The Problem area:
My area of interest falls in the hotel industry. My project could help hotels maximize profit by accurately predicting daily hotel demand and cancellations. By optimizing and accurately predicting the net rooms for a given day, a hotel more optimally maximizes revenue for projected sellout days as well as efficiently staff their hotels to improve daily operations such as housekeeping.

The User:
If a revenue management system is able to accurately predict the room demand and room cancellations for a given day, it will help a hotel have a more optimal room strategy and help operations teams properly staff their hotels on a given day or week

The Big Idea:
ML can create predictive analytics to create demand forecast patterns based on historical data, seasonality, and external factors like events. Hotels can then optimize pricing strategies and also better staff their operations teams based on the predictions. In my specific dataset, I want to be able to predict demand and cancellations for a given day.

- Data Preprocessing
    - In order to be able to predict and forecast demand and cancellations, I would first need to conduct data preprocessing and clean the data that is available.
    - I would identify and remove/impute/replace null values in my data, remove duplicate data, and correct any data errors
    - There are also several elements I would transform and create new features through the process of feature engineering for potential use during my Exploratory Data Analysis step:
        - Hotel's total capacity
        - Length of stay for each reservation
        - Rooms OTB (Rooms already sold and "on the books")
        - Remaining Supply (Rooms left to sell at the time of reservation)
        - Pickup (Number of rooms sold in the last 1, 5, 7, 15 days)
        - Booking Pace (Current year's variance compared to the same time last year for Rooms OTB, Remaining Supply, Pickup)
    - A few ideas of how I would calculate these features:
        - I would calculate the hotel's capacity by doing a value_count of the arrivalDateDayofMonth to see if there was a repeated high count for each hotel.
        - I would calculate the length of stay by combining the total StaysInWeekendNights + StaysInWeekNights
- EDA
    - In the next step of EDA, I would identify potential features that may assist in predicting demand and cancellations, such as the following:
        - Arrival date, Length of stay, Guest Type, Weekday/Weekend, lead time of bookings, market segmentation, distribution channel, reserved room type, customer type, cancelled status, and hotel capacity.

    - I would use various visualizations to identify patterns and relationships between these features and my target variable of demand and cancellations.
- Model Selection and Training
    - I would first start with linear regression models for any variables I've identified to have a linear relationship to either cancellation or demand.
    - If the linear regression models prove inconclusive or unsuccessful, then I would move onto testing more advanced models like decision trees or random forests
- Model Evaluation
    - With any of the selected models, I would access the models performance by evaluating the R-Squared value for accuracy

The Impact: Revenue management has become a growing need in the hotel industry. Gone are the days of simple spreadsheets to track what the "rate of the day" is. Hotel owners and general managers have progressively found the ROI to be high when investing in proper revenue management strategies deployed at their hotels. The goal of revenue management is "selling the right product to the right customer at the right time for the right price". A revenue managers metrics of success include the following (depending on the hotel's strategy):

- Maximizing the Revenue Per Available Room (RevPAR).
- Reaching a "perfect sell". Meaning not having any rooms go unsold.
    - Limited perishable inventory (Any rooms that are left unsold receive 0 revenue. Unlike in retail, inventory cannot be carried over the next day)
    - The two biggest contributors to reaching a perfect sale are room demand and room cancellations.

The Data:
I have identified two datasets that share the same source. The Science Direct dataset was shared in 2019 and the kaggle dataset took this data and cleaned it for "#TidyTuesday" in 2020. The data contains booking information for 2 hotels, a city hotel and a resort hotel. It provides reservation information such as: booking dates, arrival dates, number of adults/children, market segment, and if the reservation was cancelled or changed. Please see the attached excel file for all features available.

- Kaggle Data set: https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand/data
- ScienceDirect data set: https://www.sciencedirect.com/science/article/pii/S2352340918315191#ab0010

The Alternative:
Pivoting slightly but still in the hotel industry, I would be interested in developing a way to analyze hotel reviews to identify areas of opportunity to help increase guest satisfaction, guest service scores, and improve hotel operations. Some of the features of the dataset required would be the written review, user review rating, date of the review, review title, property, property address, and user name. Ideally this solution would be provided for an individual hotel to use for evaluation and prioritization of improvements they could make for themselves. However, if this

could be scaled, an enterprise or franchise company could run this across a larger portfolio/region and identify top offenders and quickly address operational issues or practices to prevent future guest dissatisfaction.

A sample dataset can be found here: https://www.kaggle.com/datasets/datafiniti/hotel-reviews