点评数据分析及推荐产品需求规格说明书

■ 目录导航

文档简介

目的

本(系列)文档主要定义点评数据分析及推荐产品的业务范围、业务场景和业务逻辑,以及基于业务需求的功能定义和页面描述。

范围

本(系列)文档的主要目的是清晰、有层次地阐述产品需求及业务逻辑,定义页面原型中各个模块的设计思路和数据关系。

优先级

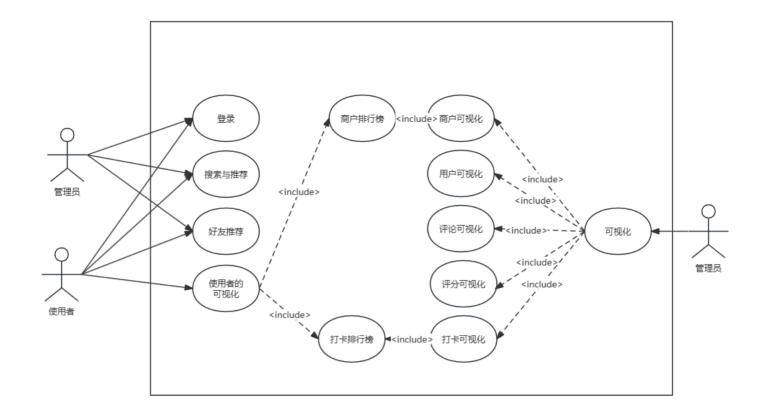
本文档主要读者为设计部门的UE、UI设计师,研发部门的前端工程师、后端工程师、算法工程师、测试工程师和项目部的QA工程师。

用户角色

角色定义

用户角色	用户描述	
使用者	查看部分模块模块,只有搜索与推荐,好友推荐和部分可视化界面,如排行榜	
管理员	查看所有模块内容,包括各种可视化界面,搜索与推荐,查看其他使用者的好友推荐	

角色用例



产品概述

业务场景

Yelp是美国著名商户点评网站,创立于2004年,囊括各地餐馆、购物中心、酒店、旅游等领域的商户,用户可以在Yelp网站中给商户打分,提交评论,交流购物体验等。在Yelp中搜索一个餐厅或者旅馆,能看到它的简要介绍以及网友的点论,点评者还会给出多少星级的评价,通常点评者都是亲身体验过该商户服务的消费者,评论大多形象细致。

本系统需要对Yelp开源的大数据进行数据分析及可视化,大数据应用开发。在Zeppelin大屏可视化数据分析,同时在VUE展示部分数据分析的可视化结果,完成用户推荐和搜索推荐功能。

系统架构

可视化界面为VUE和Zeppelin大屏可视化

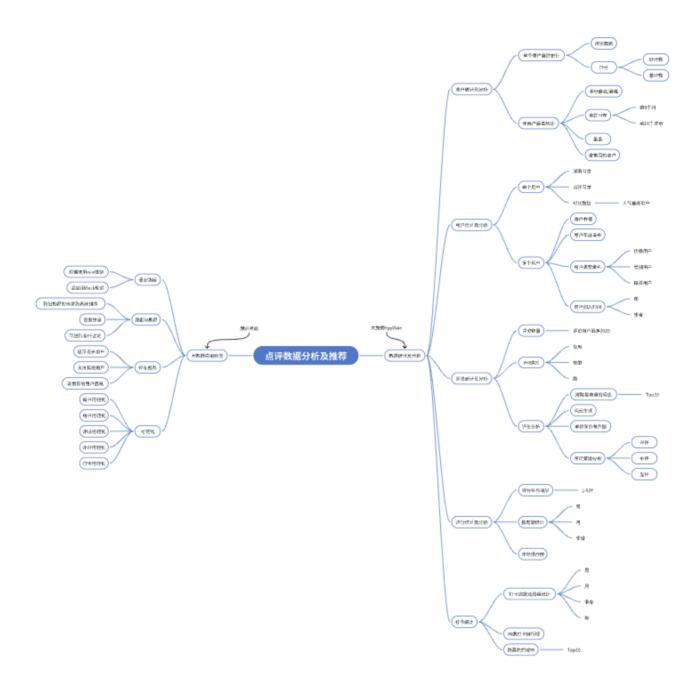
后端是flask

数据处理模块是Hadoop和Spark

数据库为集群Linux,Hive数据仓库

功能框架

■ 思维导图



功能需求规格

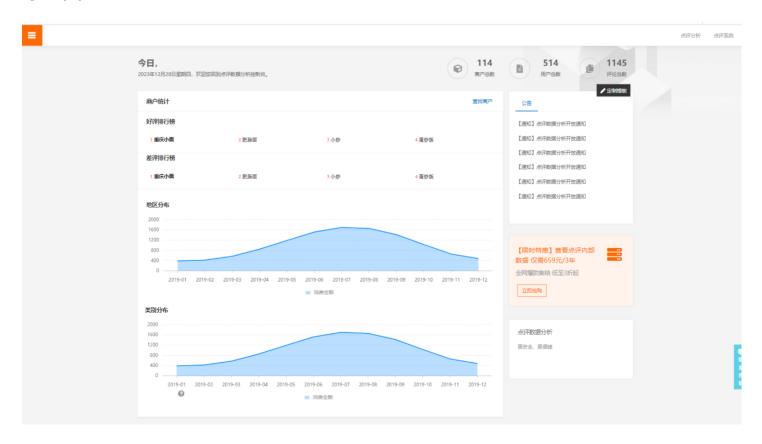
首页

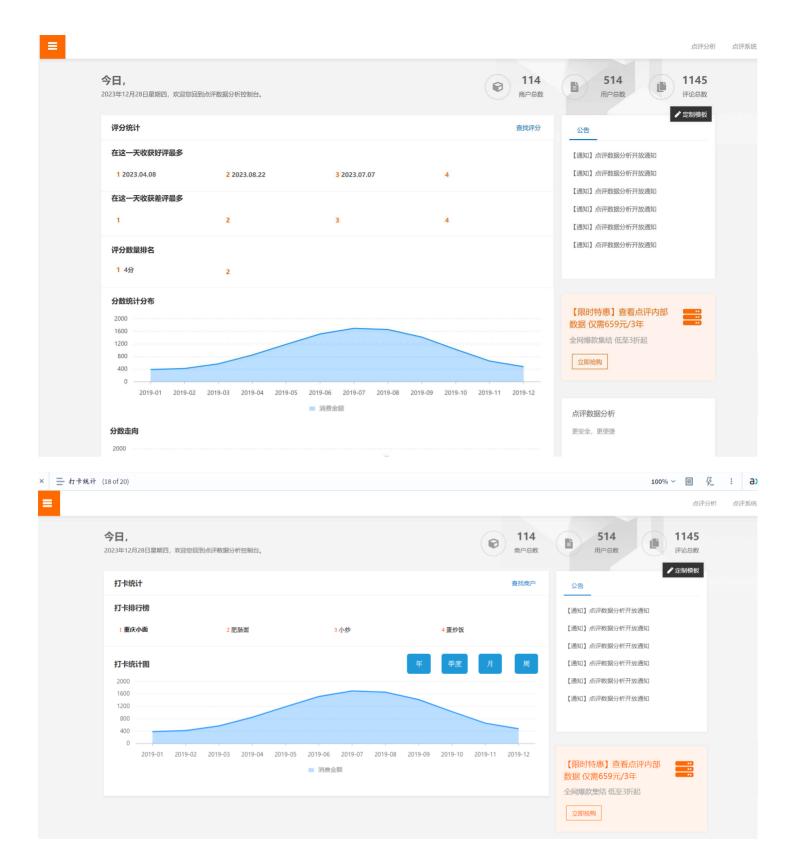


商户可视化

功能描述

单个商户属性统计,包括评论数目和评分等,按周期统计评分变化趋势和被打卡次数的变化趋势。 全商户属性统计,包括被打卡次数排行榜和评分排行榜(可以按地区分类)、评分分布情况、地区分布、商品类别分布等。

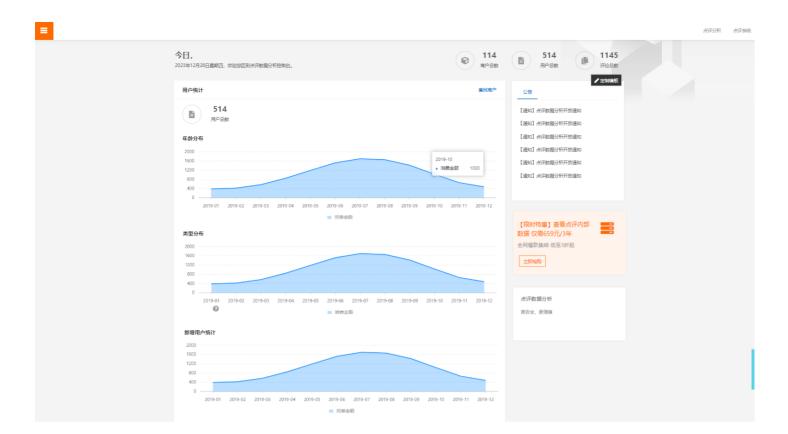




用户可视化

功能描述

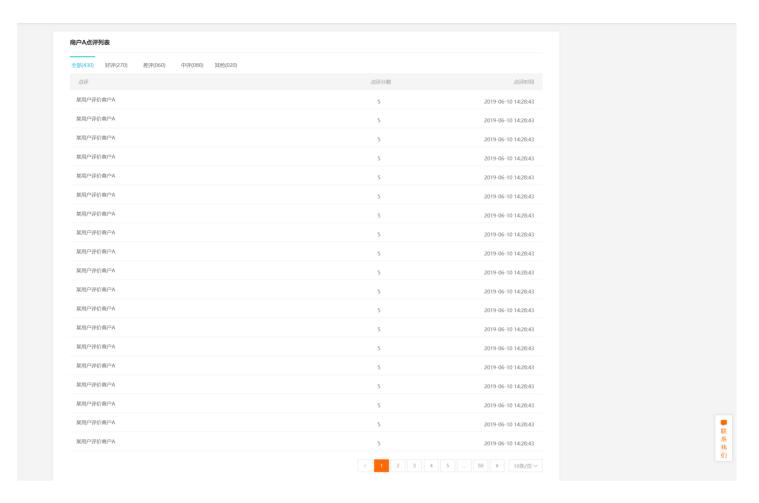
单个用户统计出简单的用户画像,包括点评消费习惯、粉丝数目等。全部用户统计,包括用户年龄分布、类型分布、加入时间分布等。



评论可视化

功能描述

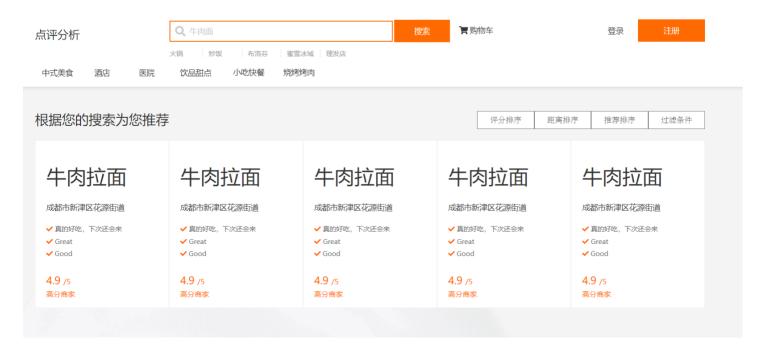
可以简要地查看用户对于商家的短评论和长评论。同时,提供筛选功能。



搜索与推荐

功能描述

获得用户的当前位置,根据点评推荐、距离、评分等条件对商家进行排序,将排序靠前的商家推荐给 用户,用户可以根据距离、评分、设施等对商家进行过滤,还可以点击商家查看商家的详细信息。可 显示推荐给用户的商家列表(页面顶部有过滤选项),商家详细信息。



好友推荐

功能描述

用户可以根据好友之间的关联性,发现有关联但还未添加的其他用户,可以将这些推荐用户添加为好友并查看其详细信息。可显示用户的当前好友列表,推荐用户列表,好友详细信息。

原型图

好友列表	
我的好友可能认识的人	
好友名称	加好友时间
上程数据	2019-06-10 14:28:43

非功能需求规格

性能

处理速度: 高性能是 Spark 的一个关键特点。非功能需求可能包括对于大规模数据集的快速处理和分析,以及对实时数据流的低延迟处理。

可伸缩性: 系统应该能够有效地处理不断增长的数据量,能够横向扩展以适应不同规模的工作负载。

容错性

故障容忍: 能够在节点故障或任务失败时保持系统的稳定性,不会导致整个分析任务失败。

数据恢复: 在发生故障时,系统能够快速地恢复并从存储中恢复数据。

安全性

数据安全: 对于敏感数据,确保数据在传输和存储过程中的安全性,可能需要加密和访问控制。

身份验证和授权: 保证只有授权用户能够访问和执行相关任务,实施合适的身份验证和授权机制。

可维护性

易于管理: 提供易于管理和监控的工具,包括日志记录、性能监测和故障排除功能。

可配置性: 允许管理员根据需要配置系统参数,以满足不同环境和要求。

可扩展性

支持插件和扩展: 允许集成新的数据源、算法或工具,以及轻松扩展系统的功能。

兼容性: 与大数据生态系统和工具(如Hadoop、Hive等)能够良好地集成。

用户友好性

易用性: 提供直观的用户界面和易于使用的 API,以简化用户的操作和开发任务。

文档和培训: 提供详细的文档和培训材料,以便用户能够更好地理解和使用系统。

资源利用效率

内存和计算资源管理: 有效地管理内存和计算资源,以最大程度地提高系统的利用率。

数据压缩: 在传输和存储中使用数据压缩,以减少网络带宽和存储空间的使用。