Transaction Analysis for Operational Optimization in Transjakarta Public Transportation System

Nouval Trezandy Lapatta¹, Andi Hendra², Raden Muhammad Firzatullah³
¹²³Informatics Engineering Studies Programme, Tadulako University, Indonesia
³Water and Land Transportation Management Programme, Inland Water and Ferries Transport Polytechnic of Palembang, Indonesia

ABSTRACT

This study presents a comprehensive exploratory data analysis (EDA) of transaction data within the Transjakarta public transportation system, focusing on routes, passenger flow, and congestion patterns. Leveraging tap-in and tap-out transaction records collected through payment cards, the research aims to provide actionable insights for optimizing operational efficiency and enhancing service quality. By analyzing transaction data, the study identifies busy routes, peak hours, and passenger distribution across different modes of transportation, including big bus (BRT), medium and big bus (non-BRT), and mini-bus (Mikrotrans). Moreover, the research investigates correlations between transaction volume and traffic congestion levels, enabling the identification of areas prone to delays and bottlenecks. The findings of this analysis offer valuable information for route planning, scheduling adjustments, and infrastructure improvements within the Transjakarta system. By optimizing resources based on data-driven insights, the study seeks to contribute to the enhancement of public transportation services, promoting accessibility, reliability, and sustainability in urban mobility.

Keywords: Transaction Analysis, Operational Optimization, Transjakarta, Public Transportation, Exploratory Data Analysis

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Public transportation plays a pivotal role in urban mobility, offering a sustainable solution to alleviate traffic congestion, reduce carbon emissions, and enhance accessibility for urban dwellers. In rapidly growing cities like Jakarta, Indonesia, the demand for efficient public transportation systems is ever-increasing to accommodate the burgeoning population and address the challenges of urbanization. Among the key players in Jakarta's public transportation landscape, Transjakarta stands out as a flagship provider, offering a diverse range of transportation modes, including big buses (BRT), medium and big buses (non-BRT), and mini-buses (Mikrotrans). Despite its significance, optimizing the operational efficiency of Transjakarta remains a critical challenge, particularly in the absence of comprehensive data analysis and actionable insights [1].

The primary problem confronting Transjakarta revolves around the lack of detailed transaction analysis, hindering the ability to make informed decisions regarding route planning, scheduling adjustments, and infrastructure improvement. While transaction data, captured through the tap-in and tap-out mechanism using payment cards, are abundant, their potential for informing strategic decision-making remains largely untapped [2]. Consequently, there is a pressing need to harness the wealth of transaction data to gain deeper insights into passenger behavior, route utilization, and traffic patterns within the Transjakarta network.
Literature on public transportation optimization and data-driven decision-making provides valuable insights into similar endeavors worldwide. Studies by [1] and [2] have highlighted the importance of leveraging transaction data for route optimization and service improvement in public transportation systems. However, few studies have specifically focused on the Transjakarta system and its unique operational challenges. This research aims to bridge this gap by conducting a comprehensive analysis of transaction data within the Transjakarta network, with the goal of providing actionable insights tailored to the local context.

The proposed approach entails leveraging advanced data analytics techniques to extract meaningful patterns and trends from the transaction data collected by Transjakarta. By employing techniques such as data mining, machine learning, and spatial analysis, this research seeks to uncover hidden correlations between transaction volume, passenger flow, and traffic congestion levels to enhance the accuracy and timeliness of the analysis [3].

The innovation of this research lies in its holistic approach to public transportation optimization within the context of Transjakarta. While existing studies have predominantly focused on individual aspects of public transportation management, such as route optimization or congestion mitigation, this research offers a comprehensive framework that integrates multiple dimensions of analysis. By synthesizing transaction data with real-time traffic information and spatial data, the study aims to provide a nuanced understanding of the dynamics shaping the Transjakarta system.

Moreover, the research contributes to the advancement of data-driven decision-making in the realm of public transportation management, particularly in emerging economies like Indonesia [4], [5]. By demonstrating the practical utility of transaction analysis for operational optimization, the study serves as a model for leveraging data analytics to address complex urban challenges. Ultimately, the new value proposition of this research lies in its potential to empower Transjakarta stakeholders with actionable insights for improving service quality, enhancing passenger experience, and fostering sustainable urban mobility.

2. METHOD
In this study, we delve into the methodology employed to conduct a comprehensive analysis of transaction data within the Transjakarta public transportation system. Leveraging a publicly available dataset and employing exploratory data analysis (EDA) techniques, we aimed to extract valuable insights into passenger behavior, route utilization, and traffic patterns [6][7].

2.1. Dataset
The dataset used in this research is transaction data of 31,730 public transportation service users by PPID Transjakarta. There are many variables in the actual master data, but this study only analyzed 24 variables as shown in Table 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Row Count</th>
<th>Null</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>transID</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>2</td>
<td>payCardID</td>
<td>31730</td>
<td>non-null</td>
<td>int64</td>
</tr>
<tr>
<td>3</td>
<td>payCardBank</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>4</td>
<td>payCardName</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>5</td>
<td>payCardSex</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>6</td>
<td>payCardBirthDate</td>
<td>31730</td>
<td>non-null</td>
<td>int64</td>
</tr>
<tr>
<td>7</td>
<td>corridorID</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>8</td>
<td>corridorName</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>9</td>
<td>direction</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
<tr>
<td>10</td>
<td>tapInStops</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>11</td>
<td>tapInStopsName</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>12</td>
<td>tapInStopsLat</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
<tr>
<td>13</td>
<td>tapInStopsLon</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
<tr>
<td>14</td>
<td>stopStartSeq</td>
<td>31730</td>
<td>non-null</td>
<td>int64</td>
</tr>
<tr>
<td>15</td>
<td>tapInTime</td>
<td>31730</td>
<td>non-null</td>
<td>datetime64</td>
</tr>
<tr>
<td>16</td>
<td>tapOutStops</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>17</td>
<td>tapOutStopsName</td>
<td>31730</td>
<td>non-null</td>
<td>object</td>
</tr>
<tr>
<td>18</td>
<td>tapOutStopsLat</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
<tr>
<td>19</td>
<td>tapOutStopsLon</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
<tr>
<td>20</td>
<td>stopEndSeq</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
<tr>
<td>21</td>
<td>tapOutTime</td>
<td>31730</td>
<td>non-null</td>
<td>datetime64</td>
</tr>
<tr>
<td>22</td>
<td>payAmount</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
<tr>
<td>23</td>
<td>weekday</td>
<td>31730</td>
<td>non-null</td>
<td>int32</td>
</tr>
<tr>
<td>24</td>
<td>travelDuration</td>
<td>31730</td>
<td>non-null</td>
<td>float64</td>
</tr>
</tbody>
</table>
2.1. Exploratory Data Analysis

Exploratory data analysis is the initial stage of data analysis that focuses on discovering patterns, relationships, and trends in the data [8], [9]. EDA involves the use of visual and statistical methods to explore and summarize the data, and to generate hypotheses that can be further tested using more advanced statistical techniques as presented in Figure 1.

![Figure 1. Stage of exploratory data analysis](image)

3. RESULTS AND DISCUSSION

In the following sections, we provide a detailed overview of the methods employed in this study, including data preprocessing steps, exploratory data analysis techniques, and statistical analysis methods. Through a systematic and transparent approach to data analysis, we aim to provide actionable insights for optimizing the operational efficiency of the Transjakarta public transportation system [10].

3.1. Peak Travel Times

The data suggests that there's a two-peaked distribution as shown in Figure 2, indicating that peak travel times on Transjakarta happen at 6 AM, then gradually decline throughout the day. The number of trips increases again in the afternoon, reaching its peak at 5 PM, then tapers off. This suggests that the increase in trips aligns with the commuting habits of Jakarta residents who leave for work around 6 AM and return home starting at 5 PM [11].

![Figure 2. Distribution of peak travel times](image)

3.2. Average Travel Durations

The average trip duration of Transjakarta users every day is shown in Figure 3. It can be seen that the duration of passenger trips is stable at 70 minutes for each trip. It can be seen that there is a slight decrease on days number 5 and 6 which indicate Saturday and Sunday. This may happen because on weekends, a number of Jakarta residents rest and choose not to travel [12].
3.3. Maximum Travel Durations

The maximum and minimum daily trip duration for Transjakarta users is shown in Figure 4. It can be seen that the maximum duration of passenger trips is stable at 175 minutes for each trip. This indicates that Transjakarta has always been a popular choice for Jakarta residents in their activities, even for long trips.

The minimum trip duration of passenger trips is stable at 20 minutes for each trip from Monday to Friday. Then there is a decrease on Saturday and Sunday. This is normal because some workers choose to rest on weekends.

3.4. Paycard Banks Distribution

Transjakarta uses a non-cash payment method (cashless) from several providers including banks as shown in Figure 5. It appears that a half of Transjakarta user prefer to use digital money from Bank DKI, a regional bank in Jakarta.
3.5. Gender Payment Cards Distribution

In Figure 6, the distribution of Transjakarta users based on gender is shown. The results are quite interesting, as 54% of passengers are women, more than men. This indicates that Transjakarta has become a public transportation mode that is perceived as safe and comfortable by female users in Jakarta.

![Figure 6. Distribution of gender payment cards](image)

3.6. Payment Amounts Distribution

Distribution of payment amounts made by Transjakarta users is presented in Figure 7. On average, passengers pay around IDR 2700. For long trips, they pay a maximum fee of IDR 20000.

![Figure 7. Distribution of payment amounts in (a) histplot and (b) boxplot](image)

3.7. Directions Distribution

There is an almost balanced distribution between direct and indirect trips as shown in Figure 8. This shows that Transjakarta remains comfortable for all types of trips.

![Figure 8. Distribution of directions](image)
3.8. Correlation Between Variables

For more in-depth analysis needs, the relationship between variables can be analyzed through a confusion matrix [13]. The confusion matrix table as shown in Figure 9.

![Confusion Matrix](image)

**Figure 9.** Confusion matrix for correlation between variables

4. CONCLUSION

This research has explored Transjakarta public transportation transaction data. There are several research results that can be used to determine company and government policies in the future. The policy in question can be in the form of regulations to improve the quality of Transjakarta services or to increase the number and loyalty of Transjakarta users. For example, increasing the safety and comfort of female passengers who make up the majority of users. Provide special discounts for DKI card users, who are the majority of passengers with 50% of all other card users. Lastly, pay special attention to the peak times for Transjakarta users at 6 WIB and 17 WIB, at which time Transjakarta passengers leave to start activities such as work and return home in the afternoon. The company must guarantee the availability of buses and the comfort and safety of passengers at that time.

REFERENCES


