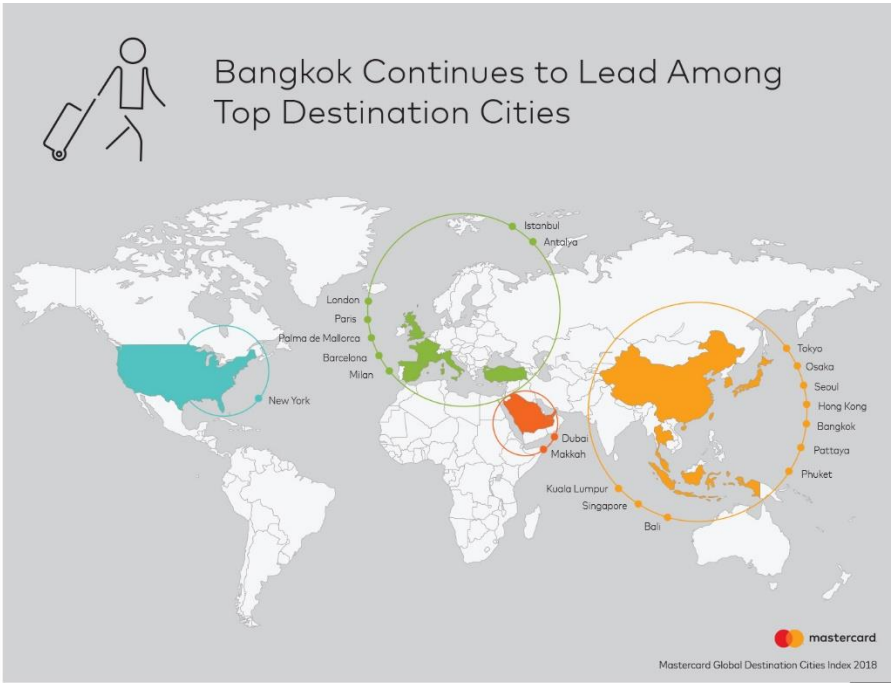
LOCATION INSIGHTS FROM TOP TRAVEL DESTINATIONS

Victor Mawusi Ayi

Travel has become an important contributor to economic growth in countries that are lucky to count some of their cities as top destinations. Some cities have had annual visitor counts in excess of 15 million and revenues in excess of 15 billion dollars.

Rank	City	Total International Visitors
01	Bangkok	20.05 mn
02	London	19.83 mn
03	Paris	17.44 mn
04	Dubai	15.79 mn
05	Singapore	13.91 mn
06	New York	13.13 mn
07	Kuala Lumpur	12.58 mn
08	Tokyo	11.93 mn
09	Istanbul	10.70 mn
10	Seoul	9.54 mn
11	Antalya	9.42 mn
12	Phuket	9.29 mn
13	Makkah	9.18 mn
14	Hong Kong	9.03 mn
15	Milan	8.81 mn
16	Palma de Mallorca	8.78 mn
17	Barcelona	8.69 mn
18	Pattaya	8.67 mn
19	Osaka	8.42 mn
20	Bali	8.30 mn



Ranking of most visited cities and the total visitor count in 2017.

Taken from *https://newsroom.mastercard.com/wp-content/uploads/2018/09/GDCI-Map.jpg*

According to the MasterCard Global Destination Index, Dubai, supposedly the highest earner from visitor spending, realized $29.70 billion from international overnight spending in 2017. That is great revenue, given that visitors also spent an average of $537 daily. The figures for other high ranking cities are equally astounding and they are expected to grow with time.

	City	VisitorCount	GrowthForecast	AvgLengthOfStay	AvgSpendPerDay
0	Bangkok	20.05 million	9.6%	4.7 nights	$173
1	London	19.83 million	3.0%	5.8 nights	$153
2	Paris	17.44 million	2.9%	2.5 nights	$301
3	Dubai	15.79 million	5.5%	3.5 nights	$537
4	Singapore	13.91 million	4.0%	4.3 nights	$286
5	New York	13.13 million	4.1%	8.3 nights	$147
6	Kuala Lumpur	12.58 million	7.5%	5.5 nights	$124
7	Tokyo	11.93 million	1.6%	6.5 nights	$154
8	Istanbul	10.70 million	19.7%	5.8 nights	$108
9	Seoul	9.54 million	6.1%	4.2 nights	$181

This is enjoyment for the most visited cities, which is expected to increase as international travel surges on the back of globalization. However, there is a larger opportunity for other cities that may have loveable resources and the potential. For these upcoming cities, a lot needs to be learned from the top cities.

The top motivations for travel relate to business and leisure. These, in turn, hinge on food, entertainment, hospitality and shopping among others. Consequently, the distribution of locations that provide these services should

churn out a lot of useful insights about top ranking cities which can be applied to other potential cities.

OBJECTIVES

This project:

a. Explores tourism location features of top 10 cities, including similarities and dissimilarities
b. Leverages location data to reveal:
    - Patterns of distribution of key services among the most visited cities in the world
    - Similarities between most visited cities in the world
    - Differences across most visited cities in the world
c. Derive insights which can be applied to improve traveler attraction and visitation rates of potential cities.

DATA REQUIREMENTS

This research is exploratory. On that premise, the data requirements focused on the ranking of the top ten cities, and location data that was just enough to have a peek into how indicators like travel spending, length of stay, among others, interact with each other and influence attraction of travelers.

Data of Top 10 cities from the MasterCard Global Destination Index

The MasterCard Global Travel Destination Index ranks the most visited cities in the world and provides data on the annual number of visits, average spending

by visitors per day, average length of stay and the growth forecasts. For the study, the 2018 index based on 2017 travel data was used.

Taken from: *https://newsroom.mastercard.com/wp-content/uploads/2017/09/GDCIBannerDPK.jpg*

Geographical Data on top ten cities

Geographical coordinates are requisite for accessing locations data. The geocoder API proved useful for grabbing geographical coordinates, using just the location name.

Location Data on Top 10 Cities using the Foursquare API

Location data from Foursquare represents the main foundation of this project. Foursquare maintains a comprehensive, frequently updated database on locations around the world.

METHODOLOGY

Tools and Methods

The major tools employed are as follows:

- Pandas for data cleaning and preparation, and descriptive statistics
- Matplotlib for Data visualization
- Folium for maps
- Scikit Learn for Machine Learning
- Scipy for Statistics
- Geopy geocoder for finding location coordinates

Data Collection and Cleaning

MasterCard Ranking Data:

The MasterCard index ranking data was obtained by scraping the press release page located at:

https://newsroom.mastercard.com/press-releases/big-cities-big-business-bangkok-london-and-paris-lead-the-way-in-mastercards-2018-global-destination-cities-index/

The data was converted into a pandas dataframe and cleaned.

Location Data:

Location data were obtained through authenticated requests to the foursquare API, using the search endpoint with the 'browse' intent.

Using filtered venue categories to query tourism related venues and services

In [123]:
```
top_cities_fsq=nearby_services(geo_data=geodata1, CATEGORYID=cat_codes)

top_cities_fsq['venues'][0:10]
```

Out[123]:
```
[['Bangkok',
  13.7538929,
  100.8160803,
  2,
  'ARL Latkrabang (A2) (ARL ลาดกระบัง)',
  13.727637765326877,
  100.74840786863355,
  'Train Station',
  '4bf58dd8d48988d129951735'],
 ['Bangkok',
  13.7538929,
  100.8160803,
  2,
  'KMITL Walking Street (สถานีคนเดิน พระจอมเกล้าฯลาดกระบัง (KMITL Walking Street))',
  13.728510044750138,
  100.77324827367164,
  'Pedestrian Plaza',
  '52e81612bcbc57f1066b7a25'],
 ['Bangkok',
  13.7538929,
  100.8160803,
  2,
```

Data Analysis

The analysis was predominantly exploratory. The objective was to explore possible trends in tourism and use outcomes to inform a more extensive research.

Analysis began with descriptive analysis. Insights were gathered from the means, median and interquartile ranges. This was helpful in finding the outliers and extreme high and low performers.

Subsequently, bar charts and scatter plots were used to visualize possible relationships between the tourism indicators captured in the index ranking:

- Average length of stay
- Average spend per day
- Visitor count

Plotting the cities on a map allowed for the visualization of the continental share of high ranking cities, and the distribution of the top destinations across the world.

Furthermore, K-means clustering was employed as part of exploratory analysis to find similarities among top cities based on just the indicators in the Mastercard index. However, it would be employed again to find similarities based on location data.

The main analysis leveraged location data using the foursquare API. The search endpoint was utilized in finding venues within an 8000 meter radius for each top city.

The search was narrowed to just venue categories which were relevant to tourism.

Filtering out foursquare categories that are linked to tourism. This was manually compiled.

```
category_dict = {'Arts & Entertainment':'4d4b7104d754a06370d81259', 'Food':'4d4b7105d754a06374d81259',
 'Nightlife & Spot':'4d4b7105d754a06376d81259', 'Outdoors & Recreation':'4d4b7105d754a06377d81259',
 'Shop & Service':'4d4b7105d754a06378d81259','Hotel':'4bf58dd8d48988d1fa931735',
 'Transportation Service':'54541b70498ea6ccd0204bff', 'Travel Lounge':'4f04b25d2fb6e1c99f3db0c0',
'Rental Car Location':'4bf58dd8d48988d1ef941735','Rest Area':'4d954b16a243a5684b65b473',
 'Food & Drink Shop':'4bf58dd8d48988d1f9941735', 'ATM':'52f2ab2ebcbc57f1066b8b56',
 'Adult Boutique':'5267e446e4b0ec79466e48c4', 'Antique Shop':'4bf58dd8d48988d116951735',
 'Arts & Crafts Store':'4bf58dd8d48988d127951735','Bank':'4bf58dd8d48988d10a951735',
 'Business Service':'5453de49498eade8af355881','Candy Store':'4bf58dd8d48988d117951735',
 'Chocolate Shop':'52f2ab2ebcbc57f1066b8b31','Clothing Store':'4bf58dd8d48988d103951735',
 'Convenience Store':'4d954b0ea243a5684a65b473','Cosmetics Shop':'4bf58dd8d48988d10c951735',
 'Costume Shop':'52f2ab2ebcbc57f1066b8b17', 'Discount Store':'52dea92d3cf9994f4e043dbb',
 'Electronics Store':'4bf58dd8d48988d122951735','Entertainment Service':'56aa371be4b08b9a8d573554',
 'Event Service':'5454152e498ef71e2b9132c6','Fabric Shop':'52f2ab2ebcbc57f1066b8b26',
 'Film Studio':'56aa371be4b08b9a8d573523','Gaming Cafe':'4bf58dd8d48988d18d941735',
 'Gift Shop':'4bf58dd8d48988d128951735', 'Massage Studio':'52f2ab2ebcbc57f1066b8b3c',
 'Outdoor Supply Store':'52f2ab2ebcbc57f1066b8b22','Outlet Mall':'5744ccdfe4b0c0459246b4df',
 'Outlet Store':'52f2ab2ebcbc57f1066b8b35','Pop-Up Shop':'52f2ab2ebcbc57f1066b8b3d',
 'Shopping Mall':'4bf58dd8d48988d1fd941735','Shopping Plaza':'5744ccdfe4b0c0459246b4dc',
 'Smoothie Shop':'52f2ab2ebcbc57f1066b8b41','Souvenir Shop':'52f2ab2ebcbc57f1066b8b1b',
 'Spa':'4bf58dd8d48988d1ed941735','Thrift / Vintage Store':'4bf58dd8d48988d101951735',
 'Travel & Transport':'4d4b7105d754a06379d81259'}
```

Using the one-hot method available for pandas dataframes, cities were matched against the number of venues for each category. The categories were further grouped under one of the sectors of tourism as follows:

- Transport – included lounges, stations etc
- Food – included venues for food and drink like restaurants
- Shopping – included shopping malls, convenience stores etc
- Lodging-included hotels, motels, etc
- Recreation – included places for relaxation and entertainment

The means for each sector was obtained for each city, standardized and used for another K-means clustering.

	City	Food	Lodging	Not Hospitality	Recreation	Shopping	Transport
0	Bangkok	0.42	0.00	0.14	0.08	0.30	0.06
1	Dubai	0.28	0.10	0.10	0.26	0.18	0.08
2	Istanbul	0.14	0.00	0.30	0.40	0.12	0.04
3	Kuala Lumpur	0.22	0.00	0.10	0.12	0.30	0.26
4	London	0.18	0.04	0.08	0.26	0.12	0.32
5	New York	0.00	0.00	0.00	0.32	0.12	0.56
6	Paris	0.10	0.00	0.14	0.48	0.12	0.16
7	Seoul	0.00	0.00	0.08	0.24	0.16	0.52
8	Singapore	0.08	0.04	0.08	0.14	0.32	0.34
9	Tokyo	0.00	0.00	0.04	0.08	0.02	0.86

The resulting clusters were explored for relevant trends and the findings are reported in the following section.

FINDINGS

The analysis yielded interesting findings, some of which align with the upheld strategic recommendations for tourism.

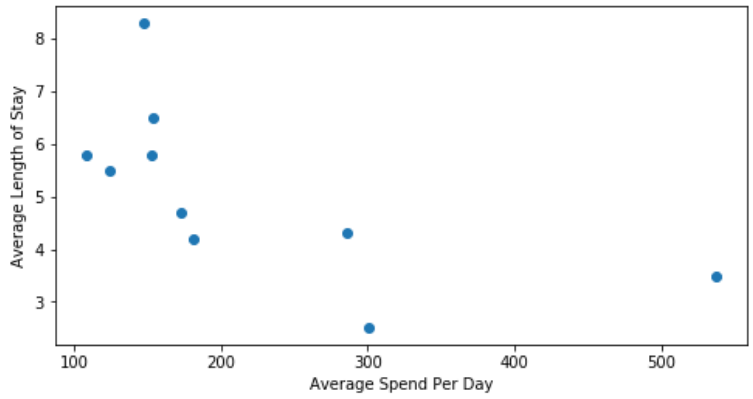
Finding 1.

Majority of the 2017 top 10 destinations were in Asia. Cities in Asia may have better advantage with improved focus on tourism development.

Finding 2.

People generally stay longer in cities where spending is less. A non-linear relationship which is statistically significant is demonstrable.

```
Pearson Correlation: Coefficient = -0.6301341589244333, p-value = 0.050840480068283116
SpearmanrResult(correlation=-0.8024353176576852, pvalue=0.005211461766706417)
```



Below is the Spearman Rank Correlation table further demonstrating the relationship between average spend per day and average length of stay.

Spearman Rank Correlation

```
top_cities.corr(method='spearman')
```

	VisitorCount	GrowthForecast	AvgLengthOfStay	AvgSpendPerDay
VisitorCount	1.000000	-0.200000	-0.243162	0.369697
GrowthForecast	-0.200000	1.000000	-0.024316	-0.381818
AvgLengthOfStay	-0.243162	-0.024316	1.000000	-0.802435
AvgSpendPerDay	0.369697	-0.381818	-0.802435	1.000000

Finding 3.

There is a positive linear relationship between lodging and average spend per day. In other words, people spend more on lodging, and cities that prioritize lodging facilities tend to be more expensive to visit.

Deriving Pearson Correlation statistics

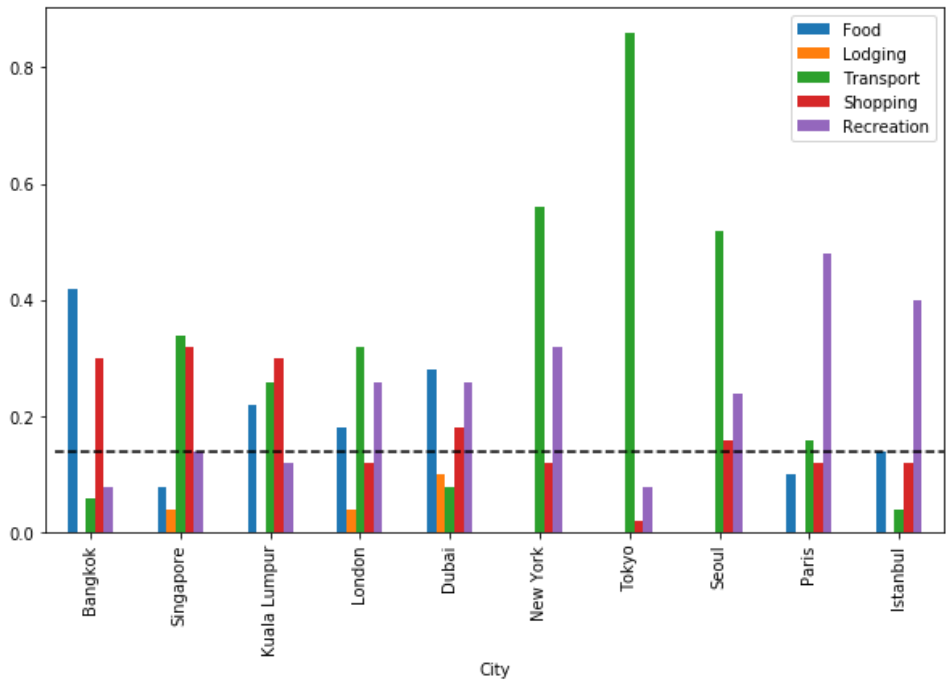
To assess linear relationships

```
fsq_cities_seccorr.corr(method='pearson')
```

	Food	Lodging	Not Hospitality	Recreation	Shopping	Transport	ClusterLabels	GrowthForecast	VisitorCount	AvgSpendPerDay	AvgLengthOfStay
Food	1.000000	0.312077	0.354425	-0.256984	0.557925	-0.747748	-0.556579	0.315073	0.652116	0.241321	-0.316207
Lodging	0.312077	1.000000	-0.111852	-0.020661	0.131929	-0.294662	-0.318697	-0.205237	0.291765	0.833254	-0.337232
Hospitality	0.354425	-0.111852	1.000000	0.393842	0.070510	-0.705358	0.302808	0.871161	-0.047014	-0.093959	-0.308268
Recreation	-0.256984	-0.020661	0.393842	1.000000	-0.429392	-0.337282	0.772373	0.215792	-0.014991	0.149161	-0.200208
Shopping	0.557925	0.131929	0.070510	-0.429392	1.000000	-0.485350	-0.805039	0.151268	0.206654	0.117023	-0.298357
Transport	-0.747748	-0.294662	-0.705358	-0.337282	-0.485350	1.000000	0.143984	-0.572308	-0.433620	-0.324106	0.517687
ısterLabels	-0.556579	-0.318697	0.302808	0.772373	-0.805039	0.143984	1.000000	0.186061	-0.351510	-0.108619	0.038046
rthForecast	0.315073	-0.205237	0.871161	0.215792	0.151268	-0.572308	0.186061	1.000000	-0.266352	-0.286297	0.061204
isitorCount	0.652116	0.291765	-0.047014	-0.014991	0.206654	-0.433620	-0.351510	-0.266352	1.000000	0.248611	-0.270630
endPerDay	0.241321	0.833254	-0.093959	0.149161	0.117023	-0.324106	-0.108619	-0.286297	0.248611	1.000000	-0.630134
ngthOfStay	-0.316207	-0.337232	-0.308268	-0.200208	-0.298357	0.517687	0.038046	0.061204	-0.270630	-0.630134	1.000000

Finding 4.

Transportation is a key sector for top cities. This is followed by recreation and shopping. Emerging cities may find it useful to develop their transportation sectors.



Finding 5.

People visit different cities for different reasons. This is evident from the different priorities of top 5 destinations. Bangkok prioritizes food and shopping, London has a diversified market and is second place, Paris tops in recreation with enviable places to relax and tour. Dubai has the best market diversity and is fourth place. Singapore prioritizes transportation and shopping. Emerging tourist cities may look at developing the right sectors which makes them unique.

City	Food	Lodging	Transport	Shopping	Recreation	ClusterLabels
Bangkok	0.42	0.00	0.06	0.30	0.08	0
Singapore	0.08	0.04	0.34	0.32	0.14	0
Kuala Lumpur	0.22	0.00	0.26	0.30	0.12	0
London	0.18	0.04	0.32	0.12	0.26	1
Dubai	0.28	0.10	0.08	0.18	0.26	1
New York	0.00	0.00	0.56	0.12	0.32	2
Tokyo	0.00	0.00	0.86	0.02	0.08	2
Seoul	0.00	0.00	0.52	0.16	0.24	2
Paris	0.10	0.00	0.16	0.12	0.48	3
Istanbul	0.14	0.00	0.04	0.12	0.40	3

Finding 6.

London and Dubai appear similar in their tourism approaches. They prioritize market diversity according to the analysis, and this may be their scoring advantage.

Balanced, market diversity

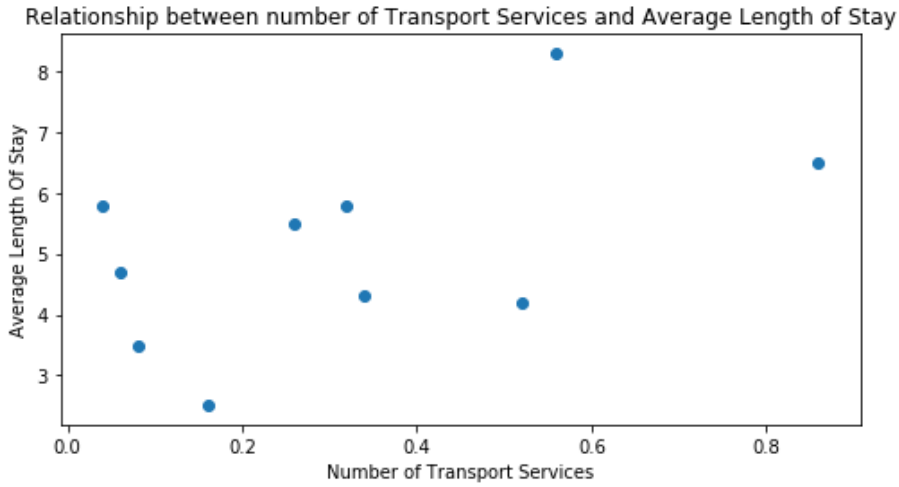
```
df_lodge = final_cities.sort_values(by=['Lodging'], ascending=False)
df_lodge.loc[df_lodge['ClusterLabels']==df_lodge.iloc[0,5]]
```

City	Food	Lodging	Transport	Shopping	Recreation	ClusterLabels
Dubai	0.28	0.10	0.08	0.18	0.26	1
London	0.18	0.04	0.32	0.12	0.26	1

Finding 7.

New York, Tokyo and Seoul appear to prioritize transportation the most amongst the top ten destination cities. Interestingly, New York and Tokyo have the highest average lengths of stay in descending order. While this seemed to suggest a relation between transportation and average length of stay, the pearson correlation is not statistically significant at p-value = 0.12536575684326665.

```
Pearson Correlation: Coefficient = 0.5176869812986228, p-value = 0.12536575684326665
SpearmanrResult(correlation=0.38905954795524134, pvalue=0.2664779212005221)
```



Finding 8.

Istanbul is similar to Paris in its sector prioritization. Given its comparable recreational offerings, Paris lovers may have a comparable experience in Istanbul. Concurrently, Istanbul's recreational offering may be the booster for it's immense tourism growth forecasted at 19.7%.

Recreation and Shopping

```
df_rec = final_cities.sort_values(by=['Recreation'], ascending=False)
df_rec.loc[df_rec['ClusterLabels']==df_rec.iloc[0,5]]
```

City	Food	Lodging	Transport	Shopping	Recreation	ClusterLabels
Paris	0.10	0.0	0.16	0.12	0.48	3
Istanbul	0.14	0.0	0.04	0.12	0.40	3

DISCUSSIONS

While several of the findings are reasonable and empirically verifiable, others are difficult to comprehend. The latter are discussed subsequently.

Discussion 1: *Longer Stay versus Transportation Services*

People stay longest in cities priotizing transportation. However, the analysis couldn't establish a meaningful relationship between length of stay and transportation services. The likely explanation would be that people stay longer in cities where they have a lot of places to tour. Transportation facilities may help enhance the touring experience. Thus, cities with good transportation and places to tour may retain visitors longer. It is also worthy of

note that New York, the city with the longest stay is the only American city featured in the top ten.

Discussion 2: *Shopping and Food may be more important than perceived*

Bangkok represents a deviation from the expected. It is expected that cities with market diversity would top for visitor count. However, Bangkok has topped the ranking for three consecutive years. Cities like Bangkok may be attractive to visitors who want to have fun. It may suggest that cities may effectively tailor tourism toward a goal - fun or business – instead of being an all rounders. Evidently, there is no solution-fits-all for tourism success.

Discussion 3: *Low spending may not necessarily boost visitor count*

From the analysis, more people visited cities where spending was average to high. Dubai, for instance, clocked in at fourth place in spite of its high average spend per day. Generally, when there is more to enjoy, spending is also high. This may suggest that people visit places where there are more avenues to enjoy. Thus if there are facilities to enjoy, visitors may not mind spending much.

CONCLUSION

Tourism and travel can be a remarkable boost to economies if tourist opportunities are effectively managed. The analysis revealed that the top cities are not entirely similar in their tourism approaches even though we could find groups amongst them.

Transportation, recreation and shopping are important sectors to develop for potential cities. Overall, cities which find a way to provide remarkable

experiences, irrespective of their sector prioritizations, may attract more visitors and accompanying revenue.