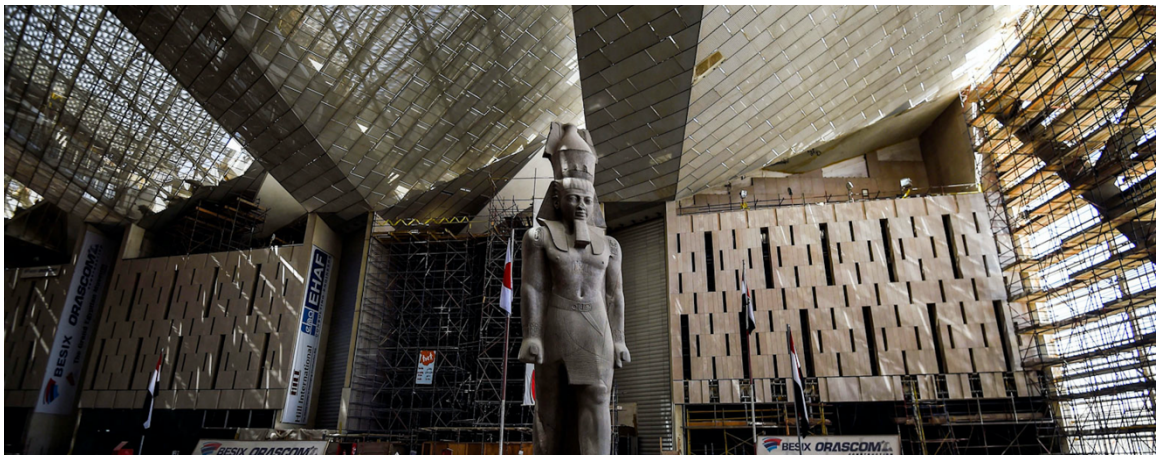
IBM DATA SCIENCE CAPSTONE PROJECT

Venues to invest in around the Grand Egyptian Museum (GEM)



By: Osama Abbas

May 2020

Introduction.

Egypt is currently building a megaproject called The Grand Egyptian Museum or GEM. The museum is projected to be the largest archeological museum in the world. The museum complex will be built on a plot of land approximately 480,000 square meters, which will be less than 2 kilometers away from the timeless wonder of the Giza Pyramids. As well as holding the entire collection of the famous king Tutankhamoun, the museum complex will be holding more than 100,000 artifacts. This will allow the museum to accommodate more than 15,000 visitors per day. While this seems like a very promising project for the Egyptian government in order to boost the tourism sector, it is also a great opportunity for investors looking to invest in what is expected to be a major touristic hub in one of the top touristic destinations in the world.

As exciting as this sounds, one cannot help but wonder, where should a potential investor put their money around such a project? In this project we will try to answer that question through a data oriented approach, by using our data science toolkit in order to acquire the data, clean the data, visualize it, and run some analysis algorithms in order to find similar museums that already exist around the world and try to get ideas of what category of venues would be the most successful around such project.

Business Problem.

The key objective of this capstone project is to try to find the best ways to invest money around a museum with such magnitude as the Grand Egyptian Museum GEM. Using our data science toolkit which includes data acquisition techniques, data wrangling, location data visualization, and machine learning algorithms such as clustering. We will be able to answer the question about what venues to invest in around a major museum.

Target audience.

This capstone project would be particularly helpful for potential investors looking to seize the opportunity to invest around the Grand Egyptian Museum megaproject. The capstone project is especially important now as the grand opening of the museum is projected to take place in early 2021. As mentioned above, the museum is likely to accommodate more than 15,000 visitors per day. Although the museum complex will include an area designated for cafes, restaurants, and a giftshop, one must also keep in mind that the museum is next to one of the seven wonders of the world. So, the tourists' itinerary will not be limited to the museum in most cases. This will certainly boost the tourist' flow to the area and its neighboring zones, potentially turning it into a hub for tourists to enjoy a full cultural experience.

Data.

In order to answer the question that we are asking, we will need the following data:

- A list of most visited museums around the world, adding to them the data of the GEM in order to study the demographic of the venues already existing around.
- Location data of these museums, including addresses and coordinates (latitude and longitude).
- Venue data, which includes the venues that are currently existing and successfully functioning around these museums and the categories of said venues.

Data Acquisition.

First, we will be acquiring the museum data from the following wikipedia page:

https://en.wikipedia.org/wiki/List_of_most_visited_art_museums

We will use BeautifulSoup module in order to scrape the table containing the data that we need. Next, we will be using Geopy to be able to acquire the location data of these museums, including their addresses and coordinates. Afterwards, using the Foursquare API we will acquire the venue data around these museums, including the venue coordinates, so we can visualize them on the map, and the venue category. This will allow us to run a cluster analysis on the demographic of the venues around the museums and cluster the museums that have similar demographic. Then we look into the venues around the museums in the same cluster as GEM, and point out the ones that are different from the ones in the area around GEM.