Documentation

1. Project Purpose
To explore, analyze, and visualize hotel booking demand, determining patterns and correlations within the data that may provide crucial insights for hotel managements.

2. Data Description + Collection
Hotel Booking Demand
Source: https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand

Amount of Data

- The CSV dataset contains 119,390 rows and 29 columns.

Value Types

- Categorical:
    - hotel type, is_canceled, meal, country, market_segment, distribution_channel, is_repeated_guest, deposit_type, agent, company, customer_type, reservation_status

- Numeric:
    - 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, previous_cancellations, previous_bookings_not_canceled, reserved_room_type, assigned_room_type, booking_changes, days_in_waiting_list, adr, required_car_parking_spaces, total_of_special_requests, reservation_status_date

Variable	Description
hotel	The type of hotel
is_canceled	Whether or not the booking was canceled
lead_time	The number of days between the booking date and the arrival date
arrival_date_year	The year of the arrival date
arrival_date_month	The month of the arrival date
arrival_date_week_number	The week number of the arrival date
arrival_date_day_of_month	The day of the month of the arrival date
stays_in_weekend_nights	The number of nights the guest stayed on a weekend
stays_in_week_nights	The number of nights the guest stayed on a weekday
adults	The number of adults in the booking

children	The number of children in the booking
babies	The number of babies in the booking
meal	The type of meal plan (breakfast included, breakfast and dinner included, etc.)
country	The country of residence of the guest
market_segment	The market segment of the guest
distribution_channel	The channel through which the booking was made
is_repeated_guest	Whether or not the guest is a repeat customer
previous_cancellations	The number of previous bookings the guest has canceled
previous_bookings_not_canceled	The number of previous bookings the guest has not canceled
reserved_room_type	The type of room the guest reserved
assigned_room_type	The type of room the guest was assigned
booking_changes	The number of times the guest changed their booking
deposit_type	The type of deposit the guest made
agent	The travel agent who booked the reservation (if applicable)
company	The company the guest works for (if applicable)
days_in_waiting_list	The number of days the guest was on a waiting list for the reservation
customer_type	The type of customer
adr	The average daily rate of the reservation
required_car_parking_spaces	The number of car parking spaces the guest requested
total_of_special_requests	The number of special requests the guest made
reservation_status	The current status of the reservation
reservation_status_date	The date the reservation status was last updated

Additional information about collection methodology should be referenced directly from Kaggle.

3. Intended Users

- This Shiny App is designed for hotel managers, business analysts, and data scientists who are looking to understand booking patterns.

4. Questions to Answer

- How do different customer segments behave regarding booking and cancellations?
- What are the patterns relating to lead time and cancellations?
- What are the busiest times of the year, and how do booking patterns change seasonally?

5. Data Insights

- **Customer Segmentation:** Business travelers show a higher rate of last-minute cancellations compared to other segments.
- **Seasonality:** Summer months display the highest booking rates, while fall shows a spike in cancellations.
- **Lead Time:** Longer lead times are correlated with a higher likelihood of cancellations.

6. Improvement Wishlist

- **Improve Data:** The data from kaggle does not include after 2018, thus it would be preferable to have most-up-to-date data for further research.
- **Enhance Visualization:** Incorporate predictive analytics visuals in the Shiny app.
- **Real-time Data Integration:** In order to take into consideration the newest data in our analysis, it would have been great if we could have Implemented functionality for real-time data updating and analysis.

7. Sources/References

- **Hotel Booking Demand**
    - https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand
    -

8. Proper Documentation

- **Tidying Data:** Missing data were removed and outliers were identified and handled using in the data cleaning part. New variables were created within the process_hotel_data function.
- **Visual Encoding:** Used color intensity to represent booking frequencies and utilized tooltips to display detailed information.
- **Interactive Elements:** Added sliders for time-based filtering and dropdowns for categorical filtering in the Shiny App.

9. Description of Design Decisions

- **What:** Analyzing and visualizing data about hotel booking and cancellations.
- **Why:** To identify patterns, correlations, and trends that could inform strategic decision-making in hotel management.
- **How:** Data is visualized using various plots, and users can interact with the data using various interactive features such as drop down or scroll bar.

Conclusion

- **Key Insights**
    - **Demand Patterns:** The analysis showed significant variability in booking demand across different seasons, customer segments, and lead times.
    - **Cancellation Insights:** A correlation was identified between lead time and cancellations, highlighting a potential area for strategic focus.

- **Recommendations**

- **Cancellation Prevention:** Implement strategies to minimize cancellations, like targeted communication or flexible booking options.

- **Targeted Marketing:** identify demand patterns and customer behaviors to develop targeted marketing strategies aimed at different customer segments and booking periods.