2.DATA

In order to collect useful data,

- 1) we need to define the basic venues for foreign tourists. I chose 8 venues.
 - Hotels
 - Restaurants (all categories)
 - Convenience stores
 - ATMs/Exchanges (only ATMs which accepts international cards)
 - Cafes
 - Parks
 - Pharmacies
 - Tourist Information Center
- 2) we need the geographical coordinates of all the stations (all the lines) in Tokyo. Most of the tourists may travel by trains or subways because traveling with cars and taxis in Tokyo are not popular options as there are limited parking spaces and using taxis is pricey. I scraped the station names from Wikipedia and obtained latitude and longitude through GeoPy library. (Table 2-1 / Map 2-1)

	station	latitude	longitude
0	Adachi-Odai Station	35.754801	139.770404
1	Aihara Station	35.606819	139.331686
2	Akabane Station	35.778139	139.720800
3	Akabane-iwabuchi Station	35.782968	139.719853
4	Akabanebashi Station	35.654987	139.743891
637	Zõshiki Station	35.550041	139.715202
638	Itabashi Kuyakushomae Station	35.751809	139.709718
639	Shimo-Shimmei Station	35.608871	139.726232
640	Shōin-Jinjamae Station	35.644143	139.655263
641	Yaguchinowatashi Station	35.562720	139.700291



Map 2-1 Map of Stations in Tokyo

642 rows × 3 columns

Table 2-1 Geographical Coordinates of Stations in Tokyo

3) Search each basic venue data 1000m radius from the all the stations using FourSquareAPI. I used 'search' endpoint, which returns maximum 50 data that matches its categoryld. Note that FourSquareAPI returns some data categorized wrongly so need to check if the data represents its category and is divided into too many small categories so I re-categorized them into one class for each. (e.g. Fast-food restaurant, Japanese restaurants, French restaurant, → 'Restaurants'). And as for the ATMs data, in Japan not all the ATMs accepts international cards. In addition to 7-Eleven and Post office (where the international cards can be used), I used the E.net website to scrape the name of the locations of their ATMs and added them to the data. To align with other

data, the stations that have more than 50 data because of this addition, I dropped the exceeded data. Now all the station names have 0 to 50 data for each basic venue. (Table 2-2)

	station	id	name	venue_lat	venue_Ing	categories	classes		
0	Adachi-Odai Station	4d21b00af7a9a1437e48389f	Ryota Kuga's Guest House	35.746466	139.771546	Hostel	Hotels		
1	Adachi-Odai Station	59a382f5cad1b628d2f030f5	東京ゲストハウス2020			Hostel	Hotels		
2	Adachi-Odai Station	53f45d47498e66d089216755	Tokyo Guest House B&B Hostel (東京ゲストハウス B&B ホステル)	35.743172	139.771286	Hostel	Hotels		
3	Aihara Station	4ce533fa5fce5481a53d5aaa	Laxio-Inn (ホテル ラクシオ・イン)	35.610112	139.344740	Hotel	Hotels		
4	Aihara Station	5b6c2cf9e65d0c002ced1699	ホテル anniversary	35.611662	139.343505	Hotel	Hotels		
139987	Yushima Station	5b0e7ba6a2a6ce002c79fc2b	Ueno Information Center	35.710779	139.775775	Tourist Information Center	Tourist Information Center		
139988	Yushima Station	4ce14895c9a0a0903596246a	Tokyo Tourist Information Center (東京観光情報センター)	35.710783	139.773482	Tourist Information Center	Tourist Information Center		
139989	Yushima Station	57368551498ed9b3ec5509db	Park Information Center (公園案內所)	35.714783	139.775900	Tourist Information Center	Tourist Information Center		
139990	Yushima Station	5e4256344c4a85000888edc8	General Information Center (総合案内所)			Tourist Information Center	Tourist Information Center		
139991	Zoshigaya Station	5bca6a8c3c858d002c6e93ac	Tobu Tourist Information Center (東武ツーリストインフォメー	35.730572	139.710577	Tourist Information Center	Tourist Information Center		
139992 rows × 7 columns									

Table 2-2 List of Venues for Tourists for each station