417804e+08	7980118.45	0.143067	3	3	12	12	3	0.091916	0.047316	0	80
1	E125	个体经营E125	9.323832e+08	9.412114e+08	14811182.88	0.142357	3	3	12	12	3	0.100641	0.045652	0	80
2	E126	个体经营E126	1.127735e+08	5.206911e+08	1271020.12	0.151771	3	3	8	12	3	0.139806	0.113133	0	80
3	E127	个体经营E127	1.646702e+06	6.519380e+08	19451364.52	0.027778	3	3	8	12	3	0.208567	0.006404	0	80
4	E128	个体经营E128	8.928033e+06	2.423902e+08	7866599.98	0.042375	3	3	12	10	3	0.169751	0.006029	0	60
...
297	E421	***保温材料有限公司	3.039381e+04	2.457008e+05	5594.88	0.021277	0	1	0	1	1	0.263158	0.178571	1	20
298	E422	***童装店	8.893900e+02	2.701831e+04	767.69	0.100000	0	2	0	3	1	0.666667	0.740741	1	20
299	E423	***通风设备有限公司	6.104614e+04	7.278640e+04	-5821.96	0.037037	0	1	0	1	1	0.380952	0.500000	1	20
300	E424	***贸易有限公司	2.543299e+05	7.689642e+04	-15362.61	0.096774	0	2	0	4	1	0.520000	0.567568	1	20
301	E425	***商贸有限公司	5.719625e+04	8.512621e+04	-2331.36	0.079365	1	2	2	2	1	0.232759	0.200000	1	20

302 rows × 15 columns

```
In [9]: def GQG(x,min,max):
        return (x-min)*0.11/(max-min)+0.04
        min_in = Company_mess["供应公司比重"].min()
        max_in = Company_mess["供应公司比重"].max()
        min_out = Company_mess["销售公司比重"].min()
        max_out = Company_mess["销售公司比重"].max()

        Company_mess['供应公司比重'] = Company_mess['供应公司比重'].apply(
            lambda x: GQG(x, min_in, max_in))
        Company_mess['销售公司比重'] = Company_mess['销售公司比重'].apply(
            lambda x: GQG(x, min_out, max_out))

        # 重命名列，列名为索引类型，为不可变类型，故需先转为可变的索引再操作
        columns = Company_mess.columns
        columns = list(columns)
        columns[columns.index('供应公司比重')] = '利率1'
        columns[columns.index('销售公司比重')] = '利率2'
        Company_mess.columns = columns

        Company_mess.to_excel("kkp.xlsx") # 不保存行索引
        Company_mess
```

Out[9]:

	企业代 号	企业名称	进项总金额	销项总金额	总税额	作废比例	进项 年数	销项 年数	进项 月数	销项 月数	公 司 规 模	利率1	利率2	信 贷 风 险	信 誉 评 级
0	E124	个体经营E124	7.855255e+08	7.417804e+08	7980118.45	0.143067	3	3	12	12	3	0.048852	0.045126	0	80
1	E125	个体经营E125	9.323832e+08	9.412114e+08	14811182.88	0.142357	3	3	12	12	3	0.049824	0.044933	0	80
2	E126	个体经营E126	1.127735e+08	5.206911e+08	1271020.12	0.151771	3	3	8	12	3	0.054187	0.052759	0	80
3	E127	个体经营E127	1.646702e+06	6.519380e+08	19451364.52	0.027778	3	3	8	12	3	0.061846	0.040381	0	80
4	E128	个体经营E128	8.928033e+06	2.423902e+08	7866599.98	0.042375	3	3	12	10	3	0.057522	0.040338	0	60

302 rows x 15 columns

Out[10]:		企业名称	信誉评级	进项总金额	销项总金额	总税额	作废比例	公司规模	利率1	利率2	进项年数	销项年数	进项月数	销项月数
0		个体经营E124	80	7.855255e+08	7.417804e+08	7980118.45	0.143067	3	0.048852	0.045126	3	3	12	12
1		个体经营E125	80	9.323832e+08	9.412114e+08	14811182.88	0.142357	3	0.049824	0.044933	3	3	12	12
2		个体经营E126	80	1.127735e+08	5.206911e+08	1271020.12	0.151771	3	0.054187	0.052759	3	3	8	12
3		个体经营E127	80	1.646702e+06	6.519380e+08	19451364.52	0.027778	3	0.061846	0.040381	3	3	8	12
4		个体经营E128	60	8.928033e+06	2.423902e+08	7866599.98	0.042375	3	0.057522	0.040338	3	3	12	10
...	
248	***汽车维修有限责任公司	60	5.487833e+05	2.065282e+06	70352.53	0.059371	2	0.057236	0.070578	1	3	1	12	
253	***文化传播有限公司	60	2.459847e+05	7.795871e+05	15428.97	0.186747	2	0.068523	0.089625	1	3	1	10	
262	个体经营E386	40	1.830971e+05	4.145631e+06	110218.57	0.033186	2	0.070239	0.052524	1	2	3	2	
264	***商贸有限公司	60	9.313358e+04	1.031505e+06	18620.88	0.059016	2	0.060049	0.070176	1	4	1	11	
271	***网络科技有限公司	60	2.318720e+05	1.900801e+06	283238.39	0.215859	2	0.069601	0.061846	2	3	4	8	

Out[11]:		企业名称	信誉评级	进项总金额	销项总金额	总税额	作废比例	公司规模	利率1	利率2	进项年数	销项年数	进项月数	销项月数
5		个体经营E129	80	7.612615e+07	3.248060e+08	40141382.32	0.093078	3	0.064319	0.057456	3	3	12	12
3		个体经营E127	80	1.646702e+06	6.519380e+08	19451364.52	0.027778	3	0.061846	0.040381	3	3	8	12
7		个体经营E131	80	1.113480e+08	2.090840e+08	18920915.48	0.081983	3	0.054075	0.070269	3	3	12	12
69		个体经营E193	80	1.708134e+07	1.223168e+08	17011504.41	0.073096	3	0.059096	0.049216	3	3	12	12
1		个体经营E125	80	9.323832e+08	9.412114e+08	14811182.88	0.142357	3	0.049824	0.044933	3	3	12	12
...	
149		个体经营E273	40	9.969873e+06	2.300473e+07	292996.65	0.018151	3	0.062875	0.114223	3	3	12	11
216		***物流有限公司	40	2.964972e+04	6.827724e+06	202311.39	0.190476	2	0.083169	0.077731	1	3	1	8
262		个体经营E386	40	1.830971e+05	4.145631e+06	110218.57	0.033186	2	0.070239	0.052524	1	2	3	2
209		***房地产营销策划有限公司	40	1.313617e+06	7.120405e+06	101148.68	0.015448	2	0.063029	0.044204	3	3	4	5
169		***物资有限公司	40	7.511440e+06	7.696186e+06	72940.81	0.079755	2	0.092318	0.042890	2	3	2	6

7/10

```
return 0.6593787589707425*log(x) + 2.188076580002009

def quit(x, pla):
    if(x["信誉评级"] == 80):
        return funA(x["利率"+str(pla)])
    elif(x["信誉评级"] == 60):
        return funB(x["利率"+str(pla)])
    else:
        return funC(x["利率"+str(pla)])

Use_Company['取消合作1'] = Use_Company.apply(lambda x: quit(x, 1), axis=1)
Use_Company['取消合作2'] = Use_Company.apply(lambda x: quit(x, 2), axis=1)
Use_Company
```

Out[12]:

	企业名称	信誉 评级	进项总金额	销项总金额	总税额	作废比例	公司 规模	利率1	利率2	进项 年数	销项 年数	进项 月数	销项 月数	取消合作 1	取消合作 2
5	个体经营E129	80	7.612615e+07	3.248060e+08	40141382.32	0.093078	3	0.064319	0.057456	3	3	12	12	0.402919	0.327437
3	个体经营E127	80	1.646702e+06	6.519380e+08	19451364.52	0.027778	3	0.061846	0.040381	3	3	8	12	0.376689	0.091512
7	个体经营E131	80	1.113480e+08	2.090840e+08	18920915.48	0.081983	3	0.054075	0.070269	3	3	12	12	0.286868	0.462117
69	个体经营E193	80	1.708134e+07	1.223168e+08	17011504.41	0.073096	3	0.059096	0.049216	3	3	12	12	0.346269	0.223879
1	个体经营E125	80	9.323832e+08	9.412114e+08	14811182.88	0.142357	3	0.049824	0.044933	3	3	12	12	0.232091	0.162963
...
149	个体经营E273	40	9.969873e+06	2.300473e+07	292996.65	0.018151	3	0.062875	0.114223	3	3	12	11	0.363834	0.757484
216	***物流有限公司	40	2.964972e+04	6.827724e+06	202311.39	0.190476	2	0.083169	0.077731	1	3	1	8	0.548276	0.503691
262	个体经营E386	40	1.830971e+05	4.145631e+06	110218.57	0.033186	2	0.070239	0.052524	1	2	3	2	0.436863	0.245229
209	***房地产营销策划有限公司	40	1.313617e+06	7.120405e+06	101148.68	0.015448	2	0.063029	0.044204	3	3	4	5	0.365451	0.131520
169	***物资有限公司	40	7.511440e+06	7.696186e+06	72940.81	0.079755	2	0.092318	0.042890	2	3	2	6	0.617096	0.111619

189 rows × 15 columns

In [13]:

```
def getEarn(x, pla):
    return x["总税额"]*(1-x["取消合作"+str(pla)])

Use_Company['期望1'] = Use_Company.apply(lambda x: getEarn(x, 1), axis=1)
Use_Company['期望2'] = Use_Company.apply(lambda x: getEarn(x, 2), axis=1)

Use_Company["利率"] = Use_Company.apply(
    lambda x: x["利率1"] if x["期望1"] > x["期望2"] else x["利率2"], axis=1)
Use_Company["取消合作"] = Use_Company.apply(
    lambda x: x["取消合作1"] if x["期望1"] > x["期望2"] else x["取消合作2"], axis=1)
Use_Company.drop('利率1',axis=1,inplace=True)
Use_Company.drop('利率2',axis=1,inplace=True)
Use_Company.drop('期望1',axis=1,inplace=True)
Use_Company.drop('期望2',axis=1,inplace=True)
Use_Company.drop('取消合作1',axis=1,inplace=True)
Use_Company.drop('取消合作2',axis=1,inplace=True)

Use_Company["年均进项"] = Use_Company["进项总金额"] / Use_Company["进项年数"]
Use_Company["年均销项"] = Use_Company["销项总金额"] / Use_Company["销项年数"]
Use_Company["月均进项"] = Use_Company["年均进项"] / Use_Company["销项月数"]
Use_Company["月均销项"] = Use_Company["年均销项"] / Use_Company["销项月数"]
Use_Company["月均税额"] = Use_Company["年均销项"] / Use_Company["进项年数"]
Use_Company.drop('进项总金额',axis=1,inplace=True)
Use_Company.drop('销项总金额',axis=1,inplace=True)
Use_Company.drop('进项年数',axis=1,inplace=True)
Use_Company.drop('销项年数',axis=1,inplace=True)
Use_Company.drop('进项月数',axis=1,inplace=True)
Use_Company.drop('销项月数',axis=1,inplace=True)

Use_Company["贷款额度"] = (Use_Company["月均进项"]*(1-Use_Company["作废比例"])/10000 + 1)*10000
Use_Company
```

Out[13]:

	企业名称	信誉 评级	总税额	作废比例	公司 规模	利率	取消合作	年均进项	年均销项	月均进项	月均销项	月均税额	贷款额度
5	个体经营E129	80	40141382.32	0.093078	3	0.057456	0.327437	2.537538e+07	1.082687e+08	2.114615e+06	9.022390e+06	3.608956e+07	1920000.0
3	个体经营E127	80	19451364.52	0.027778	3	0.040381	0.091512	5.489008e+05	2.173127e+08	4.574173e+04	1.810939e+07	7.243755e+07	50000.0

	企业名称	信誉评级	总税额	作废比例	公司规模	利率	取消合作	年均进项	年均销项	月均进项	月均销项	月均税额	贷款额度
7	个体经营E131	80	18920915.48	0.081983	3	0.054075	0.286868	3.711601e+07	6.969468e+07	3.093001e+06	5.807890e+06	2.323156e+07	2840000.0
69	个体经营E193	80	17011504.41	0.073096	3	0.049216	0.223879	5.693781e+06	4.077228e+07	4.744817e+05	3.397690e+06	1.359076e+07	440000.0
1	个体经营E125	80	14811182.88	0.142357	3	0.044933	0.162963	3.107944e+08	3.137371e+08	2.589953e+07	2.614476e+07	1.045790e+08	22220000.0
...
149	个体经营E273	40	292996.65	0.018151	3	0.062875	0.363834	3.323291e+06	7.668243e+06	3.021174e+05	6.971130e+05	2.556081e+06	300000.0
216	***物流有限公司	40	202311.39	0.190476	2	0.077731	0.503691	2.964972e+04	2.275908e+06	3.706215e+03	2.844885e+05	2.275908e+06	10000.0
262	个体经营E386	40	110218.57	0.033186	2	0.052524	0.245229	1.830971e+05	2.072816e+06	9.154853e+04	1.036408e+06	2.072816e+06	90000.0
209	***房地产营销策划有限公司	40	101148.68	0.015448	2	0.044204	0.131520	4.378724e+05	2.373468e+06	8.757448e+04	4.746936e+05	7.911561e+05	90000.0
169	***物资有限公司	40	72940.81	0.079755	2	0.042890	0.111619	3.755720e+06	2.565395e+06	6.259533e+05	4.275659e+05	1.282698e+06	580000.0

189 rows × 13 columns

4. 控制贷款金额

```
In [14]: Can_Company = pd.DataFrame(columns=Use_Company.columns)
```

```
In [15]: total_Money = 100000000
for index, row in Use_Company.iterrows():
    if (row["贷款额度"]<total_Money):
        Can_Company=Can_Company.append(row, ignore_index=True)
        total_Money -= row["贷款额度"]
    #print(row) # 输出每行的索引值
Can_Company
```

Out[15]:

	企业名称	信誉评级	总税额	作废比例	公司规模	利率	取消合作	年均进项	年均销项	月均进项	月均销项	月均税额	贷款额度
0	个体经营E129	80	40141382.32	0.093078	3	0.057456	0.327437	2.537538e+07	1.082687e+08	2.114615e+06	9.022390e+06	3.608956e+07	1920000.0
1	个体经营E127	80	19451364.52	0.027778	3	0.040381	0.091512	5.489008e+05	2.173127e+08	4.574173e+04	1.810939e+07	7.243755e+07	50000.0
2	个体经营E131	80	18920915.48	0.081983	3	0.054075	0.286868	3.711601e+07	6.969468e+07	3.093001e+06	5.807890e+06	2.323156e+07	2840000.0
3	个体经营E193	80	17011504.41	0.073096	3	0.049216	0.223879	5.693781e+06	4.077228e+07	4.744817e+05	3.397690e+06	1.359076e+07	440000.0
4	个体经营E125	80	14811182.88	0.142357	3	0.044933	0.162963	3.107944e+08	3.137371e+08	2.589953e+07	2.614476e+07	1.045790e+08	22220000.0
...
62	个体经营E128	60	7866599.98	0.042375	3	0.040338	0.069000	2.976011e+06	8.079672e+07	2.976011e+05	8.079672e+06	2.693224e+07	290000.0
63	***建筑劳务有限责任公司	60	6148429.21	0.109504	3	0.049871	0.207021	4.399134e+05	3.847502e+07	4.887926e+04	4.275003e+06	1.923751e+07	50000.0
64	***建筑劳务有限公司	60	5846710.81	0.060754	3	0.040650	0.074010	1.417051e+05	3.374687e+07	1.417051e+04	3.374687e+06	8.436717e+06	20000.0
65	***建筑工程有限公司	60	2227004.31	0.087284	3	0.040866	0.077460	7.215804e+04	1.687569e+07	1.443161e+04	3.375138e+06	5.625230e+06	20000.0
66	***网络科技有限公司	60	283238.39	0.215859	2	0.061846	0.347020	1.159360e+05	6.336002e+05	1.449200e+04	7.920003e+04	3.168001e+05	20000.0

67 rows × 13 columns

```
In [16]: Can_Company.to_excel("题目二_贷款分配.xlsx") #不保存行索引
Can_Company
```

Out[16]:

	企业名称	信誉 评级	总税额	作废比例	公司 规模	利率	取消合作	年均进项	年均销项	月均进项	月均销项	月均税额	贷款额度
0	个体经营 E129	80	40141382.32	0.093078	3	0.057456	0.327437	2.537538e+07	1.082687e+08	2.114615e+06	9.022390e+06	3.608956e+07	1920000.0
1	个体经营 E127	80	19451364.52	0.027778	3	0.040381	0.091512	5.489008e+05	2.173127e+08	4.574173e+04	1.810939e+07	7.243755e+07	50000.0
2	个体经营 E131	80	18920915.48	0.081983	3	0.054075	0.286868	3.711601e+07	6.969468e+07	3.093001e+06	5.807890e+06	2.323156e+07	2840000.0
3	个体经营 E193	80	17011504.41	0.073096	3	0.049216	0.223879	5.693781e+06	4.077228e+07	4.744817e+05	3.397690e+06	1.359076e+07	440000.0
4	个体经营 E125	80	14811182.88	0.142357	3	0.044933	0.162963	3.107944e+08	3.137371e+08	2.589953e+07	2.614476e+07	1.045790e+08	22220000.0
...
62	个体经营 E128	60	7866599.98	0.042375	3	0.040338	0.069000	2.976011e+06	8.079672e+07	2.976011e+05	8.079672e+06	2.693224e+07	290000.0
63	***建筑劳 务有限责 任公司	60	6148429.21	0.109504	3	0.049871	0.207021	4.399134e+05	3.847502e+07	4.887926e+04	4.275003e+06	1.923751e+07	50000.0
64	***建筑劳 务有限公 司	60	5846710.81	0.060754	3	0.040650	0.074010	1.417051e+05	3.374687e+07	1.417051e+04	3.374687e+06	8.436717e+06	20000.0
65	***建筑工 程有限公 司	60	2227004.31	0.087284	3	0.040866	0.077460	7.215804e+04	1.687569e+07	1.443161e+04	3.375138e+06	5.625230e+06	20000.0
66	***网络科 技有限公 司	60	283238.39	0.215859	2	0.061846	0.347020	1.159360e+05	6.336002e+05	1.449200e+04	7.920003e+04	3.168001e+05	20000.0

67 rows × 13 columns