mindX Technology School

FINAL PROJECT

X Company's
Online Hotel Search Website
In 2019

Course: **BI17**

Group: **5**

Name: **Bùi Diệp Thanh Vy**

Project Advisor: **Đặng Văn Đức**

Due Date: **26/10/2022**

CONTENTS

A. INTRODUCTION TO ABC WEBSITE:

- Company X specializes in the ABC website - a leading hotel booking search and comparison website worldwide. The Marketplace team is a large team consisting of data analysts, data scientists, data engineers, and software developers from various countries. The team's mission is to provide data analysis to support Company X's decision-making process.

B. OVERVIEW OF THE DATASET:

I. INTRODUCTION TO THE DATASET:

- The provided dataset consists of daily performance data of three advertisers on the ABC website in the year 2019 in one country (out of many countries where the ABC website operates).

- Advertisers set different prices for three different customer segments based on the "time-to-travel" (TTT), which is calculated as the number of days from the date the customer booked a hotel room to the date the customer checked-in at the booked hotel. The three customer segments based on TTT include:
  - Short: 0 – 14 days.
  - Medium: 15 – 60 days.
  - Long: hơn 60 days.

- The dataset consists of three main advertisers in this market, labeled as advertisers "A," "B," and "C." For each of the aforementioned advertisers, the provided metrics include:
  - Clicks: The number of clicks made by users on the ABC website.
  - Cost: The amount of money the advertiser has to pay to the ABC website (based on a cost-per-click model).
  - Bookings: The number of hotel bookings made by users on the ABC website.
  - Booking_rev: The amount of money spent by users on hotel bookings (essentially the gross revenue for each advertiser).

II. DATA DESCRIPTION IN PYTHON

1. Data Information:

```
#import data
df=pd.read_csv('/content/drive/MyDrive/Colab Notebooks/BI17_PYTHON/Final Project/marketplace_data_2019.csv')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 14 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   date           1095 non-null   object
 1   ttt_group      1095 non-null   object
 2   clicks_A       1095 non-null   int64
 3   clicks_B       1095 non-null   int64
 4   clicks_C       1095 non-null   int64
 5   cost_A         1095 non-null   int64
 6   cost_B         1095 non-null   int64
 7   cost_C         1095 non-null   int64
 8   bookings_A     1095 non-null   int64
 9   bookings_B     1095 non-null   int64
 10  bookings_C     1095 non-null   int64
 11  booking_rev_A  1095 non-null   int64
 12  booking_rev_B  1095 non-null   int64
 13  booking_rev_C  1095 non-null   int64
dtypes: int64(12), object(2)
memory usage: 119.9+ KB
```

(Figure 1- Data Information)

- The dataset comprises:
  - 14 columns of data: 2 columns with object data type, and 12 columns with integer data type.
  - 1095 rows of data: with no null values.

2. Creating additional columns of data:

- To facilitate data analysis and visualization, additional columns for total clicks, bookings, revenue, and cost are created.

```
df['total_clicks']=df['clicks_A']+df['clicks_B']+df['clicks_C']
df['total_bookings']=df['bookings_A']+df['bookings_B']+df['bookings_C']
df['total_cost']=df['cost_A']+df['cost_B']+df['cost_C']
df['total_revenue']=df['booking_rev_A']+df['booking_rev_B']+df['booking_rev_C']
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1095 entries, 0 to 1094
Data columns (total 18 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   date            1095 non-null   object
 1   ttt_group       1095 non-null   object
 2   clicks_A        1095 non-null   int64 
 3   clicks_B        1095 non-null   int64 
 4   clicks_C        1095 non-null   int64 
 5   cost_A          1095 non-null   int64 
 6   cost_B          1095 non-null   int64 
 7   cost_C          1095 non-null   int64 
 8   bookings_A      1095 non-null   int64 
 9   bookings_B      1095 non-null   int64 
 10  bookings_C      1095 non-null   int64 
 11  booking_rev_A   1095 non-null   int64 
 12  booking_rev_B   1095 non-null   int64 
 13  booking_rev_C   1095 non-null   int64 
 14  total_clicks    1095 non-null   int64 
 15  total_bookings  1095 non-null   int64 
 16  total_cost      1095 non-null   int64 
 17  total_revenue   1095 non-null   int64 
dtypes: int64(16), object(2)
memory usage: 154.1+ KB
```

(Figure 2 – Data information after adding columns for the total number of clicks, bookings, revenue, and cost.)

3. EDA-Exploratory Data Analysis:

a. The description of columns with integer data type:

```
[ ] #View the descriptions of columns with integer data type in the dataset
    statistic=df.describe()
    statistic.T
```

	count	mean	std	min	25%	50%	75%	max
clicks_A	1095.0	2.904898e+05	3.398313e+05	324.0	6878.5	187615.0	449134.5	1714072.0
clicks_B	1095.0	1.337045e+05	2.095589e+05	2533.0	8029.5	19198.0	214691.0	1041133.0
clicks_C	1095.0	2.204043e+05	1.323282e+05	14651.0	123992.0	191998.0	295214.0	727874.0
cost_A	1095.0	2.686955e+05	3.300333e+05	127.0	3285.5	154268.0	398446.0	1661516.0
cost_B	1095.0	1.182632e+05	1.919750e+05	1276.0	4270.0	11610.0	194146.5	956674.0
cost_C	1095.0	1.778377e+05	1.068645e+05	11849.0	99865.0	154798.0	237919.0	596599.0
bookings_A	1095.0	1.338238e+04	1.643578e+04	6.0	163.5	7697.0	19849.0	82702.0
bookings_B	1095.0	5.072310e+03	8.140734e+03	41.0	169.5	619.0	8330.5	40205.0
bookings_C	1095.0	7.633052e+03	4.189684e+03	705.0	4450.5	6875.0	10001.0	23564.0
booking_rev_A	1095.0	2.092607e+06	2.565217e+06	967.0	25464.5	1187892.0	3150881.5	13150330.0
booking_rev_B	1095.0	7.901320e+05	1.264995e+06	6152.0	26307.0	96480.0	1306487.5	6117843.0
booking_rev_C	1095.0	1.191080e+06	6.544745e+05	107349.0	691518.5	1072385.0	1565771.0	3669508.0
total_clicks	1095.0	6.445986e+05	5.512314e+05	76446.0	262531.0	446190.0	844495.5	3108833.0
total_bookings	1095.0	2.608774e+04	2.619837e+04	2157.0	7737.0	15451.0	35946.5	140061.0
total_cost	1095.0	5.647964e+05	5.207064e+05	60322.0	212506.5	364271.0	733643.5	2886016.0
total_revenue	1095.0	4.073819e+06	4.079805e+06	324738.0	1226414.0	2441476.0	5604386.0	22268231.0

(Figure 3 - The description of columns with integer data type)

- Observing the data description table, we notice that all columns have mean values greater than the median values, indicating positive skewness (left-skewed distribution). This means that the values are concentrated more towards the left side of the median.

b. The description of columns with object data type:

```
[ ] #View the descriptions of columns with object data type in the dataset
    df.describe(include='O')
```

	date	ttt_group
count	1095	1095
unique	365	3
top	2019-01-06	short
freq	3	365

(Figure 4 - The description of columns with object data type)

- The dataset provides performance data of advertisers across three customer segments. Each segment has complete information for 365 days.

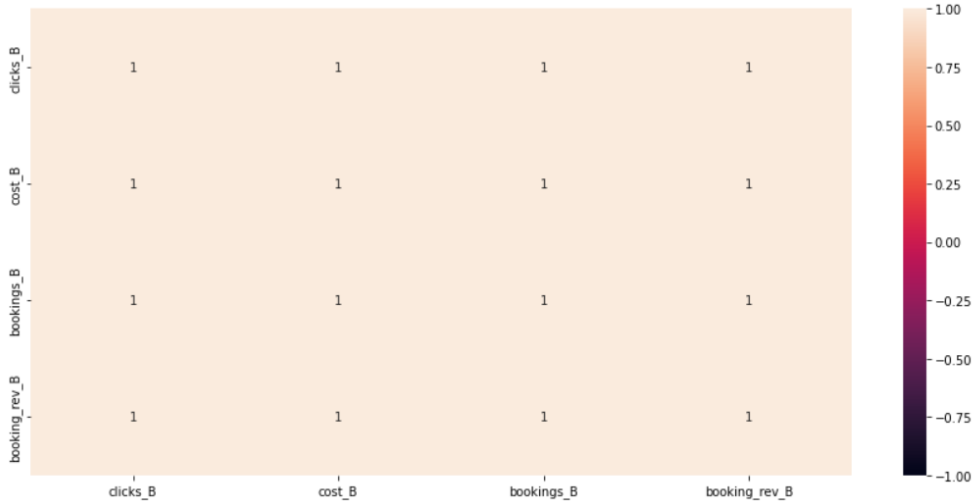
c. Correlation:

- Figure 5 is a Heat Map chart representing the correlation between the columns of data: Clicks, bookings, revenue, and cost of advertiser A. It shows a strong positive correlation with correlation coefficients equal to 1.

(Figure 5 - The correlation between the metrics of Advertiser A)

- Figure 6 is a Heat Map chart representing the correlation between the columns of data: Clicks, bookings, revenue, and cost of advertiser B. It shows a strong positive correlation with correlation coefficients equal to 1.

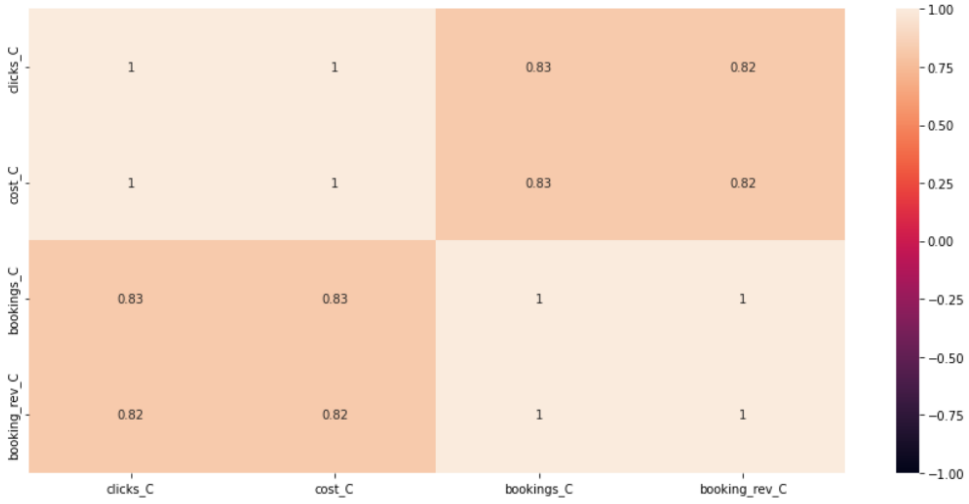
```
[53] plt.figure(figsize=(15,7))
     sns.heatmap(df[['clicks_B','cost_B','bookings_B','booking_rev_B']].corr(), annot = True, vmin = -1, vmax = 1)
     plt.show()
```



(Figure 6 - The correlation between the metrics of Advertiser B)

- Figure 7 is a Heat Map chart representing the correlation between the columns of data: Clicks, bookings, revenue, and cost of advertiser C. It shows a strong positive correlation with correlation coefficients close to 1.

```
[54] plt.figure(figsize=(15,7))
     sns.heatmap(df[['clicks_C','cost_C','bookings_C','booking_rev_C']].corr(), annot = True, vmin = -1, vmax = 1)
     plt.show()
```

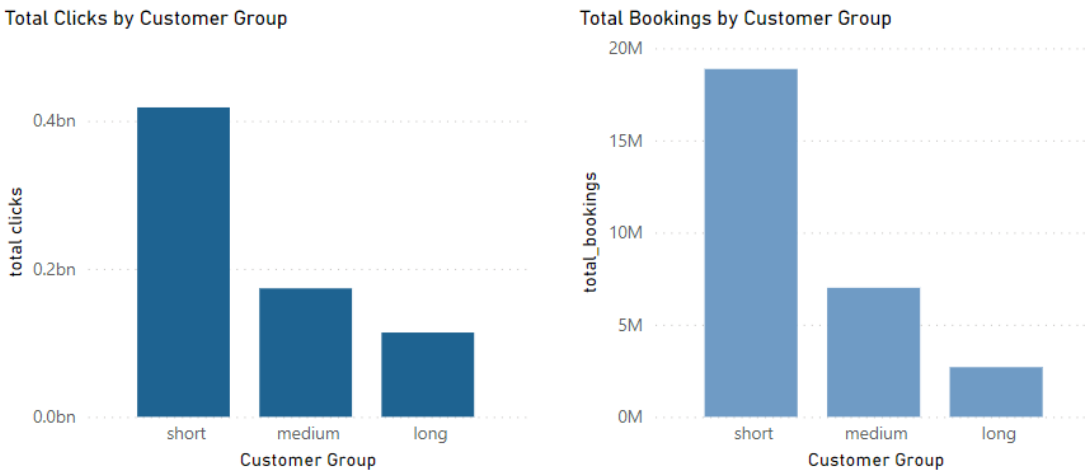


(Figure 7 - The correlation between the metrics of Advertiser C)

C. DATA ANALYSIS:

I. OVERVIEW OF BUSINESS ACTIVITIES ON ABC WEBSITE:

1. Customer Segments:



(Figure 8- The total number of clicks and bookings made on the ABC website by customer segments)

- The short-term customer segment accounts for the largest proportion of users on the ABC website, followed by the medium-term and long-term customer segments.

❖ **The short-term customer segment:**

✧ 55% of the total number of clicks and 59% of the total number of bookings on the ABC website come from advertiser A.

✧ The second position is held by advertiser B with 33% of the total number of clicks and 27% of the total number of bookings.

✧ The last position is held by advertiser C with 12% of the total number of clicks and 14% of the total number of bookings.

❖ **The medium-term customer segment:**

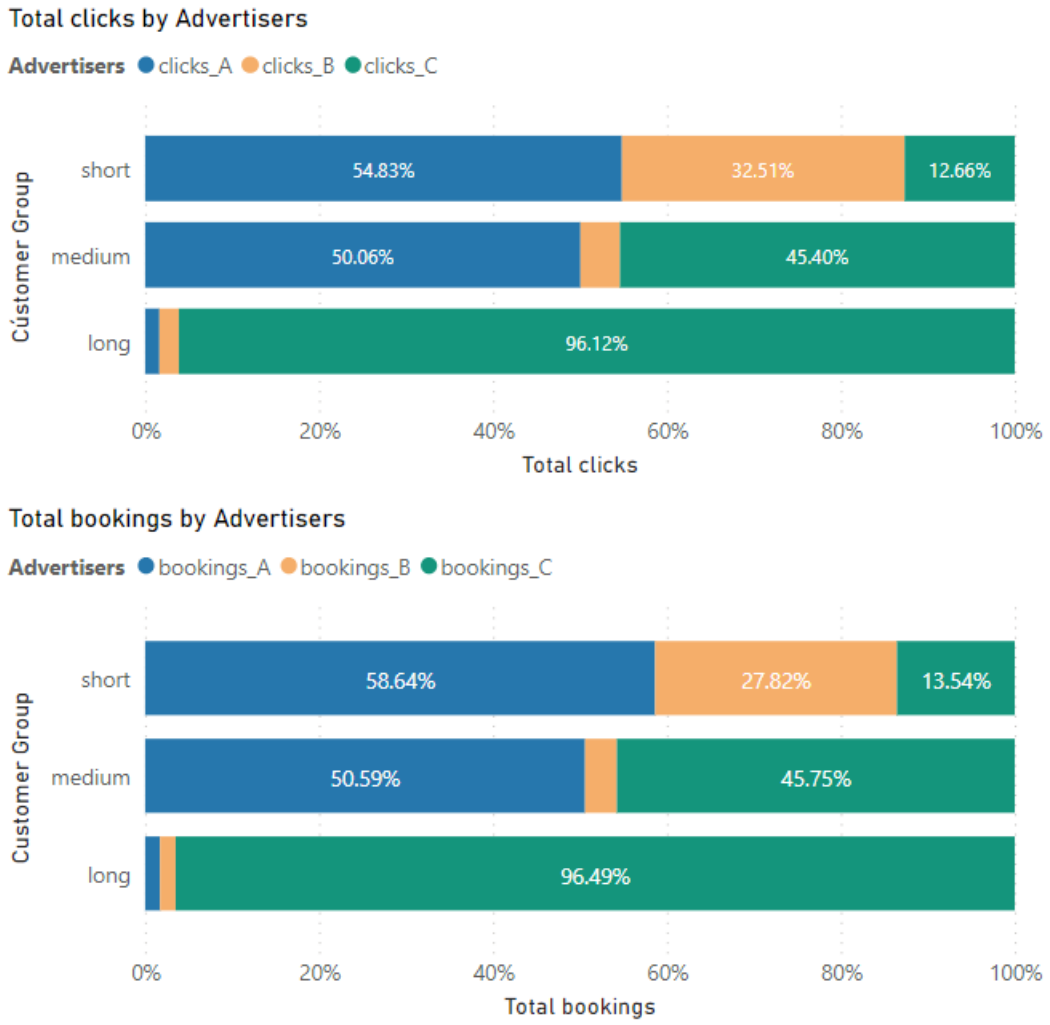
✧ 50% of the total number of clicks and the total number of bookings on the ABC Website come from advertiser A.

✧ The second position is held by advertiser C with an approximate proportion to advertiser A, which is 45%.

✧ The last position is held by advertiser B with a proportion close to 5%.

❖ **The long-term customer segment:**

- ✧ 96% of the total number of clicks and the total number of bookings come from advertiser C.
- ✧ Advertiser A and B account for an insignificant proportion.



(Figure 9- The total number of clicks and bookings for each customer segment carried out on the ABC website, categorized by advertisers.)

2. CONVERSION RATE:

a. The significance of the conversion rate metric

- One of the metrics that the Marketplace team is concerned about is the conversion rate of users, or in other words, they are interested in the ratio of users who can find suitable hotel bookings on the ABC website according to their needs. One way to determine this is by calculating the booking conversion rate, calculated by dividing the number of bookings by the number of clicks.

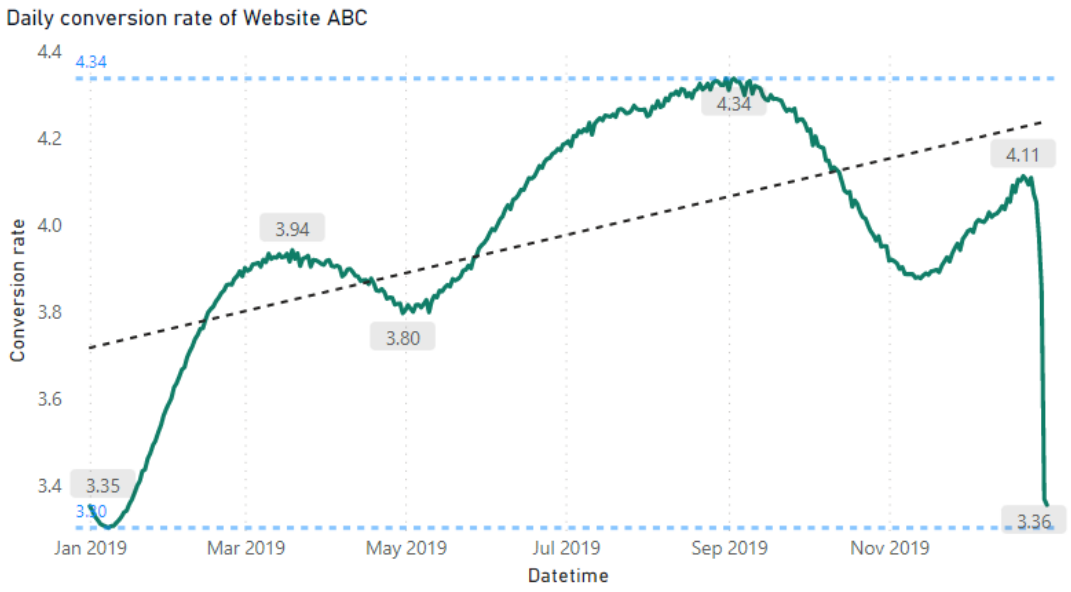
$$\text{Booking conversion} = \frac{Total\_bookings}{Total\_clicks}$$

b. The conversion rate of ABC website:

In Power BI, create a measure to calculate the conversion rate of the ABC website.

$$\text{conversion\_rate} = \frac{\text{sum('Edited Data'[total\_bookings])}}{\text{sum('Edited Data'[total\_clicks])}} \times 100$$

- Plot a graph:



(Figure 10 - The conversion rate of ABC website by day)

- Overall, the conversion rate of the ABC website in 2019 shows an increasing trend over time.
- At the beginning of January, this rate is at its lowest point of the year, at 3.3%. It then gradually increases and reaches its peak at the end of September, at 4.34%. However, the rate in the last days of December decreased to 3.36%.

c. Factors influencing the conversion rate of the ABC website::

❖ **Objective influencing factors:** These are uncontrollable factors. Hotels need to implement business strategies appropriate to these factors.

- **Market factors:**

✧ Peak seasons with favorable weather and long holidays increase demand in the hotel tourism market, contributing to an increase in hotel bookings.

✧ Conversely, during the off-peak season characterized by unfavorable weather and less favorable timing, the number of tourists decreases, impacting hotel operations.

During this time, hotels need to implement marketing strategies, price discounts, or hotel amenities to attract customers.

- **Technical infrastructure factor:**

✧ Convenient and fast road systems and transportation between hotels and tourist attractions help tourists have a good experience when choosing to travel in that country, as well as when choosing to book a room at that hotel..

✧ Furthermore, the development of the Internet system helps customers easily access hotels, leading to an increase in bookings..

- **Other objective factors:**

✧ Political, economic, and social factors: A stable political environment is the foundation for economic development. As people's living standards increase, the demand for leisure, entertainment, and vacations will also grow.

✧ Natural disasters, pandemics: A classic example is the Covid pandemic, which has brought the entire tourism industry, as well as the accommodation business sector, to a standstill, even on the brink of bankruptcy.

❖ **Subjective influencing factors:**

- **Human factor:**

✧ High ratings, positive feedback from previous customers about the hotel staff's service attitude help the hotel establish credibility and reliability with potential customers. The hotel retains existing customers while attracting new ones.

✧ Therefore, the hotel needs to excel in training a professional staff, tightly manage customer service, and establish reward policies as well as disciplinary measures for employees in a timely manner.

- **Physical infrastructure factor at the hotel:**

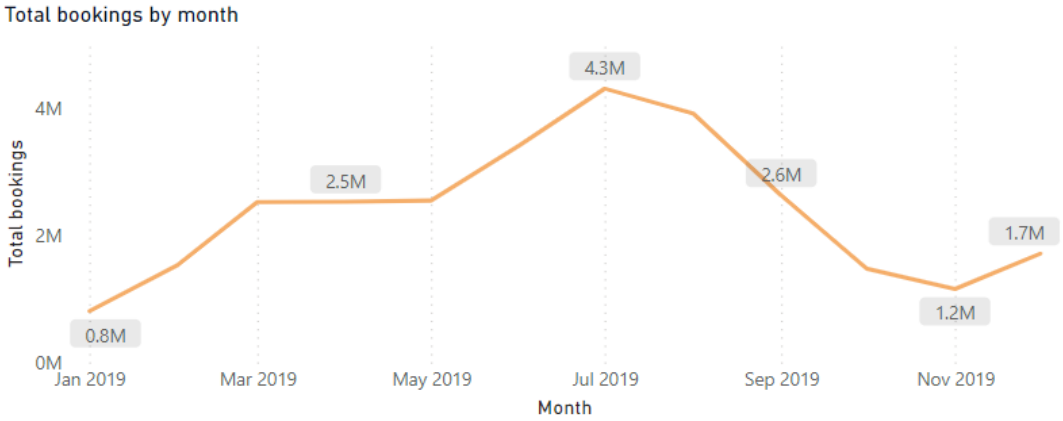
✧ Hotels must continually improve the quality and safety of rooms to create new experiences for customers and avoid boredom.

- **Business location factor:**

✧ One of the factors in deciding to book a hotel is its location in the city center, close to tourist areas and attractions.

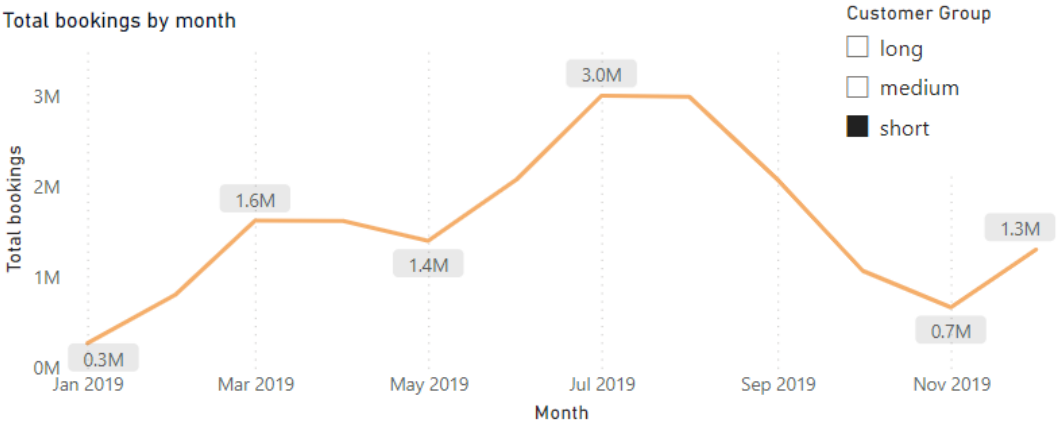
3. Analyze the dataset to determine which country it is from:

- Observing the chart showing the total number of bookings on the ABC website (Figure 11), the number of bookings is lowest in January (800,000 bookings), steadily increasing until March and stabilizing for the next 2 months (2.5 million bookings), then peaking in July (4.3 million bookings). Bookings decrease sharply until November (1.2 million bookings) and increase slightly in December.



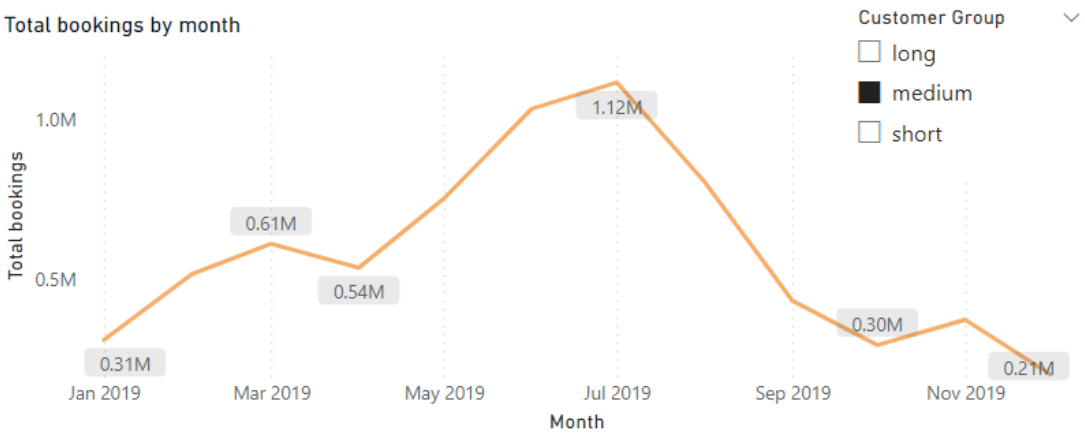
(Figure 11- The total number of bookings made on the ABC website by month)

- Observing the chart showing the total number of bookings on the ABC website for the short-term customer segment (Figure 12), the fluctuations in bookings for this customer segment show similarities with the total bookings for the year 2019.



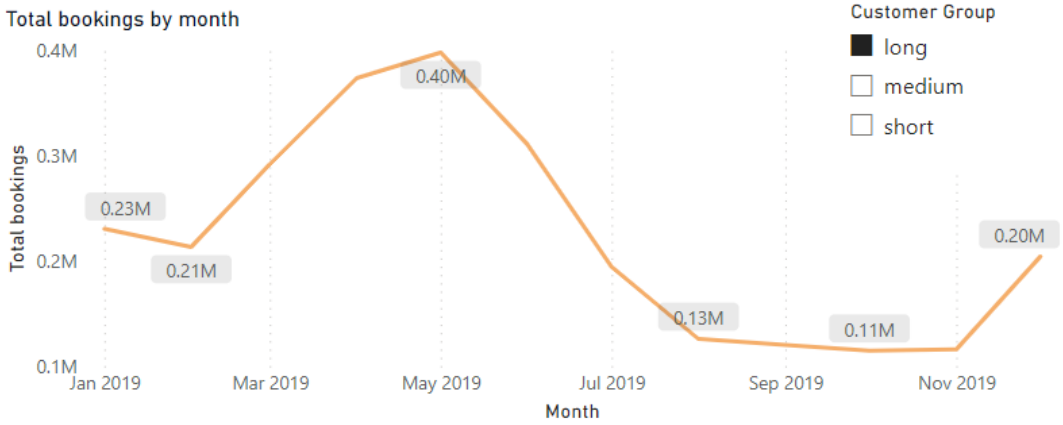
(Figure 12- The total number of bookings made on the ABC website by month for the short-term customer segment)

- Observing the chart showing the total number of bookings on the ABC website for the medium-term customer segment (Figure 13), bookings increase sharply around 15 days to 1 month before the peak season.



(Figure 13-
The total number of bookings made on the ABC website by month for the medium-term customer segment)

- And observing the chart showing the total number of bookings on the ABC website for the long-term customer segment (Figure 14), bookings increase significantly around more than one month before the peak season.



(Figure 14- The total number of bookings made on the ABC website by month for the long-term customer segment)

- With the observations above, we can see similarities with the high travel season, low travel season, and shoulder season (the intermediate season between the high travel season and the low travel season) in France::

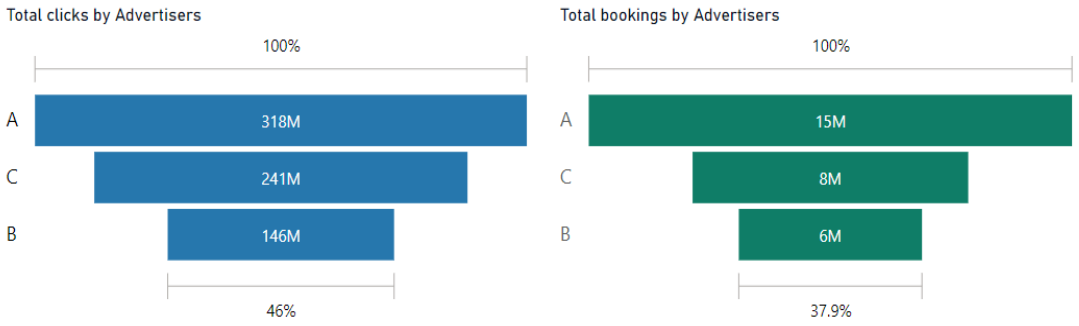
✧ The high travel season in France begins in late June and ends in the first half of September. The peak time of the high season in France is in July and August. This period coincides with the summer vacation of schools in most northern hemisphere countries, along with high temperatures. During this time, France hosts many famous festivals such as: Music Festival (La Fête de la Musique - June 21), Avignon Festival (July), French National Day (La Fête Nationale - July 14), Festival of Lights (July 19-28), Paris Plage (July to September), etc.

✧ The low travel season in France starts in November and ends in late March (the coldest months) in most major tourist cities and landmarks. While the coldest months are typically the low season, December is an exception. In December, the main destinations in France focus entirely on decorating and organizing Christmas markets. December is the peak month for retail in this continent. Tourists also flock to France during Christmas and New Year's.

✧ The shoulder season in France falls between early spring and mid to late autumn. This period is around March and April, as well as late September to early November.

II. BUSINESS EFFECTIVENESS OF ADVERTISERS A, B, C

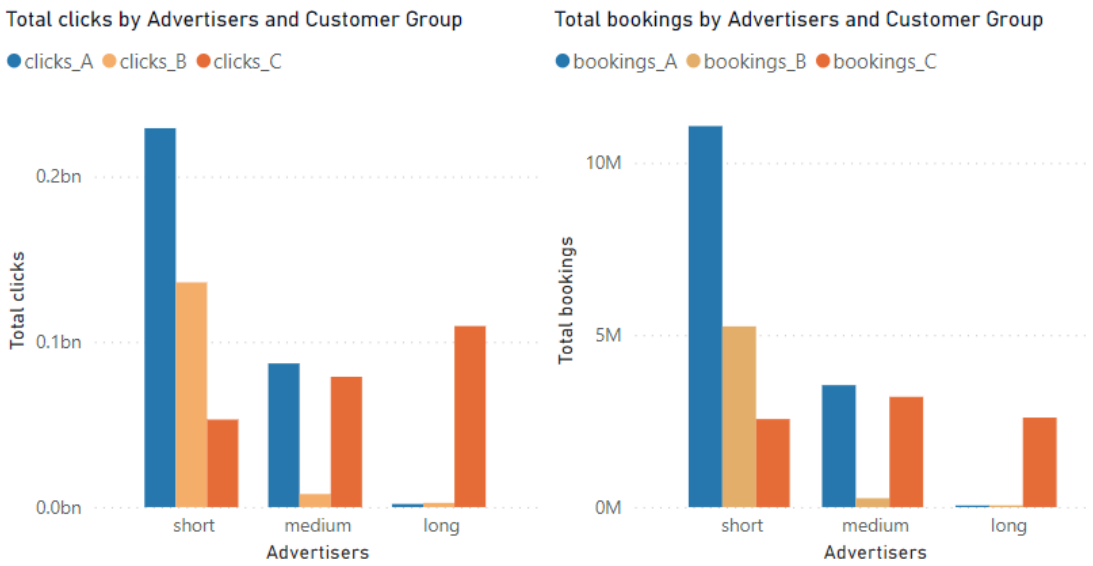
1. About advertisers A, B, C:



(Figure 15- The total number of clicks and bookings made on the ABC website by advertisers)

- In 2019, the number of clicks and bookings from advertiser A accounted for the largest proportion on the ABC website. The number of bookings from advertiser A was nearly twice that of C and 2.5 times that of B.
- The marketing strategies implemented by advertiser A on the ABC website are attracting more customers than those of advertisers B and C.

2. Potential customers for each advertiser:



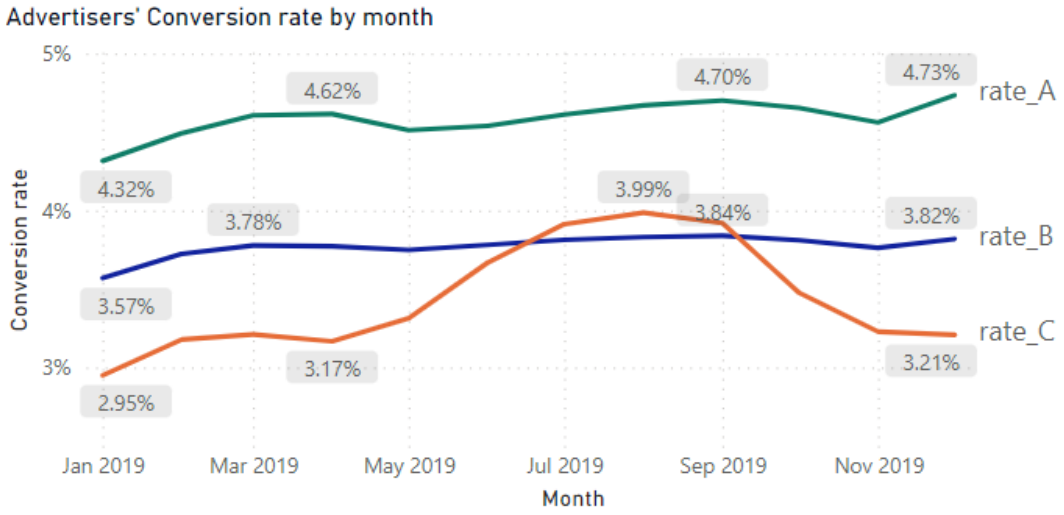
(Figure 16- The total number of clicks and bookings for each advertiser by customer segment)

- The potential customer segment for advertiser A is the short-term and medium-term customer segments. In 2019, A attracted 229 million clicks and 11 million bookings from the short-term customer segment, and 86 million clicks and 3.5 million bookings from the medium-term customer segment.

- The potential customer segment for advertiser B is the short-term customer segment, as most of the clicks and bookings in 2019 for B came from this customer segment.

- The potential customer segment for advertiser C includes all three customer segments because the clicks and bookings in 2019 from these three customer segments did not show significant differences.

3. Conversion rates of the advertisers:



(Figure 17- Conversion rates of the advertisers by day)

- Advertiser A had a higher conversion rate than advertisers B and C. The conversion rates of advertisers A and B fluctuated slightly, with a slight upward trend in 2019.
- The conversion rate of advertiser C showed a slight upward trend but fluctuated greatly from May to November.
  - ✓ In May and June, the number of clicks and bookings increased compared to the previous month, but the rate of increase in bookings was higher, so the conversion rate of C during this period also increased.
  - ✓ In July, the number of clicks decreased by 3% compared to the previous month, but the number of bookings increased by 4%, still helping to continue the increase in the conversion rate of C
  - ✓ In August, the rate of decrease in clicks was more significant than in bookings, leading to C's conversion rate reaching its peak at 3.99% (as the conversion rate is inversely proportional to the number of clicks).
  - ✓ From September to November, the rate of decrease in bookings continuously outpaced the rate of decrease in clicks, resulting in a sharp decline in C's conversion rate during this period.

month	bookings_C	clicks_C	booking_change_rate(%)	click_change_rate(%)
05	939855	28363751	12	7
06	1059214	28876794	13	2
07	1098209	28059473	4	-3
08	867998	21776036	-21	-22
09	575949	14693609	-34	-33
10	398908	11477248	-31	-22
11	362098	11216181	-9	-2

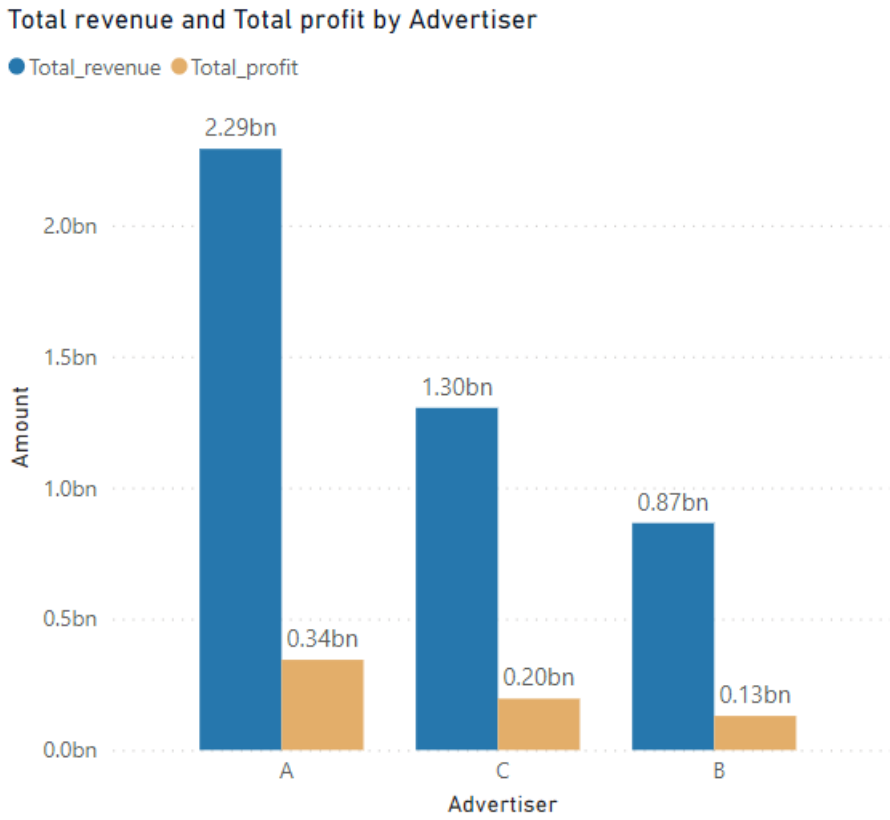
(Figure 18 - The change rate in the number of clicks and bookings for C from May to November)

4. Profit of the advertisers::

```
[ ] # Create a summary table for advertisers
    df_summary = pd.DataFrame({'Advertiser': pd.Series(dtype='str'),
                       'Total_clicks': pd.Series(dtype='float'),
                       'Total_bookings': pd.Series(dtype='float'),
                       'Total_cost': pd.Series(dtype='float'),
                       'Total_revenue': pd.Series(dtype='float')})
    df_summary.loc[len(df_summary.index)]=['A',df['clicks_A'].sum(),df['bookings_A'].sum(),df['cost_A'].sum(),df['booking_rev_A'].sum()]
    df_summary.loc[len(df_summary.index)]=['B',df['clicks_B'].sum(),df['bookings_B'].sum(),df['cost_B'].sum(),df['booking_rev_B'].sum()]
    df_summary.loc[len(df_summary.index)]=['C',df['clicks_C'].sum(),df['bookings_C'].sum(),df['cost_C'].sum(),df['booking_rev_C'].sum()]
    df_summary['Total_profit']=df_summary['Total_revenue']*0.15
    df_summary
```

	Advertiser	Total_clicks	Total_bookings	Total_cost	Total_revenue	Total_profit
0	A	318086328.0	14653703.0	294221570.0	2.291405e+09	343710763.2
1	B	146406387.0	5554179.0	129498235.0	8.651945e+08	129779174.4
2	C	241342753.0	8358192.0	194732233.0	1.304232e+09	195634819.5

(Figure 19- Profit of the advertisers)



(Figure 20- Revenue and profit of the advertisers)

- In 2019, Advertiser A had the highest revenue (over 2 billion), followed by Advertiser C (1.3 billion) and B (870 million).
- With the same profit margin of 15%, Advertiser A also had the highest profit at 340 million, followed by Advertiser C with a profit of 200 million and B with 130 million.

5. Cost-to-Revenue Ratio:

Advertiser	Total_revenue	Total_cost	Cost_Rev_Ratio
A	2,291,405,088	294,221,570	13%
B	865,194,496	129,498,235	15%
C	1,304,232,130	194,732,233	15%
Total	**4,460,831,714**	**618,452,038**	**43%**

(Figure 21 - Cost-to-Revenue Ratio of the advertisers)

With a cost-to-revenue ratio of 13%, A is the most cost-effective advertiser in generating revenue compared to the other two advertisers. This means that to generate 100 units of revenue, A needs to spend 13 units on costs, while B and C need to spend up to 15 units on costs.

D. SUMMARY:

I. General Observations:

- Overall, the conversion rate of Website ABC in 2019 showed an increasing trend over time. Advertisers effectively utilized peak travel seasons and shoulder seasons to boost the conversion rate of Website ABC in 2019 by implementing stimulating travel demand strategies and accompanying promotional policies.
- The short-term customer segment has the highest proportion of clicks and bookings among the three customer segments of Website ABC, as the marketing strategies of advertisers A, B, and C attract the majority of customers compared to other customer segments.
- The mid-term customer segment ranks second in proportion, primarily due to the strategies of advertisers A (50%) and C (45%).
- The long-term customer segment has the lowest proportion, with 96% of clicks and bookings coming from advertiser C.
- The conversion rate of advertiser A is higher than that of B, but both show a slight increasing trend with small fluctuations in 2019.
- The conversion rate of advertiser C also shows an increasing trend but fluctuates considerably, especially in the second half of 2019.
- A is the most effective advertiser on Website ABC in 2019 with the highest revenue, profit, and conversion rate of users finding suitable bookings compared to advertisers B and C. Additionally, A is also the advertiser that utilizes costs most efficiently to generate revenue.

II. Proposed Marketing Strategy Improvements for 2020:

- The objectives of the marketing strategies are to attract more customers (increase clicks) and convince them to choose their hotel for booking (increase bookings), thereby increasing revenue and profit with optimized costs

- Identifying potential customer groups will help the hotel to have the right direction in building appropriate marketing strategies, increasing the rate of customers finding suitable hotel bookings according to their needs.

1. General marketing strategy recommendations:

- **Designing Professional Hotel Website:** A significant portion of potential customers research hotel websites before making booking decisions. Therefore, hotels should focus on building brand reputation, credibility, and professionalism

through website interfaces by enhancing UX/UI to reach and engage a wide customers interested in using their services.

- **Implementing Video Marketing:** Video marketing is becoming increasingly popular due to its uniqueness, impact, and memorability, proving to be more effective in attracting and reaching customers, thereby contributing to better hotel booking rates. Additionally, hotels can collaborate with influential social media reviewers to feature in their video marketing content.
- **Understanding the actual needs of potential customer segments to develop advertising processes:** When customers desire a leisure trip, hotels should launch campaigns that inspire travel to specific countries or regions by developing travel blogs suggesting interesting tourist spots, renowned restaurants, cafes, and hotels in these areas. Once customers select a destination and start researching attractions and itineraries, this is when hotels should persuade them to choose their services through pricing policies, service quality, luxurious experiential spaces, and convenient locations to tourist spots. Importantly, your campaign should aim to attract and create desire for travel to a specific country or region.

2. Marketing strategies for each advertiser:

- Advertiser A:
  - Short-term customer segments have a maximum time-to-date of 14 days, so ensuring the availability of rooms at the time of booking will increase the likelihood of booking for customers. Additionally, providing additional information about nearby tourist destinations also enhances the decision-making process for booking rooms.
  - Mid-term customer segments have a maximum time-to-date of 60 days and are typically concerned about how the hotel handles situations where they cannot check in according to their booking schedule. Therefore, implementing policies that support rescheduling bookings will make customers feel more reassured when making reservations.
  - With the number of bookings in 2019, advertiser A could collect reviews from previous customers to improve services and offer promotional programs for repeat customers or referrals.

- Advertiser B:
  - Although having a similar potential customer base as advertiser A, advertiser B has fewer clicks, less than half of A's. B needs to review customer feedback on infrastructure and service quality for improvement. If the location is not advantageous for tourist attractions, B could renovate the hotel to create a more luxurious space and transition into a resort service.

- Advertiser C:
  - Pre-bookings made more than 15 days in advance through advertiser C represent approximately 70% of the total, indicating a limited number of available rooms. The conversion rate of C in 2019 also demonstrates that it focused its marketing campaigns during the peak tourism season, resulting in significant increases in July, August, and September. C's strength lies in attracting all three customer segments. Leveraging these advantages and expanding its scale while running continuous marketing campaigns throughout the year could further increase bookings.