The Taipei Metro, also known as the Taipei Mass Rapid Transit (MRT), began running back in 1996. It is the most important means of public transportation serving the metropolis of Taipei, Taiwan. Currently, the Metro system comprises 108 stations (123 if transfer stations are double-counted), five main routes, and two branch lines. The Taipei Metro has developed quite well, and the MRT system has become the most convenient transportation mode in Taiwan. Further, the Taoyuan Airport MRT line is in operation as well, but we must point out that it hardly falls within the scope of the present article.

There are eight terminal stations in the Taipei Metro stations, as depicted below: Minquan W. Rd., Zhongshan, Songjiang Nanjing, Nanjing Fuxing, Ximen, Daqiaotou. (In this station, we make a simplification that denotes the Huilong station and Luzhou station.) There are fifteen transfer stations in the Taipei Metro system, as depicted below: Tamsui, Xiangshan, Taipei Nangang Exhibition Center, Dingpu, Taipei Zoo, Songshan, Xindian, Nanshijiao, and Daqiaotou. (In this station, we make a simplification that denotes the Huilong station and Luzhou station.) There are fifteen transfer stations in the Taipei Metro system, as depicted below: Minquan W. Rd., Zhongshan, Songjiang Nanjing, Nanjing Fuxing, Ximen, Taipei Main Station, Zhongxiao Xinsheng, Zhongxiao Fuxing, C.K.S. Memorial Hall, Dongmen, Daan, Guting, Beitou, and Qizhang. However, our research does not include the transfer stations of two branch lines. It goes without saying that the system carries an average of millions of passengers per day. Hence, it is crucial to plan a reasonable route in the Taipei Metro system.

This paper has presented a decision model for optimizing route allocation with linear programming in the Taipei Metro system. The goal is to avoid centralization of a passenger in certain transfer stations, which means that some transfer stations are overloaded, whereas some transfer stations may be nearly empty. This study applies the open data of origin and destination station to solve this problem. Specifically, through a direct computational comparison of the objective function value and statistical analysis between our proposed model and the existing route development, it confirms that our route development would achieve a reduction of ridership transfer flow of approximately 15% in this case.

**Keywords:** Taipei Metro System, Linear Programming, Open Data Analytics.

### INTRODUCTION AND LITERATURE REVIEW

The Taipei Metro, also known as the Taipei Mass Rapid Transit (MRT), began running back in 1996. It is the most important means of public transportation serving the metropolis of Taipei, Taiwan. Currently, the Metro system comprises 108 stations (123 if transfer stations are double-counted), five main routes, and two branch lines. The Taipei Metro has developed quite well, and the MRT system has become the most convenient transportation mode in Taiwan. Further, the Taoyuan Airport MRT line is in operation as well, but we must point out that it hardly falls within the scope of the present article. There are eight terminal stations in the Taipei Metro system, as below: Tamsui, Xiangshan, Taipei Nangang Exhibition Center, Dingpu, Taipei Zoo, Songshan, Xindian, Nanshijiao, and Daqiaotou. (In this station, we make a simplification that denotes the Huilong station and Luzhou station.) There are fifteen transfer stations in the Taipei Metro system, as depicted below: Minquan W. Rd., Zhongshan, Songjiang Nanjing, Nanjing Fuxing, Ximen, Taipei Main Station, Zhongxiao Xinsheng, Zhongxiao Fuxing, C.K.S. Memorial Hall, Dongmen, Daan, Guting, Beitou, and Qizhang. However, our research does not include the transfer stations of two branch lines. It goes without saying that the system carries an average of millions of passengers per day. Hence, it is crucial to plan a reasonable route in the Taipei Metro system.

Linear Programming (LP), also called linear optimization, is a method to find the best solution in a mathematical model. It is usually applicable for the maximum profit problem or the lowest cost problem (Vanderbei, 2015). More formally, linear programming is a mathematical technique for the optimization of linear objective functions. The expression to be maximized or minimized is called the objective function. It can be expressed in the canonical form, via subject to linear inequality and linear equality constraints (Dantzig, 2016). Moreover, it can be applied to various fields of study, such as agriculture and animal husbandry. This application of linear programming aims at developing a water resource planning model that helps decision-makers increase revenue and reduce water use during the cultivation of crops (Daghighi et al., 2017). Linear Programming can even be applied to the optimization of electric vehicle charging, the optimization of cost versus life cycle assessment on drinking water production plants, the optimization of waste vehicles recycling of steel scrap and alloying, etc. (Skugor & Deur, 2018; Capitanescu et al., 2017; Ohno et al., 2017). Additionally, it can be applied to obtain an optimal framework for public transportation networks (Li et al., 2014). However, a substantial amount of research has focused on the efficiency and effectiveness of the algorithm for route planning in transportation networks. This ignores the problem of the carrying capacity (Delling et al., 2009). In accordance with this, the problem of the flow for route planning cannot be neglected.

In this paper, to consider the utility of a whole metropolitan area in Taipei, we want to avoid centralization of a passenger in certain transfer stations, which means that some transfer stations are overloaded, whereas some transfer stations may be nearly empty. Thus, we are going to apply the transportation model to the Taipei Metro system to minimize the maximum frequency of the transfer flow that may occur between terminal stations. That is, we want to minimize the maximum transfer frequency of the entire transfer station. Furthermore, to improve the quality and efficiency of the analysis and route development, we use the open data of origin and destination station to solve this problem. Section 2 summarizes and discusses the definition of the model formulation. Section 3 details the proposed model for the transfer flow algorithm and Section 4 demonstrates the solution and outcome of the proposed
model for route development in the Taipei Metro system. Finally, according to this solution, our proposed model achieved a reduction of the frequency of transfer station. It is a feasible solution for practical applications of big data analytics.

**MODEL FORMULATION**

This section summarizes the model formulation. Then the set of constraints is explained. Besides, the main idea of our proposed model is to maximize the number of people who can get in a station and exit that station of the Taipei Metro system on a single route while precluding transfer via transfer stations.

**Define**

R_{ik}: represents the k^{th} potential routes from the i^{th} initial station to j^{th} terminal station. Moreover, an initial station is not identical to a terminal station.

S_{i}: The frequency a transfer station was passed, i = 1, 2, ..., 12, namely, a transfer station of S_{1}-Minquan W. Rd., S_{2}-Zhongshan, S_{3}-Songjiang Nanjing, S_{4}-Nanjing Fuxing, S_{5}-Ximen, S_{6}-Taipei Main Station, S_{7}-Zhongxiao Xinsheng, S_{8}-Zhongxiao Fuxing, S_{9}-C.K.S. Memorial Hall, S_{10}-Dongmen, S_{11}-Daan, and S_{12}-Guting.

L_{i}: The route between transfer stations on The Taipei Metro system, i = 1, 2, ..., 19. For instance, a route that represents the transfer station between the Minquan W. Rd. station and Zhongshan station. A detailed description of L_{i} is provided below.

L_{1}: from S_{1}-Minquan W. Rd. station(O11/R13) to S_{2}-Zhongshan station(G14/R11) via the station of Shuangliang(R12).
L_{2}: from S_{1}-Minquan W. Rd. station(O11/R13) to S_{3}-Songjiang Nanjing station(G15/O08) via the station Zhongshan Elementary School(O10) and Xingtian Temple(O09).
L_{3}: from S_{2}-Zhongshan station(G14/R11) to S_{3}-Ximen station(BL11/G12) via the station of Beimen(G13).
L_{4}: from S_{2}-Zhongshan station(G14/R11) to S_{4}-Songjiang Nanjing station(G15/O08).
L_{5}: from S_{3}-Songjiang Nanjing station(G15/O08) to S_{5}-Nanjing Fuxing station(G16/BR11).
L_{6}: from S_{3}-Songjiang Nanjing station(G14/R11) to S_{6}-Taipei Main station(BL12/R10).
L_{7}: from S_{4}-Songjiang Nanjing station(G15/O08) to S_{7}-Zhongxiao Xinsheng station(BL14/BR07).
L_{8}: from S_{4}-Nanjing Fuxing station(G16/BR11) to S_{7}-Zhongxiao Fuxing station(BL15/BR10).
L_{9}: from S_{6}-Ximen station(BL11/G12) to S_{7}-Taipei Main station(BL12/R10).
L_{10}: from S_{7}-Taipei Main station(BL12/R10) to S_{8}-Zhongxiao Xinsheng station(BL14/O07) via the station of Shandao Temple(BL13).
L_{11}: from S_{7}-Zhongxiao Xinsheng station(BL14/O07) to S_{9}-Zhongxiao Fuxing station(BL15/BR10).
L_{12}: from S_{7}-Ximen station(BL11/G12) to S_{9}-C.K.S. Memorial Hall station(G10/R08) via the station of Xiaonannmen(G11).
L_{13}: from S_{8}-Taipei Main station(BL12/R10) to S_{9}-C.K.S. Memorial Hall station(G10/R08) via the station of NTU Hospital station(R09).
L_{14}: from S_{9}-Zhongxiao Xinsheng station(BL14/O07) to S_{10}-Dongmen station(R07/O06).
L_{15}: from S_{7}-Zhongxiao Fuxing station(BL15/BR10) to S_{11}-Daan station(R05/BR09).
L_{16}: from S_{9}-C.K.S. Memorial Hall station(G10/R08) to S_{12}-Dongmen station(R07/O06).
L_{17}: from S_{10}-Dongmen station(R07/O06) to S_{11}-Daan station(R05/BR09) via the station of Daan Park(R06).
L_{18}: from S_{10}-C.K.S. Memorial Hall station(G10/R08) to S_{12}-Guting station(G09/O05).
L_{19}: from S_{10}-Dongmen station(R07/O06) to S_{12}-Guting station(G09/O05).

D_{i}: An indicator variable (1 if set-up i used; otherwise, 0).

M: A small, integral constant. Given this, the model is expressed:

\[
\min \left[ \max(S_{i}, S_{j}, S_{k}, S_{l}, S_{m}, S_{n}, S_{o}, S_{p}, S_{q}, S_{r}, S_{s}, S_{t}, S_{u}, S_{v}) \right]
\]

subject to

\[
\sum_{k} R_{ik} = 5, \ \forall k
\]

\[
\sum_{i} L_{i} = 1, \ \forall i
\]

\[
\sum_{i} N_{i} \geq 2, \ \forall i
\]

\[
D_{i} \in \{0, 1\}, \ \forall i
\]

In the above formulation, the objective function is to minimize the total transfer flow, defined as the sum of the transfer station for the largest flow in the Taipei Metro system. The set-up routes are transferred according to the flow. For example, transfer stations with different transfer flows will have different set-up routes. In order to make a decision result meeting the actual circumstances better, more external and easier to be understood, the first set of constraints (1) ensures the route of five only; whereas, the second set of constraints (2) assures that the route segment can only be passed between transfer stations once at most. Moreover, for the third constraints (3), all the transfer stations will connect to at least two route segments. Finally, the fourth set of constraints (4)
imposes a fixed route-chosen indicator, $D_{i}$, associated with the potential route variable, $R_{ijk}$. In this paper, we limit our study to the application of the formulation to decrease the transfer flow of the Taipei Metro system, ruling out the problem of regionals. The definition of our model formulation is depicted in Figure 1.

The complete algorithm for the transfer flow model presented herein comprises three steps: (i) data preprocess; (ii) the generation of transfer flow routes and an initial feasible solution to the transfer flow model; and (iii) the application of simulated transfer flow to the solution of the Taipei Metro system. Moreover, to promote the proposed model efficiency and effectiveness, open datasets of origin and destination station were adopted to verify the actual transfer station flow of the Taipei Metro system. A detailed description of these four steps and datasets is provided next.

![Figure 1: Model Formulation](image)

**THE TRANSFER FLOW ALGORITHM**

The complete algorithm for the transfer flow model presented herein comprises three steps: (i) data preprocess; (ii) the generation of transfer flow routes and an initial feasible solution to the transfer flow model; and (iii) the application of simulated transfer flow to the solution of the Taipei Metro system. Moreover, to promote the proposed model efficiency and effectiveness, open datasets of origin and destination station were adopted to verify the actual transfer station flow of the Taipei Metro system. A detailed description of these four steps and datasets is provided next.
Description of datasets
The datasets of origin and destination station include the statistics of time division flow for each station in the Taipei Metro system. To decrease the frequency of the transfer station, appropriate datasets were selected based on the ridership statistics per day from the Taipei Rapid Transit Corporation. As shown in Table 1, we find the highest average ridership per day in December 2017. On the other hand, we find the largest amount of ridership, in statistical terms, at 6 p.m., as shown in Figure 2. Therefore, our proposed model has selected the datasets of origin and destination station at 6 p.m. on December 2017.

Table 1: Summary Statistics of Average Ridership per Month

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>2,080,677</td>
<td>1,953,907</td>
<td>2,172,866</td>
<td>2,022,173</td>
<td>2,096,107</td>
<td>2,050,134</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>+7.44%</td>
<td>-4.53%</td>
<td>+2.37%</td>
<td>-0.61%</td>
<td>+4.25%</td>
<td>+1.49%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2017</td>
<td>1,936,615</td>
<td>2,046,706</td>
<td>2,122,548</td>
<td>2,034,486</td>
<td>2,010,690</td>
<td>2,020,000</td>
<td>1,962,774</td>
<td>2,043,185</td>
<td>2,008,376</td>
<td>1,993,600</td>
<td>2,101,132</td>
<td>2,247,974</td>
</tr>
<tr>
<td></td>
<td>-3.93%</td>
<td>+6.21%</td>
<td>+2.71%</td>
<td>-0.49%</td>
<td>+2.81%</td>
<td>+1.49%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2016</td>
<td>2,015,843</td>
<td>1,926,929</td>
<td>2,099,816</td>
<td>2,020,689</td>
<td>1,964,780</td>
<td>1,961,747</td>
<td>1,993,495</td>
<td>2,035,594</td>
<td>2,025,883</td>
<td>2,093,201</td>
<td>2,237,913</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+4.69%</td>
<td>+14.00%</td>
<td>+3.57%</td>
<td>+5.48%</td>
<td>+1.08%</td>
<td>+3.57%</td>
<td>+1.37%</td>
<td>+7.14%</td>
<td>+0.05%</td>
<td>+0.37%</td>
<td>+4.26%</td>
<td></td>
</tr>
</tbody>
</table>

Note: the growth rate is the number of ridership statistics of this month minus the ridership statistics of that same month last year, divided by the ridership statistics of that same month last year.

Figure 2: The Average Ridership per Day in December 2017
Step 1: data preprocessing
Data preprocessing is one of the most essential steps in data mining (Famili et al., 1997; García et al., 2015). To ensure the quality of the data, we follow data preprocessing methods such as data cleaning, data integration, data transformation, and data reduction. First of all, we ignore the tuple (transfer station), when the class label is missing. Then, we remove noisy data (containing errors) and identify outliers. Second, as in data warehousing, we combine multiple tuples from open sources into a coherent data store. Third, concerning data normalization, we replace and add the columns inferred by existing attributes. Besides, we scaled the attribute values to fall within a specified range. That is, the occurrences of getting in the transfer station and exiting the transfer station are divided into the occurrences of getting in the transfer station 1, getting in the transfer station 2, exiting the transfer station 1, and exiting the transfer station 2. Moreover, we fill in the number of the two ends of the station. Finally, in order to reduce the number of attributes, if the origin station coincides with the destination station, both are excluded for the analysis. The procedure of data preprocessing is depicted in Figure 3.

Figure 3: The Data Preprocessing Methods

Step 2: the generation of transfer flow routes and an initial feasible solution to the transfer flow model
We are using an enumerative search technique to identify all possible routes ($R_{ij}$ in the model). As shown in Table 2, for example, the route from Taipei Zoo (BR01) to Nanshijiao (O01) via six ways. The first route through $L_{17}$ and $L_{19}$. The second route through $L_{17}$, $L_{16}$, and $L_{18}$. The third route through $L_{15}$, $L_{6}$, $L_{5}$, $L_{4}$, $L_{13}$, and $L_{18}$, and so on. The purpose of generating an initial feasible solution is to provide a starting solution in the transfer flow of the Taipei Metro System for route development. The initial feasible solution for our algorithm was provided by the first incumbent solution of a linear programming procedure. For this step, we query the selected routes (the set-up variables, $R_{ij}$), also matching getting in and exiting stations of the dataset in this case. Then, record the transfer flow of this case if the passenger could get in and exit the station while precluding transfer via the transfer station on the Taipei Metro system. Finally, we obtain the objective value, which is sum of the transfer flows for each recorded case (routes).

<table>
<thead>
<tr>
<th>Route</th>
<th>$L_1$</th>
<th>$L_2$</th>
<th>$L_3$</th>
<th>$L_4$</th>
<th>$L_5$</th>
<th>$L_6$</th>
<th>$L_7$</th>
<th>$L_8$</th>
<th>$L_{10}$</th>
<th>$L_{11}$</th>
<th>$L_{12}$</th>
<th>$L_{13}$</th>
<th>$L_{14}$</th>
<th>$L_{15}$</th>
<th>$L_{16}$</th>
<th>$L_{17}$</th>
<th>$L_{18}$</th>
<th>$L_{19}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR01 to O01</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Step 3: the application of simulated transfer flow to the solution of the Taipei Metro system

The steps of our final implementation of the transfer flow are outlined below on the Taipei Metro system for route development. The algorithm was implemented through linear programming using software solutions (Microsoft Excel and Lindo). Furthermore, we need to only consider the effect of those decision variables that were altered in the trial solution (with respect to the objective function). Finally, the solver found an integer solution within tolerance. All constraints are satisfied.

THE SOLUTION AND RESULT

Linear programming was applied to obtain the solution of our proposed model. This solution from linear programming has proven that the proposed model can be implemented on the Taipei Metro system for route development. Additionally, the objective function compares the original route version and the new route development version of the Taipei Metro system. The results suggest that the objective function of the original route version is equal to 383,481 values, whereas the objective function of new route development version is equal to 335,055 values (the frequency of transporting without transferring). Apparently, the new route development achieves a reduction of ridership transfer flow of approximately 15% in the Taipei Metro system. On the other hand, the results of ridership statistics per station show that the Tamsui (R28) station and the Taipei City Hall (BL18) station are most crowded at rush hour. The new route development shows that a single route has a large amount of ridership station to decrease the transfer flow of the Taipei Metro system. The route development of the Taipei Metro system is depicted in Figure 4. The description of solution as shown in Table 3.
Table 3: The Description of Solution (Route Development)

<table>
<thead>
<tr>
<th>Route (Before)</th>
<th>Route (After)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown Line from Taipei Zoo to Taipei Nangang Exhibition Center</td>
<td>from Nanshijiao to Taipei Nangang Exhibition Center</td>
</tr>
<tr>
<td>Red Line from Xiangshan to Tamsui</td>
<td>from Taipei Nangang Exhibition Center to Tamsui</td>
</tr>
<tr>
<td>Green Line from Xindian to Songshan</td>
<td>from Taipei Zoo to Xiangshan</td>
</tr>
<tr>
<td>Orange Line from Nanshijiao to Huilong or Luzhou</td>
<td>from Xindian to Huilong or Luzhou</td>
</tr>
<tr>
<td>Blue Line from Dingpu to Taipei Nangang Exhibition Center</td>
<td>from Dingpu to Songshan</td>
</tr>
</tbody>
</table>

CONCLUSIONS

This paper has presented a decision model for optimizing route allocation with linear programming in the Taipei Metro system. A linear programming model of the transfer flow problem has been developed and employed to illustrate the superior performance of our model over existing route development. Specifically, a direct computational comparison of the objective function value and statistical analysis between our proposed model and the existing route development indicates that our proposed model can provide better solutions. Further research should be conducted to analyze the maximum load of each station, especially the transfer station limit on the number of people, to determine the applicability of the route planning to other linear programming problems with decision variables.

REFERENCES