The World Tour of Lady Gaga

Qiqi Feng, Yiqing Zhu, Chufen Feng, Zhiyang Yuan

December 9, 2022

Contents

1 Introduction

Discrete optimization is an important mathematical modeling technique for tackling real-world issues. It has three main brunches: combinatorial optimization, integer programming, and constraint programming. The World Tour project is built upon the first two brunches of discrete optimalization modelling, which include the combinatorial optimization problem, the Traveling Salesman Problem (TSP), the Nearest Neighbor Search (NNS), and the integer programming resource allocation problem. The concept also incorporates the actual economic principle of inflation to plan a round-the-world tour while maximizing profits.

The goal of this project is to organize Lady Gaga's world tour concert in an efficient way to minimize traveling and staging costs and ultimately maximize overall profit, which is calculated as *Ticket Revenue - Staging Cost - Flight Cost (route).* The overall profit is therefore built from three major components of the World Tour project: total ticket revenue, staging costs, and aircraft costs. From a practical standpoint, overall profit is also defined as *Raw Profit Maximization - Route Minimization.* The cost of a flight is positively associated with the number of kilometers flown, which has implications for the route reduction issue. This problem is resolved by utilizing the TSP and NNS algorithms to locate the routes, compare the two approaches, and determine which one travels the target cities in the shortest length of the path. In terms of ticket revenue and staging costs, the first critical stage is to identify 70 high-reward locations among 91 cities chosen across six continents. The selection is executed on pulp and constructed by resource allocation problems in cities.

Overall, the flight route of the selected cities will be visualized on Google my map, and the revenue is shown by the choropleth maps in python.

2 Data

City selection of 91 cities

We first select countries in each continent by using google trends and determine a list of cities based on Lady Gaga's past Tours Concerts. Each country should have at most three cities (excluding countries in North America). Finally, we got 91 cities in total.

North America	South America	Asia	Oceania	Europe	Africa
New York	Sao Paulo	Dubai	Brisbane	London	Johannesburg
Washington D.C	Porto Alegre	Doha	Sydney	Birmingham	Cape Town
Boston	Curitiba	Mumbai	Melbourne	Düsseldorf	Cairo
Atlanta	Rio de Janeiro	Bangkok	Auckland	Hamburg	Lagos
Miami	Buenos Aires	Chiang Mai		Paris	Durban
Houston	Rosario	Hanoi		Barcelona	Nairobi
San Fanciso	Santiago	Hong Kong		Stockholm	Addis Ababa
Las Vegas	Lima	Taipei		Amsterdam	Casablanca
Los Angeles	San Juan	Hangzhou		Arnhem	
Arlinton	San José	Shanghai		Milan	
Philadelphia	Bogotá	Seoul		Antwerpen	
Detroit	Asunción	Yokohama		Zürich	
Saint Paul	Mexico City	Saitama		Istanbul	
Phoenix	Monterrey	Osaka		Oslo	
Orlando	Guadalajara	Quezon City		Herning	
Denver		Kuala Lumpur		Athens	
chicaco		Jakarta		Manchester	
Seattle		Singapore		Berlin	
nashville		Brisbane		Nice	
Toronto		Sydney		Tel Aviv	
Montreal		Melbourne		Moscow	
Vancouver		Auckland		Lisbon	
Edmonton					

Figure 1: 91 selected cities

Ticket Price

In order to take the ticket price under the same condition, we choose the world tour in 2013 which is called "The Born This Way Ball" and calculate the total revenue and total capacity for each continent to get the average ticket price. All data about revenue and capacity can be got in Wikipedia [1] . The reason why we chose this concert is because it includes most cities compared with other concerts. In order to get the ticket price for 2021, we add inflation rate [2] to the ticket price for 2013. After that, use the ticket price after inflation times with capacity to get the revenue for each city.

Ticket Revenue = ticket price in 2013 $\times$ inflation rate $\times$ venue capacity

Staging cost

The objective function of profit consists of two parts: the revenue for ticket sales, the staging cost of the concert. And the staging cost is the expense of a venue like rent fee, arrangement fee and others. Due to limited data for this part, we set the staging cost by venue capacity in different range:

Capacity	Staging Cost ($)
0-20k	80k
20-40k	150k
40-60k	250k
60-80k	400k
80k and more	550k

Table 1: Distribution of staging cost by capacity

Distance

During the process of running TSP, we need the distance between cities. It takes data by flight distance between each city.

	New York	Washington D.C	Boston	Atlanta	Miami	Houston	San Fanciso	Las Vegas	Los Angeles	Arlinton	Philadelphia	Detroit	Saint Paul
New York	0	367	300	1223	1754	2298	4162	3618	3983	2239	122	819	1655
Washington D.C	367	0	664	859	1482	1933	3893	3324	3682	1886	245	617	1461
Boston	300	664	0	1523	2025	2589	4352	3832	4202	2513	421	1018	1809
Atlanta	1223	859	1523	0	957	1120	3442	2811	3133	1177	1102	957	1459
Miami	1754	1482	2025	957	0	1537	4160	3499	3769	1804	1659	1843	2415
Houston	2298	1933	2589	1120	1537	0	2654	1988	2236	397	2177	1758	1701
San Fanciso	4162	3893	4352	3442	4160	2654	0	666	543	2357	4072	3345	2557
Las Vegas	3618	3324	3832	2811	3499	1988	666	0	380	1698	3520	2814	2091
Los Angeles	3983	3682	4202	3133	3769	2236	543	380	0	1987	3883	3185	2471
Arlinton	2239	1886	2513	1177	1804	397	2357	1698	1987	0	2121	1587	1371
Philadelphia	122	245	421	1102	1659	2177	4072	3520	3883	2121	0	739	1585
Detroit	819	617	1018	957	1843	1758	3345	2814	3185	1587	739	0	850
Saint Paul	1655	1461	1809	1459	2415	1701	2557	2091	2471	1371	1585	850	0
Phoenix	3465	3148	3701	2555	3173	1642	1047	411	596	1397	3360	2589	2054
Orlando	1519	1219	1804	650	309	1367	3936	3282	3569	1585	1416	1540	2108
Denver	2616	2337	2823	1930	2750	1421	1556	1011	1367	1032	2522	1807	1094
chicaco	1177	929	1385	951	1902	1508	2985	2447	2817	1283	1076	367	561
Seattle	3897	3711	4016	3511	4384	3047	1092	1394	1535	2671	3827	3101	2251
nashville	1232	873	1517	344	1297	1078	3168	2555	2893	1016	1086	734	1119
Toronto	589	556	718	1190	1985	2086	3636	3127	3501	1930	553	345	1092
Montreal	537	489	409	1600	2261	2574	4086	3604	3981	2435	608	853	1529
Vancouver	3941	3772	4045	3618	4507	3204	1288	1597	1739	2822	3883	3158	2311
Edmonton	3285	3168	3352	3175	4116	3020	1871	1919	2186	2623	3243	2553	1748
Sao Paulo	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000
Porto Alegre	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000	30000

Figure 2: distance of all cities

The full edition of flight distance see Appendix.

3 Computation

PuLP

In selecting the 70 high-rewarded cities, the project sets the constraints to the 91 original cities and applies *PuLP* to examine the problem. The constraints include the raw profit and city numbers. Each x_i represents the cities, the goal of the function is to select 70 cities based on the raw profit.

$$\text{Raw profit} = \text{Ticket Revenue } r_i - \text{Staging Cost } c_i$$

$$\text{Maximize}: \ (r_i - c_i)x_i$$

$$\textstyle\sum_{i=1}^{91} x_i = 70, \quad i = 1, \ldots, 91;$$

$$x_i \in \{0, 1\}$$

Therefore, the project is selecting the highest 70 raw profit cities, with the result of 0 for the non-chosen city and 1 for chosen city. Finally, the total raw profit can be calculated by the result matrix below multiple by the raw profit of each city.

```
[1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.
 1. 0. 1. 0. 1. 1. 1. 1. 0. 1. 0. 1. 1. 0. 1. 0. 1. 0. 1. 1. 0. 1. 0. 0.
 1. 0. 0. 1. 0. 1. 1. 0. 1. 0. 0. 0. 1. 0. 0. 1. 0. 1. 1.]
```

```
np.dot(profit,x)
```

```
232586315.0
```

Figure 3: Final result of PuLP

At the end of integer programming, the project approached the total raw profit of the World Tour, total accounts to 232586315.0 dollar, which is a big step toward the final profit. Then, the project visualizes all chosen cities' raw profits, the Visual Interaction Diagram can be seen in Appendix Raw Profit Visualization.

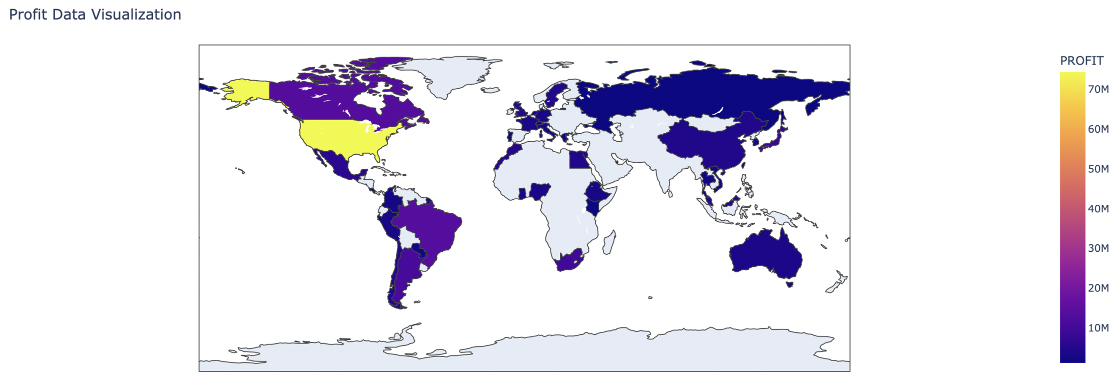

Figure 4: Visualization of profit

Traveling Salesman Problem (TSP)

An online open source Python tool called *Google OR* [3], which shows very high executing efficiency in handling large size matrices, is used in our Travelling Salesman Problem. The algorithm requires a distance matrix C which represents all distance between cities to cities. It is cumbersome to compute all distance as there are 2415 ($\sum_{n=1}^{69} n$) operations. Fortunately, this can be reduced to simply collect the distance between cities within the same continent, and for the rest part of matrix C, we assume that the distance is 30,000 km except for few critical cities.

$$C = \begin{bmatrix} C_{NA} & \cdots & \cdots & \cdots & \cdots \\ \cdots & C_{SA} & \cdots & \cdots & \cdots \\ \cdots & \cdots & C_{AF} & \cdots & \cdots \\ \cdots & \cdots & \cdots & C_{EU} & \cdots \\ \cdots & \cdots & \cdots & \cdots & C_{AS} \end{bmatrix}$$

The number $30,000$ was chosen because half the circumference of the earth is about 20,000km, which means that the actual distance between any cities on the earth must be less than 30,000. That means this route is too costly to be chosen to the Google OR.

For the critical cities, or we say entry and exit of the continent, we write their distance as 0 km into C so that this must be the shortest transcontinental route.Taking Europe as an example, we set the distance from all European cities to any Asian cities to 30,000 km except Athens to Doha, which is 0. Once the total distance is created from code, adding distance in Table 1 back to obtain the final total result of TSP.

Starting City (City Index)	Destination (City Index)	Distance (km)
Houston (5)	Monterry (36)	665
Rio De Janeiro (26)	Cape Town (39)	6062
Cassablanca (45)	Lisbon (69)	586
Athens (66)	Doha (47)	2968
Singapore (57)	Melburn (59)	6005
Sydney (58)	Los Angeles (8)	12072

Table 2: Chosen routes between continents

Firstly import matrix from csv: By running the code we have the following result:

```
matrix = np.loadtxt('Big data.csv', delimiter=',')
print(matrix)
```

```
[[    0.   367.   300. ... 30000. 30000. 30000.]
 [  367.     0.   664. ... 30000. 30000. 30000.]
 [  300.   664.     0. ... 30000. 30000. 30000.]
 ...
 [30000. 30000. 30000. ...     0.  2530.  1477.]
 [30000. 30000. 30000. ...  2530.     0.  3907.]
 [30000. 30000. 30000. ...  1477.  3907.     0.]]
```

Figure 5: distance matrix of TSP read by Panda

```
Objective: 68474 miles

Route for vehicle 0:
 0 -> 10 -> 1 -> 4 -> 14 -> 3 -> 18 -> 9 -> 5 -> 36 -> 37 -> 35
-> 32 -> 31 -> 33 -> 30 -> 29 -> 28 -> 27 -> 34 -> 24 -> 25 ->
23 -> 26 -> 39 -> 42 -> 38 -> 43 -> 44 -> 40 -> 41 -> 46 -> 45
-> 69 -> 67 -> 64 -> 65 -> 60 -> 63 -> 61 -> 62 -> 68 -> 66 ->
47 -> 53 -> 54 -> 55 -> 52 -> 51 -> 50 -> 49 -> 48 -> 56 -> 57
-> 59 -> 58 -> 8 -> 13 -> 7 -> 6 -> 17 -> 21 -> 22 -> 15 -> 12
-> 16 -> 11 -> 19 -> 20 -> 2 -> 0
```

Figure 6: Result from TSP

Finally, adding the distance between continent with result together, we have the final total distance to be 96,832km.

Nearest Neighbor Search (NNS)

In contrast to TSP, we also include the nearest neighbor algorithm to calculate the best route. After converting the latitude and longitude of each city into coordinates, the cities can plotted on an x-y plane.At the same time, it produce a 70×70 distance matrix. Under the nearest neighbor algorithm, all paths between cities are available. The distance matrix will be:

```
[[  0.           3.52796629   3.37623988 ...  81.32325763 112.63237968
   64.89723306]
 [  3.52796629   0.           6.90354999 ...  84.4356166  115.88555087
   67.89785176]
 [  3.37623988   6.90354999   0.         ...  78.33253568 109.49868138
   62.02636901]
 ...
 [ 81.32325763  84.4356166   78.33253568 ...   0.          32.65793498
   17.14298072]
 [112.63237968 115.88555087 109.49868138 ...  32.65793498   0.
   49.76263423]
 [ 64.89723306  67.89785176  62.02636901 ...  17.14298072  49.76263423
    0.        ]]
```

Figure 7: distance matrix of NNS

After running the nearest neighbor code, it results in the following route, the nodes on the left represent cities in North and South America:

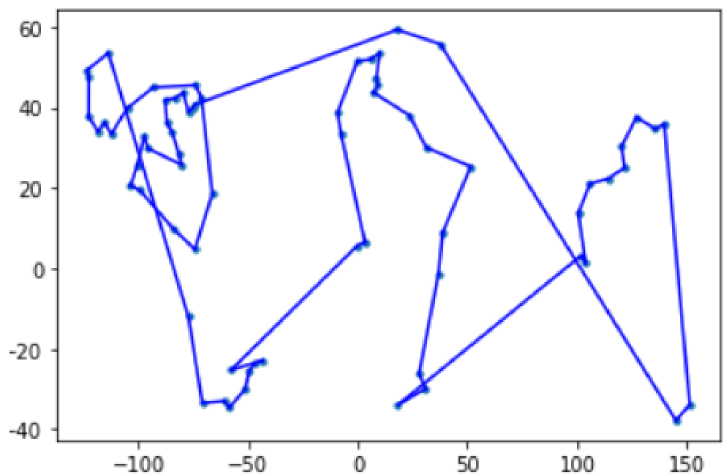

Figure 8: Result of NNS

The Nearest Neighbor code also produces the total distance of 112908 km of the optimal route, and the visiting order of cities:

$0 \rightarrow 10 \rightarrow 1 \rightarrow 19 \rightarrow 11 \rightarrow 16 \rightarrow 18 \rightarrow 3 \rightarrow 14 \rightarrow 4 \rightarrow 5 \rightarrow 9 \rightarrow 36 \rightarrow 37 \rightarrow 35 \rightarrow 32 \rightarrow 33 \rightarrow 31 \rightarrow 2 \rightarrow 20 \rightarrow 12 \rightarrow 15 \rightarrow 13 \rightarrow 7 \rightarrow 8 \rightarrow 6 \rightarrow 17 \rightarrow 21 \rightarrow 22 \rightarrow 30 \rightarrow 29 \rightarrow 28 \rightarrow 27 \rightarrow 24 \rightarrow 25 \rightarrow 23 \rightarrow 26 \rightarrow 34 \rightarrow 46 \rightarrow 41 \rightarrow 45 \rightarrow 69 \rightarrow 60 \rightarrow 63 \rightarrow 61 \rightarrow 65 \rightarrow 64 \rightarrow 67 \rightarrow 66 \rightarrow 40 \rightarrow 47 \rightarrow 44 \rightarrow 43 \rightarrow 38 \rightarrow 42 \rightarrow 39 \rightarrow 56 \rightarrow 57 \rightarrow 48 \rightarrow 49 \rightarrow 50 \rightarrow 51 \rightarrow 52 \rightarrow 53 \rightarrow 55 \rightarrow 54 \rightarrow 58 \rightarrow 59 \rightarrow 68 \rightarrow 62$

4 Presentation of result

After running the TSP code and the NNS code, we found two paths. By plotting these cities on google my maps to show the exact paths obtained from these two methods. Note that these two paths both start in New York.

Figure 9: Visualization on world map

Under the Travel salesman method, we set up some restriction to the route between cities on different continents. For example, the permitted route from North America to South America is from Houston to Monterrey, and the permitted route from South America to Africa is from Rio de Janeiro to Cape Town. The path shows that it will first visit some cities in the united states and then directly visit all cities in South America, Africa, Europe, Middle Asia, east Asia, north Asia, and Australia, and the remaining cities in North America. And finally, back to New York. The total distance of this tour is 99466 km in map (the distance by running TSP code is 96167km)

Under the nearest neighbor method, we compute the optimal tour based on the city's latitude and longitude. It first visits 12 cities in the united states and then goes to South America. After visiting some cities in South America, the tour returns to North America

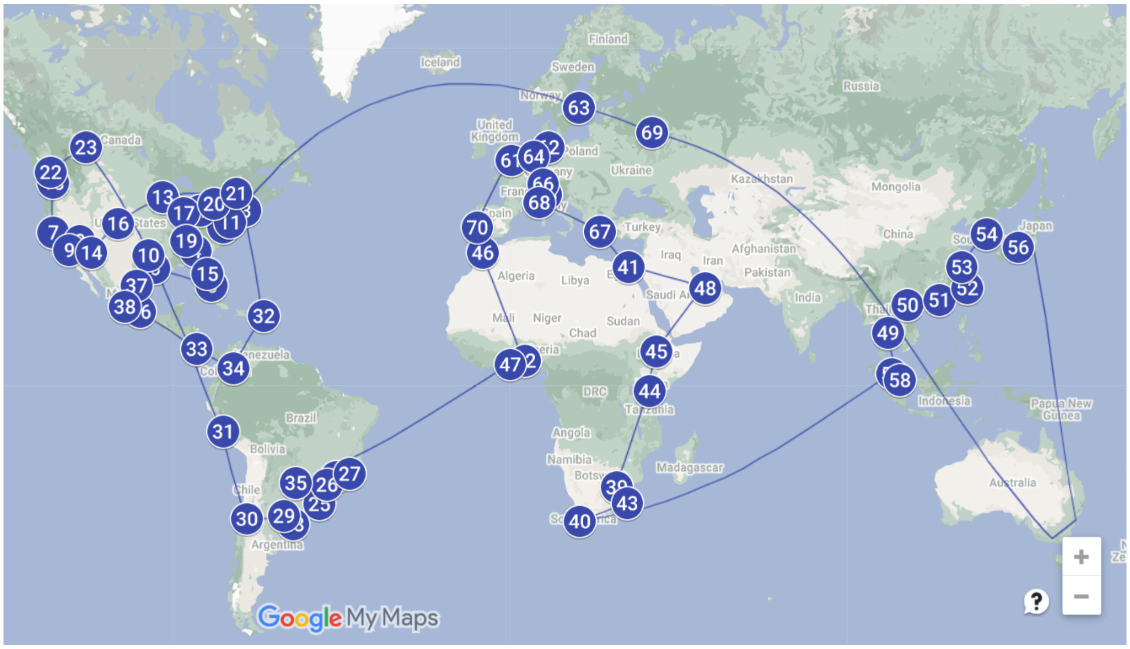

Figure 10: Visualization of NNS

again. Cities in the same continent or area might not be visited continuously, such as city 21 Montreal in North America and city 69 Moscow. We can see that the tour goes from Melbourne to Moscow rather than from a Europe city. The total distance of this tour is 112951km in map (112908 km by running nearest neighbor code) And Compared to the tour from the travel salesman method, one significant difference is that some routes are across other tours.

According to the route comparison of the Travelling Salesman problem and the Nearest Neighbor Method, The project eventually chose the shortest path for the globe trip, which has a total distance of 96167 Km. In terms of the traveling cost, it depends on the total traveling distance for approaching the final total profit.

In this step, the objective function would be:

$$\text{Total profit} = \text{Raw Profit} - \text{Traveling cost}$$
$$\text{Raw profit} = \text{Ticket Revenue } r_i - \text{Staging Cost } c_i$$
$$\text{Traveling cost} = \text{Fuel price } p \ * \text{Traveling Distance } d$$

$$\text{Maximize}: \ (r_i - c_i)x_i - p * d$$
$$\textstyle\sum_{i=1}^{91} x_i = 70, \quad i = 1, \ldots, 91;$$
$$x_i \in \{0, 1\}$$

By the report *The Cost Of Flying [4] : What Airlines Have To Pay To Get You In The Air*, the project defines the fuel fee by an example of the average price for fuel for a London to New York flight as $33,411, and the traveling distance is 5576km. Therefore, the fuel price p is calculated by $33,411/5576$, about 5.99 dollars per kilometer.

The final result of the total profit for the designed World tour is $232586315.0 by using the TSP route and the selected 70 high-reward cities crossing six continents.

5 Result Analysis

These two methods give us two different paths. And the difference in the total distances between these two paths is more than 10,000 kilometers.

Computations of the optimal path. The TSP method is finding the optimal route by comparing all routes one by one. Any two routes will not be crossed. While under the nearest neighbor method, some routes are crossed. For example, the route from city 5(Miami) to city 6(Houston) and the route from city 23(Edmonton) and city 31(Lima). This is because the path will only choose the closest city as the next destination. It will also result in some routes being much longer than others routes such as the route from city 60 to city 69, which across three continents, Oceania, Asia, and Europe.

Data bias. Both methods require a 70 times 70 distance matrix to find the optimal route. There are some restrictions on the route connecting two different continents. Under the TSP computation, the distance matrix can be regarded as a composition of four different distances matrix. The distance between cities in different continents is either 0 or 30000. If it the distance is 0 it means that this route must be included in the optimal route. The distance between cities in the matrix might differ from the exact distance in reality. For the nearest neighbor method, the latitude and longitude of each city is converted to two-dimensional coordinate. After the data transformation, a new distance matrix is created and input this matrix to the nearest neighbor code to get the tour. Hence, data using in these two methods might be biased.

Theoretical tour or Realistic tour. Under the TSP computation, it also considers the situation that some routes do not exist in reality. And the distances of these routes are 30000km so that the code will not select these routes as one of the routes in the optimal tour. The nearest neighbor method will automatically compare all possible routes from city i to other cities and selects the nearest city as the next destination from city i. Hence, the optimal tour may consist of some routes that do not exist in reality under nearest neighbor method.

91 cities were selected initially with approximately 21 cities in each continent, and 70 cities remained after performing the *PuLP* algorithm based on the profit earning ability. It is interesting to notice that the cities that were not selected are all from Asia Pacific and Europe. Specifically, the amounts of city in the Asia Pacific and Europe reduced from 22 to 13 and 10 respectively. By looking at these two continents we found some very similar characteristics: both two continents have the lowest average ticket price; their venue capacity no bigger than 21k (Oceania will be discussed separately as it behalf as outliers). As we are maximizing the total profit earned (that is Revenue from ticket sales – Staging cost), when the staging cost is constant, lower capacity and fares will accelerate the rate of decline in total revenue, ultimately leading to a lack of competitiveness in the city's profits.

For the first factor, the ticket price in each continent is calculated from the *Born This Way* Concert in 2014 as well as total inflation rate in 8 years. Compared to North America as the majority source of Lady Gaga's fans, which usually have the most amounts of concert stations, crew members tend to choose more economically developed cities in other continents. The stable political situation in Eurasia has resulted in lower inflation rates than in Latin America and Africa, which are relatively volatile political, over 8 years. For example, assuming the exchange rate to USD remains same, although the average ticket price in South Africa was almost same with Eurasia, remaining at $60, the aggregate inflation rate in Africa has exceeded 50% in eight years, compared to about 10% in the latter. In general, higher inflation will inevitably lead to higher total returns, given the same capacity. Venue capacity strongly correlated to the total revenue. The lower the total seats the lower the total revenue. The capacity in all 21 cities from Eurasia have no more than 21k. Although all cities in the Oceania have venues with more than 30k capacity, their too low entrance fees (40$) make them uncompetitive.

There are many other factors need to be considered in order (Limitations) to improve our accuracy include:

- *More factors in the PuLP algorithm.* This includes cost of crew members, hotel fee and cost of renting equipment etc.;

- *More potential cities as options.* Selecting 70 out of 91 cities isn't really a big scale and resulted in some bias (i.e. no city in North America is crossed out at all);

- *Estimating travelling cost based on real flight route.* In this project we assume the cost is directly proportional to the price and set a fixed price $a /km. This may lead to errors as consumption level vary greatly from region to region;

- *More accurate staging cost.* This is for professional purpose. It is hard to access the real staging cost of each venue since the data is not public;

- *Popularity.* This project is completed entirely based on economic factors. We are not considering the popularity of Lady Gaga in each area;

- *Ticket Sales.* We are assuming that all tickets in each venue are sold out.

- *Data References.* The past data are cited from Wikipedia- world tour of Lady Gaga page. We cannot guarantee their accuracy.

6 Appendix

General List of Data

1. Big Data Worldwide (Distance matrix)

2. City list and coordinates (city, city index, longitude, latitude)

3. Ticket Revenue and Staging Cost

Algorithm

1. *puLP*: pulp world tour.ipynb

   Data File: pulp world tour.csv

2. Travelling Salesman Problem: TSP.ipynb

   Data File: Big Data TSP.csv

3. Nearest Neighbor Method: Nearest.ipynb

   Data File: nearest nodes.csv

Data Visualization

1. TSP Travelling Map

   Google my map file: TSP world map

2. Nearest Neighbor Method Map

   Google my map file: Nearest Neighbor map

3. Raw Profit Visualization: Profit Data Visualization.ipynb

   Data File: profit.csv

7 Reference

References

[1] Born this way ball.(2022, December 05), Retrieved December 8, 2022, from https://en.wikipedia.org/wiki/Born This Way Ball

[2] The long term perspective on markets. (n.d.),Retrieved December 8, 2022, Retrieved December 8, 2022, from https://www.macrotrends.net/

[3] Traveling salesperson problem, OR-tools,google developers. (n.d.),Retrieved December 8, 2022, fromhttps://developers.google.com/optimization/routing/tsp

[4] Hayward, J., Ahlgren, L.(2022, June 30), The cost of flying: What airlines have to pay to get you in the Air, Retrieved December 8, 2022, from https://simpleflying.com/the-cost-of-flying/