

Hotel Booking Demand with Economic Indicators

▼ About Dataset

- This dataset contains 119390 observations for a City Hotel and a Resort Hotel. Each observation represents a hotel booking between the 1st of July 2015 and 31st of August 2017. I have coupled each month of booking with leading economic indicators for the United States for that timeframe using the FRED.
- While this dataset does not specify the location of the hotels, I have gone with the assumption of both being located in the United States for ease of applying economic data for learning purposes

Acknowledgments

- The data is originally from the article Hotel Booking Demand Datasets, written by Nuno Antonio, Ana Almeida, and Luis Nunes for Data in Brief, Volume 22, February 2019. The original cleaned version was uploaded to Kaggle by by Thomas Mock and Antoine Bichat. I used that dataset to couple with various indicators:
- GDP
- CPI
- CPI for Hotels
- Fuel Prices
- Disposable income per capita
- Inflation
- Consumer Sentiment
- GINI Indicator

```
!pip install pyspark
#installing the pyspark
```

```
Collecting pyspark
  Downloading pyspark-3.4.1.tar.gz (310.8 MB)
    310.8/310.8 MB 2.1 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.10/dist-packages (from pyspark) (0.10.9.7)
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) ... done
  Created wheel for pyspark: filename=pyspark-3.4.1-py2.py3-none-any.whl size=311285387 sha256=de30b62e3fa35b42f4ff330c0af120f0f700
  Stored in directory: /root/.cache/pip/wheels/0d/77/a3/ff2f74cc9ab41f8f594dabf0579c2a7c6de920d584206e0834
Successfully built pyspark
Installing collected packages: pyspark
Successfully installed pyspark-3.4.1
```

```
# Importing libraries
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql import functions as f
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

spark = SparkSession.builder.appName("Hotel Booking").getOrCreate()

file = "/content/drive/MyDrive/Data Analysis Files/hotel_bookings_raw.csv"

original_datasets = spark.read.csv(
    file,
    header=True,
    inferSchema=True
)

df = original_datasets.alias("copy")

df.show()
```

hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_i
Resort Hotel	0	342	2015	July	27		1
Resort Hotel	0	737	2015	July	27		1
Resort Hotel	0	7	2015	July	27		1
Resort Hotel	0	13	2015	July	27		1
Resort Hotel	0	14	2015	July	27		1
Resort Hotel	0	14	2015	July	27		1
Resort Hotel	0	0	2015	July	27		1
Resort Hotel	0	9	2015	July	27		1
Resort Hotel	1	85	2015	July	27		1
Resort Hotel	1	75	2015	July	27		1
Resort Hotel	1	23	2015	July	27		1
Resort Hotel	0	35	2015	July	27		1
Resort Hotel	0	68	2015	July	27		1
Resort Hotel	0	18	2015	July	27		1
Resort Hotel	0	37	2015	July	27		1
Resort Hotel	0	68	2015	July	27		1
Resort Hotel	0	37	2015	July	27		1
Resort Hotel	0	12	2015	July	27		1
Resort Hotel	0	0	2015	July	27		1
Resort Hotel	0	7	2015	July	27		1

only showing top 20 rows



```
df.count()
```

```
119390
```

```
df.columns
```

```
['hotel',
 'is_canceled',
 'lead_time',
 'arrival_date_year',
 'arrival_date_month',
 'arrival_date_week_number',
 'arrival_date_day_of_month',
 'stays_in_weekend_nights',
 'stays_in_week_nights',
 'adults',
 'children',
 'babies',
 'meal',
 'country',
 'market_segment',
 'distribution_channel',
 'is_repeated_guest',
 'previous_cancellations',
 'previous_bookings_not_canceled',
 'reserved_room_type',
 'assigned_room_type',
 'booking_changes',
 'deposit_type',
 'agent',
 'days_in_waiting_list',
 'customer_type',
 'adr',
 'required_car_parking_spaces',
 'total_of_special_requests',
 'reservation_status',
 'reservation_status_date',
 'MO_YR',
 'CPI_AVG',
 'INFLATION',
 'INFLATION_CHG',
 'CSMR_SENT',
 'UNRATE',
 'INTRSRT',
 'GDP',
 'FUEL_PRCS',
 'CPI_HOTELS',
 'US_GINI',
 'DIS_INC']
```

```
len(df.columns)
```

```
43
```

```
print(f"The Datasets is having {df.count()} rows and {len(df.columns)} columns.")
```

```
The Datasets is having 119390 rows and 43 columns.
```

```
df.select([f.sum(f.col(i).isNull().cast('int')).alias(i) for i in df.columns]).show()
```

hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights
0	0	0	0	0	0	0	0	0

```
df.printSchema()
```

```
root
|-- hotel: string (nullable = true)
|-- is_canceled: integer (nullable = true)
|-- lead_time: integer (nullable = true)
|-- arrival_date_year: integer (nullable = true)
|-- arrival_date_month: string (nullable = true)
|-- arrival_date_week_number: integer (nullable = true)
|-- arrival_date_day_of_month: integer (nullable = true)
|-- stays_in_weekend_nights: integer (nullable = true)
|-- stays_in_week_nights: integer (nullable = true)
|-- adults: integer (nullable = true)
|-- children: string (nullable = true)
|-- babies: integer (nullable = true)
|-- meal: string (nullable = true)
|-- country: string (nullable = true)
|-- market_segment: string (nullable = true)
|-- distribution_channel: string (nullable = true)
|-- is_repeated_guest: integer (nullable = true)
|-- previous_cancellations: integer (nullable = true)
|-- previous_bookings_not_canceled: integer (nullable = true)
|-- reserved_room_type: string (nullable = true)
|-- assigned_room_type: string (nullable = true)
|-- booking_changes: integer (nullable = true)
|-- deposit_type: string (nullable = true)
|-- agent: string (nullable = true)
|-- days_in_waiting_list: integer (nullable = true)
|-- customer_type: string (nullable = true)
|-- adr: double (nullable = true)
|-- required_car_parking_spaces: integer (nullable = true)
|-- total_of_special_requests: integer (nullable = true)
|-- reservation_status: string (nullable = true)
|-- reservation_status_date: string (nullable = true)
|-- MO_YR: string (nullable = true)
|-- CPI_AVG: string (nullable = true)
|-- INFLATION: string (nullable = true)
|-- INFLATION_CHG: string (nullable = true)
|-- CSMR_SENT: string (nullable = true)
|-- UNRATE: string (nullable = true)
|-- INTRSR: string (nullable = true)
|-- GDP: string (nullable = true)
|-- FUEL_PRCS: string (nullable = true)
|-- CPI_HOTELS: string (nullable = true)
|-- US_GINI: string (nullable = true)
|-- DIS_INC: string (nullable = true)
```

```
[i for i,v in df.dtypes if v in ['string']]
```

```
['hotel',
 'arrival_date_month',
 'children',
 'meal',
 'country',
 'market_segment',
 'distribution_channel',
 'reserved_room_type',
 'assigned_room_type',
 'deposit_type',
 'agent',
 'customer_type',
 'reservation_status',
 'reservation_status_date',
 'MO_YR',
 'CPI_AVG',
 'INFLATION',
 'INFLATION_CHG',
 'CSMR_SENT',
 'UNRATE',
 'INTRSR',
 'GDP',
 'FUEL_PRCS',
 'CPI_HOTELS',
 'US_GINI',
 'DIS_INC']
```

```
df.select([i for i,v in df.dtypes if v in ['string']]).show()
```

hotel	arrival_date_month	children	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	de
Resort Hotel	July	0	BB	PRT	Direct	Direct	C	C	
Resort Hotel	July	0	BB	PRT	Direct	Direct	C	C	
Resort Hotel	July	0	BB	GBR	Direct	Direct	A	C	
Resort Hotel	July	0	BB	GBR	Corporate	Corporate	A	A	
Resort Hotel	July	0	BB	GBR	Online TA	TA/TO	A	A	
Resort Hotel	July	0	BB	GBR	Online TA	TA/TO	A	A	
Resort Hotel	July	0	BB	PRT	Direct	Direct	C	C	
Resort Hotel	July	0	FB	PRT	Direct	Direct	C	C	
Resort Hotel	July	0	BB	PRT	Online TA	TA/TO	A	A	
Resort Hotel	July	0	HB	PRT	Offline TA/TO	TA/TO	D	D	
Resort Hotel	July	0	BB	PRT	Online TA	TA/TO	E	E	
Resort Hotel	July	0	HB	PRT	Online TA	TA/TO	D	D	
Resort Hotel	July	0	BB	USA	Online TA	TA/TO	D	E	
Resort Hotel	July	1	HB	ESP	Online TA	TA/TO	G	G	
Resort Hotel	July	0	BB	PRT	Online TA	TA/TO	E	E	
Resort Hotel	July	0	BB	IRL	Online TA	TA/TO	D	E	
Resort Hotel	July	0	BB	PRT	Offline TA/TO	TA/TO	E	E	
Resort Hotel	July	0	BB	IRL	Online TA	TA/TO	A	E	
Resort Hotel	July	0	BB	FRA	Corporate	Corporate	A	G	
Resort Hotel	July	0	BB	GBR	Direct	Direct	G	G	

only showing top 20 rows

```
df.groupBy(f.col('CPI_AVG')).count().show()
```

CPI_AVG	count
243.892	4826
244.193	4634
244.243	4038
238.034	3615
237.733	5742
237.652	4482
237.001	275
244.006	4858
238.017	3077
237.761	3062
245.183	3417
236.222	151
235.342	44
240.222	4552
244.004	5006
242.026	5021
244.163	4060
241.176	4993
242.637	4338
246.435	393

only showing top 20 rows

```
df = df.withColumn("CPI_AVG", f.col("CPI_AVG").cast("double"))
df = df.withColumn("INFLATION_CHG", f.col("INFLATION_CHG").cast("double"))
df = df.withColumn("CSMR_SENT", f.col("CSMR_SENT").cast("double"))
df = df.withColumn("UNRATE", f.col("UNRATE").cast("double"))
df = df.withColumn("INTRSRT", f.col("INTRSRT").cast("double"))
df = df.withColumn("GDP", f.col("GDP").cast("double"))
df = df.withColumn("FUEL_PRCS", f.col("FUEL_PRCS").cast("double"))
df = df.withColumn("CPI_HOTELS", f.col("CPI_HOTELS").cast("double"))
df = df.withColumn("US_GINI", f.col("US_GINI").cast("double"))
df = df.withColumn("DIS_INC", f.col("DIS_INC").cast("double"))
df = df.withColumn("INFLATION", f.col("INFLATION").cast("double"))
df = df.withColumn("children", f.col("children").cast("double"))
df = df.withColumn("agent", (f.when(f.col("agent")= "NULL", "0").otherwise(f.col("agent"))).cast('double'))
```

```
[i for i,v in df.dtypes if v in ['string']]
```

```
['hotel',
'arrival_date_month',
'meal',
'country',
'market_segment',
'distribution_channel',
'reserved_room_type',
'assigned_room_type',
'deposit_type',
'customer_type',
'reservation_status',
```

```
'reservation_status_date',  
'MO_YR']
```

```
df.select([i for i,v in df.dtypes if v in ['string']]).show()
```

hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	deposit_type
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	GBR	Direct	Direct	A	C	No Deposit
Resort Hotel	July	BB	GBR	Corporate	Corporate	A	A	No Deposit
Resort Hotel	July	BB	GBR	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	BB	GBR	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	FB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	HB	PRT	Offline TA/TO	TA/TO	D	D	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	E	E	No Deposit
Resort Hotel	July	HB	PRT	Online TA	TA/TO	D	D	No Deposit
Resort Hotel	July	BB	USA	Online TA	TA/TO	D	E	No Deposit
Resort Hotel	July	HB	ESP	Online TA	TA/TO	G	G	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	E	E	No Deposit
Resort Hotel	July	BB	IRL	Online TA	TA/TO	D	E	No Deposit
Resort Hotel	July	BB	PRT	Offline TA/TO	TA/TO	E	E	No Deposit
Resort Hotel	July	BB	IRL	Online TA	TA/TO	A	E	No Deposit
Resort Hotel	July	BB	FRA	Corporate	Corporate	A	G	No Deposit
Resort Hotel	July	BB	GBR	Direct	Direct	G	G	No Deposit

only showing top 20 rows

```
from pyspark.sql import types as t
```

```
[i for i,v in df.dtypes if v in ['string']]
```

```
['hotel',  
'arrival_date_month',  
'meal',  
'country',  
'market_segment',  
'distribution_channel',  
'reserved_room_type',  
'assigned_room_type',  
'deposit_type',  
'customer_type',  
'reservation_status',  
'reservation_status_date',  
'MO_YR']
```

```
df.select([i for i,v in df.dtypes if v in ['string']]).show()
```

hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	deposit_type
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	GBR	Direct	Direct	A	C	No Deposit
Resort Hotel	July	BB	GBR	Corporate	Corporate	A	A	No Deposit
Resort Hotel	July	BB	GBR	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	BB	GBR	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	FB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	HB	PRT	Offline TA/TO	TA/TO	D	D	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	E	E	No Deposit
Resort Hotel	July	HB	PRT	Online TA	TA/TO	D	D	No Deposit
Resort Hotel	July	BB	USA	Online TA	TA/TO	D	E	No Deposit
Resort Hotel	July	HB	ESP	Online TA	TA/TO	G	G	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	E	E	No Deposit
Resort Hotel	July	BB	IRL	Online TA	TA/TO	D	E	No Deposit
Resort Hotel	July	BB	PRT	Offline TA/TO	TA/TO	E	E	No Deposit
Resort Hotel	July	BB	IRL	Online TA	TA/TO	A	E	No Deposit
Resort Hotel	July	BB	FRA	Corporate	Corporate	A	G	No Deposit
Resort Hotel	July	BB	GBR	Direct	Direct	G	G	No Deposit

only showing top 20 rows

```
df.select("reservation_status_date").show()
```

```
+-----+  
|reservation_status_date|
```

```
+-----+
|          7/1/2015|
|          7/1/2015|
|          7/2/2015|
|          7/2/2015|
|          7/3/2015|
|          7/3/2015|
|          7/3/2015|
|          7/3/2015|
|          5/6/2015|
|          4/22/2015|
|          6/23/2015|
|          7/5/2015|
|          7/5/2015|
|          7/5/2015|
|          7/5/2015|
|          7/5/2015|
|          7/2/2015|
|          7/2/2015|
|          7/5/2015|
+-----+
only showing top 20 rows
```

```
df.groupBy(f.col("reservation_status_date")).count().show()
```

```
+-----+
|reservation_status_date|count|
+-----+
|          11/14/2015|    80|
|          6/29/2016|   113|
|          7/31/2016|   125|
|          8/21/2016|   139|
|          12/8/2016|   153|
|          3/20/2017|   194|
|          2/27/2015|     1|
|          8/31/2015|   116|
|          8/9/2016|   121|
|          11/24/2016|   172|
|          5/7/2017|   146|
|          11/26/2015|   135|
|          8/28/2016|   134|
|          2/14/2017|   121|
|          4/10/2017|   164|
|          6/11/2016|    89|
|          8/31/2016|   155|
|          11/30/2016|   137|
|          7/20/2017|   104|
|          6/15/2015|    53|
+-----+
only showing top 20 rows
```

```
df = df.withColumn("reservation_status_date", f.to_date(df["reservation_status_date"], "M/d/yyyy"))
```

```
df.select("reservation_status_date").show()
```

```
+-----+
|reservation_status_date|
+-----+
|          2015-07-01|
|          2015-07-01|
|          2015-07-02|
|          2015-07-02|
|          2015-07-03|
|          2015-07-03|
|          2015-07-03|
|          2015-07-03|
|          2015-05-06|
|          2015-04-22|
|          2015-06-23|
|          2015-07-05|
|          2015-07-05|
|          2015-07-05|
|          2015-07-05|
|          2015-07-05|
|          2015-07-05|
|          2015-07-02|
|          2015-07-02|
|          2015-07-05|
+-----+
only showing top 20 rows
```

```
[i for i,v in df.dtypes if v in ['string']]
```

```
['hotel',
 'arrival_date_month',
 'meal',
 'country',
 'market_segment',
 'distribution_channel',
 'reserved_room_type',
 'assigned_room_type',
 'deposit_type',
 'customer_type',
 'reservation_status',
 'MO_YR']
```

```
df.select([i for i,v in df.dtypes if v in ['string']]).show()
```

hotel	arrival_date_month	meal	country	market_segment	distribution_channel	reserved_room_type	assigned_room_type	deposit_type
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	GBR	Direct	Direct	A	C	No Deposit
Resort Hotel	July	BB	GBR	Corporate	Corporate	A	A	No Deposit
Resort Hotel	July	BB	GBR	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	BB	GBR	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	BB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	FB	PRT	Direct	Direct	C	C	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	A	A	No Deposit
Resort Hotel	July	HB	PRT	Offline TA/TO	TA/TO	D	D	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	E	E	No Deposit
Resort Hotel	July	HB	PRT	Online TA	TA/TO	D	D	No Deposit
Resort Hotel	July	BB	USA	Online TA	TA/TO	D	E	No Deposit
Resort Hotel	July	HB	ESP	Online TA	TA/TO	G	G	No Deposit
Resort Hotel	July	BB	PRT	Online TA	TA/TO	E	E	No Deposit
Resort Hotel	July	BB	IRL	Online TA	TA/TO	D	E	No Deposit
Resort Hotel	July	BB	PRT	Offline TA/TO	TA/TO	E	E	No Deposit
Resort Hotel	July	BB	IRL	Online TA	TA/TO	A	E	No Deposit
Resort Hotel	July	BB	FRA	Corporate	Corporate	A	G	No Deposit
Resort Hotel	July	BB	GBR	Direct	Direct	G	G	No Deposit

only showing top 20 rows



```
[i for i,v in df.dtypes if v in ['int', 'double']]
```

```
['is_canceled',
 'lead_time',
 'arrival_date_year',
 'arrival_date_week_number',
 'arrival_date_day_of_month',
 'stays_in_weekend_nights',
 'stays_in_week_nights',
 'adults',
 'children',
 'babies',
 'is_repeated_guest',
 'previous_cancellations',
 'previous_bookings_not_canceled',
 'booking_changes',
 'agent',
 'days_in_waiting_list',
 'adr',
 'required_car_parking_spaces',
 'total_of_special_requests',
 'CPI_AVG',
 'INFLATION',
 'INFLATION_CHG',
 'CSMR_SENT',
 'UNRATE',
 'INTRSRT',
 'GDP',
 'FUEL_PRCS',
 'CPI_HOTELS',
 'US_GINI',
 'DIS_INC']
```

```
df.select([i for i,v in df.dtypes if v in ['int', 'double']]).show()
```

is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_n
0	342	2015	27	1	0	
0	737	2015	27	1	0	
0	7	2015	27	1	0	
0	13	2015	27	1	0	
0	14	2015	27	1	0	
0	14	2015	27	1	0	

0	0	2015	27	1	0
0	9	2015	27	1	0
1	85	2015	27	1	0
1	75	2015	27	1	0
1	23	2015	27	1	0
0	35	2015	27	1	0
0	68	2015	27	1	0
0	18	2015	27	1	0
0	37	2015	27	1	0
0	68	2015	27	1	0
0	37	2015	27	1	0
0	12	2015	27	1	0
0	0	2015	27	1	0
0	7	2015	27	1	0

only showing top 20 rows

```
[i for i,v in df.dtypes if v in ['date']]
```

```
['reservation_status_date']
```

```
df.select([i for i,v in df.dtypes if v in ['date']]).show()
```

```
+-----+
|reservation_status_date|
+-----+
|      2015-07-01|
|      2015-07-01|
|      2015-07-02|
|      2015-07-02|
|      2015-07-03|
|      2015-07-03|
|      2015-07-03|
|      2015-07-03|
|      2015-05-06|
|      2015-04-22|
|      2015-06-23|
|      2015-07-05|
|      2015-07-05|
|      2015-07-05|
|      2015-07-05|
|      2015-07-05|
|      2015-07-05|
|      2015-07-02|
|      2015-07-02|
|      2015-07-05|
+-----+
```

only showing top 20 rows

```
df.printSchema()
```

```
root
|-- hotel: string (nullable = true)
|-- is_canceled: integer (nullable = true)
|-- lead_time: integer (nullable = true)
|-- arrival_date_year: integer (nullable = true)
|-- arrival_date_month: string (nullable = true)
|-- arrival_date_week_number: integer (nullable = true)
|-- arrival_date_day_of_month: integer (nullable = true)
|-- stays_in_weekend_nights: integer (nullable = true)
|-- stays_in_week_nights: integer (nullable = true)
|-- adults: integer (nullable = true)
|-- children: double (nullable = true)
|-- babies: integer (nullable = true)
|-- meal: string (nullable = true)
|-- country: string (nullable = true)
|-- market_segment: string (nullable = true)
|-- distribution_channel: string (nullable = true)
|-- is_repeated_guest: integer (nullable = true)
|-- previous_cancellations: integer (nullable = true)
|-- previous_bookings_not_canceled: integer (nullable = true)
|-- reserved_room_type: string (nullable = true)
|-- assigned_room_type: string (nullable = true)
|-- booking_changes: integer (nullable = true)
|-- deposit_type: string (nullable = true)
|-- agent: double (nullable = true)
|-- days_in_waiting_list: integer (nullable = true)
|-- customer_type: string (nullable = true)
|-- adr: double (nullable = true)
|-- required_car_parking_spaces: integer (nullable = true)
|-- total_of_special_requests: integer (nullable = true)
|-- reservation_status: string (nullable = true)
|-- reservation_status_date: date (nullable = true)
|-- MO_YR: string (nullable = true)
```



```
|-- CPI_AVG: double (nullable = true)
|-- INFLATION: double (nullable = true)
|-- INFLATION_CHG: double (nullable = true)
|-- CSMR_SENT: double (nullable = true)
|-- UNRATE: double (nullable = true)
|-- INTRSRRT: double (nullable = true)
|-- GDP: double (nullable = true)
|-- FUEL_PRCs: double (nullable = true)
|-- CPI_HOTELS: double (nullable = true)
|-- US_GINI: double (nullable = true)
|-- DIS_INC: double (nullable = true)
```

df.show()

hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_i
Resort Hotel	0	342	2015	July	27		1
Resort Hotel	0	737	2015	July	27		1
Resort Hotel	0	7	2015	July	27		1
Resort Hotel	0	13	2015	July	27		1
Resort Hotel	0	14	2015	July	27		1
Resort Hotel	0	14	2015	July	27		1
Resort Hotel	0	0	2015	July	27		1
Resort Hotel	0	9	2015	July	27		1
Resort Hotel	1	85	2015	July	27		1
Resort Hotel	1	75	2015	July	27		1
Resort Hotel	1	23	2015	July	27		1
Resort Hotel	0	35	2015	July	27		1
Resort Hotel	0	68	2015	July	27		1
Resort Hotel	0	18	2015	July	27		1
Resort Hotel	0	37	2015	July	27		1
Resort Hotel	0	68	2015	July	27		1
Resort Hotel	0	37	2015	July	27		1
Resort Hotel	0	12	2015	July	27		1
Resort Hotel	0	0	2015	July	27		1
Resort Hotel	0	7	2015	July	27		1

only showing top 20 rows

