

悉尼大学 | 跨学科项目组 | 2019年7月

粤港澳大湾区公司画像

赵子晴 Alex
张艺璇 Nikki
邓佳琪 Linda

胡宇皓 Mirope
汤深浅 Eric
陈思 Reya

研究方向

PROBLEM STATEMENT

定性转定量

将定性研究的结果转化成定量数据

风险因子重要性

选出对风险有影响的风险因子并根据重要性排序

风险因子关联性

考虑风险因子关联性

风险画像可视化

选择合理的可视化手段使风险分析结果一目了然

关键建议

1. 数据处理



1. 风险因子分层
2. 定性转定量

2. 模型构建

1. 随机森林模型
2. 分析风险因子重要性

3. 评测结果可视化

1. 多维画像模型
2. 应用情景



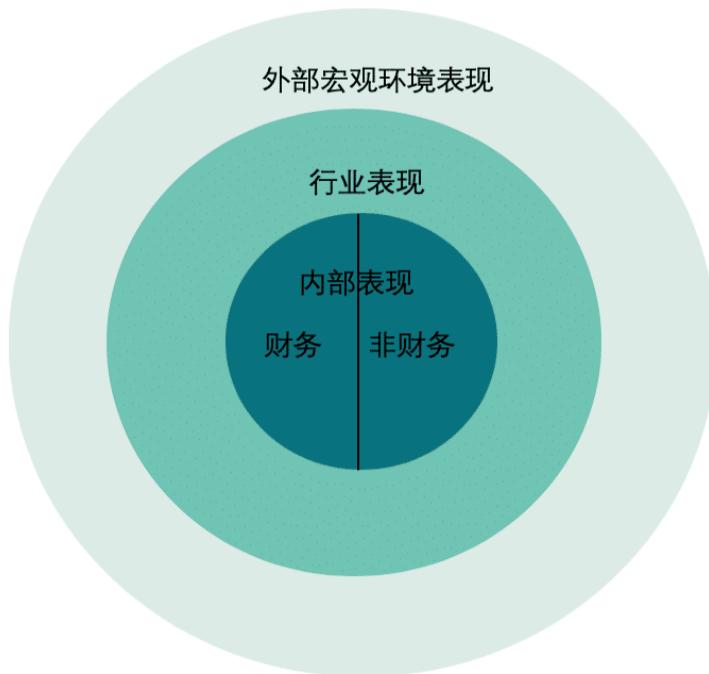
数据收集



建议 1.1 选择和排列风险因子

SELECTING AND CATEGORISING RISK FACTORS

风险因子分层



宏观环境

大湾区是否已建成营商友好型环境

行业环境

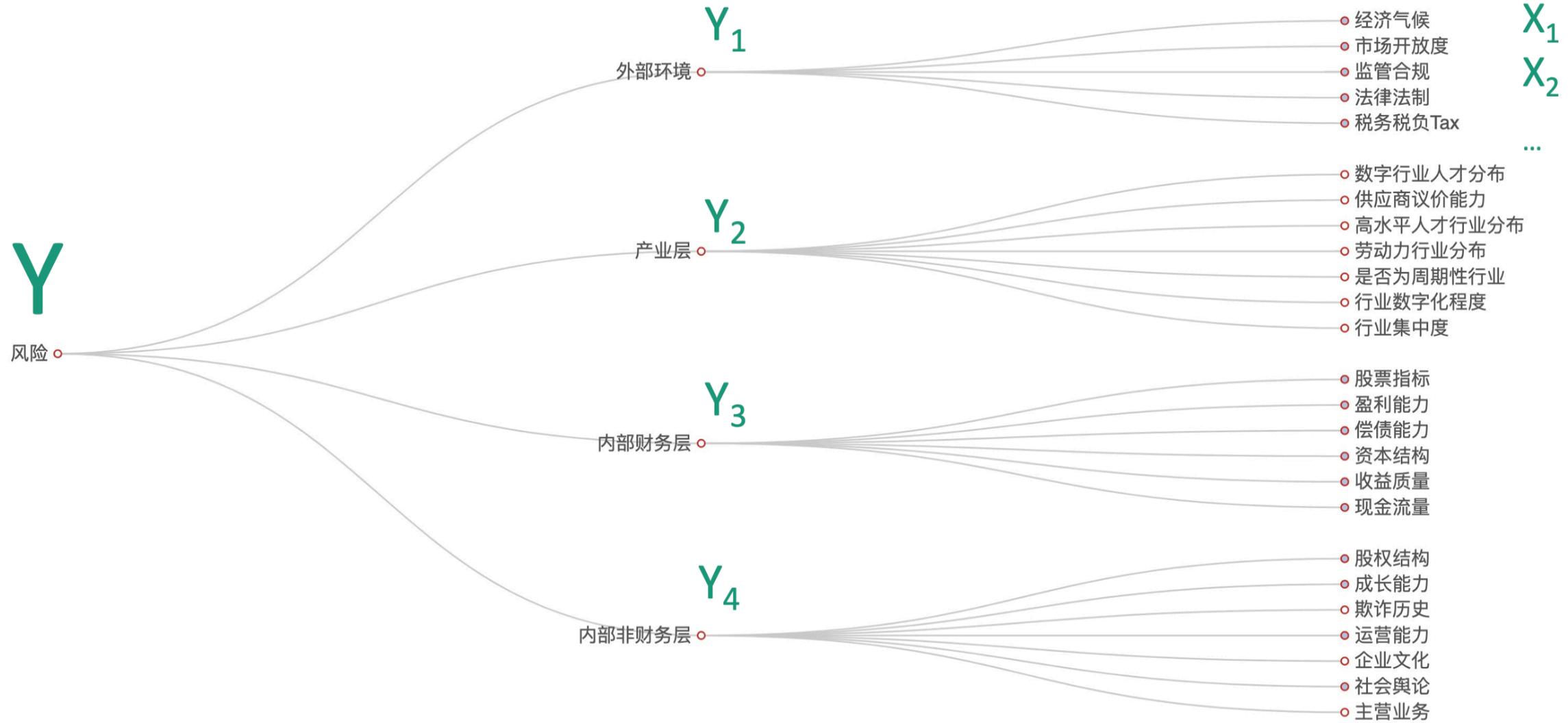
大湾区重点行业（软件服务 电子制造 交通运输）环境分析

内部非财务

股权结构 企业成长 主营业务 企业形象 处罚历史

内部财务

财务比率研究



数据处理

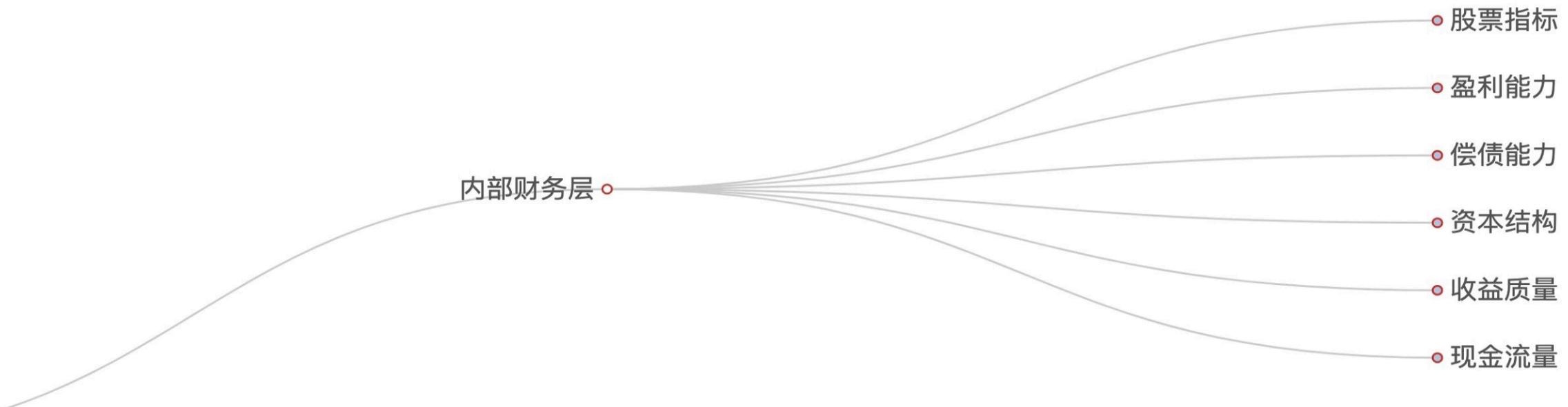


建议 1.2 定性转定量

QUALITATIVE TO QUANTITATIVE
TRANSFORMATION



内部财务风险因子



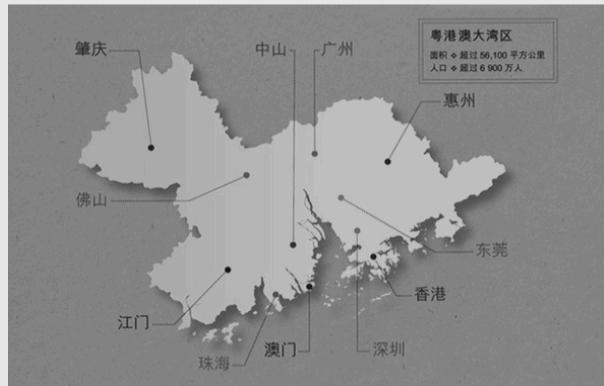
192 家深交所上市公司

105 项财务指标

请点击电脑上的红点查看具体指标

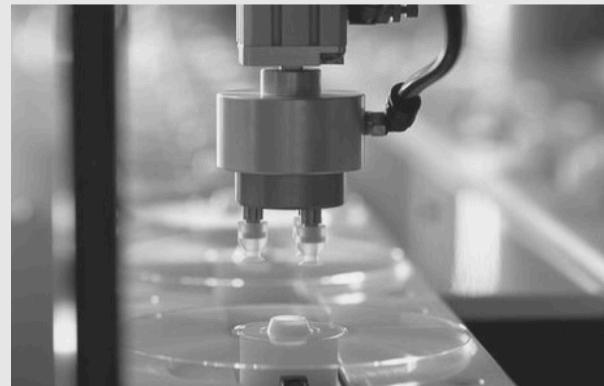


需要定性转定量的三个方面



宏观环境分析

经济和政策



行业环境分析

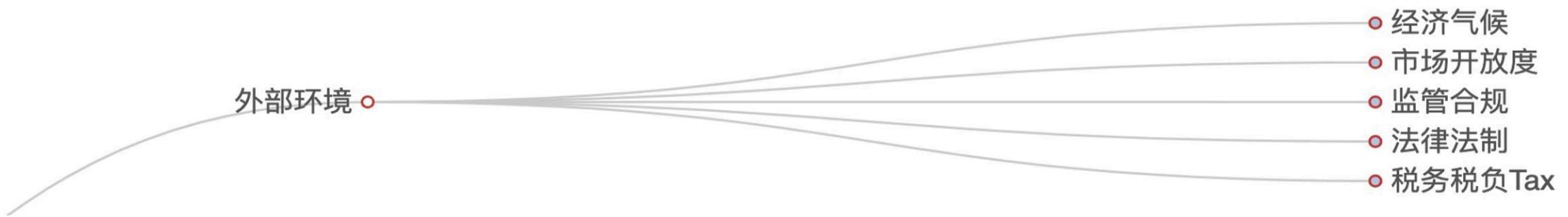
交通运输 | 电子制造 |
信息软件服务



内部非财务分析

股权结构 | 创新能力 |
公司形象

外部环境因子



5个方面衡量商业友好程度

13类宏观环境指标

请点击电脑上的红点查看具体指标



三大机构指数



Doing Business and Related Research

Each year *Doing Business* highlights important new work that speaks to a variety of issues facing the private sector. Brief summaries of this work—more than 100 academic papers and reports—are available by *Doing Business* topic and related policy spaces. If you would like to receive additional papers that are not already listed on the website please email your request to the Doing Business team. We would gladly share the relevant literature. The list of papers is updated periodically by the Doing Business team and only reputable research institutions are included on the website.

世界银行 全球营商环境指数

通过文献研究和数据收集对世界各国营商便利程度打分。我们取用了其中对于贸易，企业管理，资产保护，合同执行和税负五个角度的分数



EVENT International Economies

January 25, 2019

Freedom and Prosperity: The 2019 Index of Economic Freedom



HERITAGE FOUNDATION
经济自由指数

华盛顿保守派权威智库，坚信小型政府、自由经济是商业繁荣的本源。我们取用了大部分指数

WORLD JUSTICE PROJECT
全球法制指数

考察全球法制程度。好的法制环境对商业确定性和营商成本有根本性影响。我们去用了其中对于法制和法规确定性的分数。

政策研究

财税[2019]31号：财政部 税务总局关于粤港澳大湾区个人所得税优惠政策的通知

粤港澳大湾区优惠政策是针对在大湾区工作的境外(含港澳台)高端人才和紧缺人才，由广东省和深圳市按内地与香港个人所得税税负差额给予免税补贴

此政策下，人员流动优惠惠及三地，税收惠及广东省一地，所以三地人员流动优惠政策根据有无原则各打一分，税收优惠广东一地打一分

假设国际知识产权保护条约为基准，专利权的保护为20年(TRIPS)，商标的保护为10年(WIPO 马德里系统)。用粤港澳三地对专利权和商标的保护年限除以上基准，得出三地的得分。

人员流动政策优惠

税收政策优惠

知识产权保护

Number	Factor	H-C-M Avg.
Economic Climate		
Ex 1.1	National economic growth position	0.73
Ex 1.2	Global economic growth position	1.60
Market Openness		
Ex 2.1	Trade flow Heritage Foundation - Trade Freedom Doing business - Trade Across Borders Adjusted	0.85
Ex 2.2	Capital flow Heritage Foundation – Investment Freedom Heritage Foundation – Financial Freedom Chinn Ito Adjusted	0.64
Ex 2.3	Labour mobility Heritage Foundation – Labour Freedom Policy research Adjusted	1
Regulatory Environment		
Ex 3.1	Regulatory certainty Rule of Law Index - Open Government - Publicised law & govt data Rule of Law Index - Open Government - Right to information Rule of Law Index - Regulatory Enforcement Adjusted	0.63
Ex 3.2	Rule of law Heritage Foundation – Government integrity Rule of Law index - Constraint on Government Power Rule of Law index - Absence of Corruption Adjusted	0.52
Ex 3.3	Regulation on Corporate Governance Doing Business - Protection of minority shareholder Adjusted	0.63



Legal System		
Ex 4.1	Tangible property rights Heritage Foundation - Property Right Doing business - Registering property Adjusted	0.69
Ex 4.2	Intangible property rights (terms of protection) Patent Trademark Adjusted	0.95
Ex 4.3	Dispute resolution Heritage Foundation – Judicial effectiveness Doing Business - Enforcing contract Adjusted	0.70
Tax		
Ex 5.1	Tax incentive Regulation research	0.33
Ex 5.2	Tax burden Heritage Foundation – Tax Burden Doing Business - Paying Tax Adjusted	0.802

行业环境因子

产业层

- 数字行业人才分布
- 供应商议价能力
- 高水平人才行业分布
- 劳动力行业分布
- 是否为周期性行业
- 行业数字化程度
- 行业集中度

3个大湾区核心产业

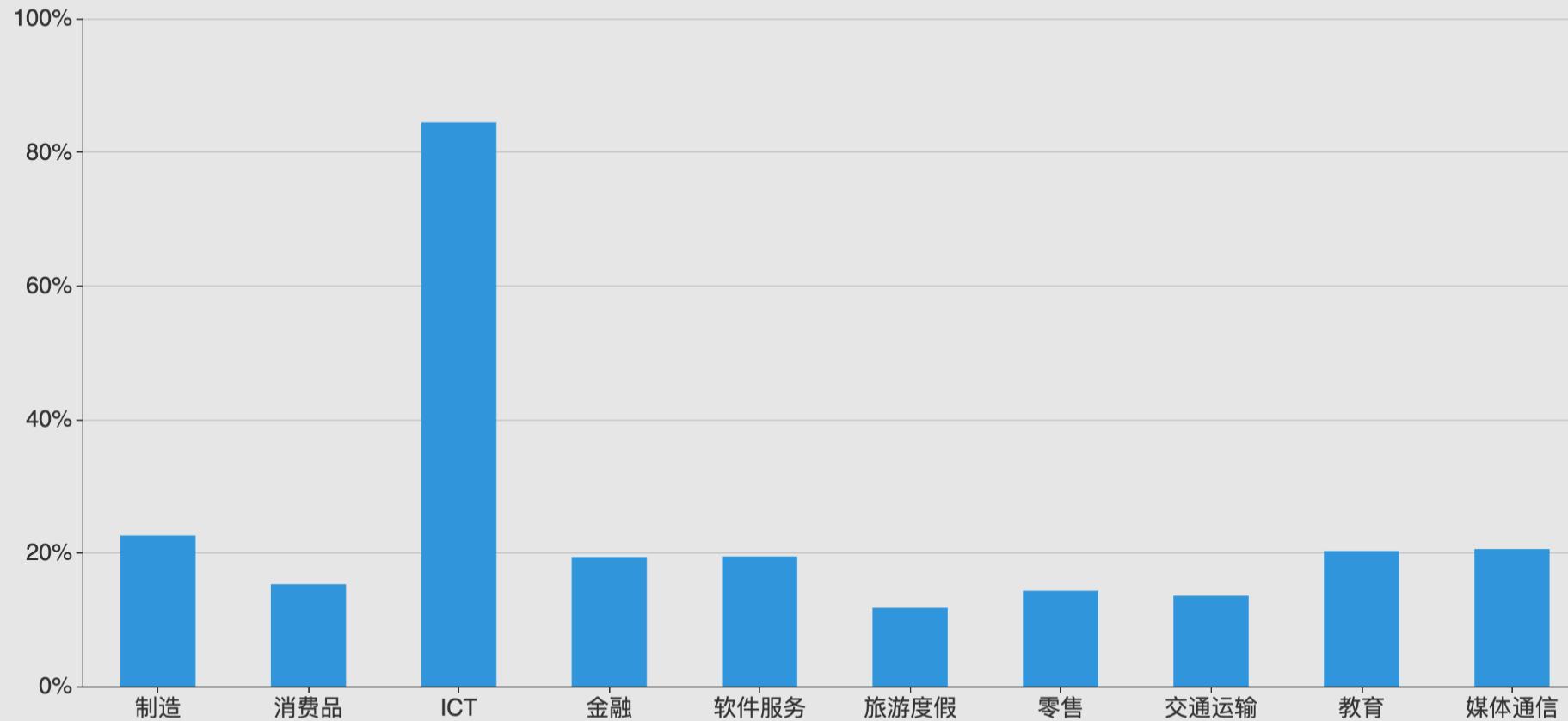
9类行业指标

请点击电脑上的红点查看具体指标



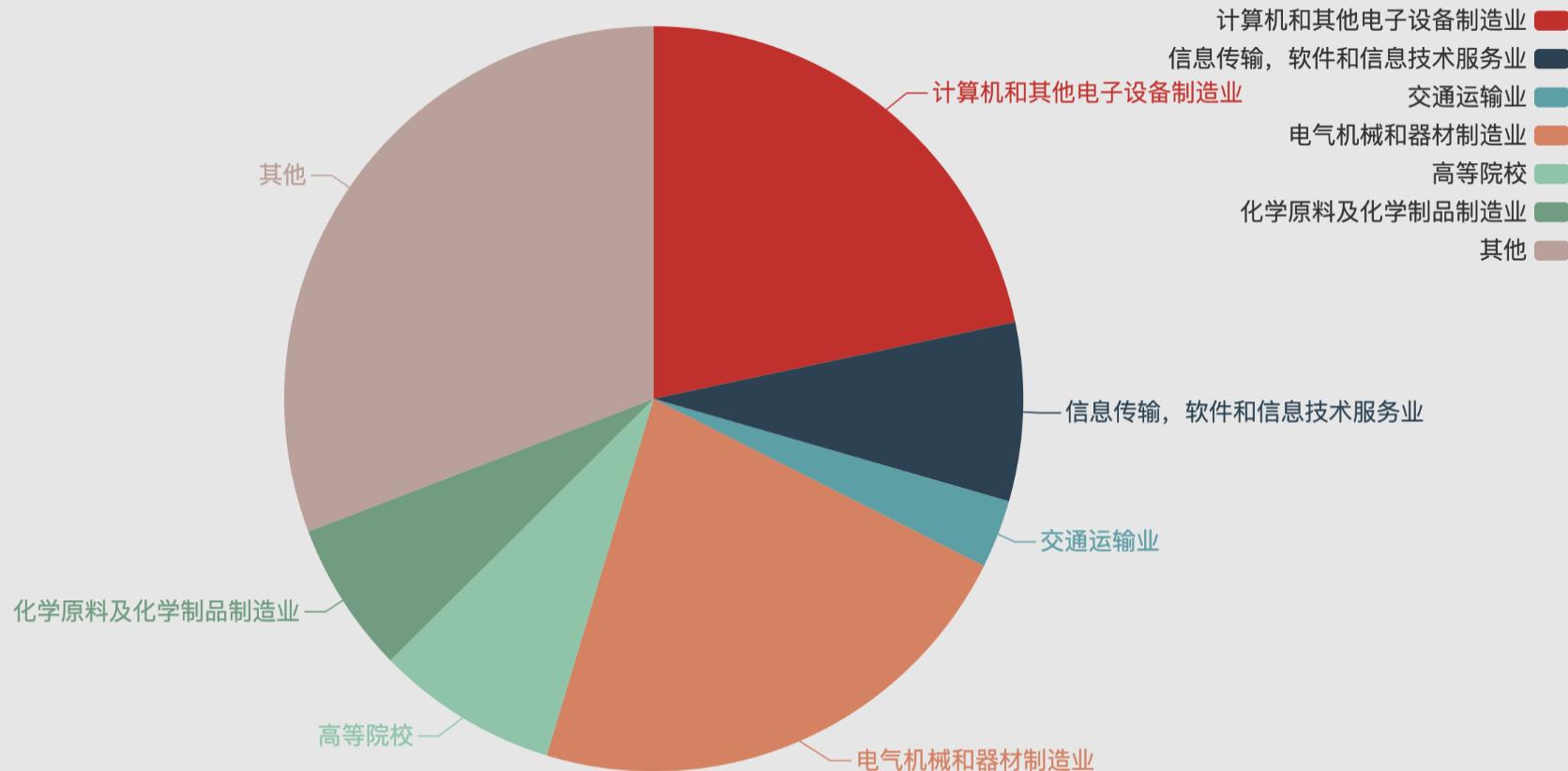
大湾区主要行业人才数字化程度

来源:《粤港澳大湾区数字经济与人才发展研究报告 (2019.2)》



粤港澳大湾区创新机构行业分布

来源:《粤港澳大湾区协同创新发展报告(2018)》



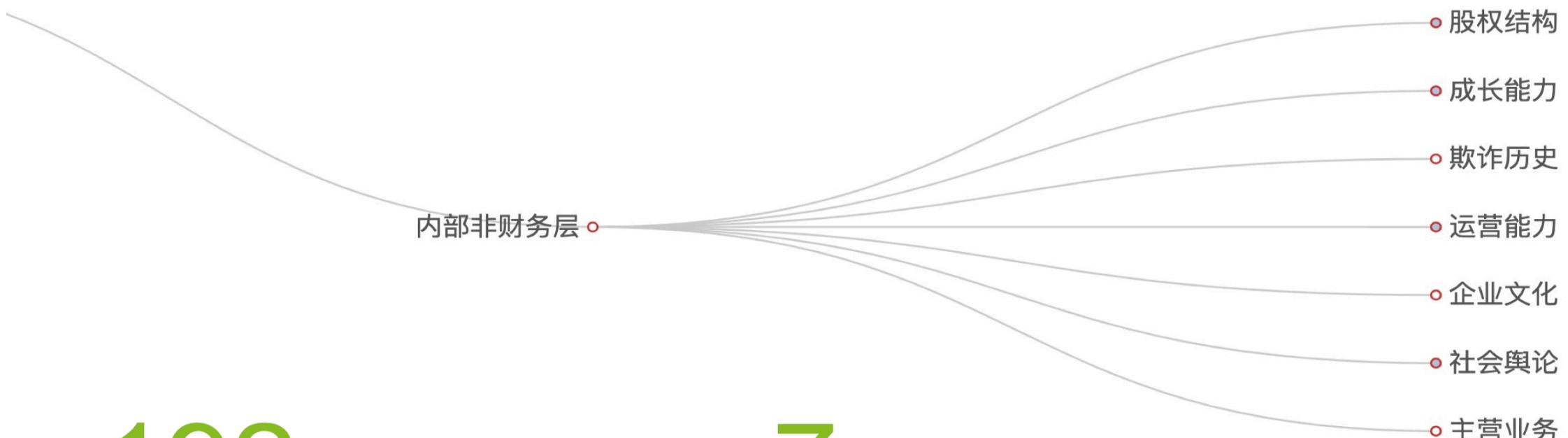
行业集中度

	CR ₈ 值 (%)
寡占II型（高集中寡占）	CR ₈ ≥ 85
寡占III型（中上集中寡占）	75 ≤ CR ₈ < 85
寡占IV型（中下集中寡占）	45 ≤ CR ₈ < 75
寡占V型（低集中寡占）	40 ≤ CR ₈ < 45
竞争型（原子型）	CR ₈ < 40

	交通运输	电子制造	软件服务
CR ₈	0.3991	0.4970	0.3287

来源：西安交通大学经济与金融学院

内部非财务因子



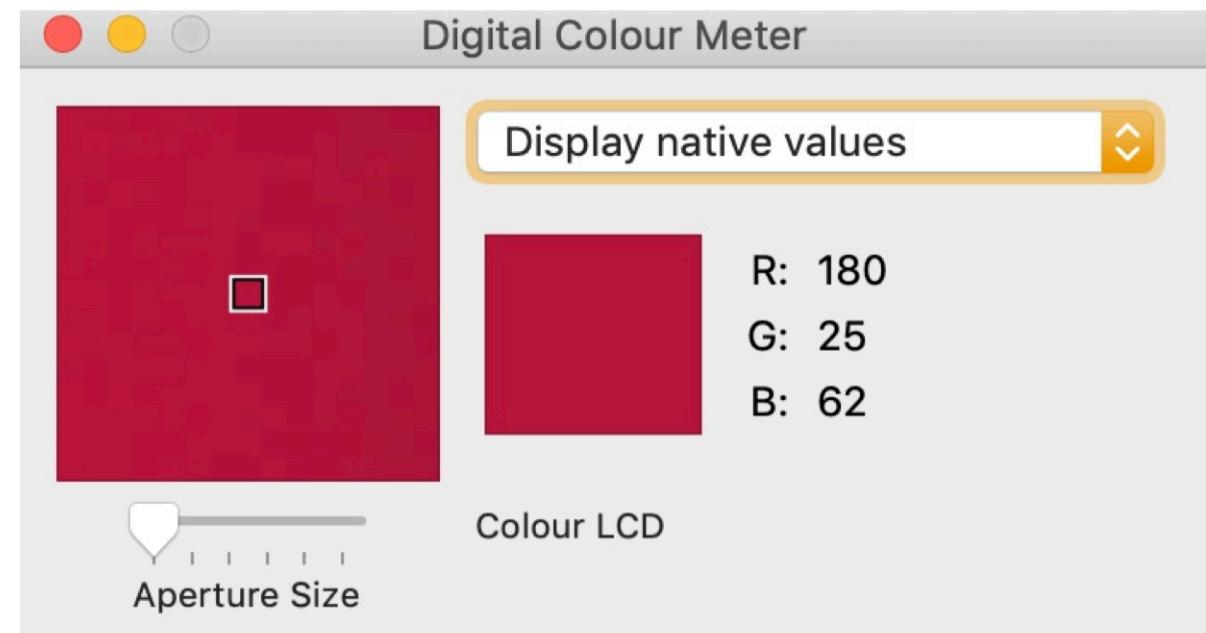
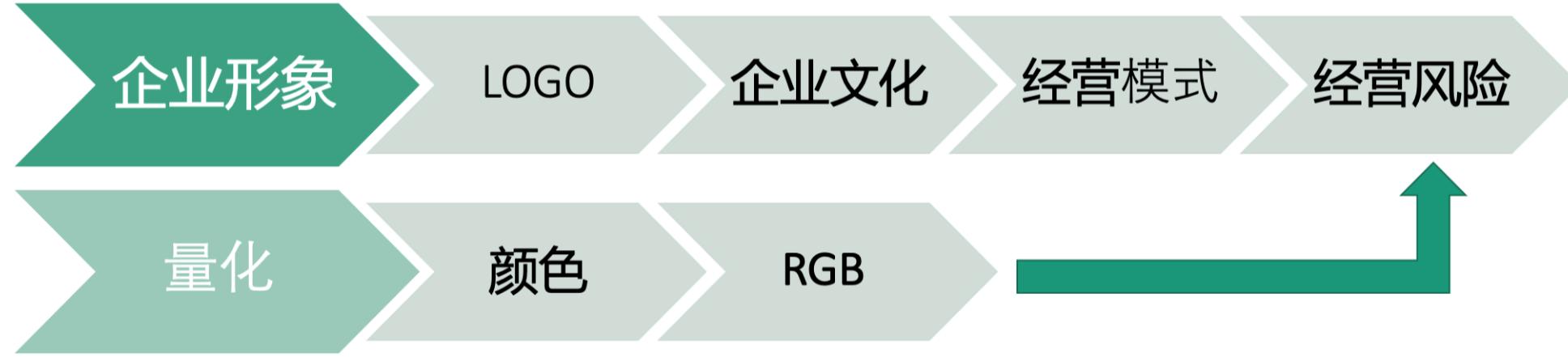
192家深交所上市公司

7类非财务指标

请点击电脑上的红点查看具体指标







模型构建



建议 2.1 画像模型

PROFILING MODEL

随机森林回归模型

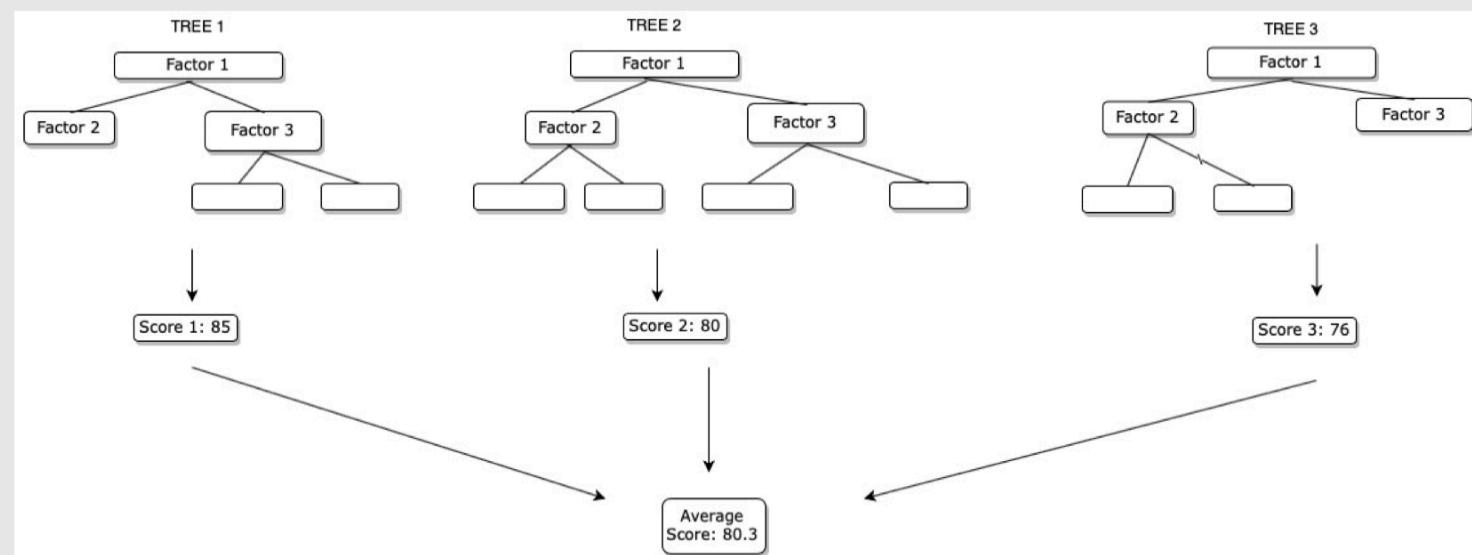
Random Forest Algorithm (Biermann, 2001)

基于多个决策树的建立:

- 从训练集中，选取N个公司样本
- 从n个风险因子中，随机抽取p个因子($p < N$)
- 根据评价标准MSE，选取最佳因子(Risk factors)和分割点(Slit-point)

通过调整参数，对树的形状进行“修剪”

平均“每棵树”的预测结果



为何选择随机森林回归模型

可解释性:

- 机器学习黑箱问题
- 随机森林可视化风险因子重要性 “factor importance” (Friedman, 2001)
- 集成学习算法可视化风险因子间交互作用 “interaction effect” (Friedman, 2008)

准确性: Random Forest is good at tackling overfitting (Breiman, 2001)

- 测试集回测结果:
 - 以多元线性回归模型为基准，随机森林算法可降低50%预测错误率
 - 和同样具有因子筛选功能的Lasso回归模型比较，随机算法错误率低于Lasso

主要发现



建议 2.2 重要性分析

RISK FACTOR SIGNIFICANCE

Business Layers

Step1: Import Factors Data & Rename column names

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.pipeline import Pipeline,make_pipeline
from sklearn.model_selection import cross_val_score, GridSearchCV, KFold
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor, ExtraTreesRegressor
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import RobustScaler, StandardScaler
from sklearn.metrics import mean_squared_error
from scipy.stats import skew
from sklearn.decomposition import PCA, KernelPCA
from sklearn.inspection.partial_dependence import partial_dependence, plot_partial_dependence
from sklearn.metrics import r2_score
from collections import defaultdict
import warnings
warnings.filterwarnings("ignore", category=FutureWarning, module="sklearn", lineno=196)
sns.set_palette("BuGn_r")
```

Jupyter Notebook 与 数据库

1.1: Data Cleaning and Average Filling

```
In [2]: #Import data "new_data_two"
new_data_two = pd.read_csv('Raw_Data/Financial_new_two.csv')
new_data_two = pd.DataFrame(new_data_two)
new_data_two.shape
```

Out[2]: (192, 125)

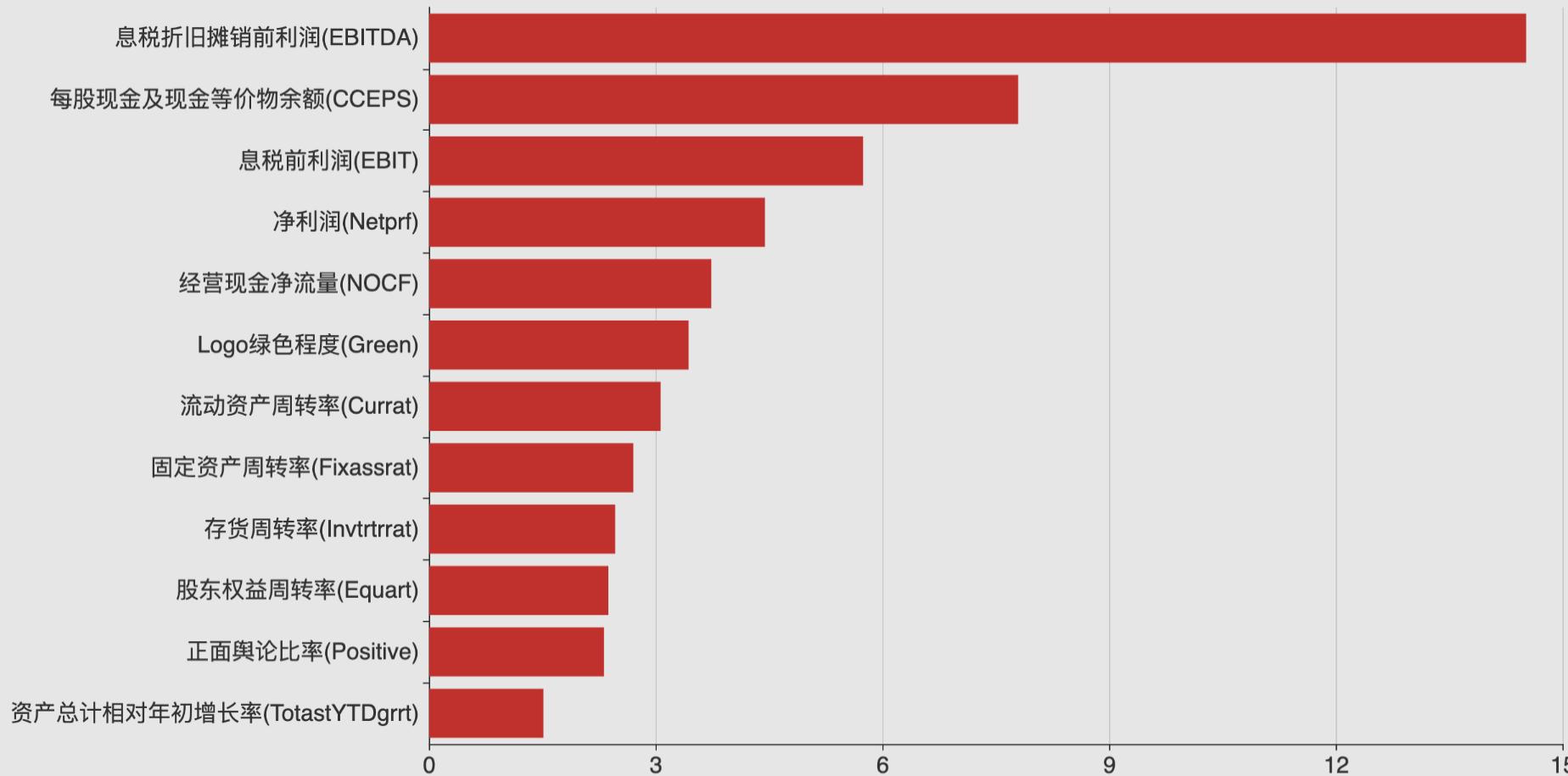
```
In [3]: #Rename column names and average filling NaN
column_name=new_data_two.columns.str
columns_name = column_name.split("_")
replace = []
for items in columns_name:
    replace.append(items[1])

new_data_two.columns = replace
for rows in new_data_two:
    null=new_data_two[rows].isnull()
    countnull=null.sum()
    if countnull>=20:
        new_data_two = new_data_two.drop(rows, 1)

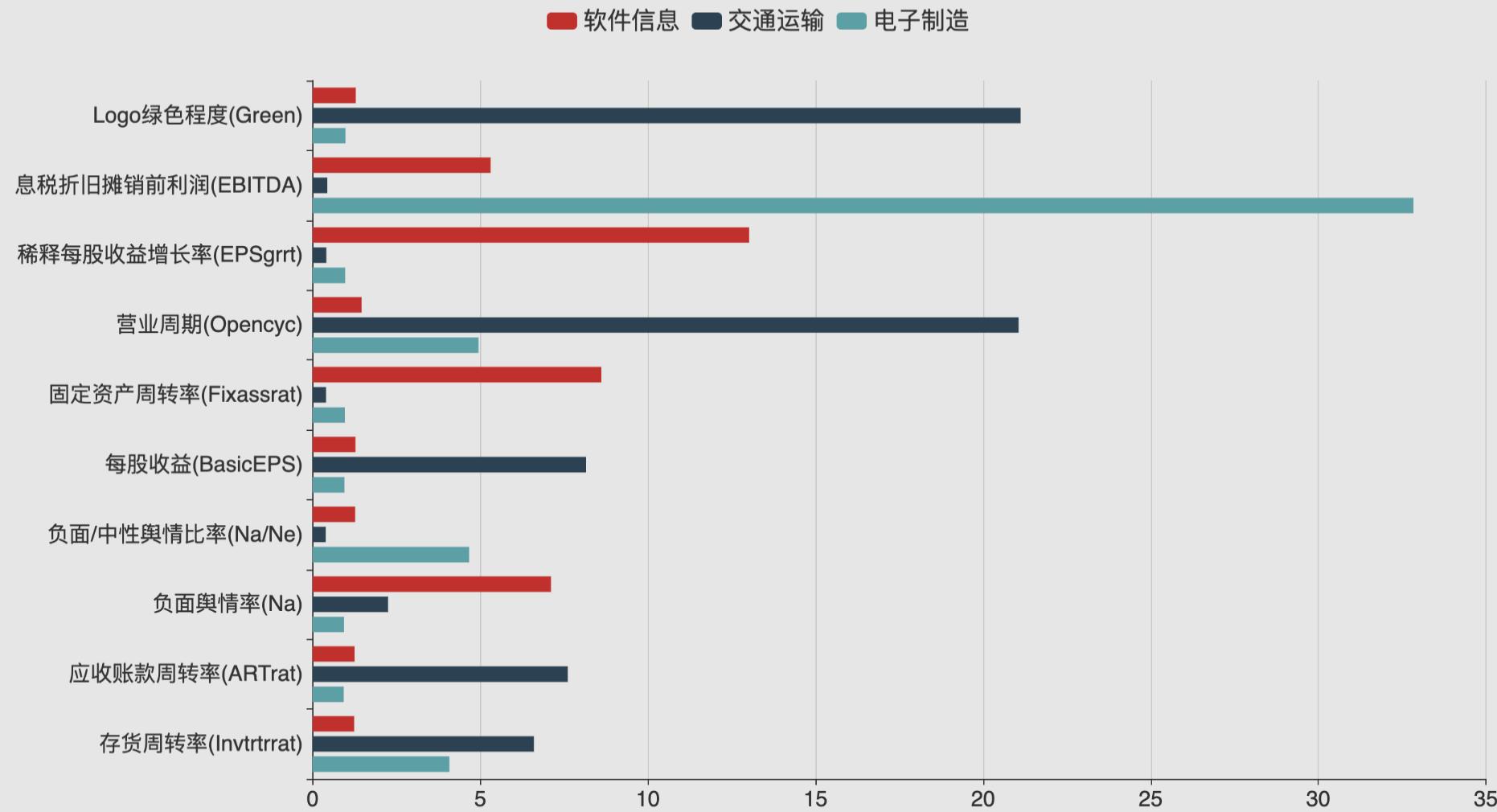
    else:
        fill = new_data_two[rows].mean() * 7
```

三个产业的因子重要性

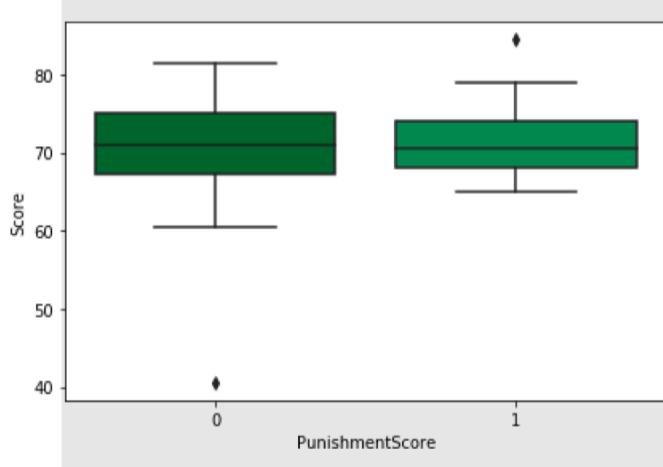
数据来源于192个上市公司



三个产业的因子重要性对比

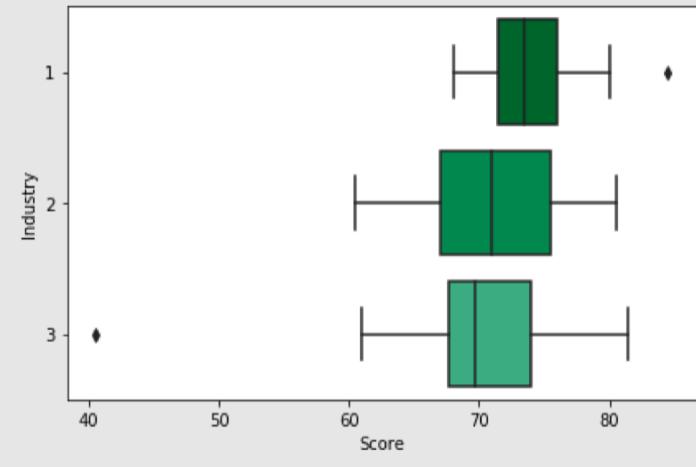


主要发现



证监会处罚与分值

法规和政治压力将导致高风险公司改变其内部监管行为，以避免潜在的高成本股东诉讼 (Trueman 1997)

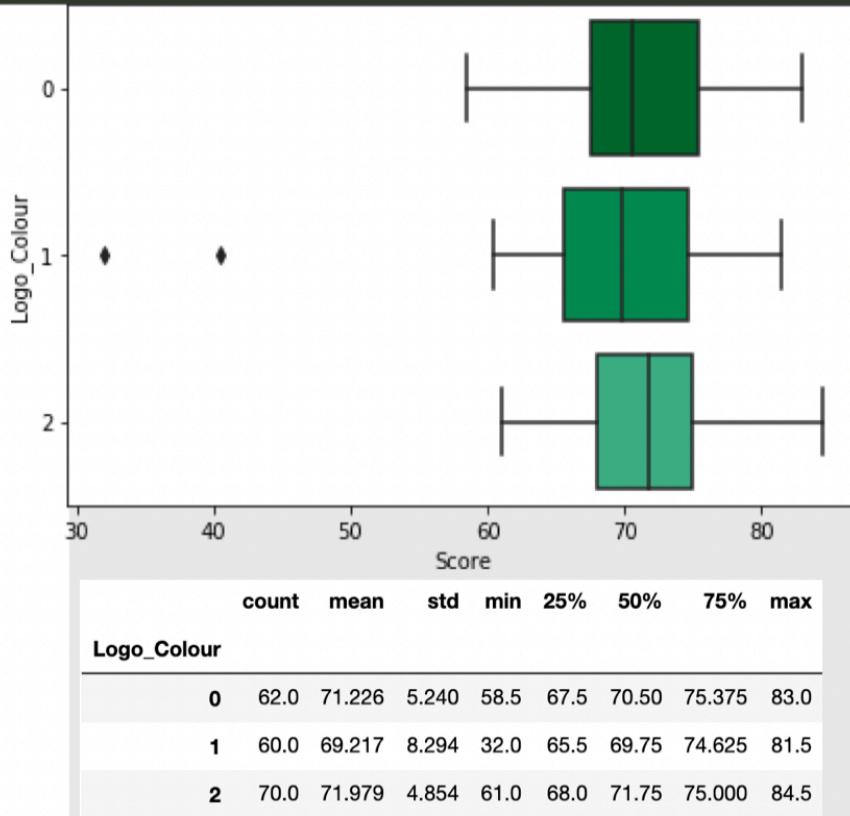


从行业特点来看，交通运输、仓储业属于低风险、稳定发展的行业 (Lijuan 2011)

新兴产业具有科技创新活动高投入、高风险、强信息不对称等特点 (Weihong 2013)

行业与分值

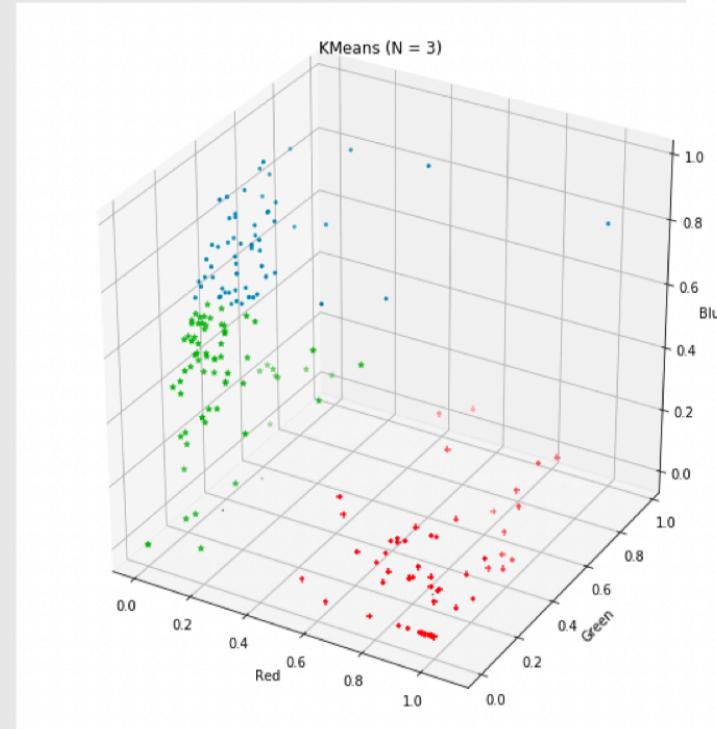
主要发现



公司标志与分值

“著名的公司非常重视选择和设计他们的标志颜色，因为它们是公司使用的所有颜色中必不可少的最重要的，直接代表公司形象和文化”
(Wang 2009)

绿色标志使得消费者一目了然地明确哪些产品有益于环境和健康，而通过消费者的选择和市场竞争，可以引导企业自觉调整产业



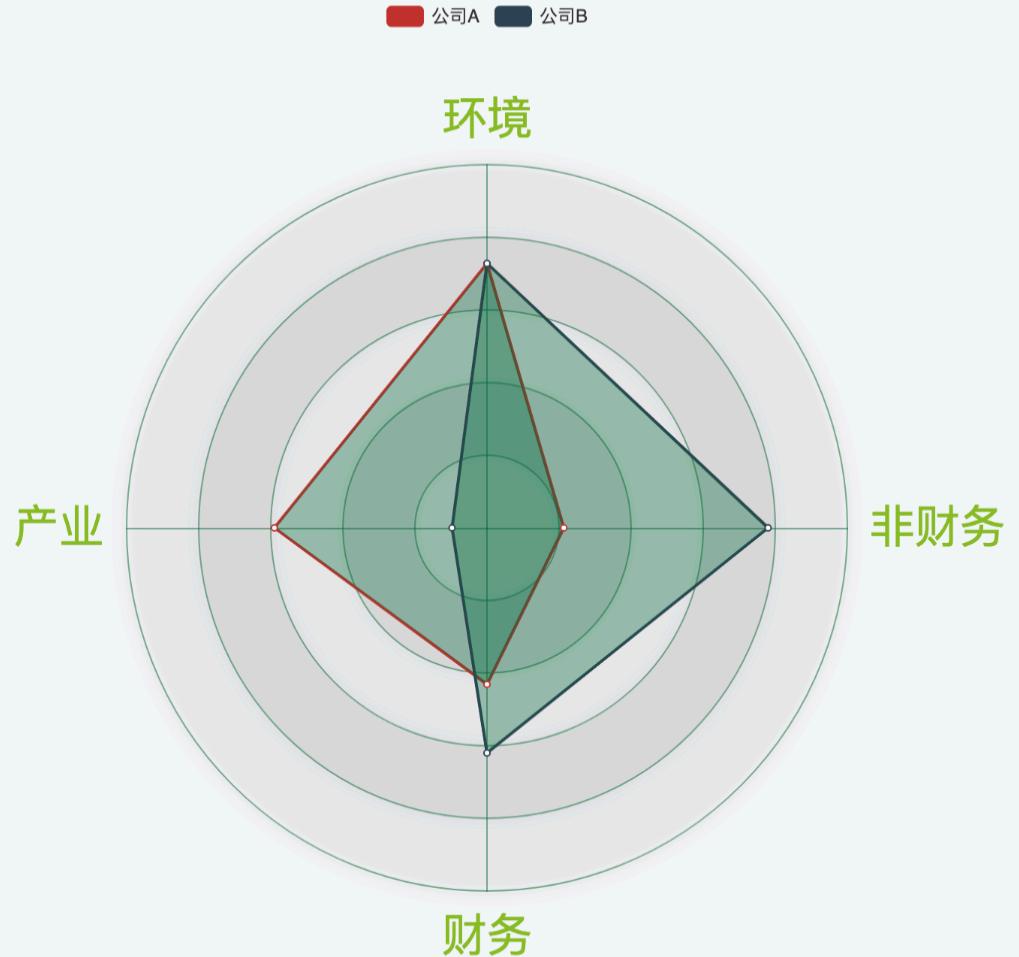
可视化



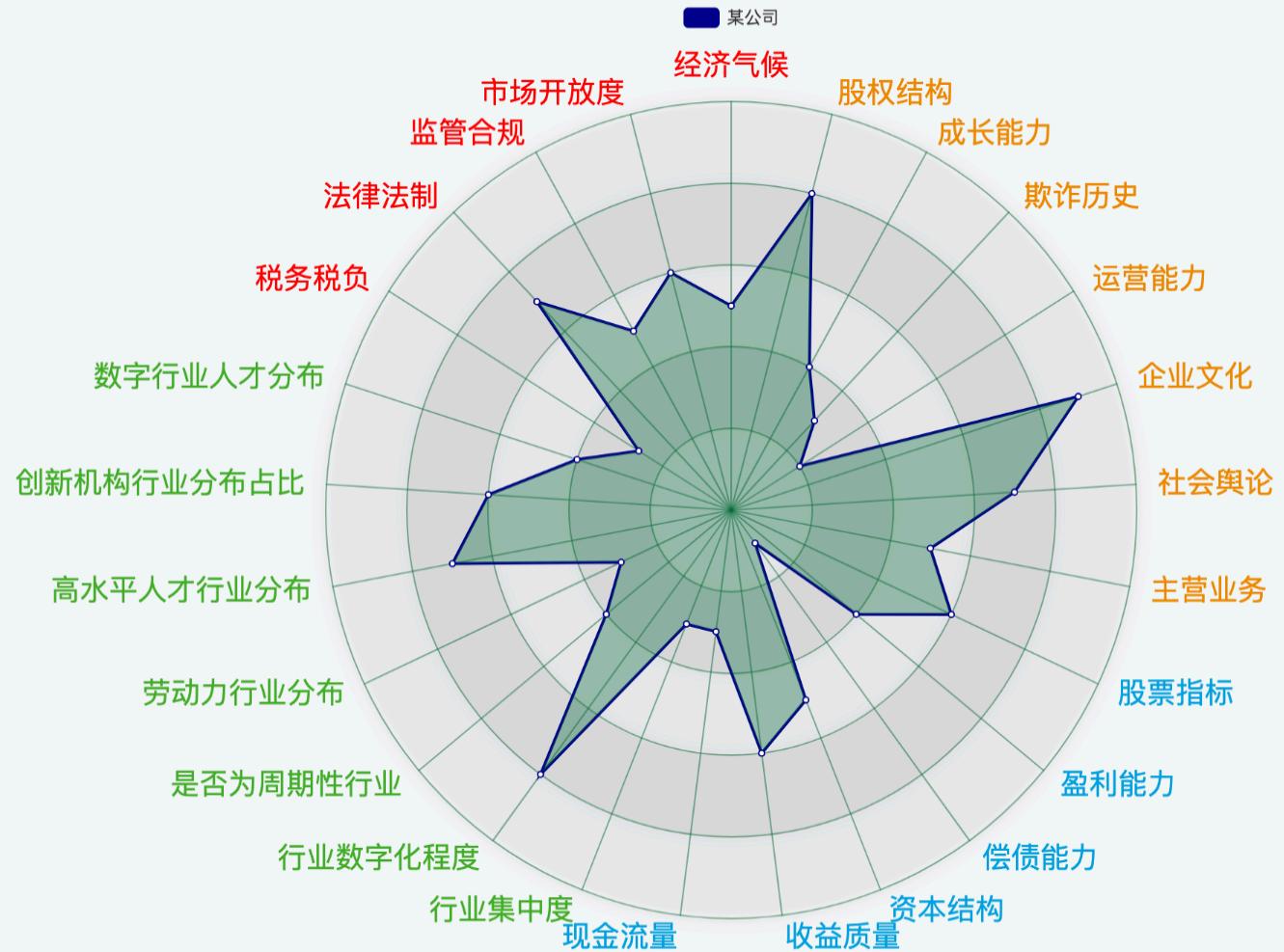
建议 3.1 & 3.2
多维画像模型可视化

ANALYSIS RESULT VISUALISATION

四个维度的公司画像模型



25个维度的公司画像模型



应用场景

The image displays three screenshots of the Deloitte Greater Bay Area Company Profiling platform, illustrating its features for company analysis.

- Screenshot 1: Company Profile (顺丰控股股份有限公司 - 002352)**
 - External:** Includes the SF logo and company name.
 - Industry:** External macro-environmental表现 (0.73), domestic economic growth position (0.73), global economic growth position (1.6), industry performance (0.0219), financial industry distribution (0.0478), and non-financial industry distribution (0.1357).
 - Financial:** Internal financial metrics like turnover rate (118.0665), cash flow debt ratio (0.2057), EPS (1.03), sales net margin (4.9089), and negative news sentiment (0.2238).
 - Non-financial:** Internal non-financial metrics like equity concentration (0.861646), revenue growth (414.31%), and accounts receivable days (29).
- Screenshot 2: Data Visualization**
 - Metrics:** A circular progress bar showing 72% METRIC.
 - Bar Charts:** A bar chart titled "Title" comparing "Returning" (blue) and "New" (red) metrics across dates from January 11 to January 15, with values ranging from 0.0 to 1.6k.
 - Line Charts:** A line chart titled "PROGRESS" showing a fluctuating trend line over the same period.
- Screenshot 3: Geographic Map**
 - A map of the Guangdong-Hong Kong-Macao Greater Bay Area showing major cities like Guangzhou, Dongguan, Foshan, and Shenzhen, along with provincial boundaries and road networks.

模型验证



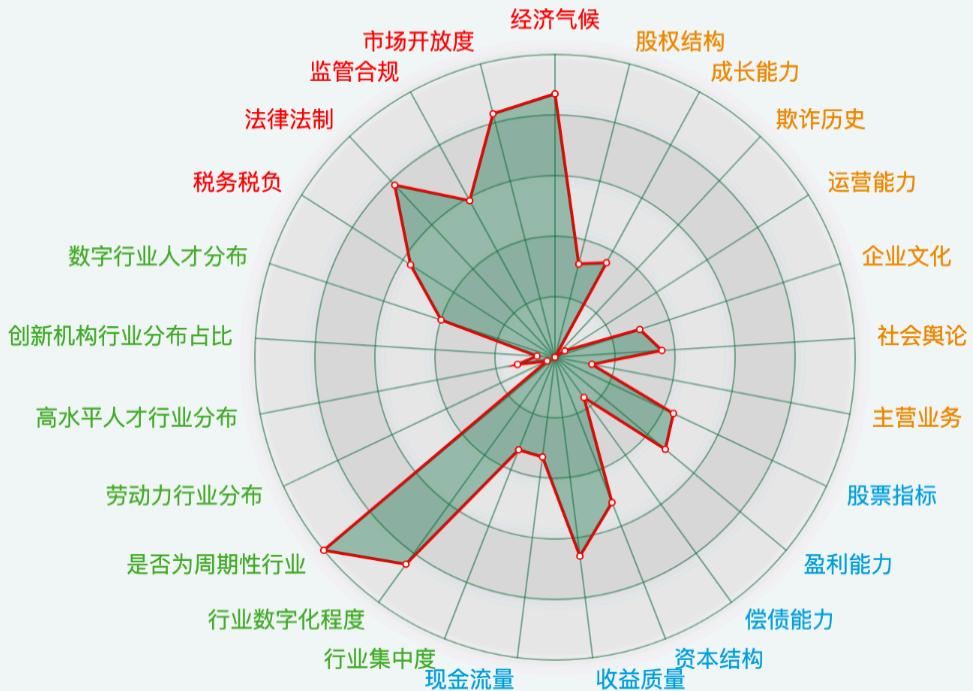
个案研究

CASE STUDY

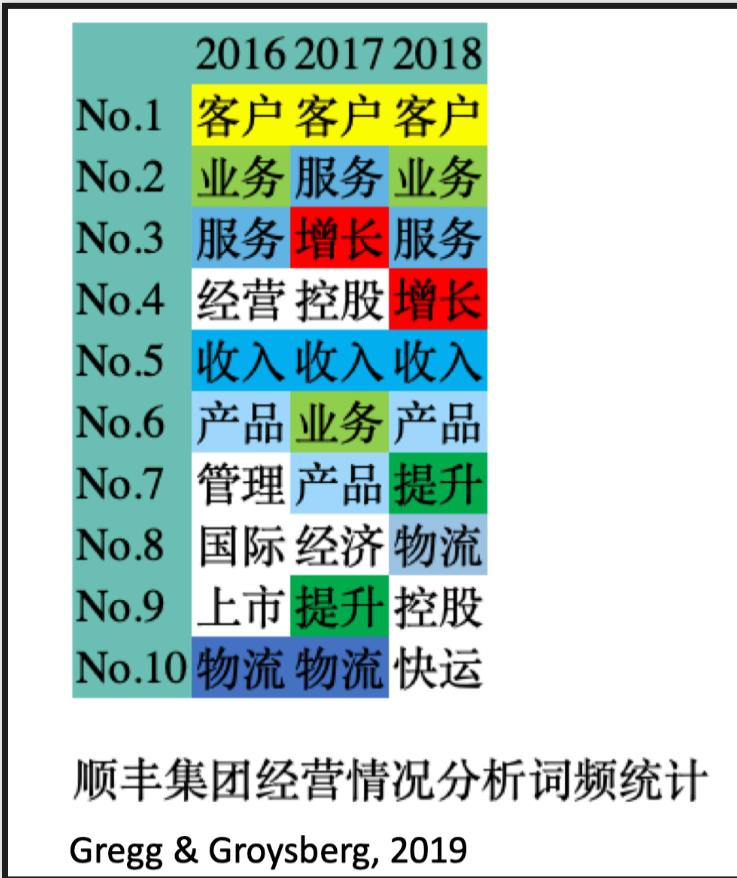
北讯集团股份有限公司 —— 系统得分: 32



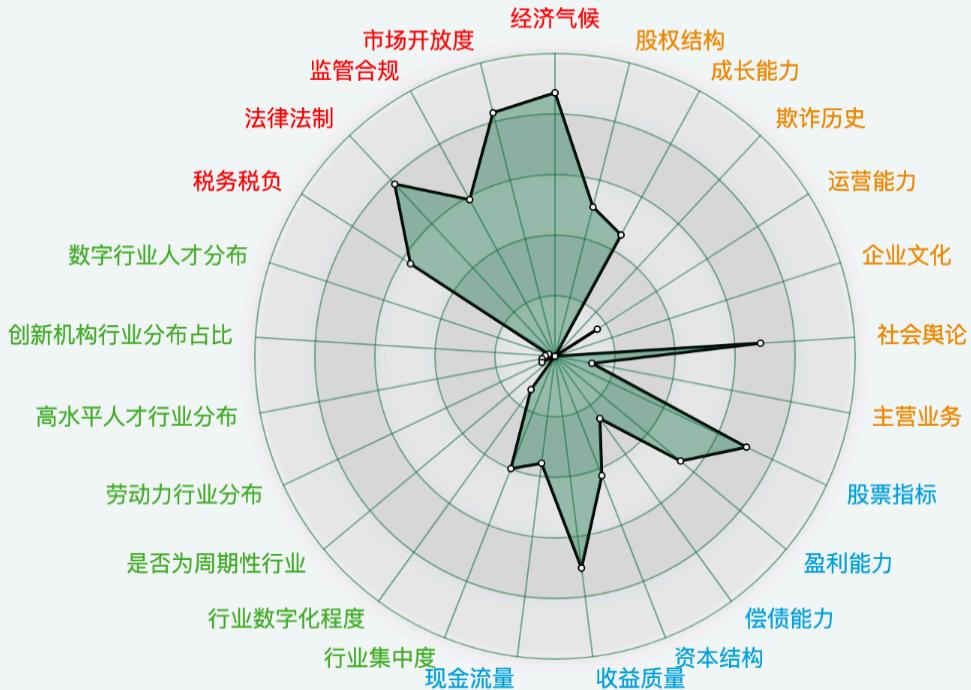
北讯25个维度的画像



顺丰控股股份有限公司 —— 系统得分: 84.5



顺丰25个维度的画像



局限性

数据不完整，数据清洗难度较大

- 部分数据缺失需要手动填充数据

数据量局限

- 模型预测准确度
- 筛选因子波动较大
- 外部风险缺乏对照组

数据处理方式局限性

- 行业环境风险因子对公司影响程度各不相同，但由于资料有限，我们给同一行业的公司打分相同
- 同理外部环境风险因子对公司影响程度各不相同，但由于资料有限，我们假设大湾区商业环境给地区内所有公司的风险相同

建议总结



实施时间轴

IMPLEMENTATION TIMELINE

模型未来发展

优化打分模板制定

- 不断优化定性转定量的打分机制，挖掘维度更广的风险因子，从而达到提前预警风险

考虑交互式作用

- 考虑Friedman的理论用集成学习的方法可视化交互作用
- 可以更精确把控公司风险结构

考虑做出完整系统的客户端

- 带入5G大数据时代，增强模型维度，复杂性，进而增强模型精确度

短期

中期

长期



非常感谢德勤
提供的学习机会
我们受益匪浅

Thank you.

附录



研究附录

APPENDIX

Number	Factor	Benchmark	HK-data	HK-mark	CHN-data	CTN-data	CTN-mark	MFM-data	MFM-mark	H-C-M Avg.
Economic Climate										
Ex 1.1	National economic growth position	6.60%	3.02%	0.46	6.60%	6.80%	1.03	4.70%	0.71	0.73
Ex 1.2	Global economic growth position	3.03%	3.02%	1.00	6.60%	6.80%	2.24	4.70%	1.55	1.60
Market Openness										
Ex 2.1	Trade flow									
	Heritage Foundation - Trade Freedom		95	0.95	73	73	0.73	90	0.9	
	Doing business - Trade Across Borders		93.56	0.9356	69.91	69.91	0.6991			
	Adjusted		94.28	0.9428	71.455	71.455	0.71455	90	0.9	0.85
Ex 2.2	Capital flow									
	Heritage Foundation – Investment Freedom	100	90	0.9	25	25	0.25	85	0.85	
	Heritage Foundation – Financial Freedom	100	90	0.9	20	20	0.2	70	0.7	
	Chinn Ito		1	1	0.165808752	0.165808752	0.165808752			
	Adjusted		60.333333333	0.933333333	15.05526958	15.05526958	0.205269584	77.5	0.775	0.64
Ex 2.3	Labour mobility									
	Heritage Foundation – Labour Freedom		89.2	0.892	64.2	64.2	0.642	50	0.5	
	Policy research	Y/N		1	1		1	1	1	
	Adjusted				1			1	1	1
Regulatory Environment										
Ex 3.1	Regulatory certainty									
	Rule of Law Index - Open Government - Publicised law & govt data		0.72	0.72	0.45	0.45	0.45			
	Rule of Law Index - Open Government - Right to information		0.76	0.76	0.56	0.56	0.56			
	Rule of Law Index - Regulatory Enforcement		0.82	0.82	0.48	0.48	0.48			
	Adjusted		0.766666667	0.766666667	0.496666667	0.496666667	0.496666667			0.63
Ex 3.2	Rule of law									
	Heritage Foundation – Government integrity		83.8	0.838	49.1	49.1	0.491	33.2	0.332	
	Rule of Law index - Constraint on Government Power		0.65	0.65	0.33	0.33	0.33			
	Rule of Law index - Absence of Corruption		0.84	0.84	0.55	0.55	0.55			
	Adjusted			0.776			0.457		0.332	0.52
Ex 3.3	Regulation on Corporate Governance									
	Doing Business - Protection of minority shareholder		76.67	0.7667	48.33	48.33	0.4833			
	Adjusted			0.77135			0.4833			0.63

Legal System											
Ex 4.1 Tangible property rights											
	Heritage Foundation - Property Right			93.3	0.933	49.9	49.9	0.499	60	0.6	
	Doing business - Registering property			73.54	0.7354	76.15	76.15	0.7615			
	Adjusted			83.42	0.8342	63.025	63.025	0.63025	60	0.6	0.69
Ex 4.2 Intangible property rights (terms of protection)											
	Patent	20	20	1	20	20	1	20	1		
	Trademark	10	10	1	10	10	1	7	0.7		
	Adjusted				1			1		0.85	0.95
Ex 4.3 Dispute resolution											
	Heritage Foundation – Judicial effectiveness			75.3	0.753	75.2	75.2	0.752	60	0.6	
	Doing Business - Enforcing contract			69.13	0.6913	78.23	78.23	0.7823			
	Adjusted			72.215	0.72215	76.715	76.715	0.76715	60	0.6	0.70
Tax											
Ex 5.1 Tax incentive											
	Regulation research	Y/N		0			1		0		0.33
Ex 5.2 Tax burden											
	Heritage Foundation – Tax Burden			93.1	0.931	70.4	70.4	0.704	77.1	0.771	0.802
	Doing Business - Paying Tax			98.82	0.9882	62.9	62.9	0.629			
	Adjusted				0.9596			0.6665		0.771	0.80

The Most Commonly Used Words in Amazon's Letters to Shareholders

An analysis of the top words used per year shows
that "customer(s)" and "Amazon" lead the pack.

	No. 1	No. 2	No. 3	No. 4	No. 5
1997	customers	investments	continue	Amazon	market
1998	customers	Amazon	build	working	new
1999	customers	Amazon	business	store	continue
2000	customer	company	Amazon	sales	time
2001	customers	cash	flows	shares	company
2002	customer	price	Amazon	books	experience
2003	customer	term	cost	long	Amazon
2004	cash	flows	earnings	capital	free
2005	decisions	customers	prices	make	term
2006	businesses	new	Amazon	grow	culture
2007	book	Kindle	reading	change	tools
2008	customer	Amazon	Kindle	new	sellers
2009	customers	goals	Amazon	experience	new
2010	data	Amazon	services	systems	technology
2011	Amazon	authors	get	KDP	publishing
2012	customers	Amazon	price	Prime	proactively
2013	Amazon	customers	team	Prime	new
2014	customers	Prime	Amazon	AWS	sellers
2015	customers	Amazon	service	sellers	one
2016	customer	decision	day	process	Amazon
2017	Amazon	high	standards	Prime	customers

Note: "KDP" stands for Kindle Direct Publishing,
and "AWS" stands for Amazon Web Services.

Source: Textual analysis of Amazon's letters to shareholders,
1997–2017, by Tricia Gregg and Boris Groysberg

Reference

Hong Kong Patents Ordinance (Chapter 514), S39

Hong Kong Trade Marks Ordinance (Chapter 559) S49

澳门《工業產權法律制度》第97/99/M號法令 第一百零三條

澳门《工業產權法律制度》第97/99/M號法令 第二百一十八條

《中华人民共和国公司法》第一百零三条

<https://hbr.org/2019/05/amazons-priorities-over-the-years-based-on-jeff-bezos-letters-to-shareholders>

Hickson, D. J. (1971). 'A Strategic Contingencies' Theory of Intraorganizational Power', *Administrative Science Quarterly*, Vol. 16, No. 2, Jun., 1971, pp. 216-229. Johnson Graduate School of Management, Cornell University.

KPMG (2018). Navigating the Greater Bay Area: Second Annual Survey on Key Drivers for Success. KPMG, HSBC and HKGCC.

Li, M. & Zhang, M. (2019). 股权结构、高管特征对环境信息披露的影响研究. 会计之友 2019 年第 12 期.

Siddiqi, N. (2016). Introduction. In *Intelligent Credit Scoring* (pp. 1–18). Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119282396.ch1>

Wang, L (2019). 世界一流湾区的发展经验:对推动我国大湾区建设的启示与借鉴. 银行家. 2019年06期. ISSN : 1671-1238

Paunov, C. (2016). Corruption's asymmetric impacts on firm innovation. *Journal of Development Economics*, 118(C), 216–231. <https://doi.org/10.1016/j.jdeveco.2015.07.006>

Johnson, S., McMillan, J., & Woodruff, C. (2002). Courts and Relational Contracts. *Journal of Law, Economics, and Organization*, 18(1), 221–277. <https://doi.org/10.1093/jleo/18.1.221>