Exploring London Tube Neighborhood using web and Foursquare location data: Data Science Real Life

Data Preparation

Scraping London Underground Station data from Wikipedia

First used List of London Underground stations page from Wiki to scrap the table to create a dataframe

	Station	Photograph	Line(s)[*]	Local authority	Zone(s) [†]	Opened[4]	Main lineopened	Other name(s)[note 2]	Usage[5]
1	Acton Town	NaN	DistrictPiccadilly	Ealing	3	1 July 1879	NaN	Mill Hill Park: 1879–1910	5.99
2	Aldgate	NaN	Metropolitan[a]Circle	City of London	1	18 November 1876	NaN	NaN	8.47
3	Aldgate East	NaN	Hammersmith & City[d]District	Tower Hamlets	1	6 October 1884resited 31 October 1938	NaN	Commercial Road: Proposed before opening	13.17
4	Alperton	NaN	Piccadilly[h]	Brent	4	28 June 1903	NaN	Perivale-Alperton: 1903–10	2.82
5	Amersham	NaN	Metropolitan	Chiltern	9	1 September 1892	NaN	Amersham: 1892–1922Amersham & Chesham Bois: 19...	2.10

After little manipulation of the data, the dataframe is obtained as below

	Station	Lines	Local authority	Zones
1	Acton Town	DistrictPiccadilly	Ealing	3
2	Aldgate	MetropolitanCircle	City of London	1
3	Aldgate East	Hammersmith & CityDistrict	Tower Hamlets	1
4	Alperton	Piccadilly	Brent	4
5	Amersham	Metropolitan	Chiltern	9

Next used Geo details of London Underground stations to get the Coordinates and created the below dataframe from the table

	Name	Latitude	Longitude	Line
1	Acton Town	51.502500	-0.278126	District, Piccadilly
2	Acton Central	51.50883531	-0.263033174	London Overground
3	Acton Central	51.50856013	-0.262879534	London Overground
4	Aldgate	51.51394	-0.07537	Metropolitan
5	Aldgate East	51.51514	-0.07178	District, Hammersmith & City

Joined the above 2 dataframes to obtain the dataframe as below

	Station	Lines	Local authority	Zones	Name	Latitude	Longitude	Line
0	Acton Town	DistrictPiccadilly	Ealing	3	Acton Town	51.502500	-0.278126	District, Piccadilly
1	Aldgate	MetropolitanCircle	City of London	1	Aldgate	51.51394	-0.07537	Metropolitan
2	Aldgate East	Hammersmith & CityDistrict	Tower Hamlets	1	Aldgate East	51.51514	-0.07178	District, Hammersmith & City
3	Alperton	Piccadilly	Brent	4	Alperton	51.54097	-0.30061	Piccadilly
4	Amersham	Metropolitan	Chiltern	9	Amersham	51.67435	-0.60732	Metropolitan

Dropped the common columns to obtain the final dataframe for analysis

	Station	Lines	Local authority	Zones	Latitude	Longitude
0	Acton Town	DistrictPiccadilly	Ealing	3	51.502500	-0.278126
1	Aldgate	MetropolitanCircle	City of London	1	51.51394	-0.07537
2	Aldgate East	Hammersmith & CityDistrict	Tower Hamlets	1	51.51514	-0.07178
3	Alperton	Piccadilly	Brent	4	51.54097	-0.30061
4	Amersham	Metropolitan	Chiltern	9	51.67435	-0.60732

After little more playing around with pandas, could get one well-arranged dataframe as below where duplicate records and removed and Latitude & Longitude values were standardized

	Station	Lines	Local authority	Zones	Latitude	Longitude
0	Acton Town	DistrictPiccadilly	Ealing	3	51.5025	-0.278126
1	Aldgate	MetropolitanCircle	City of London	1	51.51394	-0.07537
2	Aldgate East	Hammersmith & CityDistrict	Tower Hamlets	1	51.51514	-0.07178
3	Alperton	Piccadilly	Brent	4	51.54097	-0.30061
4	Amersham	Metropolitan	Chiltern	9	51.67435	-0.60732

Foursquare Location Data

Foursquare outlines high-level venue categories with sub-categories, used Foursquare API to explore food venues surrounding each station using **Food sub-category id (4d4b7105d754a06374d81259)**. Queried 100 popular spots in a 1000m radius around each station. This radius was chosen because 1000m is a reasonable walking distance within each station.

Created a function which returns all the Food Venues by category for the neighborhood passed as an argument. The call returns a JSON file is converted into a dataframe

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Zones
0	Acton Town	51.5025	-0.278126	The Apple Tree Cakes	51.503174	-0.280474	Coffee Shop	3
1	Acton Town	51.5025	-0.278126	WP Fish X Chips	51.502580	-0.281240	Fish & Chips Shop	3
2	Acton Town	51.5025	-0.278126	London Transport Museum Depot	51.504175	-0.280622	Museum	3
3	Acton Town	51.5025	-0.278126	Gunnersbury Park Museum	51.499661	-0.286223	Museum	3
4	Acton Town	51.5025	-0.278126	Acton Centre	51.506608	-0.266878	Gym / Fitness Center	3

Since I am interested in Restaurants & Bars as popular spots, filtered the data frame with **Venue Category** containing the word **Restaurant** and **Bar**

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Zones
13	Acton Town	51.5025	-0.278126	casereccio	51.502436	-0.281674	Italian Restaurant	3
16	Acton Town	51.5025	-0.278126	The Corner Terrace	51.509839	-0.286459	Eastern European Restaurant	3
18	Acton Town	51.5025	-0.278126	Amigo's Peri Peri	51.508396	-0.274561	Fast Food Restaurant	3
28	Acton Town	51.5025	-0.278126	Persian Nights	51.508529	-0.282383	Middle Eastern Restaurant	3
29	Acton Town	51.5025	-0.278126	North China Restaurant	51.508251	-0.277435	Chinese Restaurant	3

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Zones
35	Acton Town	51.5025	-0.278126	The Chatsworth	51.508373	-0.276460	Cocktail Bar	3
53	Aldgate	51.51394	-0.07537	Discount Suit Company	51.516705	-0.075506	Cocktail Bar	1
96	Aldgate	51.51394	-0.07537	Kill The Cat	51.518658	-0.071516	Beer Bar	1
98	Aldgate	51.51394	-0.07537	BrewDog	51.509948	-0.080977	Beer Bar	1
105	Aldgate	51.51394	-0.07537	citizenM 7th Floor Skybar	51.510049	-0.076615	Hotel Bar	1