



MONASH
BUSINESS
SCHOOL

Expert advice from experts

Marie Curie

Nobel Prize, PhD

Pierre Curie

Nobel Prize, PhD

Report for
Acme Corporation

16 October 2020

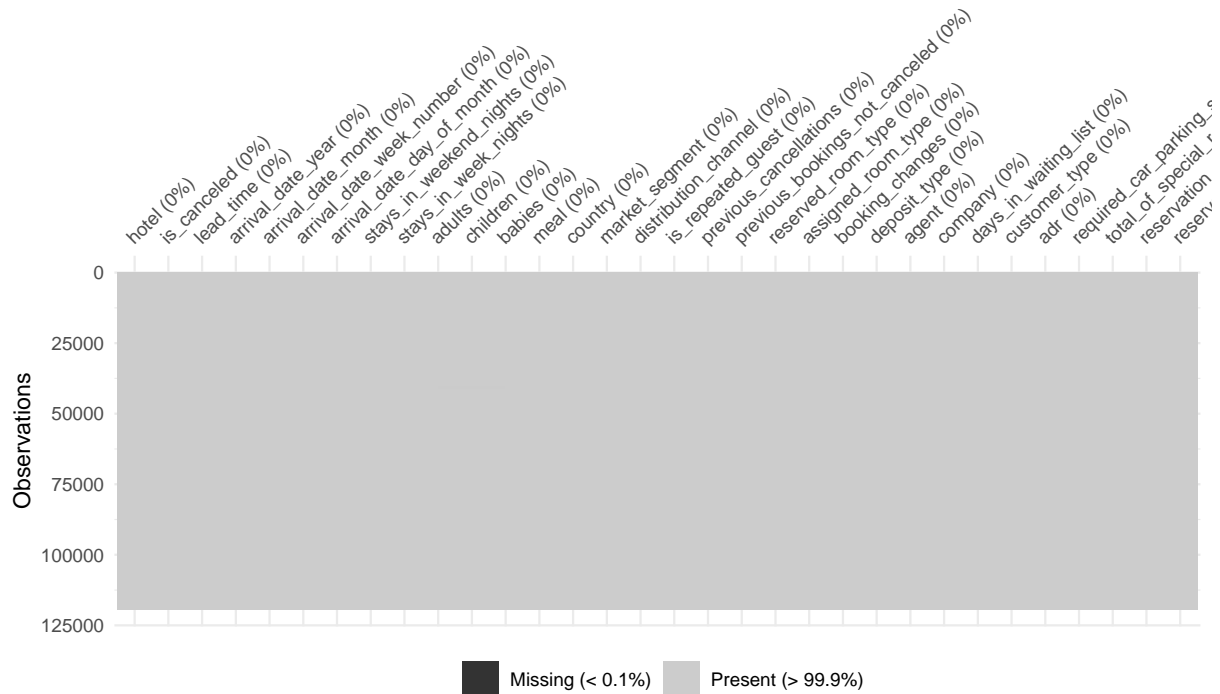
**Department of
Econometrics &
Business Statistics**

☎ (03) 9905 2478
✉ BusEco-Econometrics@monash.edu

ABN: 12 377 614 012

```
Miss_tidy <- vis_miss(hotels, warn_large_data = FALSE) # Check missing value
```

```
Miss_tidy
```



```
hotel <- na.omit(hotels)
```

```
hotel <- hotel %>% mutate(
  arr_date_month = case_when( # change Month to the numeric
    arrival_date_month=="January" ~ 01,
    arrival_date_month=="February" ~ 02,
    arrival_date_month=="March" ~ 03,
    arrival_date_month=="April" ~ 04,
    arrival_date_month=="May" ~ 05,
    arrival_date_month=="June" ~ 06,
    arrival_date_month=="July" ~ 07,
    arrival_date_month=="August" ~ 08,
    arrival_date_month=="September" ~ 09,
    arrival_date_month=="October" ~ 10,
    arrival_date_month=="November" ~ 11,
    arrival_date_month=="December" ~ 12),
```

```

country = str_replace(country, pattern = "cn", replacement = "chn"), # replace China pattern
deposit = case_when(deposit_type == "No Deposit"~0, # Change deposit as binary r
                    deposit_type == "Non Refund"~1,
                    deposit_type == "Refundable"~2),
arrival_time=paste(arrival_date_year,
                  arrival_date_month,
                  arrival_date_day_of_month,
                  sep="-")
sec_mute <- hotel %>% mutate(Country = countrycode(sourcevar = country, # Accord
                  origin = "iso3c",
                  destination = "country.name"),
                  reservation_status_date=ymd(reservation_status_date),
                  arrival_time=ymd(arrival_time))

```

```
Miss <- sapply(hotels, function(x) sum(is.na(x))) # Check missing value
```

```
Miss
```

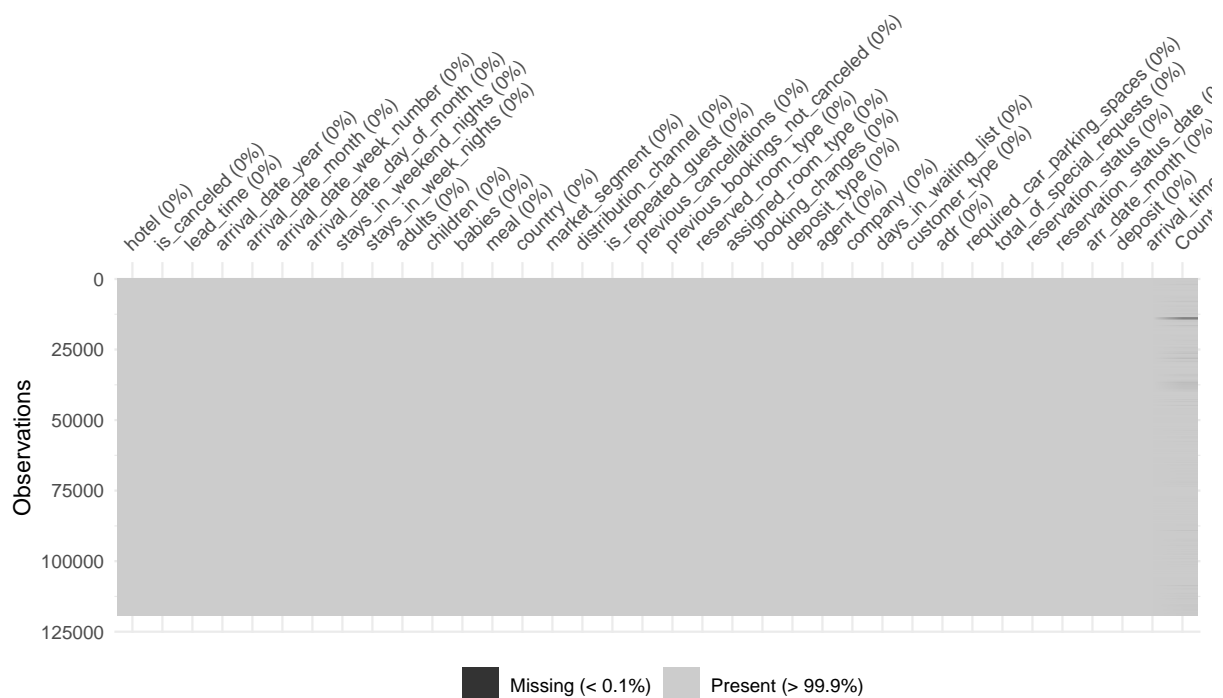
```

##          hotel          is_canceled
##          0          0
##      lead_time      arrival_date_year
##          0          0
##      arrival_date_month      arrival_date_week_number
##          0          0
##      arrival_date_day_of_month      stays_in_weekend_nights
##          0          0
##      stays_in_week_nights          adults
##          0          0
##          children          babies
##          4          0
##          meal          country
##          0          0
##      market_segment      distribution_channel
##          0          0
##      is_repeated_guest      previous_cancellations

```

```
##                                0                                0
## previous_bookings_not_canceled reserved_room_type
##                                0                                0
## assigned_room_type booking_changes
##                                0                                0
## deposit_type agent
##                                0                                0
## company days_in_waiting_list
##                                0                                0
## customer_type adr
##                                0                                0
## required_car_parking_spaces total_of_special_requests
##                                0                                0
## reservation_status reservation_status_date
##                                0                                0
```

```
vis_tidy <- vis_miss(sec_mute, warn_large_data = FALSE)
vis_tidy
```



```
sec_mute <- sec_mute %>% dplyr::select(-market_segment,  
                                     -distribution_channel,  
                                     -deposit_type,  
                                     -agent,  
                                     -company)
```

```
final <- na.omit(sec_mute) # Delete the missing country observations
```