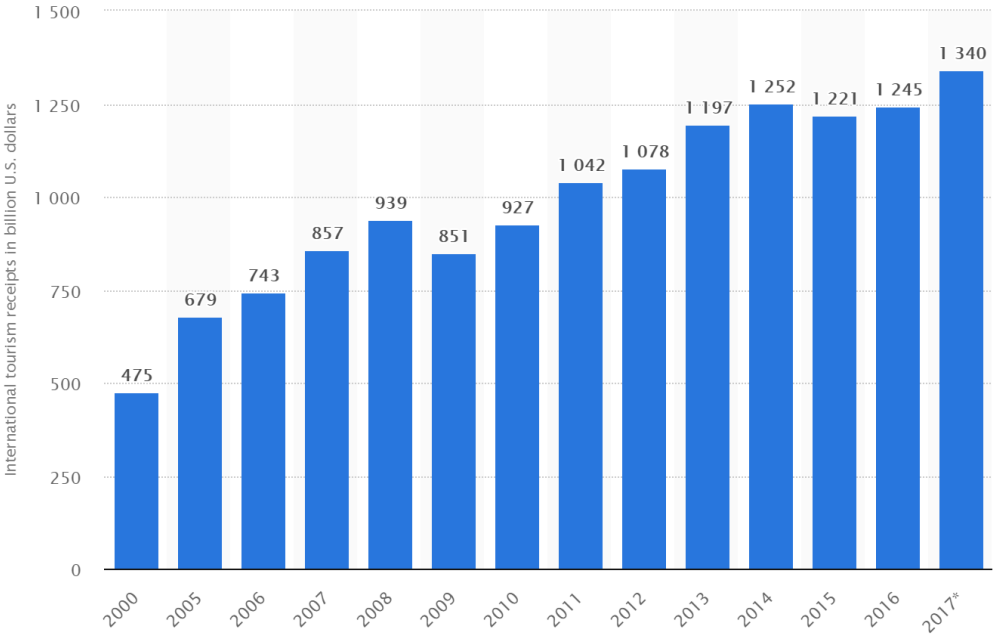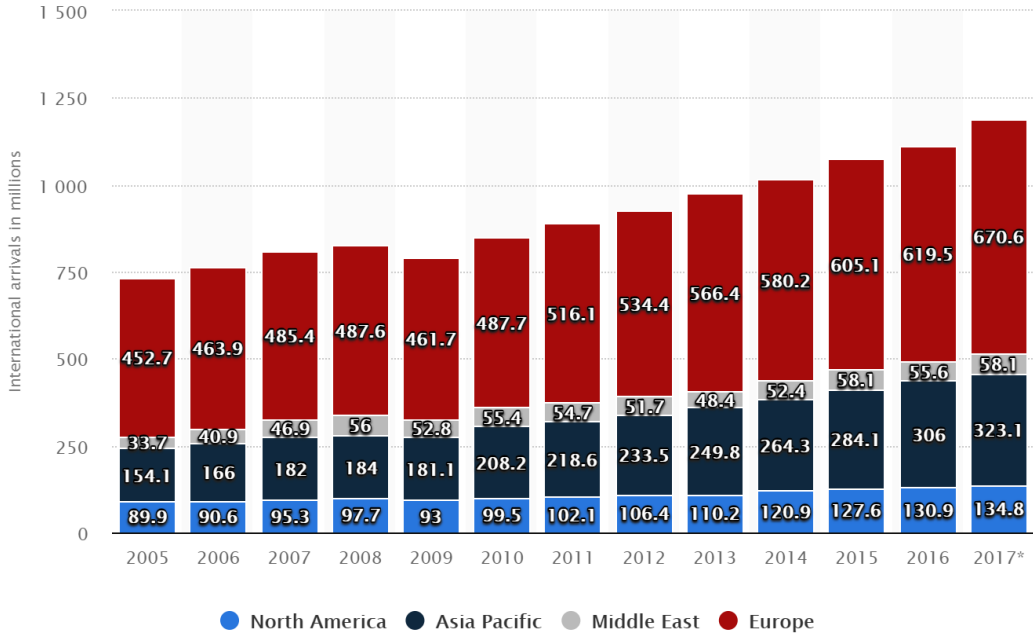# Segmenting World Tourist Cities

Applied Data Science Capstone Project

Background

Global International Tourism Revenue & Destinations

Tourism spend has grown steadily throughout the past 2 decades. The increasing thirst for traveling may be good opportunities for countries looking for an alternative growth engine.
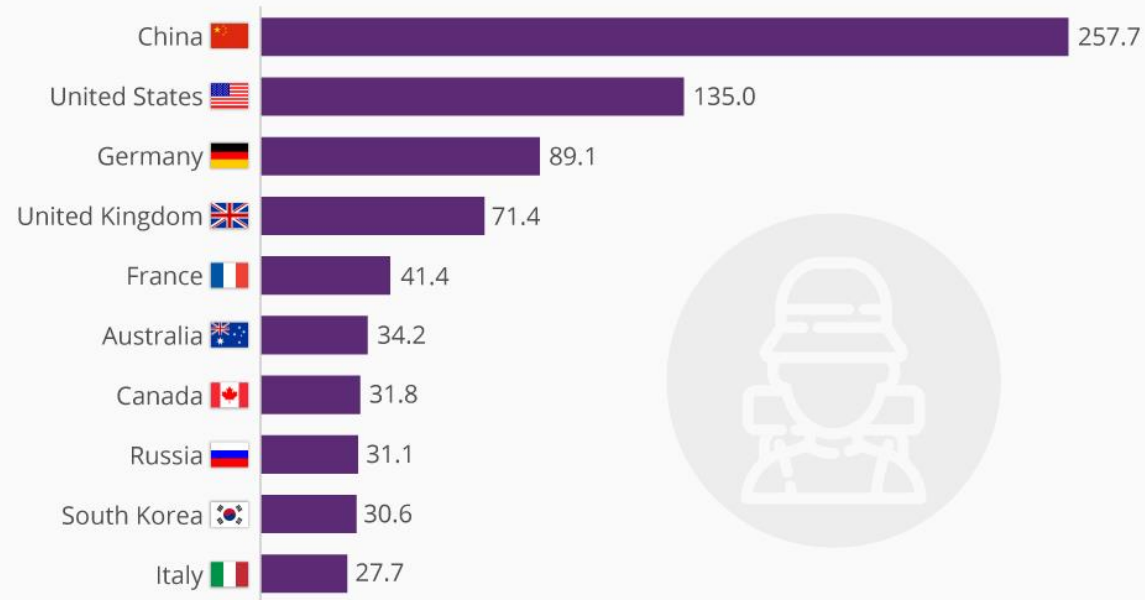
All regions exhibit a growth of arrivals, implying that outbound travel preferences may not only be tied to where in the world the destination is.

International tourism spend per country

Chinese tourists spent a quarter of a trillion dollars abroad
International tourism expenditure in 2017 (billion U.S. dollars)

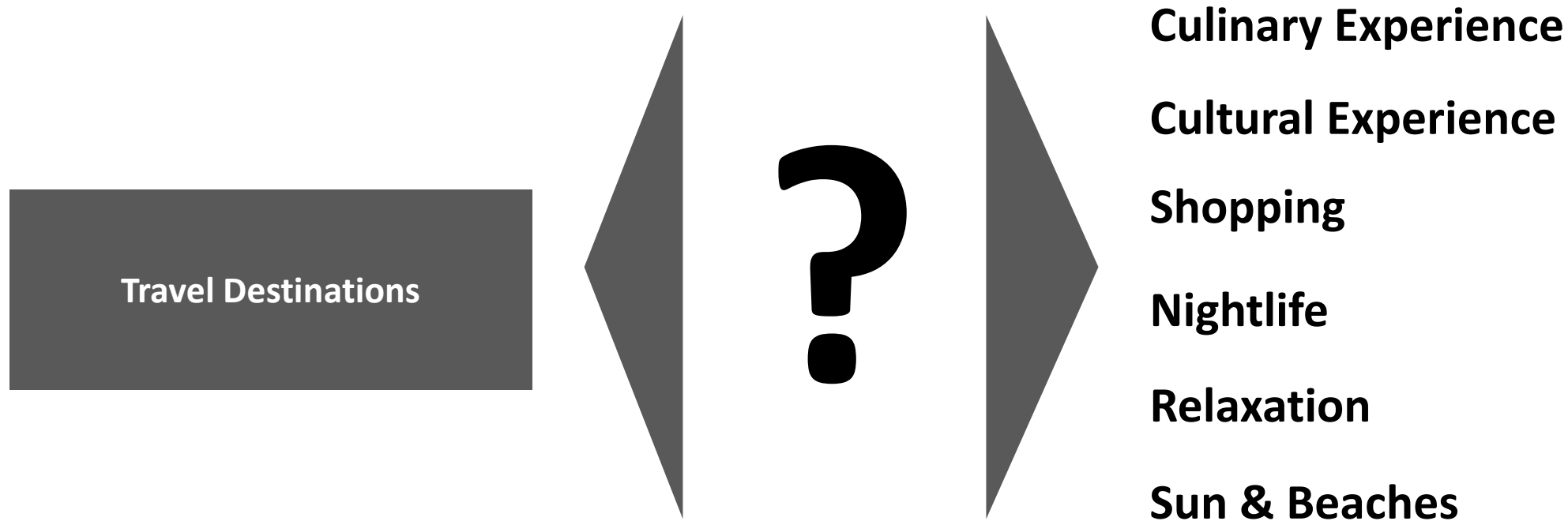| Country	Spend
China	257.7
United States	135.0
Germany	89.1
United Kingdom	71.4
France	41.4
Australia	34.2
Canada	31.8
Russia	31.1
South Korea	30.6
Italy	27.7

@StatistaCharts   Source: World Tourism Organization (UNWTO)

statista

With spend reaching astronomical levels, it is not only important to tap into the growth from a nation / city /private enterprise perspective, but also for the traveller to be able to get value out of the money spent. This very often boils down to the unique preference of each individual and what experiences he/she prefers when going abroad.

Choosing the destination based on experience

Travel Destinations

?

Culinary Experience

Cultural Experience

Shopping

Nightlife

Relaxation

Sun & Beaches

A traveller may be clueless when shifting through seemingly almost endless lists of travel destinations. This may be especially true when going abroad for the first few times. To alleviate this, we propose to **consolidate cities into clusters with similar characteristics to proxy for overall experience**. These experience segments will be **presented to undecided travellers to choose from to either book a ticket to or do more research on.**

Data Sources

City Sample

Rank	City	Country	Population (Metropolitan Area)
1	Tokyo	Japan	38,001,000
2	Delhi	India	25,703,168
3	Shanghai	China	23,740,778
4	Sao Paulo	Brazil	21,066,245
5	Mumbai	India	21,042,538
6	Mexico City	Mexico	20,998,543
7	Beijing	China	20,383,994
8	Osaka	Japan	20,237,645
9	Cairo	Egypt	18,771,769

The world's 150 largest cities are scraped from worldatlas.com to give us the city data sample of possible destinations.

Foursquare Places API – Venues Details

Venue details are extracted for each of our 150 cities from the Foursquare Places API to give an idea of what the destination has to offer.

Dataset

Data Set Descriptives

	City	Venue	Venue Category
Count	12213	12213	12213
Unique	150	10965	511
Frequency	100	186	1046

Our venues data set numbers 12213 in total, divided into our 150 cities. Venue Category is our main variable of interest as we will be classifying cities and what experiences it will have using the same.

Venues distribution

Hotel	1046	Plaza	151
Coffee Shop	633	Seafood Restaurant	144
Café	450	Theater	143
Park	380	Japanese Restaurant	126
Shopping Mall	333	Steakhouse	123
Italian Restaurant	268	Chinese Restaurant	117
Ice Cream Shop	252	History Museum	115
Bakery	240	Burger Joint	115
Restaurant	234	Dessert Shop	111
Pizza Place	223	BBQ Joint	100
Bar	183	Art Museum	99
Indian Restaurant	173	Sandwich Place	99
Historic Site	162	French Restaurant	99
Bookstore	156	Art Gallery	97
Fast Food Restaurant	153	Scenic Lookout	95
		...	

The most common venues seem to be in line with what one would expect from large cities. The high number of hotels seem to imply that our sample of cities are suitable travel destinations. Moving on we see diverse set of eateries, with other cultural categories such as historic sites, theatres and art galleries sprinkled in for good measure.

Top venue categories per city – Base for clustering

|	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abidjan	Nightclub	Ice Cream Shop	African Restaurant	Shopping Mall	Hotel	Italian Restaurant	Bakery	Lounge	Pizza Place	Café
1	Addis Ababa	Hotel	Italian Restaurant	Ethiopian Restaurant	Restaurant	Pizza Place	Nightclub	Turkish Restaurant	Coffee Shop	Greek Restaurant	Grocery Store
2	Ahmadabad	Café	Indian Restaurant	Fast Food Restaurant	Hotel	Restaurant	Coffee Shop	Tea Room	Multiplex	Dessert Shop	Bakery
3	Aleppo	Mountain	RV Park	Market	General College & University	Forest	Café	Moving Target	Fast Food Restaurant	Field	Fish & Chips Shop
4	Alexandria	Coffee Shop	Café	Restaurant	Seafood Restaurant	Hotel	Juice Bar	Bakery	Sandwich Place	Bar	Historic Site
5	Ankara	History Museum	Theater	Dance Studio	Art Gallery	Seafood Restaurant	Café	Bookstore	Historic Site	Pizza Place	Dessert Shop
6	Athens	Bar	Mexican Restaurant	Pizza Place	Grocery Store	Music Venue	Liquor Store	Coffee Shop	New American Restaurant	Fast Food Restaurant	American Restaurant
7	Atlanta	Trail	Park	Brewery	Italian Restaurant	Southern / Soul Food Restaurant	Pizza Place	Ice Cream Shop	Mexican Restaurant	American Restaurant	Bar
8	Baghdad	Café	Shopping Mall	Hotel	Fast Food Restaurant	Middle Eastern Restaurant	Ice Cream Shop	Bakery	Burger Joint	Fried Chicken Joint	Coffee Shop
9	Bangalore	Hotel	Indian Restaurant	Pub	Ice Cream Shop	Lounge	Breakfast Spot	Bakery	Café	Burger Joint	Brewery
10	Bangkok	Coffee Shop	Hotel	Park	Shopping Mall	Thai Restaurant	Noodle House	Japanese Restaurant	Golf Course	Bookstore	Asian Restaurant

Each city record is assigned their 10 most common venue categories to be used as our input to the clustering algorithm.

Clustering & Results
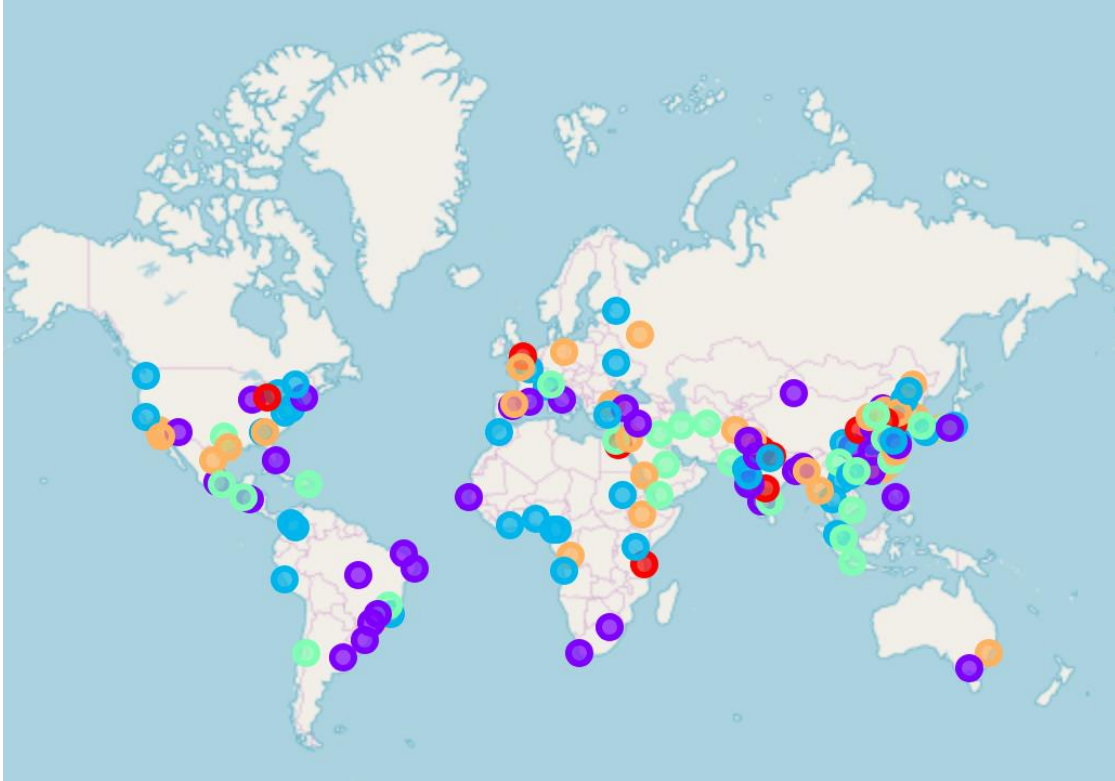
Machine Learning Algorithm

A k-Means clustering algorithm with k=5 is used to segment our data set. 5 is chosen to balance the need for a minimum number of segment options against a possible dilution of results if too many segments are created.
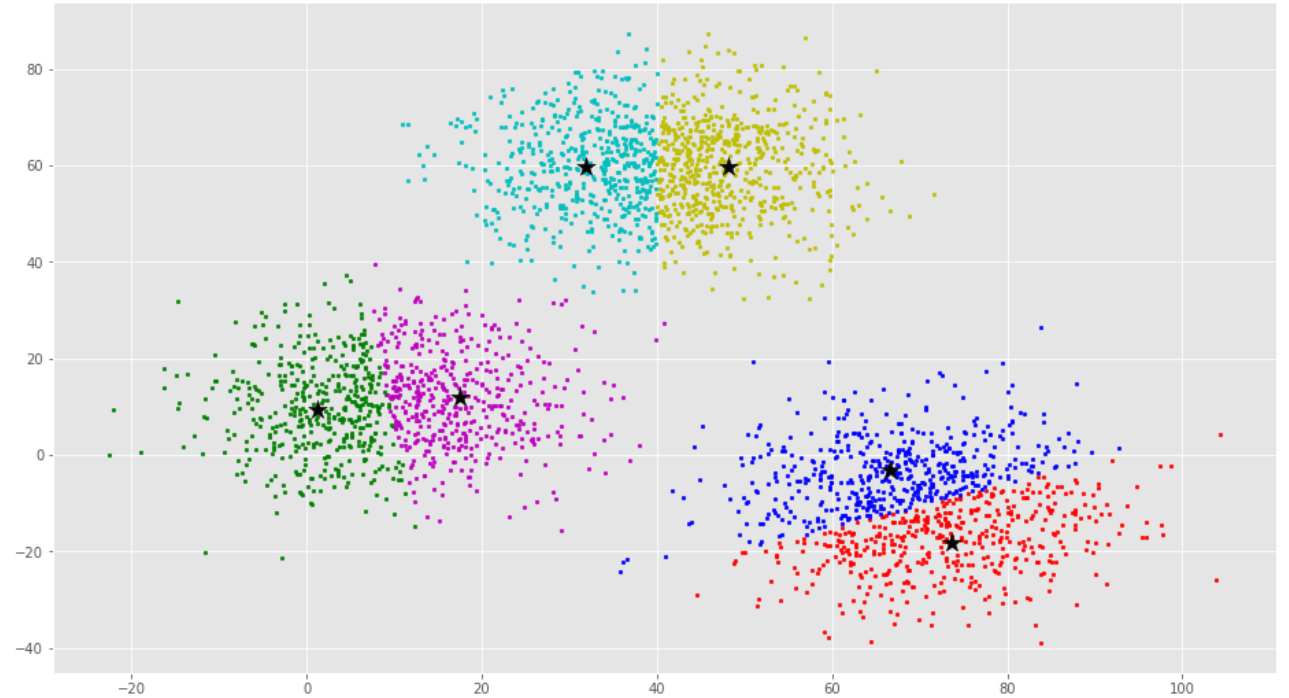
City Clustering

Applying the k-Means algorithm results in the division of clusters as per the illustration to the left.

The clusers are seemingly evenly spread out throughout the world which is desirable if we want to have an international outlook.

Given this clustering, we have looked into

City experience classifications 1/2

Shopping & Hangout	The first segment has been named City Shopping & Hangouts due to the high frequency of shopping malls in combination with cafes and coffee shops in the top 10 most common venues lists. This segment of cities suits those looking for a convenient experience with lots of well needed opportunities for breaks between the last shop to the next while carrying around on bags.
Hangout, Dining & Nightlife	This segment is mainly characterised by the high frequency of cities with high densities of eateries of all kinds. Coupled together with the relatively high occurrence of bars in the top 10 venue types, these cities provide a well-rounded experience focusing more on the culinary and social aspects.
Green City Hangout & Dining	With a high rate of parks and eateries, these cities provide visitors with a city stay complemented with plenty dining opportunities in scenic environments thanks to the high frequency of parks.

City experience classifications 2/2

Shopping & Hangout	The Shopping, Sweets & Nightlife segment provides an experience focused on indulging yourself that new jacket while enjoying local pastries to satisfy your sweet tooth. Once ready, there is also ample of opportunities to grab a fast food burger or fried chicken before heading out into the night.
Green City, Shopping & Nightlife	Want to enjoy the shopping & bubbling nightlife of cities without losing touch with nature? Look no more! This segment appeals to those who wants to strike a balance between city life and greenery.

We have reached our end goal of providing prospective travellers 5 distinct city travel experiences to choose from when picking their next destination. For details on how these segments were created and their detailed characteristics, please refer to the full report section **3.1 – City Segments**.

Using the created Segments

User profile: Travelers

With our constructed segments travellers can choose to use the insights from 2 different perspectives:

Travel Suggestions
Travellers can choose from a list of desired city trip experienced as defined in our previous Results section. Once chosen, the user will then then get a list of cities to choose from that fall into the preferred category before proceeding to booking the ticket or choosing to do more research. In the same way, users can also enter a city name and then be returned with what kind of experience that can be expected from going there. This is handy when the user receives a recommendation without too many specifics on what to expect.

Finding similar destinations
In the cases where a user wants to have an experience similar to a past trip but to a new destination, he/she can enter the city name and be returned with the full list of cities contained within the cluster of belonging.

User Profile: City Planners, developers & entrepreneurs

In the same manner but opposite direction for travellers, city planners, developers and entrepreneurs can make use of the insights to identify business opportunities to cover the weak points of the cities. Is a city strong in shopping and dining but lack green areas? Build more parks! Does the city die after 7PM after the last shop closes? Set up bars!

Discussion

Room for improvement

- Methodology Refinement

- Sample Data Limitations

- Sample Selection Bias - Destinations

- Omitted features bias

- Features data refinement & transformations
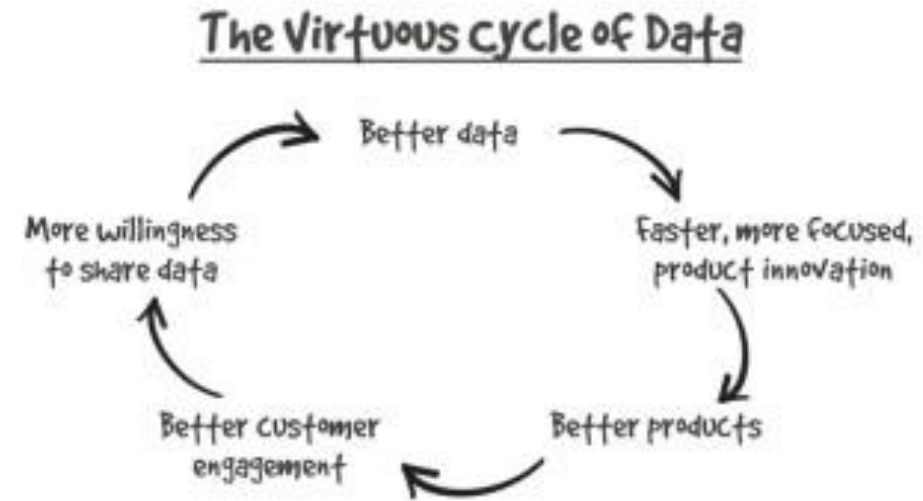
Though we have completed our objective of grouping cities into manageable experience clusters, there is always room of improvement to further enhance our product to give a better customer experience. Examples of these are mentioned in the list above, which is **discussed more in detail in the full report in section 4. Discussion.**