Stop Attractiveness Scoring Using Public Transport Ridership Data

Ege Altıok
Graduate School of Natural and Applied Sciences
Dokuz Eylül University

I. Introduction

Public transport is an essential component of daily life, which is why continuous efforts are made to make it more sustainable, accessible, and affordable. Despite these efforts, many cities struggle with people preferring private to public transportation, causing a variety of problems.

Recent advancements in data collection, especially with Automatic Fare Collection (AFC) systems and Smart Card Data (SCD), have been a key factor to analyse and understand passenger behaviour to optimise public transportation. Smart card transactions reveal when and where passengers board or alight, offering insights into travel patterns and network dynamics.

This study aims to develop an **Attractiveness Scoring (AS)** framework using origin–destination (OD) ridership records, encompassing multiple lines. Unlike optimisation-based works that focus on improving route planning and network efficiency, this approach prioritises human-oriented understanding of service accessibility by analysing how and why people choose specific stops. By doing so, it aims to help passengers make more informed travel decisions and improve their public transport experience.

II. Literature Survey

Earlier studies on public transport optimisation [2] emphasised algorithmic efficiency in route planning through Dijkstra-based methods and transfer-minimisation strategies. While these approaches were effective in improving network performance, they largely treated passengers as static objects rather than active decision-makers. Most existing studies therefore focused on route-level optimisation or network-wide accessibility, overlooking how individual stops influence passengers' travel choices.

More recent work, such as Luo et al. [3], introduced joint passenger flow prediction models to understand multimodal travel interactions. This study showed the potential of using large-scale Automatic Fare Collection (AFC) data for uncovering hidden mobility patterns.

However, the literature remains limited when it comes to analysing stop attractiveness from a passenger perspective. This study fills this research gap by analysing passenger preferences through ridership data, focusing on three key metrics: boarding frequency, effective destination diversity, and attraction-weighted accessibility scores.

These metrics provide a quantitative framework for understanding stop attractiveness from a passenger-centric perspective. Ultimately, the focus shifts from *"How to make a route more optimal?"* to *"What makes each stop attractive to passengers?"*, offering transparency into the decision-making factors that shape public transit usage.

III. Methodology

A. Data Source

This study uses the hourly ridership origin–destination (OD) records from the Bay Area Rapid Transit (BART) system in California, United States [4]. Each record represents the number of passengers travelling from a *source* station to a *destination* station within a specific hour. The dataset is provided in a comma-separated values (CSV) format with the following fields:

Variable	Description	Value
date	service date	YYYY-MM-DD
hour	hour of day	integer (0 – 23)
source	source station	4-letter code
destination	destination station	4-letter code
passengers	number of passengers	integer ...

The station reference table reports the selected attributes used to represent each *stop* in the analysis. It includes a unique station identifier, descriptive naming fields, and geographic coordinates, which together support consistent station matching across datasets and enable spatial computations.

Variable	Description	Value
code	station code (unique)	2-letter uid
name	station name	up to 100 chars
abbreviation	station abbreviation	4-letter code
latitude	station latitude	float
longitude	station longitude	float

B. Data Preparation

The data preparation methodology consists of three principal stages that transform the raw origin–destination (OD) records into an analysable format.

- **Filtering and Cleaning**: The dataset is first cleaned by removing invalid entries, such as records where the source and destination stops are identical.
- **Stop-Level Aggregation**: Records are grouped by source stop to compute stop-level indicators, including total boardings and destination-choice dispersion (effective destination diversity).
- **Day-Level Aggregation**: The stop-level indicators are aggregated by date to facilitate day-level and time-series analyses.

C. Attractiveness Scoring Framework

For each stop S_i, the **Attractiveness Score (AS)** is computed as a weighted aggregation of **three** metrics derived from the aggregated OD records:

$$AS_i = w_1 \cdot \mathit{Board}_i + w_2 \cdot \mathit{EffDst}_i + w_3 \cdot \mathit{Access}_i$$

Where:

- Board_i: Number of boardings at the stop S_i, normalised to $[0, 1]$:

$$B_i = \sum_{\text{source}=S_i} passengers$$

$$Board_i = \frac{B_i - \min(B)}{\max(B) - \min(B)}$$

- EffDst_i: The **effective destination diversity**, which shows how widely passengers distribute their destination choices (i.e. many meaningful options rather than rarely used destinations). First, define aggregated OD flow $F_{ij} = \sum passengers$ for trips from S_i to S_j, and the destination choice probability

$$p_{ij} = \frac{F_{ij}}{\sum_j F_{ij}} \quad \text{for } \sum_j F_{ij} > 0$$

Then compute Shannon entropy H_i [1] and its "effective number of destinations" (Hill number of order 1):

$$H_i = -\sum_j p_{ij} \ln(p_{ij}), \qquad D_i = \exp(H_i)$$

Finally, normalise to $[0, 1]$:

$$\mathit{EffDst}_i = \frac{D_i - \min(D)}{\max(D) - \min(D)}$$

- Access_i: An **attraction-weighted option index** that values access to highly demanded destinations. Let destination attractiveness be the inbound volume:

$$A_j = \sum_{\text{destination}=S_j} passengers$$

Let $dist_{ij}$ be the geographic distance between stops S_i and S_j (computed from stop coordinates), and let the distance-decay function be:

$$f(dist_{ij}) = \frac{1}{1 + dist_{ij}}$$

Then:

$$Acc_i = \sum_{j \neq i} A_j \cdot f(dist_{ij}),$$

$$Access_i = \frac{Acc_i - \min(Acc)}{\max(Acc) - \min(Acc)}$$

- w_1, w_2, w_3 : Weight coefficients to be adjusted.

Note: `TransferFrequency` is not considered in this setting because the OD-hour aggregated format does not provide enough information to reliably identify individual transfer events or sequential legs of a single passenger journey.

D. Django Web App Implementation

The interactive environment to display results is built with the Django framework (Python 3.13), following a Model-View-Template (MVT) pattern to manage large-scale transit datasets. The source code is publicly available on GitHub [5].

E. Frontend and Interactive Visualisation

The presentation layer is a single-page application using **Leaflet.js** to achieve an interactive map display.

- **Geospatial Visualisation**: Stops are rendered as interactive markers on top of their corresponding coordinates on the map.
- **Dynamic Parameter Selection**: The interface features several input fields, allowing users to choose the database table holding the records, start and end dates to filter records within an interval, and a "Weight Splitter" component to assign coefficients for scoring metrics.
- **Client-Side State Management**: Upon receiving an API response, the system updates the UI with the relevant data received.

F. Backend and API Endpoints

The data is stored in a relational database schema of two primary entities: *'Stations'*, containing geographic metadata (coordinates and unique identifiers) of the stops, and *'YearlyUsage'*, containing hourly origin-destination (OD) passenger records.

A custom management script ('load_csv.py') handles the insertion of high-volume CSV data into relevant databases with transactional batch processing.

The system uses a REST API that serves as the bridge between the database and the visualisation layer. The key endpoints include:

- **"GET /api/stations/"**: Returns geospatial metadata of stations in JSON format for marker placement.
- **"GET /api/data/"**: Accepts parameters ('model', 'start _date', 'end_date') to fetch the data from the given model to return records between the start and end dates.
- **"GET /api/station-scores/"**: Accepts parameters ('start _date', 'end_date') and user-defined weights (w_1, w_2, w_3) to return normalised attractiveness metric scores.

IV. Discussion

This section discusses the strengths and limitations of the Attractiveness Scoring (AS) framework, reflects on the challenges encountered, and mentions how alternative choices might influence the outcomes.

A. Strengths and Weaknesses of the Proposed Approach

One of the primary strengths of the proposed framework is its complete reliance on observed behaviours. By using large-scale OD ridership data, the method remains data-driven and reproducible. Additionally, the modular design of the scoring components provides transparency on how of each scoring metric contributes to the attractiveness score.

However, one key weakness of the AS framework is that there is no direct way of knowing the true attractiveness scores. Instead, attractiveness is inferred through a weighted combination of multiple behavioural indicators. For this reason, the choice of weights may introduce an element of subjectivity that can influence stop rankings, unless a more objective method for determining them is provided.

B. Challenges Encountered and Mitigation Strategies

The most significant challenge was the data availability, as most public transport systems do not make their automated fare collection or smart card transaction data publicly accessible. As a result, each 'station' was considered as 'stop' and the stop-level analysis was constrained to station-level for this study.

Another challenge was handling the scale of the dataset, as year-long hourly OD records can be computationally expensive to process. This was addressed by limiting the analysis window for score calculation to a maximum of seven days, however, this limit is not essential to the framework and can be increased in systems with greater computational capacity.

C. Impact of Alternative Methods

Alternative methodological choices could lead to different interpretations of attractiveness.

For example, clustering-based approaches may group stops with similar usage profiles without producing an actual ranking. Similarly, machine learning models could be used to predict attractiveness from external features, but it would also reduce the transparency and interpretability of those features.

Compared to these alternatives, the proposed AS framework prioritises explainability at the expense of predictive performance and relational complexity.

V. Conclusion

This study was done to analyse public transport usage from the perspective of passengers on stop-level attractiveness rather than network-wide optimisation. Using large-scale hourly origin–destination ridership data from the San Francisco Bay Area Rapid Transit (BART) system, an Attractiveness Scoring (AS) framework was developed to determine how appealing individual stops are to passengers based on the observed transit data.

A. General Evaluation of the Project

Overall, the project successfully achieved its primary objective. The framework was able to transform raw public transportation data into meaningful stop-level indicators. By combining boarding demand, effective destination diversity, and accessibility-related measures into a single score, the system provides a data-driven way to compare stops.

B. Main Contributions and Achievements

The main contribution of this work lies in shifting the analytical focus from system optimisation to passenger choice and stop attractiveness. Unlike traditional studies that prioritise timetable efficiency or route planning, this research introduces a composite attractiveness score based purely on passenger behaviour.

Key achievements include:

- The formulation of an Attractiveness Scoring framework which merges multiple behavioural indicators into a single metric.
- A data preparation pipeline for handling large-scale public transit datasets.
- Providing insights into how transit stops differ in volume of use, diversity, and reach of destinations they serve.

C. Suggestions for Future Work

Several improvements could further enhance this research. First, the weighting scheme of the AS framework could be optimised or learned automatically using machine learning instead of being user-defined.

Then, having additional contextual data such as transfer-trip frequency, socioeconomic indicators, residential density, or intermodal connectivity could improve the explanatory power.

Finally, validating the attractiveness scores against survey data or passenger satisfaction metrics would significantly improve the interpretation of the results.

D. Real-World Applicability

The proposed system has practical relevance for both authorities and the public transport users. Attractiveness scores can support decisions for infrastructural upgrades, service prioritisation, and targeted interventions to encourage public transport usage.

From a passenger-oriented perspective, the framework can be integrated into mobile apps like "Google Maps" or "İzmirimKart" to help users have a better understanding of public transportation usage.

The approach is also transferable to other cities and transit systems where ridership data is available, making it a flexible tool for public transport analysis.

VI. References

[1] L. Jost, "Entropy and diversity," *Oikos*, vol. 113, no. 2, pp. 363–375, 2006. DOI: https://doi.org/10.1111/j.2006.0030-1299.14714.x [Online]. Available: https://nsojournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.2006.0030-1299.14714.x

[2] A. Bozyigit, G. Alankuş, and E. Nasibov, "A public transport route recommender regarding the least transfer," *Sigma Journal of Engineering and Natural Sciences*, vol. 9, pp. 437–446, 2018.

[3] D. Luo, J. Wang, W. Lu, W. Ding, X. Yan, and H. Yang, "Learning dynamic interaction for multimodal transportation systems: A joint passenger flow prediction approach based on spatiotemporal hypergraph convolution networks," *Transportation Research Part C: Emerging Technologies*, vol. 178, p. 105 257, 2025, ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2025.105257

[4] *Bay area rapid transport ridership data*, Accessed: 2025-12-28. [Online]. Available: https://www.bart.gov/about/reports/ridership

[5] *Stop attractiveness scoring*. [Online]. Available: https://github.com/Apothemia/Stop-Attractiveness-Scoring/tree/main/djangoapp