**Answers for Pacaso Data Scientist Challenge**

**Scheduling Analysis**

*(2) Do later co-owners within the same home have different usage patterns than earlier*

*ones?*

**Yes.** According to the definitions I created, which are the following:

1. **Earlier Co-Owners**: These are the co-owners who booked their ownership earlier compared to others within the same property. They are identified as having a rank equal to or less the median rank of all co-owners for that property.
2. **Later Co-Owners**: These are the co-owners who booked their ownership later compared to others within the same property. They are identified as having a rank greater than the median rank of all co-owners for that property.

With that said, there was a difference between usage patterns between earlier and later co-owners. These are the findings:

1. **Total Stays**: Later co-owners have **more stays** on average (165.66) compared to earlier co-owners (138.17).
2. **Average Stay Duration**: Later co-owners have **shorter stays** on average (6.44 days) compared to earlier co-owners (11.03 days).
3. **Total Days**: Later co-owners spend **more total days** (1043.34) in the home compared to earlier co-owners (917.34).

Basically, later co-owners tend to visit more frequently but for shorter durations, leading to a higher total number of days spent at home. I also created three statistical tests and visualized the data to test my assumptions more rigorously and even investigated this at a seasonal level:

1. The statistical tests reveal significant differences in the usage patterns between earlier and later co-owners. The Chi-Square Test for seasonality shows a statistically significant difference in monthly stay preferences (Chi2 = 68.99, P < 0.0001), with earlier co-owners favoring stays earlier in the year (e.g., January to March) and later co-owners preferring December.
2. The T-Test for stay durations (T = 4.25, P < 0.0001) indicates that earlier co-owners tend to have longer average stays compared to later co-owners.
3. The Mann-Whitney U Test (U = 69985842.50, P < 0.0001) further confirms the significant differences in stay duration distributions, offering robust validation against non-normal data.
4. A heatmap visualization supports these findings, highlighting the clear distinctions in seasonal preferences and usage patterns between the two groups.

*(5) Typically, the end of a Pacaso stay requires a one-day block for cleaning. For example, if*

*a stay has a Sunday departure, the next possible arrival date would be the following*

*Monday. If Pacaso moved to same-day turnaround, how much additional calendar*

*opportunity would this create for co-owners?*

Broadly speaking, I found that there would be 11849 days (7870 days if it is just 1 day gap) for the same day or more turnaround available. The logic is predicated on departure\_date and arrival\_date, and the cleaning block was 1 day or more. I did include a top 10 properties graph, a regional graph, and a graph with top 10 properties with additional calendar day also look at different visual perspectives.

Below are the findings:

1. The Pacific Southwest has significantly more additional days compared to other regions, indicating a higher potential for optimization in this area.
2. Properties with IDs like 192, 193, and 194 have the most additional days, indicating they experience frequent gaps in scheduling that could be filled. This indicates that the distribution can be found at specific property levels.

**Scheduling Optimization**

*How would you determine whether the time in each Pacaso is distributed equitably*

*across the co-owners? Consider especially that sometimes of year or holidays like*

*Thanksgiving and 4th of July may have higher demand than others.*

Overall, the code evaluates the equity of time distribution among co-owners of shared properties by first calculating the duration of each stay and normalizing it by the number of shares each owner holds. It identifies whether stays occur during high-demand periods, such as holidays, by mapping stay dates to predefined holiday ranges. The normalized utilization ratios for each owner are then aggregated by ownership ID and holiday category.

To measure equity, the code computes the Gini coefficient for utilization ratios, quantifying inequality in how time is allocated among co-owners. This analysis provides insights into whether co-owners are using the shared property equitably across all time periods, with a particular focus on holidays where demand may be higher.

Below are the findings:

1. 5 Holidays were created: 4th of July, Christmas, New Years Day, Spring Break and Thanksgiving. A graph was created to display the Gini Coefficients.
2. There was an interest in 4th of July and specific co-0wners who utilized the property more than others during high demand periods (for instance, ownership ID 1086 and 1289)
3. There was a very simple suggestion of a policy limit. Albeit, this is not being advocated as a policy, more than to display that there are specific locations, customers, and times where traffic could be regulated for fairness.