TASK-0:

HOTEL DEMAND PREDICTION

BY: Harsh Makwana

Abstract:

In tourism industry it is prudent for hotel business owners to know about customer behaviour regarding hotel booking which could have several parameters for it to matter.Based on prediction of demand of customers of hotel bookings,Hotel owners can optimise their business of service and operations management to adequately fulfil the customer demand while maximising their profits.

In this report we will look at how can we predict customer demand of hotel rooms based on several variables using several well known ML techniques.

Problem Statement:

In the tourism industry,hotels would like to predict customer demand for room bookings and cancellation rates of customers based on various factors that exist such as time of arrival , type of rooms etc.

Knowing the distribution of room demand based on time would help hotels to adjust themselves accordingly based on customer needs and knowing cancellation rates based on various factors would help hotels determine why a customer in general would cancel tickets,which would help hotels calculate their expected revenue with less errors and randomness.

Also knowing what type of customers would bring personal vehicles would help hotels to determine parking space required for a hotel to manage.

Business Needs Assessment:

Running a hotel business profitably is very hard.There are unforeseen risks involved.Hence we plan operations using our prediction.

Hotels ideally would perform systematic analysis of customer demand ,behaviour and preference and would align their operations accordingly.

Forecasting demand which uses historical data helps hotels to manage their inventory accordingly.Also these forecasts help us estimate profit,revenue and other important values.

Mainly there are 3 types of forecasting:

(i) Operational forecasting is about predicting how much manpower is required to manage operational tasks effectively.

(ii) Financial forecasting is about predicting future expected revenue and profits.

(iii) Revenue management forecasting is done to predict rise or fall in demand as to help stay business profitable.

Target Specification:

Hotel Business would target variables which would help them in getting fair and accurate assessment of demand some of which are:

(i) From which country most of guests are coming from?
(ii) Busiest time of year
(iii) Distribution of stay of people in rooms.
And so on.

These values would help us in determining operational and financial requirements of a hotel.

External Search:

Dataset was obtained from research paper by NunoAntonio, Anade Almeida,Luís Nunes and Kaggle :

https://www.sciencedirect.com/science/article/pii/S2352340918315191

https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand

Benchmarking alternate products:

Tourism and travel industries primarily do forecasting based on PNR(Passenger Name Record) data obtained from the aviation industry.

However, the utility of this data is limited as we lack industry specific data on tourism and hence this model uses Hotel booking data which is industry specific.

Applicable Regulations:

All the standard regulations of developing an application apply of its use within given jurisdiction such as:

IP Rights ,Business Registration, Copyrights etc..

In addition to that, since this product acquires data from various hotels it should also follow the standard regulatory norm of collection of data in accordance with the privacy laws of the country.

Applicable Constraints:

Maintenance of servers for collection of data and computational tasks is the main constraint in this product.

Business Opportunity:

Many hotel owners would like to know and forecast room demand bookings to ascertain operational and financial tasks.An analytics firm could collect and gather data about similar hotels to do prediction and help hotels predict their room demand and cancellation rates to make more prudent decisions.

Concept Development:

To forecast customer demand of rooms and parking spaces application would gather data of hotels from various sources and then it uses various supervised learning ML algorithms to predict peak room bookings in a year and parking space required based on several hotel's data in similar situations.

Code Implementation:

Implementation of this code is given in my github :
https://github.com/harsh392/EDA-Hotel-demand-booking

countries from which guests come from?

```python
country_wise_guests = df[df['is_canceled'] == 0]['country'].value_counts().reset_index()
country_wise_guests.columns = ['country', 'No of guests']
country_wise_guests
```
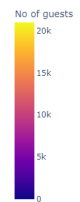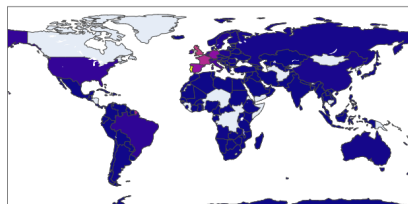
	country	No of guests
0	PRT	20977
1	GBR	9668
2	FRA	8468
3	ESP	6383
4	DEU	6067
...	...	...
161	AIA	1
162	SDN	1
163	BHS	1
164	MAC	1
165	NPL	1

166 rows × 2 columns

```python
basemap = folium.Map()
guests_map = px.choropleth(country_wise_guests, locations = country_wise_guests['country'],
                           color = country_wise_guests['No of guests'], hover_name = country_wise_guests['country'])
guests_map.show()
```
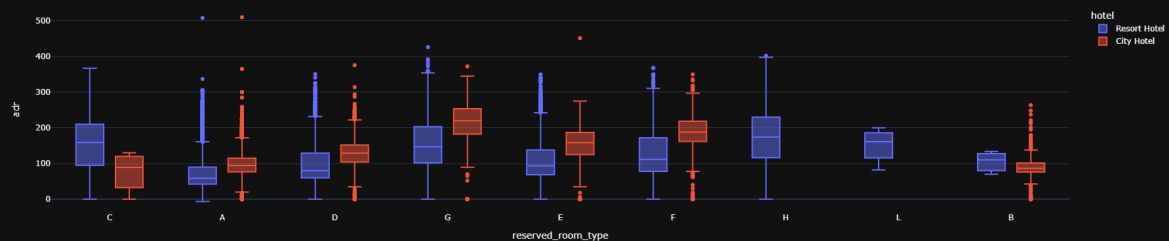

How much guest pay per night

```python
data = df[df['is_canceled'] == 0]

px.box(data_frame = data, x = 'reserved_room_type', y = 'adr', color = 'hotel', template = 'plotly_dark')
```

Variation of Hotel room prices by month

```python
data_resort = df[(df['hotel'] == 'Resort Hotel') & (df['is_canceled'] == 0)]
data_city = df[(df['hotel'] == 'City Hotel') & (df['is_canceled'] == 0)]

resort_hotel = data_resort.groupby(['arrival_date_month'])['adr'].mean().reset_index()

city_hotel=data_city.groupby(['arrival_date_month'])['adr'].mean().reset_index()

final_hotel = resort_hotel.merge(city_hotel, on = 'arrival_date_month')
final_hotel.columns = ['month', 'price_for_resort', 'price_for_city_hotel']
final_hotel
```

	month	price_for_resort	price_for_city_hotel
0	April	75.867816	111.962267
1	August	181.205892	118.674598
2	December	68.410104	88.401855
3	February	54.147478	86.520062
4	January	48.761125	82.330983
5	July	150.122528	115.818019
6	June	107.974850	117.874360
7	March	57.056838	90.658533
8	May	76.657558	120.669827
9	November	48.706289	86.946592
10	October	61.775449	102.004672
11	September	96.416860	112.776582

```python
import sort_dataframeby_monthorweek as sd

def sort_month(df, column_name):
    return sd.Sort_Dataframeby_Month(df, column_name)

final_prices = sort_month(final_hotel, 'month')
final_prices
```

	month	price_for_resort	price_for_city_hotel
0	January	48.761125	82.330983
1	February	54.147478	86.520062
2	March	57.056838	90.658533
3	April	75.867816	111.962267
4	May	76.657558	120.669827
5	June	107.974850	117.874360
6	July	150.122528	115.818019
7	August	181.205892	118.674598
8	September	96.416860	112.776582
9	October	61.775449	102.004672
10	November	48.706289	86.946592
11	December	68.410104	88.401855

```python
plt.figure(figsize = (17, 8))

px.line(final_prices, x = 'month', y = ['price_for_resort','price_for_city_hotel'],
        title = 'Room price per night over the Months', template = 'plotly_dark')
```
Python

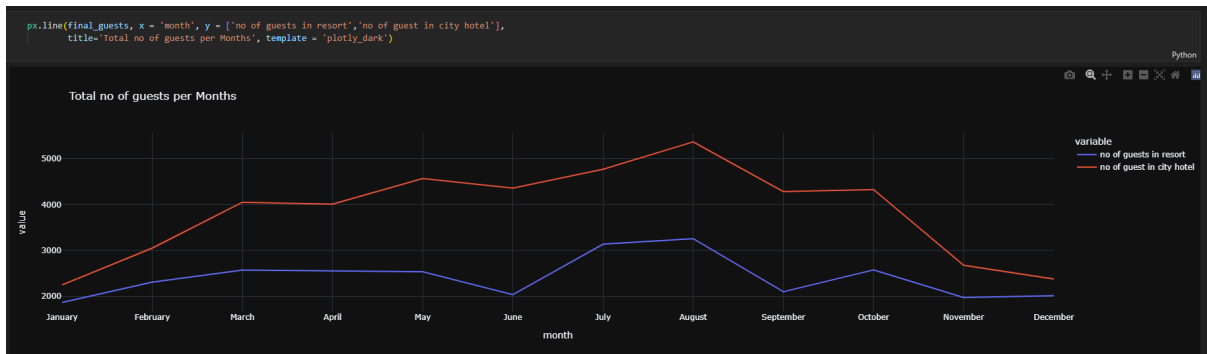**Room price per night over the Months**

Busy months

```python
resort_guests = data_resort['arrival_date_month'].value_counts().reset_index()
resort_guests.columns=['month','no of guests']

city_guests = data_city['arrival_date_month'].value_counts().reset_index()
city_guests.columns=['month','no of guests']

final_guests = resort_guests.merge(city_guests,on='month')
final_guests.columns=['month','no of guests in resort','no of guest in city hotel']

final_guests = sort_month(final_guests,'month')
final_guests
```

	month	no of guests in resort	no of guest in city hotel
0	January	1866	2249
1	February	2308	3051
2	March	2571	4049
3	April	2550	4010
4	May	2535	4568
5	June	2037	4358
6	July	3137	4770
7	August	3257	5367
8	September	2102	4283
9	October	2575	4326
10	November	1975	2676
11	December	2014	2377

```python
px.line(final_guests, x = 'month', y = ['no of guests in resort','no of guest in city hotel'],
        title='Total no of guests per Months', template = 'plotly_dark')
```

Total no of guests per Months

```python
filter = df['is_canceled'] == 0
data = df[filter]

data['total_nights'] = data['stays_in_weekend_nights'] + data['stays_in_week_nights']

stay = data.groupby(['total_nights', 'hotel']).agg('count').reset_index()
stay = stay.iloc[:, :3]
stay = stay.rename(columns={'is_canceled':'Number of stays'})

px.bar(data_frame = stay, x = 'total_nights', y = 'Number of stays', color = 'hotel', barmode = 'group',
       template = 'plotly_dark')
```

How long do people stay at hotels?

Final Product Prototype:

Final product would ascertain for a particular hotel to get to know its type (on which category would it belong to) and perform predictive analysis to determine room and space demand for a hotel in a given period of time to help hotels with financial and operational costs.

And further we could implement ml classification algorithms to determine rate of cancellations of new bookings.
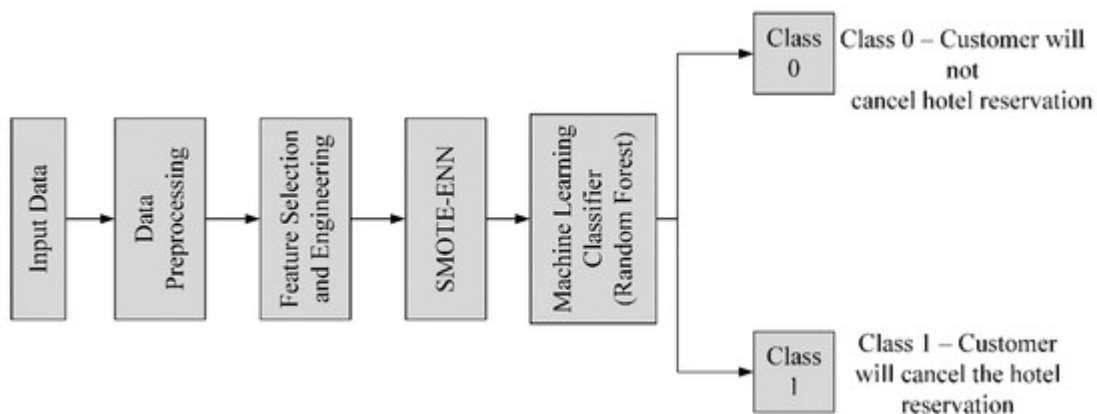
Conclusion:

We have use various data visualisation tools to predict the room bookings demand and parking space demand for a given type of hotel in a given time of year.Furthermore we can extend our implementation to predict the chance whether the new booking would be cancelled in future or not using various Ml algorithms (like Random algorithms).