2021年全国职业院校技能大赛

高职组

“大数据技术与应用”

赛项赛卷（GZ-2021041-02卷）

任

务

书

参赛队编号：

背景描述

据央视财经报道，2020年我国O2O市场规模突破万亿元，O2O市场存在着巨大的潜力。特别是餐饮和外卖行业，占据市场较大份额，并且业务增长迅速。截至2020年底，全国外卖总体订单量已超过171.2亿单，同比增长7.5%，全国外卖市场交易规模达到8352亿元，同比增长14.8%。我国外卖用户规模已接近5亿人，其中80后、90后是餐饮外卖服务的中坚消费力量，消费者使用餐饮外卖服务也不再局限于传统的一日三餐，下午茶和夜宵逐渐成为消费者的外卖新宠。为把握这一商业机遇，ChinaSkills公司计划进驻外卖平台市场，现需对大规模成熟外卖平台进行详细评估调研，采集多方多维度数据，寻找行业痛点，摸清市场需求，以技术为手段为投资保驾护航。

为完成该项工作，你所在的小组将应用大数据技术，以Python、Java、Scala作为整个项目的基础开发语言，基于大数据平台综合利用MapReduce、Spark、MySQL、Scrapy、Flask、ECharts等，对数据进行获取、处理、清洗、挖掘、分析、可视化呈现，力求实现对公司未来的重点战略方向提出建议。

你们作为该小组的技术人员，请按照下面任务完成本次工作，并编制综合报告。

# 模块A：Hadoop平台及组件的部署管理（15分）

**环境说明：**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 主机名 | 类型 | 用户 | 密码 |
| 1 | master | 主节点 | root | 123456 |
| 2 | slave1 | 从节点 | root | 123456 |
| 3 | slave2 | 从节点 | root | 123456 |
| 补充说明：主节点MySQL数据库用户名/密码：root/123456  相关软件安装包在/chinaskills目录下  所有模块中应用命令必须采用绝对路径 | | | | |

### 任务一：Hadoop 全分布部署管理

本环节需要使用root用户完成相关配置，安装Hadoop需要配置前置环境。命令中要求使用绝对路径，具体部署要求如下:

1. 将/chinaskills下的JDK包解压到/usr/local/src路径，将完整命令复制粘贴到对应报告中;
2. 修改/root/.bash\_profile文件，设置JDK环境变量，并使环境变量只对当前root用户生效将环境变量配置内容复制粘贴至对应报告中;
3. 从master复制上面步骤配置的JDK环境变量文件到slave1、slave2节点，命令和结果复制粘贴在对应报告中;
4. 配置SSH密钥登录，实现从master登录到slave1，将登录命令和结果复制粘贴在对应报告中;
5. 将配置文件hadoop-env.sh变更内容复制粘贴在对应报告中；
6. 将配置文件core-site.xml变更内容复制粘贴在对应报告中；
7. 初始化Hadoop环境namenode，将命令及结果复制粘贴在对应报告中；
8. 查看master及slave1节点jps进程，将其命令及结果复制粘贴至对应报告中。

### 任务二：Flume部署管理

1. 设置Flume环境变量，并使环境变量只对当前root用户生效，将变量内容复制并粘贴到对应报告中。
2. 修改并配置flume-env.sh文件，将修改内容复制并粘贴到对应报告中。
3. 查看Flume启动的conf-file文件内容，将内容复制并粘贴到对应报告中；
4. 启动Flume传输Hadoop日志，将Flume启动命令以及查看HDFS上生成的Hadoop日志文件结果复制并粘贴到对应报告中（若文件数量大于10则只截取10条）。并查看HDFS中/tmp/flume目录下生成了文件，将内容复制并粘贴到对应报告中。

### 任务三：Zookeeper部署管理

1. 解压Zookeeper安装包到“/usr/local/src”路径，并修改解压后文件夹名为zookeeper，将修改命令及结果复制粘贴至对应报告中；
2. 设置ZOOKEEPER\_HOME环境变量，并使环境变量只对当前用户生效，命令及环境变量内容复制粘贴至对应报告中；
3. 配置“zoo.cfg”配置文件，将文件变更内容复制粘贴至对应报告中；
4. 修改myid配置文件，将文件变更内容复制粘贴至对应报告中；
5. 启动每个虚拟机上的Zookeeper节点，启动完成之后查看每个节点的zkServer服务状态，命令及结果复制粘贴至对应报告中。

# 模块B：数据采集与处理 （20分）

**项目背景说明**

1. 查看餐饮外送统计平台网站源码结构。
2. 打开网站，在网页中右键点击检查，或者F12快捷键，查看源码页面；
3. 检查网站：浏览网站源码查看所需内容。
4. 从餐饮外送统计平台中采集需要数据，按照要求使用Python语言编写代码工程，获取指定数据项，并对结果数据集进行必要的数据处理。请将符合任务要求的结果复制粘贴至对应报告中。

具体步骤如下：

1. 创建工程项目：C:\food\_delivery\_crawl
2. 构建采集请求
3. 按要求定义相关字段
4. 获取有效数据
5. 将获取到的数据保存到指定位置
6. 对数据集进行基础的数据处理

至此已从餐饮外送统计平台中获取所需数据，并完成了必要的基础的数据处理。

1. 自行创建Scrapy工程项目food\_delivery\_crawl，路径为C:\ food\_delivery\_crawl按照任务要求从餐饮外送统计平台中获取数据。
2. 每条数据记录请以单独一行保存，信息存储格式为key：value。文件保存路径为：C:\output。

示例：

{" rest\_id ": "\*\*\*", " rest\_name ": "\*\*\*",……}，

{" rest\_id ": "\*\*\*", " rest\_name ": "\*\*\*",……}，

……

### 任务一：爬取配送平台数据

自行创建Scrapy工程编写爬虫代码，爬取“配送平台数据”页面相关数据，通过爬虫代码分页爬取，将使用re解析分页链接的程序源代码复制粘贴至对应报告中。

### 任务二：爬取店铺运营数据

自行创建Scrapy工程编写爬虫代码，爬取“店铺运营数据”页面相关数据，通过爬虫代码分页爬取，将使用re解析分页链接的程序源代码复制粘贴至对应报告中。

### 任务三：爬取指定文件

运行代码，爬取“配送平台数据”与“店铺运营数据”页面相关数据分别至range\_audited.json、grey\_value.json文件。查看文件并填写采集到的记录条数，并将答案复制粘贴至对应报告中。

示例格式：

range\_audited.json行数为：

grey\_test.json行数为：

### 任务四：属性判断

审查爬取的range\_audited数据，判断属性“申请递交时间”与“申请时间”是否为重复属性。如果为重复属性，则删除“申请时间”，并输出前10条数据记录；如果不是重复属性，请输出数据集中数值不相同的记录条数。并将答案复制粘贴至对应报告中。

### 任务五：数据探索

针对爬取的grey\_test数据，利用DataFrame.describe方法探索数据基本情况，将输出结果复制粘贴至对应报告中。

### 任务六：缺失值统计

根据任务4的输出结果，grey\_test数据中哪一个属性缺失值最多？请将正确答案复制粘贴至对应报告中。

示例格式：

缺失值最多的属性为：

# 模块C：数据清洗与挖掘分析（25分）

**项目背景说明**

餐饮外卖平台的核心价值体现在配送，而配送的价值则依赖于商家与客户的双向选择。外卖平台通常会通过内容激活消费者和商家两个群体的活跃度。消费者会参考平台展示的内容选择商家，商家也会以消费者评价与平台统计数据为依据调整策略，由此再吸引更多的用户下单、评论、形成正向循环。保证配送的时效与品质是从优化用户体验的角度，吸引更多的用户参与，进而带动商家不断入驻。由此，商家、消费者、骑手在平台上形成越来越多的真实可靠的数据，帮助消费者更好的做出消费决策，同时促进商家提高服务质量。而平台通过数据，不断调整优化服务，从而不断提升这种多边网络效应。提升网络效应的直接结果就是用户和商家规模大幅提升，进而形成规模效应——降低获客成本、提高效益，并且不断提升自己的行业壁垒。

为探索各大外卖平台的市场策略与经营模式，现从及平台获取到了原始数据集，为保障用户隐私和行业敏感信息，已经对数据脱敏。数据脱敏是指对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护。在涉及客户安全数据或一些商业性敏感数据的情况下，对真实数据进行改造并提供测试使用，如身份证号、手机号等个人敏感信息都需要进行数据脱敏。本题已将脱敏后的数据存放于“C:\数据源”。

### 任务一：数据清洗

**任务背景：**

数据源为众多网站及平台的数据汇总，且为多次采集的结果，在整合多来源数据时可能遇到数据重复，或数据拼接导致的属性列缺失或冗余等情况。请根据任务具体参数要求，针对原始数据集进行清洗，并写入指定的数据库或数据文件，复制并保存结果。

**任务描述：**

数据源文件存放于路径“C:\数据源”，请按照如下要求编写Python程序对数据进行清洗，并将结果保存至对应报告中。

1. 分析“C:\数据源”路径中json格式数据文件。
2. 针对属性列“商家id”排查并删除异常数据条目。
3. 针对属性列缺失或冗余的样本进行删除。
4. 运行python程序，保存清洗后的结果数据集，并将答案复制粘贴至对应报告中。

**具体任务要求**：

1、编写python程序读取相关数据文件，包含“city\_name，location，latitude，longtitude，rest\_type，platform\_A\_restid，A\_rst\_name，A\_day\_30\_cnt，platform\_B\_restid，B\_rst\_name，B\_day\_30\_cnt”字段，在PyCharm控制台打印输出数据集样本条数。将打印语句复制粘贴至对应报告中。

示例格式：

===数据集初始样本条数为\*\*\*条===

2、针对属性列“platform\_A\_restid”或“platform\_B\_restid”重复的样本，请以多条样本记录的“A\_day\_30\_cnt”或“B\_day\_30\_cnt”属性均值作为该属性的值，并删除多余样本。请在PyCharm控制台打印输出删除的样本条数。将打印语句复制粘贴至对应报告中。

示例格式：

===因重复样本记录，删除样本条数为\*\*\*条===

3、编写python程序，删除数据源中缺失值大于3个字段的数据记录，打印输出删除条目数，将运行结果复制粘贴至对应报告中

打印语句格式如下：

=== “删除缺失值大于3个的字段的数据条数为\*\*\*条”===

4、数据记录以[latitude，longtitude]降序排列，保存至路径“C:\数据源“中，并命名为diliveryoutput1.json。通过Python程序查看文件前5条记录。将程序输出复制粘贴至对应报告中。

### 任务二：数据挖掘分析

**子任务1**

**任务背景：**

聚类分析又称群分析，它是研究分类问题的一种统计分析方法，同时也是数据挖掘的一个重要算法。聚类分析是由若干模式组成的。通常，模式是一个度量的向量，或者是多维空间中的一个点。聚类分析以相似性为基础，同一个聚类簇中的模式之间具有相似性，不同聚类簇之间具有相异性。在商业上，聚类可以帮助平台市场分析人员对一定地理范围内的商家进行合理的管辖区域划分，或在经营数据中区分出不同的商家群体，并提取每一类商家的经营模式。它作为数据挖掘中的一个模块，可以作为一个单独的工具以发现数据中的深层的信息，并且提取出每一类样本的特点，或者把注意力放在某一个特定的类上以作进一步的分析；同时，聚类分析也可以作为数据挖掘算法中其他分析算法的一个预处理步骤。本题数据请采用任务数据清洗的输出文件diliveryoutput1.csv。

聚类分析成功地应用于心理、经济、社会、管理、医学、地质、生态、地震、气象、考古、企业决策等方面，同时也是划分市场的有效工具，可用于寻找新的潜在市场、选择实验的市场，并且可作为多元分析的预处理。由于加入外卖平台的商家数量急剧上升，现ChinaSkill公司希望对北京地区各个商圈商家再次进行汇总划分，请根据各商家的地理位置（经纬度）进行聚类统计，并将结果保存至对应报告中。

**任务描述**：

请以数据清洗任务结果数据集diliveryoutput1.json作为输入数据源，按照如下要求Python程序实现对数据的分析挖掘，并将结果保存。

1. 解析diliveryoutput1.json文件。
2. 提取商家相关属性数据。
3. 针对商家地理位置进行聚类划分。
4. 查看聚类结果。

**具体任务要求**：

1、读取diliveryoutput1.json，抽取北京地区商家数据记录，查看数据记录条数。请在PyCharm控制台打印输出样本条数，并将打印语句复制粘贴至对应报告中。

示例格式：

===北京地区样本条数为\*\*\*条===

2、根据北京地区商家的经纬度属性，对商家进行k-means聚类，聚类数设为5，迭代次数为2000次，请在PyCharm控制台以打印语句输出聚类中心、每个类的商家数，以及该类所包含的商圈，并将打印语句复制粘贴至对应报告中。

示例格式：

=cluster 0: 聚类中心为[\*\*，\*\*]，商家数为\*\*\*个，包含商圈：【\*，\*】=

=cluster 0: 聚类中心为[\*\*，\*\*]，商家数为\*\*\*个，包含商圈：【\*，\*】=

……

**子任务2**

**任务背景：**

灰度测试起源于软件测试策略，指在同一个时间段内，存在两个不同的应用版本，一个版本叫做黑色版本，而另一个版本叫做白色版本。通过观测两个同时存在的版本的表现来调整黑色版本和白色版本的比例，如果一切顺利，渐渐地把所有用户的应用从黑色版本过渡到白色版本。而这种通过共存黑白版本的手段进行测试的过程就叫做灰度测试或灰度发布。

将灰度测试应用于商业中，可以帮助市场管理团队快速试验并发现问题，在新的市场策略大规模推向全部区域用户之前，及时修正问题，在极大程度上减少了不必要的风险。因此灰度测试对商业决策是非常必要的。而小规模的灰度测试，不但满足了一部分人抢先体验的愿望，同时也可以发现新决策中不容易发现的各种问题，并收集到真正的用户体验。现ChinaSkill公司为估算市场容量、探求外卖平台运力负荷拐点，进行了一轮灰度测试，请根据灰度测试前后的数据，完成任务要求的相关分析统计，并将结果保存至对应报告中。

**任务描述**：

数据源文件存放于路径C:\数据源，请按照如下要求编写Python程序对数据进行分析，并将结果保存至对应报告中。

1. 分析“C:\数据源”路径中json格式数据文件。
2. 针对指定属性列灰度测试前后的数据进行对比分析。
3. 运行python程序，输出分析结果，并将答案复制粘贴至对应报告中。

**具体任务要求**：

1、编写python程序读取相关数据文件中：网格id，推单数9，有效完成率9，超时率9，推单数8，有效完成率8，超时率8字段，字段分别表示灰度测试对象的网格id，以及灰度测试前（某月8日）、后（某月9日）的平台外卖销售表现。请在PyCharm控制台以打印语句输出数据记录条数，并将打印语句复制粘贴至对应报告中。

示例格式：

===文件数据记录条数为\*\*\*条===

2、请统计灰度测试前后，平台推单数(该网格下所有的推单数总和)增加10%以上的网格，并以增加幅度降序排列。请在PyCharm控制台以打印语句输出增加幅度最高的5个网格相关信息，并将打印语句复制粘贴至对应报告中。

示例格式：

===1. 网格id：\*\*\*， 推单数增幅\*\*\*%===

===2. 网格id：\*\*\*， 推单数增幅\*\*\*%===

……

3、请统计灰度测试前后，有效完成率没有降低、或降低不超过3%，同时灰度测试后有效完成率保持在95%及以上的网格信息。请在PyCharm控制台以打印语句输出符合条件的网格数，及该结果在总网格数中的占比，并将打印语句复制粘贴至对应报告中。

示例格式：

===符合条件的网格数为\*\*个， 在总网格数中占比\*\*\*%===

# 模块D：数据可视化（20分）

MySQL数据库中的相关数据集包含了城市、地点、商家id、网格id、餐品种类、标品属性等多项基础信息字段。请使用Flask框架，结合Echarts完成下列任务。

数据库账号: root 密码：123456

自行创建代码工程路径为C:\food\_delivery

### 任务一：双折线图呈现平台销量

**任务背景：**

市场份额亦称“市场占有率”。指某企业的销售量（或销售额）在市场同类品类中所占比重。反映企业在市场上的地位。通常市场份额越高，竞争力越强。市场占有率一般有3种基本测算方法：（1)总体市场份额，指某企业销售量在整个行业中所占比重。（2)目标市场份额，指某企业销售量在其目标市场，即其所服务的市场中所占比重。（3)相对市场份额，指某企业销售量与市场上最大竞争者销售量之比，若高于1，表明该企业其为这一市场的领导者。请按任务指定要求，输出不同平台商家销售分析相关图例。

**任务描述：**

请根据数据库中相关数据表的数据，分别统计A平台与B平台30天销量最高的10个商家的销量，并以共享y轴的双折线图呈现。

**具体任务要求：**

1、根据相关数据表中city\_name，location，latitude，longitude，rest\_type，platform\_A\_restid，A\_rst\_name，A\_day\_30\_cnt，platform\_B\_restid，B\_rst\_name，B\_day\_30\_cnt等字段，分别统计A平台与B平台30天销量最高的10个商家及销量，在PyCharm控制台按照“30天销量”降序打印输出商家id，商家所属平台，及30天销量。

示例格式：

==1: “\*\*\*\*”，Platform-A, 销量为\*\*\*===

==2: “\*\*\*\*”，Platform-A, 销量为\*\*\*===

……

==10: “\*\*\*\*”，Platform-A, 销量为\*\*\*===

==1: “\*\*\*\*”，Platform-B, 销量为\*\*\*===

==2: “\*\*\*\*”，Platform-B, 销量为\*\*\*===

……

==10: “\*\*\*\*”，Platform-B, 商家数为\*\*\*个===

2、使用Flask框架，结合Echarts，完成可视化输出。要求图标题为“各平台30天销量最高10大商家”,横坐标1(下方)为平台A商家id，横坐标2(下方)为平台B商家id（倾斜显示、互不遮挡），纵坐标为商家销量，以销量降序排列，红色折线标识Platform-A商家，蓝色折线标识Platform-B商家。将可视化结果截图并保存（截图需包含浏览器地址栏）。

### 任务二：气泡图呈现各商圈的商家数量

**任务背景：**

商圈，指某商场以其所在地为原点，沿着一定的方向和距离扩展，吸引顾客的辐射范围。简单地说，就是来店顾客所居住或工作的区域范围。无论餐厅规模大小，其销售覆盖区域总是有一定的地理范围。这个地理范围就是以某商场为中心，向四周辐射到可能来店消费的顾客居住地或工作地。请按任务指定要求，输出相关图例。

**任务描述：**

请根据数据库相关数据集中city\_name，location，latitude，longitude，rest\_type，platform\_A\_restid，A\_rst\_name，A\_day\_30\_cnt，platform\_B\_restid，B\_rst\_name，B\_day\_30\_cnt等字段，明晰地理位置与商铺聚集程度之间的关系。请以经度为横坐标，纬度为纵坐标，绘制商家数量气泡图，并以该地理位置的商家数量/10作为气泡半径。

**具体任务要求：**

1、提取表格相关字段，在控制台按照“商家数量”降序排列，打印输出商圈名称及包含的商家数量。

示例格式：

==1: 商圈 \*\*\*\*=商家数为\*\*\*个===

==2: 商圈 \*\*\*\*=商家数为\*\*\*个===

……

2、使用Flask框架，结合Echarts，完成气泡图输出。要求标题为“商家聚集地理位置展示”，横坐标为经度，纵坐标为维度，以该地理位置的商家数量/10作为气泡半径，绘制气泡图。将可视化结果截图并保存（截图需包含浏览器地址栏）。

### 任务三：柱状图呈现A平台商家分布

**任务背景：**

外卖平台为更多消费者和商户提供了数字化的生活方式和经营模式，推动供需两端稳步增长。同时数字化趋势也将带来更多机会，外卖行业是具备成长性和长远回报的新业务领域，需要坚持长期积极投入。城市平台签约商户数量直观地体现了该平台在一定区域内的投入力度和发展现状。请按任务指定要求，输出不同城市商家统计相关图例。

**任务描述：**

请根据相关数据表中的数据，统计A平台在各城市的商家数量。

**具体任务要求：**

1、根据相关数据表中city\_ name，location，latitude，longitude，rest\_type，platform\_A\_restid，A\_rst\_name，A\_day\_30\_cnt，platform\_B\_restid，B\_rst\_name，B\_day\_30\_cnt等字段，统计每个城市所包含的A平台商家数量，在控制台按商家数量降序，打印输出前10个城市名称及对应的商家数量。

示例格式：

==1. 城市:\*\*\*，商家数量为\*\*\*===

==2. 城市:\*\*\*，商家数量为\*\*\*===

……

2、使用Flask框架，结合Echarts，完成柱状图输出。要求标题为“平台A商家数量排名前十座”,横坐标为城市名称，纵坐标为该城市平台A商家数量，以商家数量降序排列。将可视化结果截图并保存（截图需包含浏览器地址栏）。

### 任务四：柱状堆叠图呈现审核通过比例

**任务背景：**

外卖的配送范围一般由外卖平台的业务人员为商家设定，商家也可以根据实际情况向平台申请修改配送范围。通常配送范围并不是毫无棱角的圆形，因为这种划分并没有因地制宜，是不合理的。最大配送距离指的是配送范围中距离商家最远的直线距离，最小配送距离则是配送范围中距离商家最近的直线距离。在申请配送范围修改时，审核员将考虑实际道路限制、综合运力成本、顾客等待时长等因素，对配送范围变更的申请进行考量，最终做出通过/拒绝/驳回的结果判定。请根据相关数据集，按任务指定要求，输出范围审核相关分析图例。

**任务描述：**

请根据数据库相关数据表中的数据，统计审核通过与拒绝的范围申请记录数量，并以柱状堆叠图表达。

**具体任务要求：**

1、根据相关表格中“id，Request\_id，Walle\_id，Retailer\_id，retailer\_name，retailer\_address，etailer\_location，City\_id，City\_name，Grid\_id，Carrier\_id，Team\_id，Applicant\_id，Applicant\_name，first\_auditor\_role，first\_auditor\_candidate\_ids，first\_auditor\_id，first\_auditor\_name，second\_auditor\_role，second\_auditor\_candidate\_ids，second\_auditor\_id，second\_auditor\_name，status，max\_distance\_before\_edit，min\_distance\_before\_edit，max\_distance\_after\_edit，min\_distance\_after\_edit，area\_before\_edit，area\_after\_edit，created\_at，updated\_at，申请时间，创建时间”等字段，汇总各二级审核人审批的申请数量，并统计其中通过与拒绝的条数。在PyCharm控制台打印输出审核记录数最多的10位审核人id，审批总数，以及其中通过与拒绝的条数，按审批总数降序排列。

示例格式：

==1.二级审核人id:\*\*\*, 审批数量：\*\*\*条，其中通过\*\*条，拒绝\*\*条===

==2.二级审核人id:\*\*\*, 审批数量：\*\*\*条，其中通过\*\*条，拒绝\*\*条===

……

2、使用Flask框架，结合Echarts，完成柱状堆叠图输出。要求标题为“审核通过与拒绝对比”，横坐标为审核人id，纵坐标为审核记录数。画出审核数量最多的10位二级审核人记录审批情况，通过记录数蓝（蓝色表示）上方叠加拒绝和驳回记录数（红色表示），以审批总数降序排列。将可视化结果截图并保存（截图需包含浏览器地址栏）。

### 任务五：散点地图呈现各城市商家数量分布

**任务背景：**

近年来，我国外卖行业发展迅速，互联网餐饮外卖市场规模逐渐扩大，外卖平台已覆盖全国所有省份。2019全年外卖交易超7274亿元，截至2020年3月，我国网上外卖及手机网上外卖用户渗透率已达44%。2020年底，全国外卖总体订单量已超过171.2亿单。为纵观全国范围内外卖平台签约商家分布情况，请根据相关数据集，按任务指定要求，输出相关分析图例。

**任务描述：**

请根据相关数据表中的数据，统计不同城市商家数量，并散点地图呈现。

**具体任务要求：**

1、根据相关表格，参考字段同可视化3、（任务数字编号），统计不同城市商家数量。请在PyCharm控制台以商家数量**降序**打印输出城市名称，商家数量。

示例格式：

==1.城市: \*\*\*，商家\*\*\*个===

==2.城市: \*\*\*，商家\*\*\*个===

……

2、使用Flask框架，结合Echarts，完成散点地图输出。要求标题为“各城市商家数量散点地图”。将可视化结果截图并保存（截图需包含浏览器地址栏）。

### 任务六：条形图呈现各网格商家数量分布

**任务背景：**

网格化营销是近年来新兴的一种营销管理模式，已被广泛地运用于市场精准营销。在市场营销中，采取地图营销、网格管理、精准策略等，可将客户的心理需求与其日常生活紧密地联系起来，巧妙运用网格化管理和营销地图，以此来实现客户的营销精准度与价值提升，提高单一客户贡献值和营销效率。其核心旨在帮助企业快速响应市场需求，为用户提供更加便捷的、专业化服务，并使企业的资源分配以客户为中心，以市场变化为导向，固本强基、开拓市场、提升效益，有助于平台有效地制定绩效战略。请根据相关数据集，按任务指定要求，输出相关分析图例。

**任务描述：**

请根据数据库相关数据表中的数据，统计不同网格的商家数量，并以条形图表达。

**具体任务要求：**

1、根据相关表格中“id，Request\_id，Walle\_id，Retailer\_id，retailer\_name，retailer\_address，etailer\_location，City\_id，City\_name，Grid\_id，Carrier\_id，Team\_id，Applicant\_id，Applicant\_name，first\_auditor\_role，first\_auditor\_candidate\_ids，first\_auditor\_id，first\_auditor\_name，second\_auditor\_role，second\_auditor\_candidate\_ids，second\_auditor\_id，second\_auditor\_name，status，max\_distance\_before\_edit，min\_distance\_before\_edit，max\_distance\_after\_edit，min\_distance\_after\_edit，area\_before\_edit，area\_after\_edit，created\_at，updated\_at，申请时间，创建时间”等字段，统计北京地区不同网格所包含的商铺数。在控制台以商铺数降序打印输出前10位网格id，网格名称，商铺数。

示例格式：

==1 网格id：1111，网格名称：\*\*\*\*，商铺数为\*\*===

==2 网格id：1111，网格名称：\*\*\*\*，商铺数为\*\*===

……

2、使用Flask框架，结合Echarts，完成条形图输出。输出商铺数前10位的网格名称及商铺数量，其中标题为“商家数量Top10网格”，横坐标为商家数量，纵坐标为网格名称。将可视化结果截图并保存（截图需包含浏览器地址栏）。

# 模块E：综合分析（15分）

通过模块B的网站分析及数据爬取、模块C的数据清洗与挖掘分析及模块D的数据可视化呈现，我们已经清晰的了解了餐饮外卖平台业务背景及相关数据，在综合理解外卖业务数据的基础上，根据任务要求进行分析，并编写分析报告。

请根据任务要求，分析以下内容，并编写分析报告。分析影响商家申请的范围变更审核结果的因素，以及全国范围内商家分布情况，并对平台经营提出几点建议。

分析报告要求：

### 任务一：分析范围变化对审核结果的影响

结合平台相关数据文件，以折线图说明申请前后的范围变化对审核结果有怎样的影响？当“申请前/后最大配送距离”或“申请前/后最小配送距离”变更幅度（下文简称为“申请前/后距离变更幅度”）为45%，55%，65%，75%，85%，95%时，审核记录的通过率分别是多少？请以折线图表达，横坐标为申请前/后距离变更幅度，纵坐标为审核记录通过率，并以文字说明折线图的含义。

### 任务二：外卖平台推广分析

结合模块D可视化分析中对各城市商家数量的统计结果，外卖平台在全国范围内的推广情况。分别以文字描述和图例进行说明。

### 任务三：平台规划建议

请结合平台业务背景及相关分析结论，对平台未来规划提出建议（不少于3条建议）。

# 附录：补充说明

数据集中涉及字段及中文说明

|  |  |
| --- | --- |
| 原字段 | 中文释义 |
| city\_name | 城市 |
| location | 商圈 |
| latitude | 纬度 |
| longitude | 经度 |
| rest\_type | 商家所属平台 |
| platform\_A\_restid | 商家id（A平台） |
| A\_rst\_name | 店铺名称（A平台） |
| A\_day\_30\_cnt | 30天销量（A平台） |
| platform\_B\_restid | 商家id（B平台） |
| B\_rst\_name | 店铺名称（B平台） |
| B\_day\_30\_cnt | 30天销量（B平台） |
| 推单数-9 | 推单数（9日） |
| 有效完成率-9 | 有效完成率（9日） |
| 超时率-9 | 超时率（9日） |
| 推单数-8 | 推单数（8日） |
| 有效完成率-8 | 有效完成率（8日） |
| 超时率-8 | 超时率（8日） |
| 灰度餐厅 | 灰度餐厅 |
| id | 标识id |
| Request\_id | 请求id |
| Walle\_id | （平台）商铺id |
| Retailer\_id | （配送系统）商铺id |
| retailer\_name | 商铺名称 |
| retailer\_address | 商铺地址 |
| retailer\_location | 位置POI编码 |
| City\_id | 城市id |
| City\_name | 城市名称 |
| Grid\_id | 网格id |
| Carrier\_id | 渠道经理 |
| Team\_id | 渠道小组 |
| Applicant\_id | 申请人id |
| Applicant\_name | 申请人 |
| first\_auditor\_role | 一级审批角色 |
| first\_auditor\_candidate\_ids | 候选审批人id |
| first\_auditor\_id | 审核人id |
| first\_auditor\_name | 审批人名 |
| second\_auditor\_role | 二级审批角色 |
| second\_auditor\_candidate\_ids | 审核小组成员 |
| second\_auditor\_id | 二级审核人id |
| second\_auditor\_name | 二级审核人名 |
| status | 申请状态 |
| max\_distance\_before\_edit | 申请前最大配送距离 |
| min\_distance\_before\_edit | 申请前最小配送距离 |
| max\_distance\_after\_edit | 申请后最大配送距离 |
| min\_distance\_after\_edit | 申请后最小配送距离 |
| area\_before\_edit | 申请前配送面积 |
| area\_after\_edit | 申请前配送面积 |
| created\_at | 申请递交时间 |
| updated\_at | 审核完成 |
| 申请时间 | 申请时间 |
| 创建时间 | 创建时间 |
| 餐厅id | 餐厅id |
| 餐厅名称 | 餐厅名称 |
| 所属城市 | 所属城市 |
| 营业时长 | 营业时长 |
| 餐厅状态 | 餐厅状态 |
| 是否托管 | 是否托管 |
| 总单量 | 总单量 |
| 总单量增长率 | 总单量增长率 |
| 有效订单量 | 有效订单量 |
| 有效订单增长率 | 有效订单增长率 |
| 订单配送成功率 | 订单配送成功率 |
| 超时订单率 | 超时订单率 |
| 无效订单率 | 无效订单率 |
| 平均预计送达时长 | 平均预计送达时长 |
| 平均实际配送时长 | 平均实际配送时长 |
| 当前配送面积 | 当前配送面积 |
| 面积变更值 | 面积变更值 |
| 网格id | 网格id |
| 网格名称 | 网格名称 |
| 战营 | 战营 |
| 餐厅id | 餐厅id |
| retailer\_id | retailer\_id |
| 餐厅名称 | 餐厅名称 |
| 城市 | 城市 |
| 商户业务包 | 商户业务包 |
| 配送范围 | 配送范围 |
| 客单价 | 客单价 |
| 推单数 | 推单数 |
| 接单数 | 接单数 |
| 有效完成单数 | 有效完成单数 |
| 投诉率 | 投诉率 |
| 异常率 | 异常率 |
| 欺诈单数 | 欺诈单数 |
| 拒单数 | 拒单数 |
| 商户取消数 | 商户取消数 |
| 客户取消数 | 客户取消数 |
| 系统取消数 | 系统取消数 |
| 配送取消异常数 | 配送取消异常数 |
| 整体时长 | 整体时长 |
| 接单时长 | 接单时长 |
| 到店时长 | 到店时长 |
| 取餐时长 | 取餐时长 |
| 送达时长 | 送达时长 |
| 商户投诉数 | 商户投诉数 |
| 用户投诉数 | 用户投诉数 |
| 差评数 | 差评数 |
| 好评数 | 好评数 |
| 评价数 | 评价数 |
| 最远订单距离 | 最远订单距离 |
| 该订单整体时效 | 该订单整体时效 |
| 该订单接单时效 | 该订单接单时效 |
| 该订单到店时效 | 该订单到店时效 |
| 该订单取餐时效 | 该订单取餐时效 |
| 该订单送达时效 | 该订单送达时效 |
| 该订单评价 | 该订单评价 |
| 网格ID | 网格ID |
| 网格名称 | 网格名称 |
| 城市 | 城市 |
| 战团 | 战团 |
| 餐厅ID | 餐厅ID |
| 近7天平台单量 | 近7天平台单量 |
| 近7天推单 | 近7天推单 |
| 餐厅名 | 餐厅名 |
| 餐厅地址（取餐地址） | 餐厅地址（取餐地址） |
| 餐品种类 | 餐品种类 |
| 标品属性 | 标品属性 |
| 全推/选推 | 全推/选推 |