SiteSage Project Proposal

Guan Zheng Huang
Yaowen Liu
word-count: 1174

Introduction

Entrepreneurs in traditional markets often struggle with evaluating optimal business locations, a challenge that can significantly impact success. Expert analysis during site selection can mitigate these risks, but such services are often inconsistent, slow and expensive. SiteSage addresses this need by automating evaluation of user-provided locations based on their business description. It offers staged assessments and produces evaluation reports that enable comparative analysis across sites. SiteSage enables informed decision-making by replacing subjective judgments with data-driven insights, obtained through real-time, open-internet access.

Background

Store location selection is a critical determinant of retail performance, directly influencing customer accessibility, brand visibility, and often accounting for millions of dollars in annual revenue [1]. With the growing availability of geospatially affiliated data and technologies, such as google map and yelp reviews. Many aspects of site selection evolved from an experience-driven process into a data-driven discipline [2].

Despite these advancements, site selection remains a human-intensive process, requiring extensive data cleaning, criterion design, model construction, and expert interpretation. Decision-makers often tailor analytical criteria to specific industries with limited scalability and automation.

Recent research suggests that large language models (LLMs) demonstrate promising geospatial and logistical understanding ability required to conduct site selection [3]. Building on this potential, we hope an agentic solution can derive more adaptive and explainable decisions.

Constraints

- Geographical: The experiment is limited to Shanghai, China for ease of data collection.
- Target: Only coffee shops are analyzed in testing due to data limitation.
- Data Source: The agent can access public data and APIs only.
- Scope: Sole evaluating customer traffic. Cost of any kind is not considered due to lack of data and knowledge drift.

Data and Labels

General: Accessed by agent on the run.

- AMap LBS API: Including location encoding, nearby search, static map, etc. [4]
- Projected population, age and sex data in China: Downloaded from https://hub.worldpop.org/
- Open Internet: All public internet non-structured data.

Evaluation: Ground-truth evaluation.

- Rating information for coffee shops from Dianping. [5]

Architecture of the system

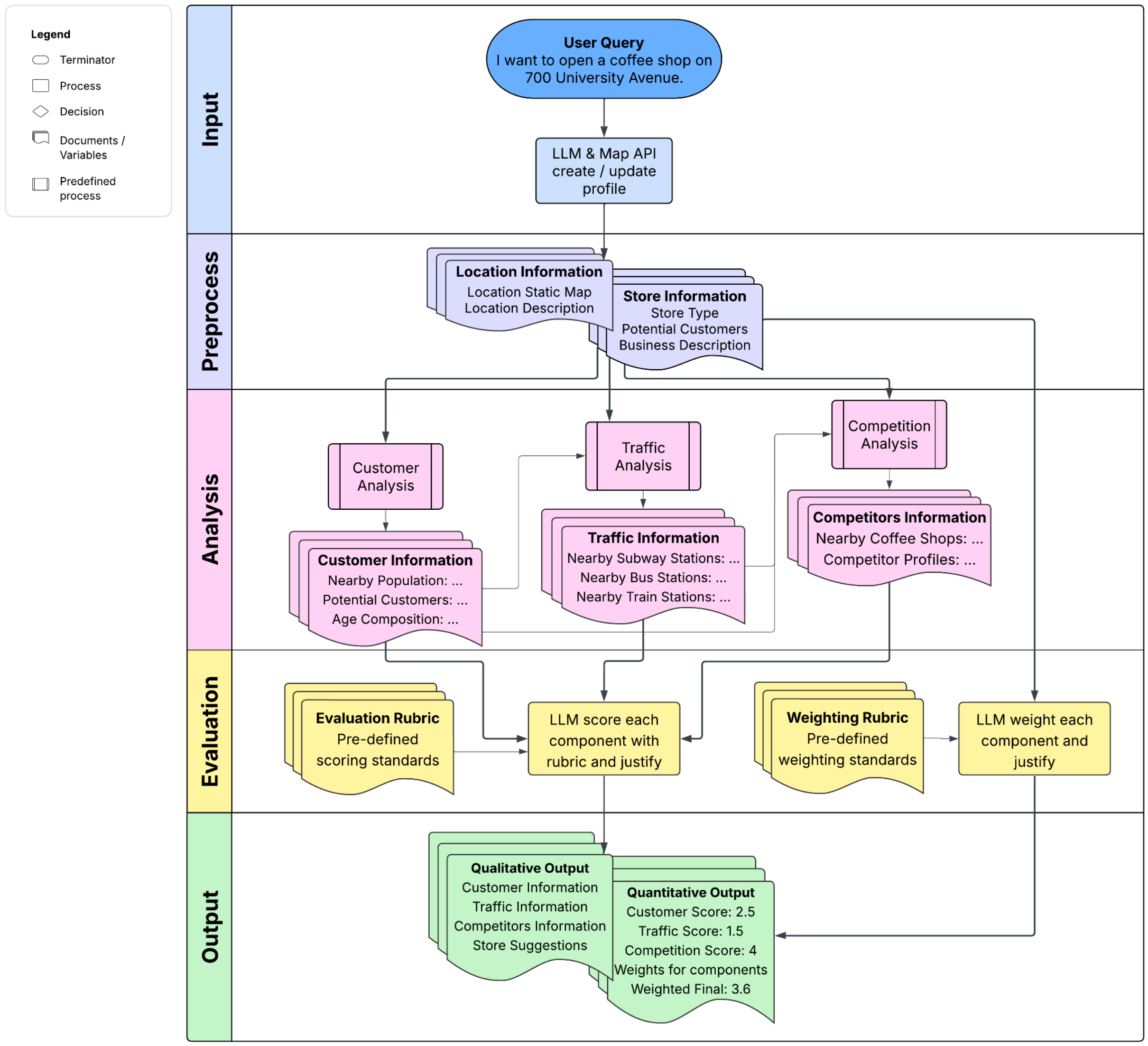


Figure 1. Workflow of SiteSage

As shown in Figure 1, SiteSage begins with a **user query** specifying a candidate location and store concept (e.g., “Open a coffee shop at 700 University Ave”). The tooled **LLM** (illustrated in Figure 2, with only map API access) parses the query and generate structured data in two-part:

- **Location Information**: static map and location context representing general knowledge around the site
- **Store Information**: store type, possible customers, and business description

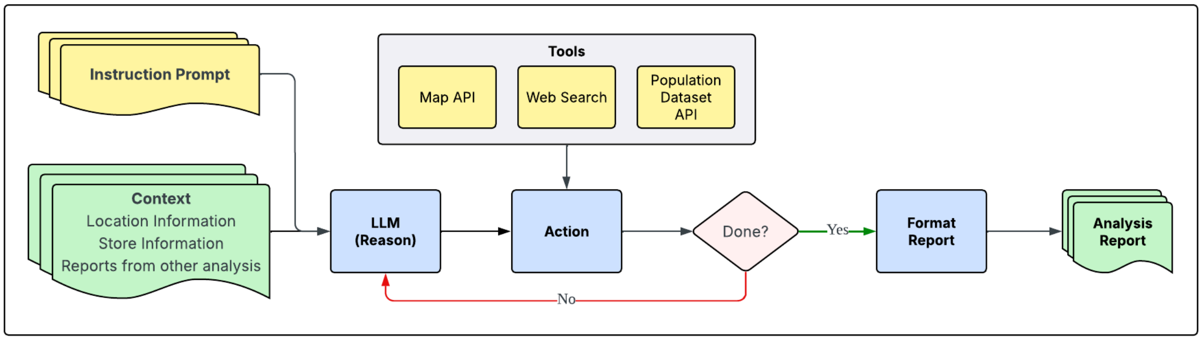


Figure 2. ReAct (Reasoning and Acting) pipeline in analysis modules

The system proceeds through three connected analysis modules: **Customer, Traffic,** and **Competition** (please reference Design Justification). Each follows the ReAct paradigm (Figure 2) to produce domain-specific facts, such as nearby population statistic, transit accessibility, and competitor density using map services, population datasets, and web searches. The modules are contextually linked: results from customer analysis are required for traffic analysis, and both feed into competition analysis. Each module operate independently based on report produced by the previous module. This design reduce context load, improve focus and allow more granular prompt control.

$$final\ score = a*customer\ score + b*traffic\ score + c*competition\ score$$

Formula 1. Calculating final score

In the location **Evaluation** stage, a predefined **scoring rubric** guides the LLM to convert the finding from each module into separate **area scores** with justifications. A separate **weight designing rubric** allows the LLM to derive weights from the store description. For example, assigning lower weight to traffic factors for a take-out store. A final score is achieved by weighted suming the area-wise score with weight (Formula 1).

Finally, the system outputs a two-part report:

(i) **Qualitative** section: explains the customer, traffic, and competition contexts with key findings. As well as reason for weights.

(ii) **Quantitative** section: presents numerical scores for each dimension, enabling cross-location comparison.

Design Justification

The design is inspired by Huff's gravity model [6], which defines a location's attractiveness based on its **endogenous** attributes (e.g., floor area) and **exogenous** context (e.g., traffic, competition), as well as the distance between customers and the site. To practice the idea, during the analysis, we designed 3 area to analyze:

- **Customer Information**: Potential customers, sources, geographic distribution, and age/gender stratification.
- **Traffic Information**: Nearby parking, public transport stations, traffic nodes, flow intensity, visiting cost, and store visibility.
- **Competition Information**: Nearby competitors, their profiles, and customers' visiting costs.

Metric of Success

To evaluate SiteSage's performance, we adopt a **comparative visitor-traffic ranking** approach. The system's predictions are assessed based on its ability to correctly determine which of two candidate locations will attract higher visitor traffic. For ground truth, we use the ratio of online store reviews within the past 30 days as a proxy for visitor volume. This ratio is compared with the corresponding ratio calculated from SiteSage's predicted scores for the two site. It should be noted the number of reviews can be influenced by external factors in many ways (ie: promotions, target customers). We will select similar size store with similar ratings to reduce bias. We deem the prototype successful if it achieves a result better than random guess, with the following quantitative targets:

- **Accuracy > 0.5**: ratio of pairs successfully ranked
- **MAE <= 0.33**: predicted score mean absolute error with ground truth assuming normal distribution of rating data

Plan

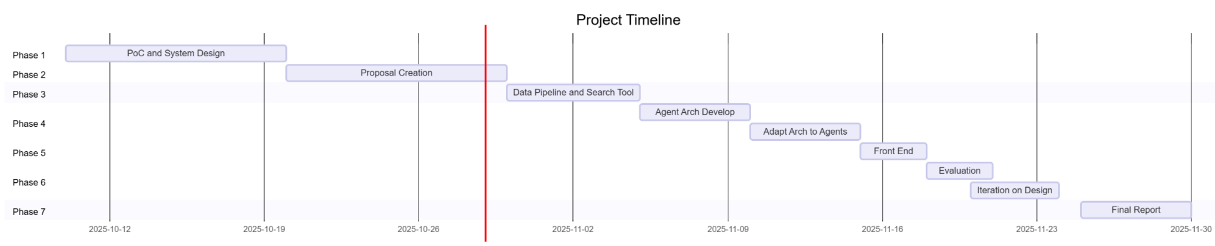
Our development will follow a waterfall model, guided by a Kanban board hosted in GitHub Projects. The choice of the waterfall model is justified by the small team size, short duration, and prototype nature of the project. We will hold in-person pair-coding sessions twice a week to work and address any blockers together.

Expected Tech Stack:

- Code Management: Git and GitHub
- DevOps: GitHub Project Board and GitHub Actions
- Front-End Hosting: GitHub Pages
- Server: Azure Function
- LLM: GPT-5 from OpenAI

Timetable:

Since both developers will work together through paired programming, we will not distinguish tasks among individuals. This approach is justified and expected to work well in a small, close-knit team.



Date	**Task**
Oct 10-20	PoC and system design
Oct 20-30	Proposal creation
Oct 30-Nov 04	Data pipeline construction and Internet search tool development
Nov 05-09	Agent architecture implementation and tool integration
Nov 10-14	Adapt the architecture to agents for each stage
Nov 15-17	Front-end development

Date	Task
Nov 18-20	Final testing and evaluation
Nov 20-24	Iteration on design
Nov 25-29	Final report

Risks

ID	Risk Name	Cause	Impact	Mitigation
H1	Internet Data Drift	Data available is not recent	Prediction may reference outdated results	Suggest recent result in prompt, rule-based filter.
H2	Data Availability	Limited open data access	Incomplete analysis	Limit project scope, provide api access to the data source.
H3	Generalizat ion	Diverse business models	Difficulty to tune model across domains	Acceptable in prototype, more flexible structure if developed product
H4	LLM Abilities	Intrinsic model constraints	Potential performance issues	Select more powerful models, evaluation to detect issues.
H5	Hallucinati on	LLM limitation	Misleading results	More powerful models, RAG and force reference generation.
H6	User input accuracy	The context provided is not detailed	Poor search and result performance, too general.	Educate users on effective input or provide more granular-entry fields.

References

[1] KPMG, “Retail site selection: Using data and analytics to make better location decisions,” KPMG Global Retail Insights Report, 2021.
[2] J. Lu, X. Zheng, E. Nervino, Y. Li, Z. Xu and Y. Xu, “Retail store location screening: A machine learning-based approach,” Journal of Retailing and Consumer Services, vol. 77, no. C, Mar. 2024, Art. 103620. doi: 10.1016/j.jretconser.2023.103620.
[3] M. Koda, Y. Zheng, R. Ma, M. Sun, D. Pansare, F. Duarte and P. Santi, “LocationReasoner: Evaluating LLMs on real-world site selection reasoning,” arXiv preprint arXiv:2506.13841, Jun. 2025.
[4] Amap, “Web服务 API 概述 – 高德开放平台,” 高德地图开放平台, 3 Aug. 2023. [Online]. Available: https://lbs.amap.com/api/webservice/summary.
[5] Dianping, “大众点评,” 大众点评网. [Online]. Available: https://www.dianping.com/.
[6] D. L. Huff, “A probabilistic analysis of shopping center trade areas,” *Land Economics*, vol. 39, no. 1, pp. 81–90, Feb. 1963.

Mermaid

```
---
config:
  theme: mc
---
gantt
    title Project Timeline
    dateFormat YYYY-MM-DD
    section Phase 1
    PoC and System Design    :a1, 2025-10-10, 10d
    section Phase 2
    Proposal Creation        :a2, 2025-10-20, 10d
    section Phase 3
    Data Pipeline and Search Tool  :a3, 2025-10-30, 6d
    section Phase 4
    Agent Arch Develop  :a4, 2025-11-05, 5d
    Adapt Arch to Agents       :a5, 2025-11-10, 5d
    section Phase 5
    Front End               :a6, 2025-11-15, 3d
    section Phase 6
    Evaluation       :a7, 2025-11-18, 3d
    Iteration on Design                :a8, 2025-11-20, 4d
    section Phase 7
    Final Report                       :a9, 2025-11-25, 5d
```

Placeholder for evaluation:
The primary metric for evaluating SiteSage's effectiveness is the accuracy of **visitor forecasts**. Accurate predictions are crucial as they directly impact the business's ability to make informed decisions about location viability. This can be measured by comparing predicted visitor numbers with historical data from similar locations, reported using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). This metric provides a concrete, quantifiable reflection of the system's performance. We deem the prototype successful if it achieves a result better than our base model, which is random guess.

Detail of background

Modern location analytics integrates diverse datasets, such as population distribution and points of interest (POIs) with geographic information systems (GIS) and machine learning models to predict revenue potential and evaluate the advantages of candidate sites[2]