

Web Scraping and EDA on Booking.com

About Tourism in INDIA

- The Indian travel and tourism sector contributed nearly \$194 billion, around 6.8% of GDP, during calendar year 2019. Out of 185 economies, India stood 10th in terms of the size of travel and tourism spend in 2019.
- In 2019, the domestic tourists arriving to the state of Goa accounted for approximately 7.1 million, while the foreign tourist arrivals accounted for over 0.9 million.

About Booking.com

- Founded in 1996 in Amsterdam, Booking.com has grown from a small Dutch startup to one of the world's leading digital travel companies.
- Part of Booking Holdings Inc. (NASDAQ: BKNG), Booking.com's mission is to make it easier for everyone to experience the world.
- Booking.com enables properties all over the world to reach a global audience and grow their businesses.
- Booking.com is available in 43 languages and offers more than 28 million total reported accommodation listings, including over 6.2 million listings alone of homes, apartments and other unique places to stay.
- CEO-Glenn Fogel

Problem Statement-

- Assuming that a client(Tour Planners) are planning for designing a tour package for the new year celebrations 2021 at Goa.
- And in process of designing the package, there is a need to understand impact of price, location, rating and many other factors.

Tools & Techniques used for Analysis-

- Python
- JupyterNotebook
- BeautifulSoup
- Urllib
- Visualization-Seaborn, Plotly
- Missing value treatment
- RegEx

Observations drawn form analysis

- 19.3 % of the hotels are located on 'Calangute' beach. Next comes the 'Baga' beach holding 10.4 % of the hotels.
- The proportion of hotels providing 'FREE Cancellation' on hotel bookings are least compared to one that are charging some charges.
- The median of the price is '12k'. And the minimum price is '3k', 25 % of the hotels have price of '8k', 75 % of the hotels have price of '17k', the maximum price is '31k'. And the 'Price' holds some outliers, they holds prices of range '33k-319k'.
- There is high scatteredness at the 'ReviewRate' 7.9 to 8.0
- The four locations that holds high review rate on an average are-Palolem-8.7(7.14,8.15,9.3,9.8),Anjuna-8.2(7.8,8.825,9.6),Vagator-8.15(6.9,7.1,8.9,9.1),Patnem-8.15(7.8,7.9,8.5,8.8).
- The hotels in Madgaon (4k,5k,6k,9k,33k), Salcete (5k,7k,9k), Vagator (4k,6k,8k,12k,19k,40k),Palolem (3k,5k,8k,10k,12k,31k) beaches have the least median value which allows us to conclude that hotels here are relatively cheapest to stay for.
- The number of rooms available based on location are as follows- Calangute-99,Baga-74,Candolim-53,Morjim-47,Vagator-44,Arambol-42,Anjuna-37,Palolem-27,etc...
- The hotels in Madgaon(4660k),Salcete(7550k),Vagator(14.515k),Palolem(8960k)- in terms of total expenditure.
- First Class(>100000 rupees),Business Class(50000-100000 rupees),Economy Class(<=50000 rupees).Out of total 269 hotels. Economy Class-260,First Class-1,Business Class-8.

Challenges-

- Scraping the data from windowless browser.
- Limiting to only few factors of study.
- COVID-19 pandemic, which make it difficult to understand some instances.

Final Summary-

- After performing some limited analysis, it is found that the locations-**Palolem** and **Valagate** are best with respect to price, reviewrate, hotel rooms available, Cancellation Policy,etc.
- It is recommended that further a greater analysis needs to be carried on taking into consideration some of the other factors such as hygiene, COVID-precautions, Food facility,etc.