**Business establishments in the City of Dubai**

**1 Introduction**

1.1 Background

A global leader in innovation as a hyper-connected pro-business hub between East & West, providing unrivaled access to the world's highest potential economies, Dubai is proven to deliver an efficient, secure and future forward network ecosystem for accelerated growth. Dubai is internationally recognized as the leading financial and trading center at the heart of the world's Islamic economy.

1.2 Problem

Dubai receives hundreds of thousand tourists each year. Demand placed on the city can be understood by examining the businesses already established within the city. The objective of this report is to gain insights into the business dynamics of the city.

**2. Data acquisition and cleaning**

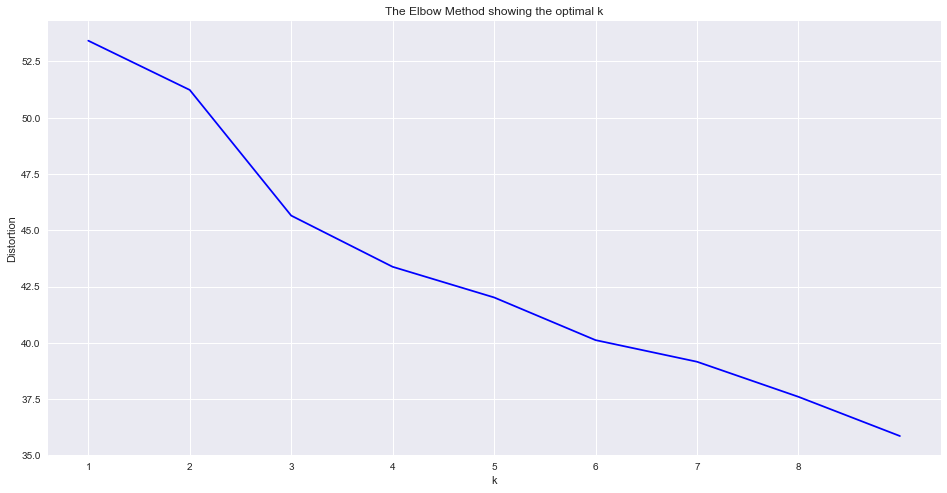
2.1 Data source

The Neighborhood and Coordinates was gathered from Smart Dubai ([www.smartdubai.ae](http://www.smartdubai.ae)). Smart Dubai is the government office charged with facilitating Dubai’s citywide smart transformation, to empower, deliver and promote an efficient, seamless, safe and impactful city experience for residents and visitors. Venue details were extracted from Foursquare; features include venue name, venue category, and venue category. Venue categories were scraped from the Foursquare API documentation to create category classes (<https://developer.foursquare.com/docs/build-with-foursquare/categories/>).

**3. Exploratory Data Analysis**

To derive a more meaningful understanding of the business geography of Dubai, K-means clustering algorithm was used. K-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.

Clustering was applied initially to the Venue categories, to grasp the zoomed-in business pattern in the various neighborhoods. The elbow method was employed in selecting the number of clusters to be used on the K-means algorithm. As can be seen from the graph below, there are two elbow points at 3 and 6 k.



Running the K-means algorithm at 3 clusters produced poor results. The analysis was re-run using 6 clusters. The output is as below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Clusters** | **Most common category** | **Number of occurrences** | **% of occurrences** |
| Cluster 1 | Coffee shops | 7 | 46.67% |
| Cluster 2 | Restaurants | 12 | 44.44% |
| Cluster 3 | Coffee shops | 520 | 30.34% |
| Cluster 4 | Convenience store | 1 | 100% |
| Cluster 5 | Auto garages | 7 | 100% |
| Cluster 6 | Campgrounds | 5 | 100% |

As can be seen from the table above, Cluster 1 & 3 have coffee shops as the most common category, with the 520 occurrences in Cluster 3.

**4. Further Exploratory Analysis**

To find the broader business dynamic in the city, the venue categories need to be reclassified. The 931 categories of venues were grouped into 10 classes:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Venue Classes** | **No of Venue categories** | **Example categories** |
| 1 | Arts & Entertainment | 64 | Aquarium, Circus, Museum |
| 2 | College & University | 34 | Law School, Medical School |
| 3 | Events | 12 | Trade fair, Parade, Conference |
| 4 | Food | 348 | Cafe, Food court, Pizza place |
| 5 | Nightlife spot | 24 | Nightclub, Beach bar, Lounge |
| 6 | Outdoor & Recreational | 107 | Gym, Yoga Studio, Garden |
| 7 | Professional & Other places | 107 | Business Center, Church, Office |
| 8 | Residence | 5 | Home, Trailer park, Condo |
| 9 | Shop & Services | 173 | Shoe store, Bank, IT Services |
| 10 | Travel & Transport | 53 | Airport, Hostel, Resort |

K-means algorithm was used on the reclassified venues with 3 clusters. The output is as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Clusters** | **Most common Class** | **No. of class occurrences** | **Most common Category** | **No. of category occurrences** |
| Cluster 1 | Outdoors & Recreation | 59 | Beach | 8 |
| Cluster 2 | Shop & Service | 228 | Market | 12 |
| Cluster 3 | Food | 1421 | Hotel | 119 |

As can be seen from the table above, Cluster 3 has Food as the most common business class with 1421 occurrences and Hotels being the most common category in that class. Cluster 2 has Shop & Services as the most common business class with 228 occurrences and Market being the most common category in that class. Cluster 1, has Outdoor & Recreation as the most common class and Beach as the most common category in that class.

**5. Conclusion**

Dubai being a tourist destination, it is not surprising to find lot of Food, shopping and outdoor activities. This exploratory analysis shows the demand within the city and the various industries providing for it.