Hotel Data Analysis with Group by and Probabilities

Reshu Gupta

2024-01-30

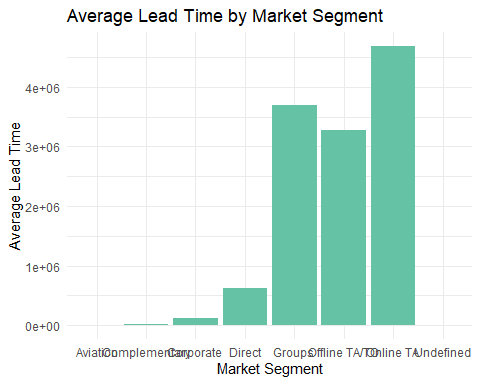
1. Initial setup and Configure the data set.
2. Install and import library dplyr and ggplot2.
3. Load the data set file in variable h\_data (read.csv()) which will be used in through out the program.
4. Convert reservation\_status\_date in data set as Date.

Group 1 : Based on lead\_time 1. Convert reservation\_status\_date of data set to date mm/dd/YYYY. 2. Groups the data by ‘market\_segment’ using the group\_by function. 3. Calculated the mean value of lead time for each market segment which is group by summarize function. 4. Computes the count of observations in each group using the n function. 5. Filters the grouped data to identify the group with the minimum count (n). 4. Assign a special tag to the smallest group Lowest Probability Group. 5. Merge back to the original data frame.

#Convert reservation\_status\_date to date mm/dd/YYYY according to R.  
h\_data$reservation\_status\_date <- as.Date(h\_data$reservation\_status\_date, format="%m/%d/%Y")  
grouped\_market\_segment <- h\_data %>%  
group\_by(market\_segment) %>%  
summarize(average\_lead\_time = mean(lead\_time),  
 n = n()) %>%  
 ungroup()  
# Assign a special tag to the smallest group  
smallest\_group\_market\_segment <- grouped\_market\_segment %>%  
 group\_by\_all() %>%  
 filter(n == min(n)) %>%  
 ungroup() %>%  
 mutate(special\_tag = "Lowest Probability Group")  
  
# Merge back to the original data frame  
hotel\_data\_with\_tag\_market\_segment <- h\_data %>%  
 left\_join(smallest\_group\_market\_segment, by = "market\_segment")

## Including Plot #1

1. Creates a bar plot using ggplot to visualize the average lead time for each market segment.
2. The x-axis represents the ‘Market Segment’, and the y-axis represents the ‘Average Lead Time’.
3. Bars are colored with the fill Green.
4. The plot has a minimal theme.
5. Prints the generated plot.



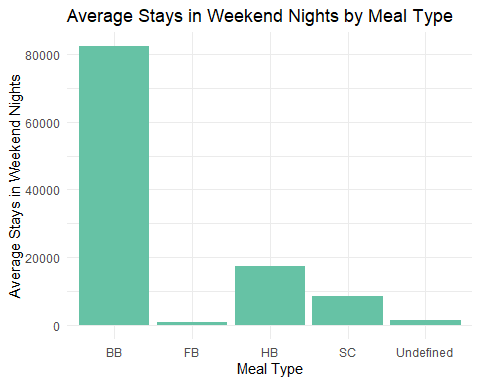
Insight : This analysis aims to identify the market segment with the lowest probability, based on the average lead time. The bar plot visually represents the average lead time for each market segment, and the special tag is used to highlight the group with the lowest probability.

#Grouping 2: Based on meal column 1. Groups the data by the meal column using the group\_by function. 2. Calculates the mean of the stays\_in\_weekend\_nights variable within each meal group using the summarize function. 3. Computes the count of observations in each group using the n function. 4. Filters the grouped data to identify the group with the minimum count (n). 5. Assign a special tag to the smallest group- Lowest Probability Group. 6. Merges the data frame with the special tags back to the original data frame using the meal column.

grouped\_meal <- h\_data %>%  
 group\_by(meal) %>%  
 summarize(average\_stays\_in\_weekend\_nights = mean(stays\_in\_weekend\_nights),  
 n = n()) %>%  
 ungroup()  
# Assign a special tag to the smallest group  
smallest\_group\_meal <- grouped\_meal %>%  
 group\_by\_all() %>%  
 filter(n == min(n)) %>%  
 ungroup() %>%  
 mutate(special\_tag = "Lowest Probability Group")  
  
# Merge back to the original data frame  
hotel\_data\_with\_tag\_meal <- h\_data %>%  
 left\_join(smallest\_group\_meal, by = "meal")

## Including Plot #2

1. Creates a bar plot using ggplot to visualize the average stays in weekend nights for each meal type.
2. The x-axis represents the ‘Meal Type’, and the y-axis represents the ‘Average Stays in Weekend Nights’.
3. Bars are colored with the fill Green Color.
4. The plot has a minimal theme.

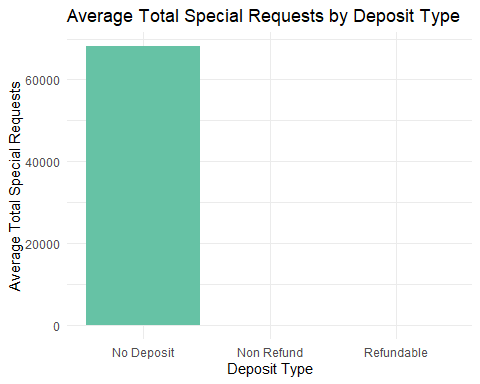
 Insight : The analysis aims to identify patterns or differences in the average stays in weekend nights based on different meal types. The visualization provides a clear comparison of the average stays in weekend nights across different meal types, allowing for insights into potential relationships or variations associated with meal choices.

#Grouping 3: Group by deposit\_type and summarize total\_of\_special\_requests. 1. Groups the data by the deposit\_type column using the group\_by function. 2. Calculates the mean of the total\_of\_special\_requests variable within each deposit\_type group using the summarize function. 3.Computes the count of observations in each group using the n function. 4. Filters the grouped data to identify the group with the minimum count (n). 5. Assign a special tag to the smallest group- Lowest Probability Group. 6. Merges the data frame with the special tags back to the original data frame using the ‘deposit\_type’ column.

grouped\_deposit\_type <- h\_data %>%  
 group\_by(deposit\_type) %>%  
 summarize(average\_total\_of\_special\_requests = mean(total\_of\_special\_requests),  
 n = n()) %>%  
 ungroup()  
  
#Number\_total\_of\_special\_requests <-mean(total\_of\_special\_requests)  
  
# Assign a special tag to the smallest group  
smallest\_group\_deposit\_type <- grouped\_deposit\_type %>%  
 group\_by\_all() %>%  
 filter(n == min(n)) %>%  
 ungroup() %>%  
 mutate(special\_tag = "Lowest Probability Group")  
  
# Merge back to the original data frame  
hotel\_data\_with\_tag\_deposit\_type <- h\_data %>%  
 left\_join(smallest\_group\_deposit\_type, by = "deposit\_type")

## Including Plot #3

1.Creates a bar plot using ggplot to visualize the average total special requests for each deposit type. 2.The x-axis represents the ‘Deposit Type’, and the y-axis represents the ‘Average Total Special Requests’. 3.Bars are colored with the fill “#66c2a5”. 4.The plot has a minimal theme.



Insight: The analysis aims to explore the relationship between the deposit type and the average total special requests made by guests.The visualization provides a comparison of the average total special requests across different deposit types, allowing for insights into potential patterns or differences associated with deposit preferences. This analysis can help understand if there are specific deposit types that are associated with a higher or lower average total of special requests, providing valuable insights for hotel management or decision-makers.

Thank You