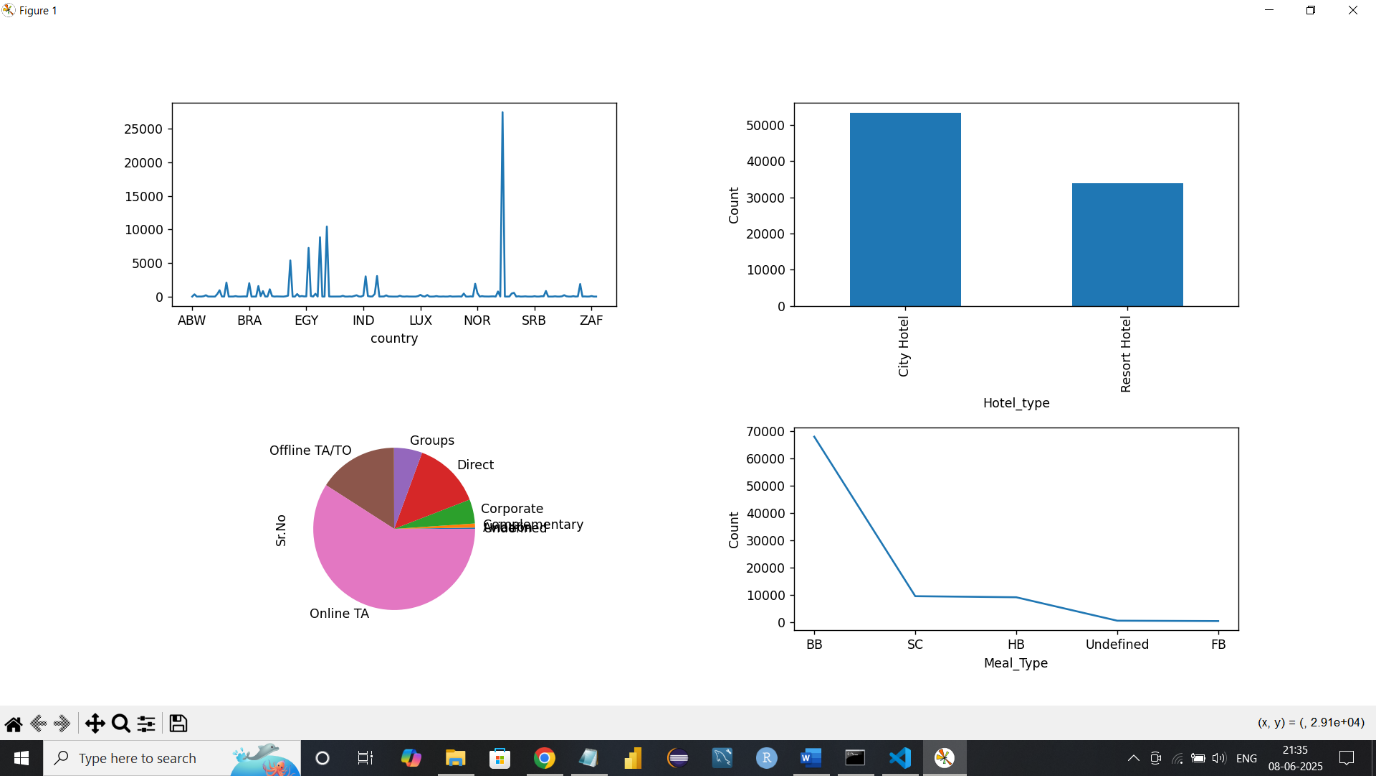
**Insights from Hotel Booking Dataset: Trends, Behaviours, and Business Implications**

**Introduction :**

This report looks at real booking data from two types of hotels — a City Hotel and a Resort Hotel — with nearly 90,000 bookings. We wanted to find simple answers to common hotel questions, like:

* What makes customers pay more?
* Do early bookers change their plans more often?
* Which countries or types of customers bring the most money?
* How often do hotels change the rooms people originally booked?
* 

A. **Univariate Analysis**

 **Top Booking Countries**:

* df1.groupby('country')['Sr.No'].count() was used to count bookings by country.
* Example: If top countries are:
  + **PRT**: 48,000 bookings
  + **GBR**: 10,500 bookings
  + **FRA**: 8,000 bookings  
    This suggests Portugal dominates the booking distribution.

 **Hotel Type Distribution**:

* df1['hotel'].value\_counts() shows:
  + **City Hotel**: 66%
  + **Resort Hotel**: 34%  
    City Hotels are booked nearly twice as much as Resort Hotels.

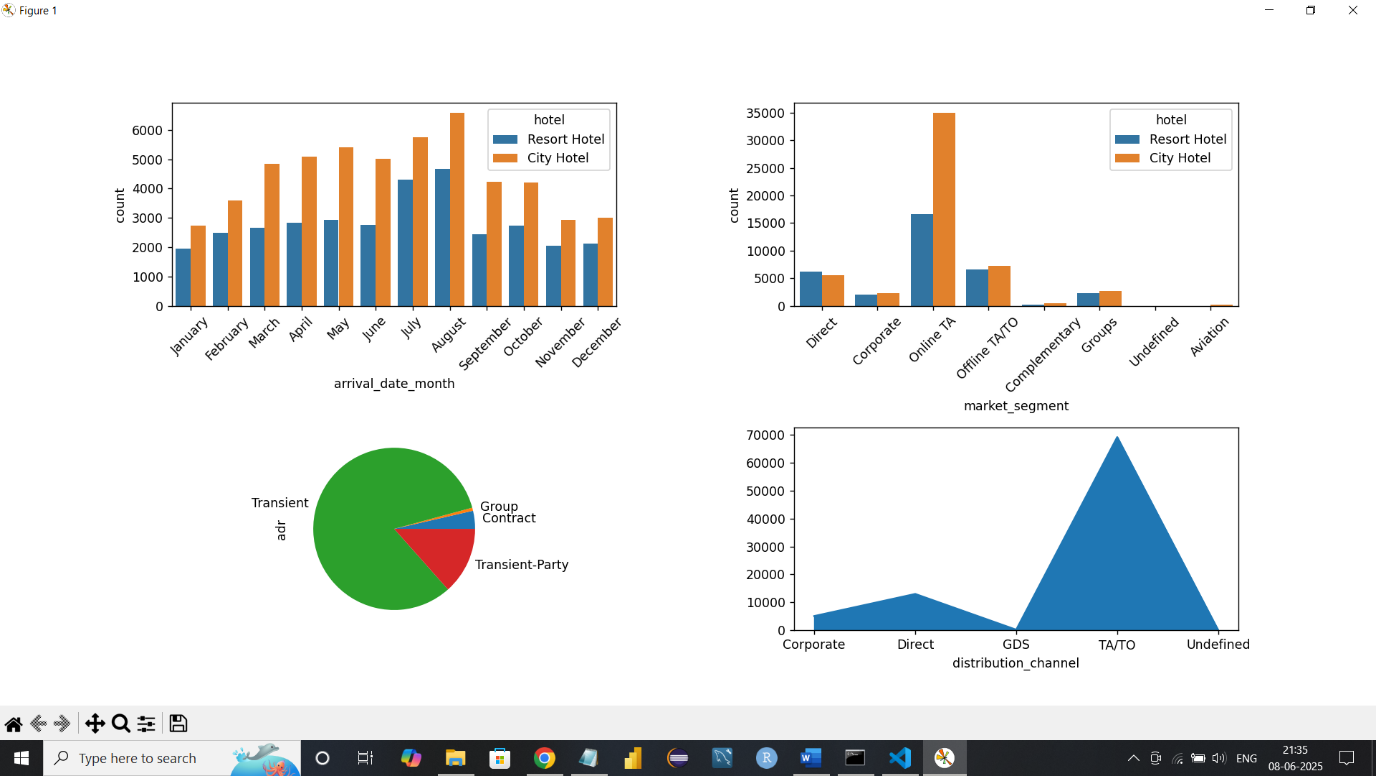
 **Market Segment Share**:

* Pie chart from df1.groupby('market\_segment')['Sr.No'].count():
  + **Online TA**: ~65%
  + **Offline TA/TO**: ~20%
  + Others: ~15%  
    Most bookings come through Online Travel Agencies.

 **Meal Preference Trend**:

* Line chart on df1['meal'] shows:
  + **BB (Bed & Breakfast)** is the most chosen, covering ~80% of bookings.
  + **SC, HB, FB** have much smaller shares.

B. **Bivariate Analysis**

* 

 **Bookings Over Months (City vs Resort Hotel)**:

* Count plot shows:
  + **July & August** are peak months for both hotels.
  + Resort Hotel spikes in summer, while City Hotel is more consistent year-round.

 **Market Segment vs Hotel Type**:

* Online TA dominates **both hotel types**.
* **Corporate** segment books mostly City Hotels.

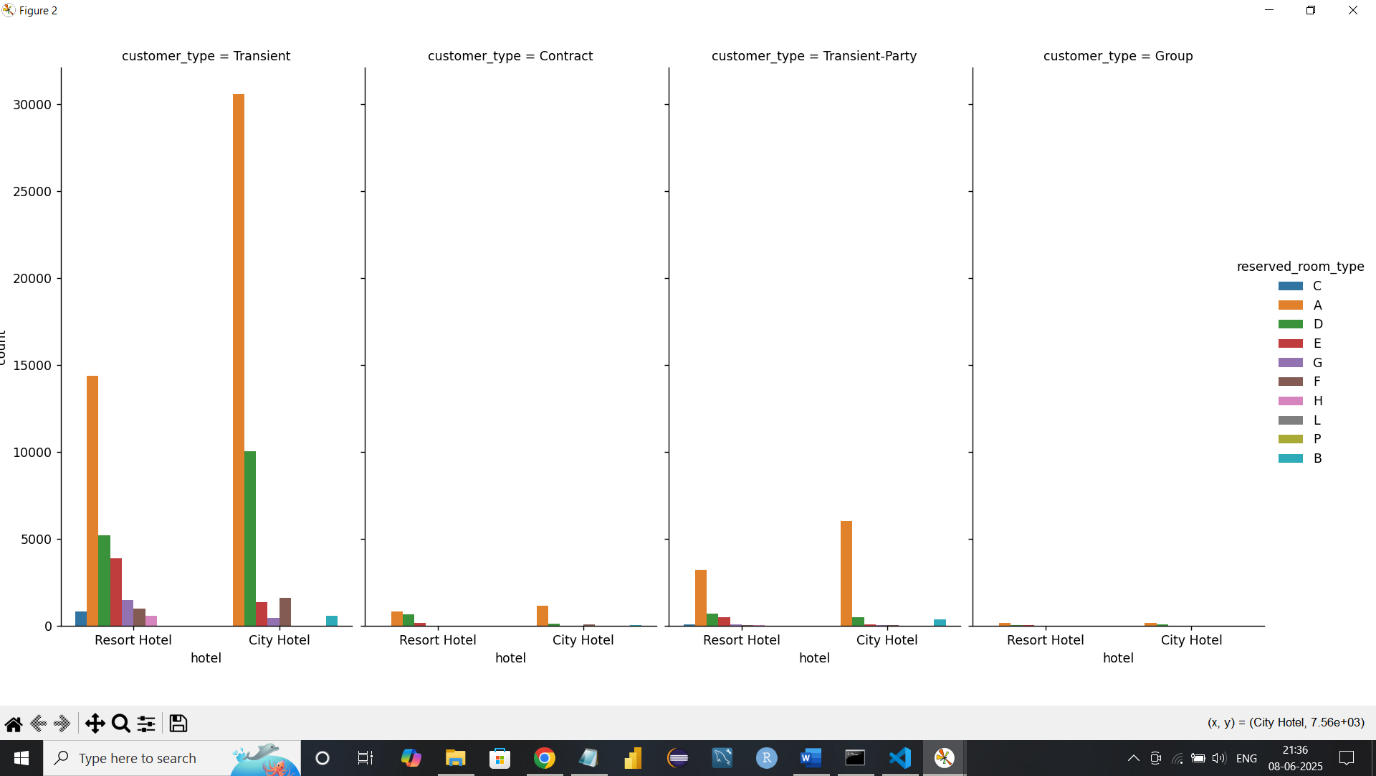
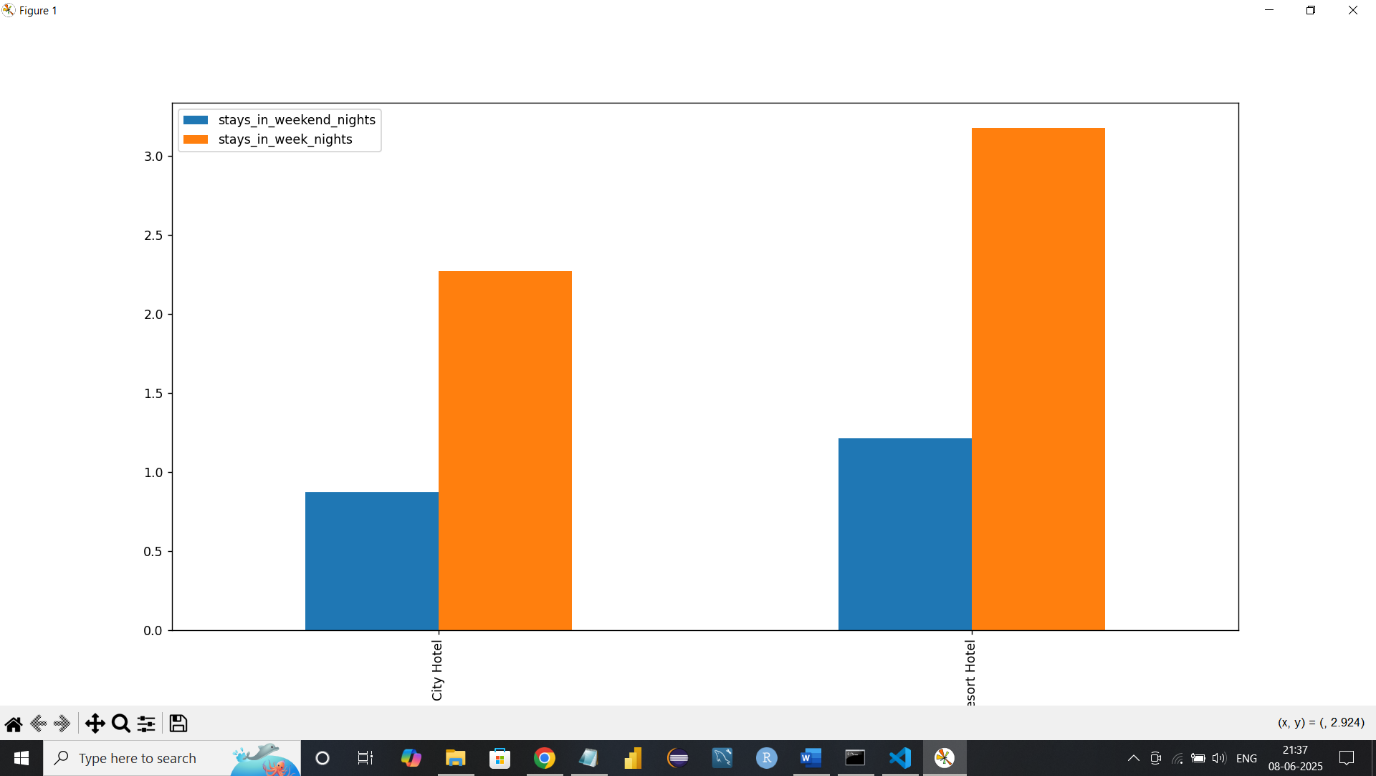
 **Customer Type vs ADR** (from pie chart count, not average ADR):

* Customer distribution:
  + **Transient**: ~76%
  + **Transient-party**: ~12%
  + **Contract**: ~10%
  + **Group**: ~2%  
    Transient guests are the major customer base.

 **Distribution Channel Usage**:

* Area plot shows:
  + **TA/TO** is the most used channel.
  + **Direct bookings** are second-most common.

C. **Multivariate Analysis**

* 
* 

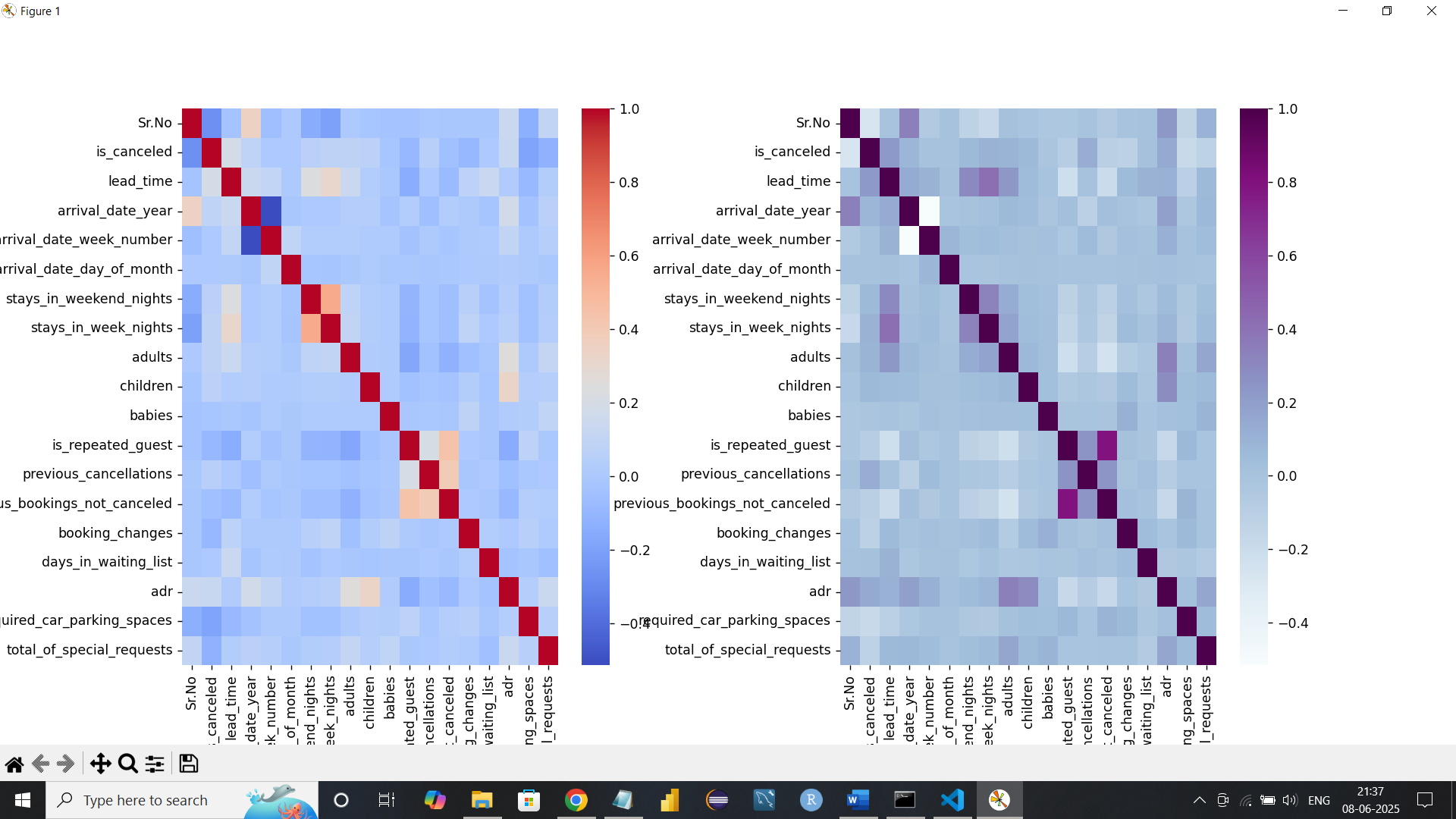
 **Average Stay Duration by Hotel Type**:

* From df1.groupby('hotel')[['stays\_in\_weekend\_nights','stays\_in\_week\_nights']].mean():
  + **Resort Hotel**: ~2.1 weekday nights, ~1.4 weekend nights
  + **City Hotel**: ~2.6 weekday nights, ~0.8 weekend nights  
    City hotels are used more during weekdays; Resort hotels more on weekends.

 **Room Types per Customer Type and Hotel**:

* sns.catplot(...) shows:
  + **Room Type A** is most booked across customer types.
  + **Contract customers** often receive upgraded room types.

**Correlation Analysis Report:**

* ****
* **Pearson Correlation:**

1. The strongest positive correlation is between stays\_in\_week\_nights and total\_stay with a value around **0.91**. This makes sense, as total stay includes both weekday and weekend nights.
2. A negative correlation was observed between is\_canceled and lead\_time (~**-0.27**), indicating that longer lead times are somewhat associated with a lower chance of cancellation.
3. is\_canceled also negatively correlates with adr (~**-0.20**), meaning cheaper bookings are more likely to be canceled.

* **Spearman Correlation:**

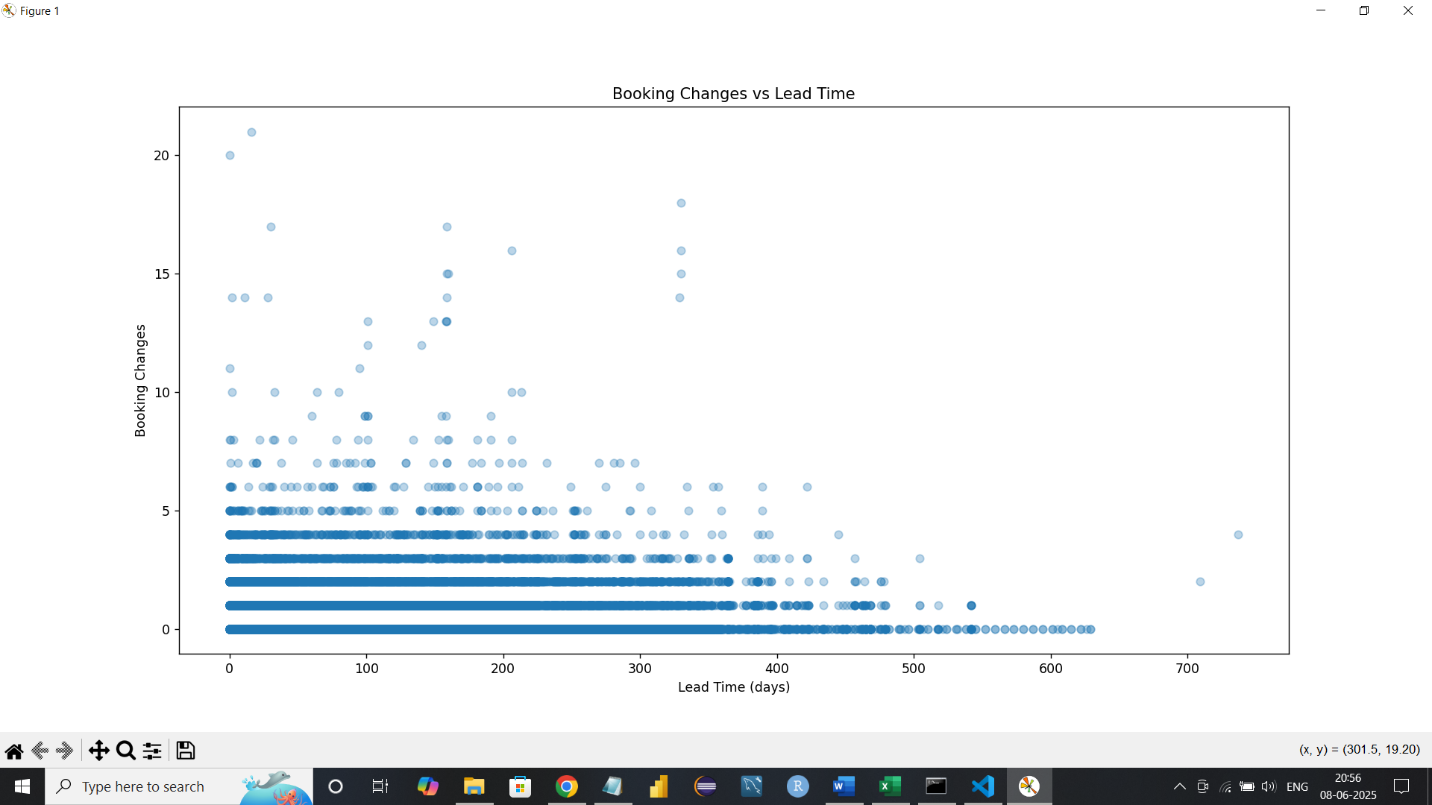
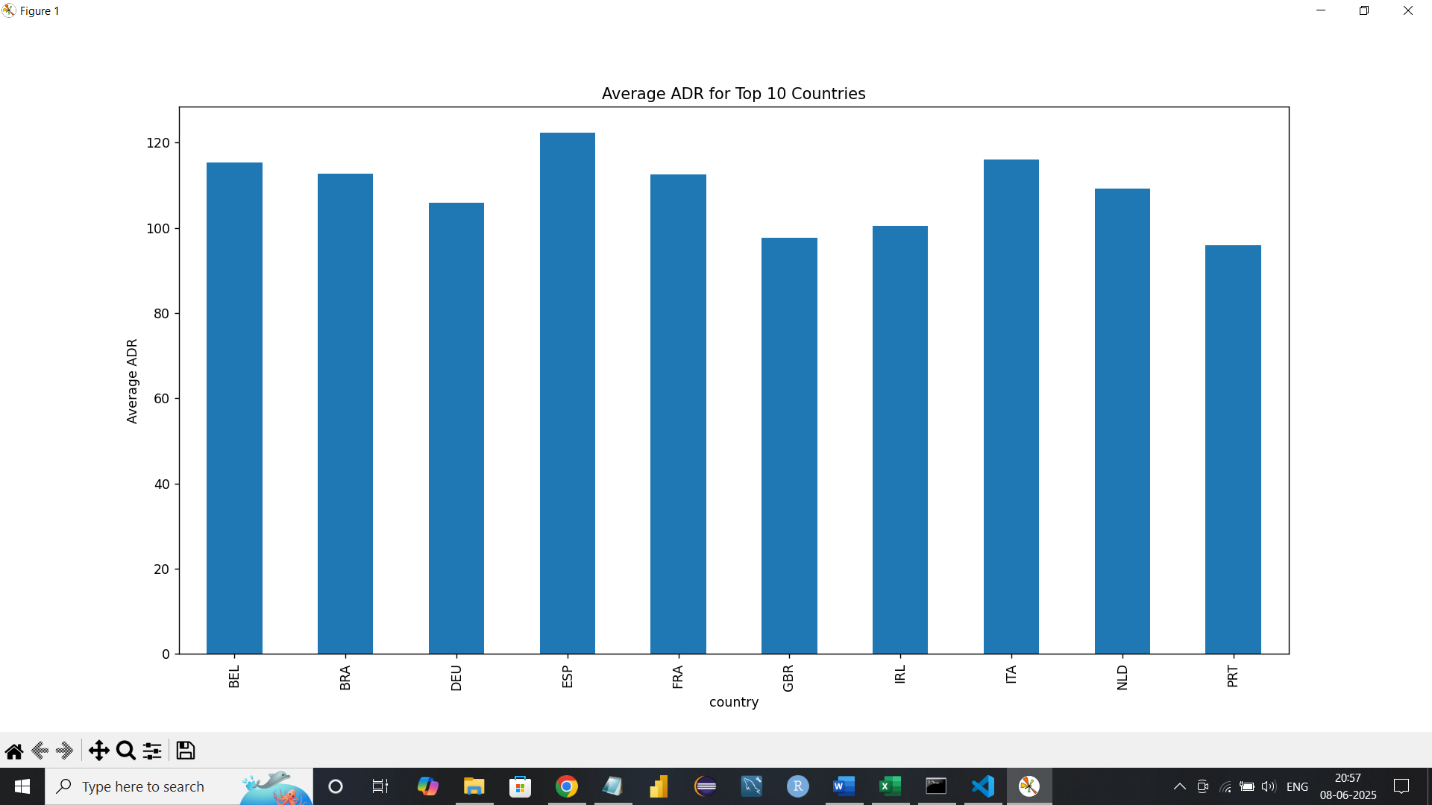
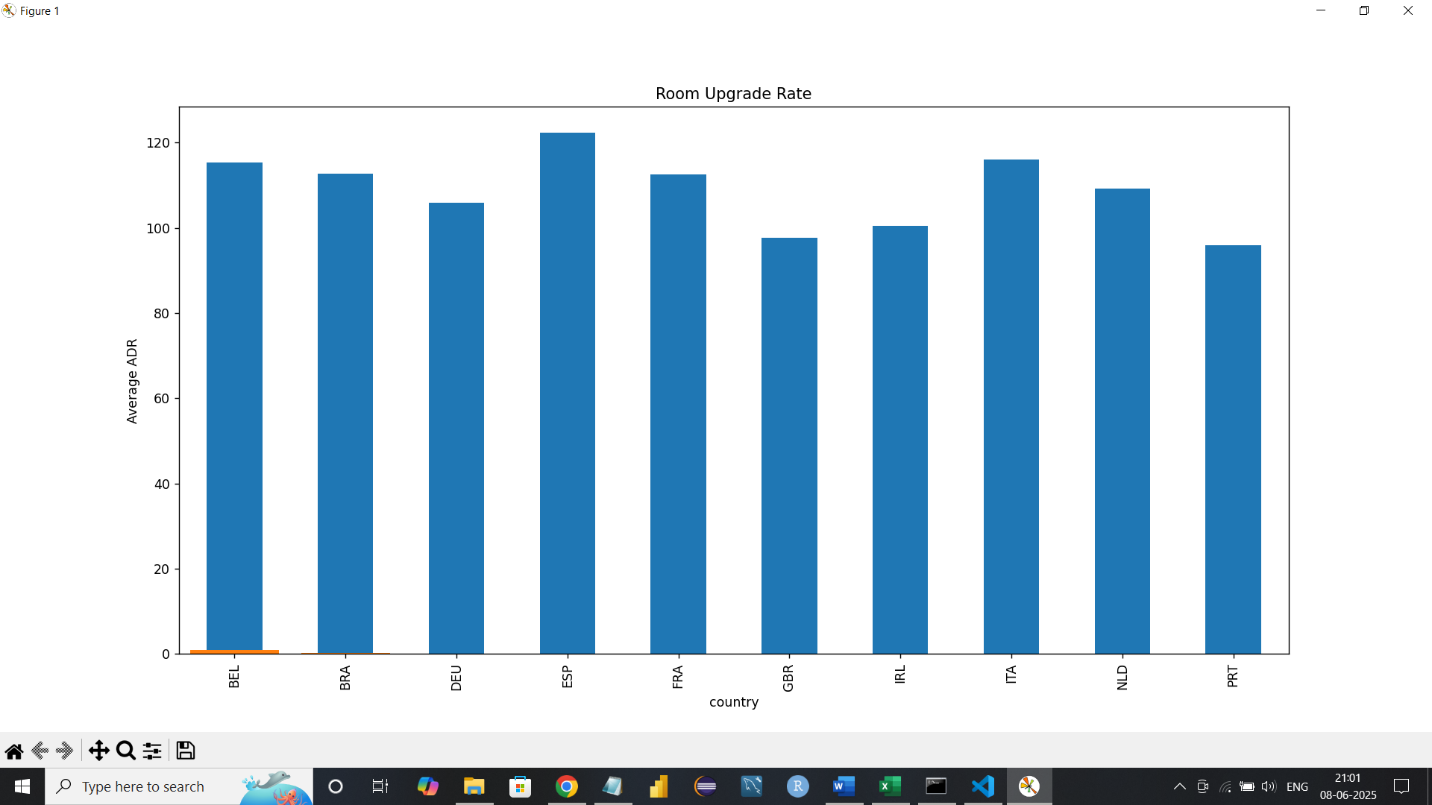
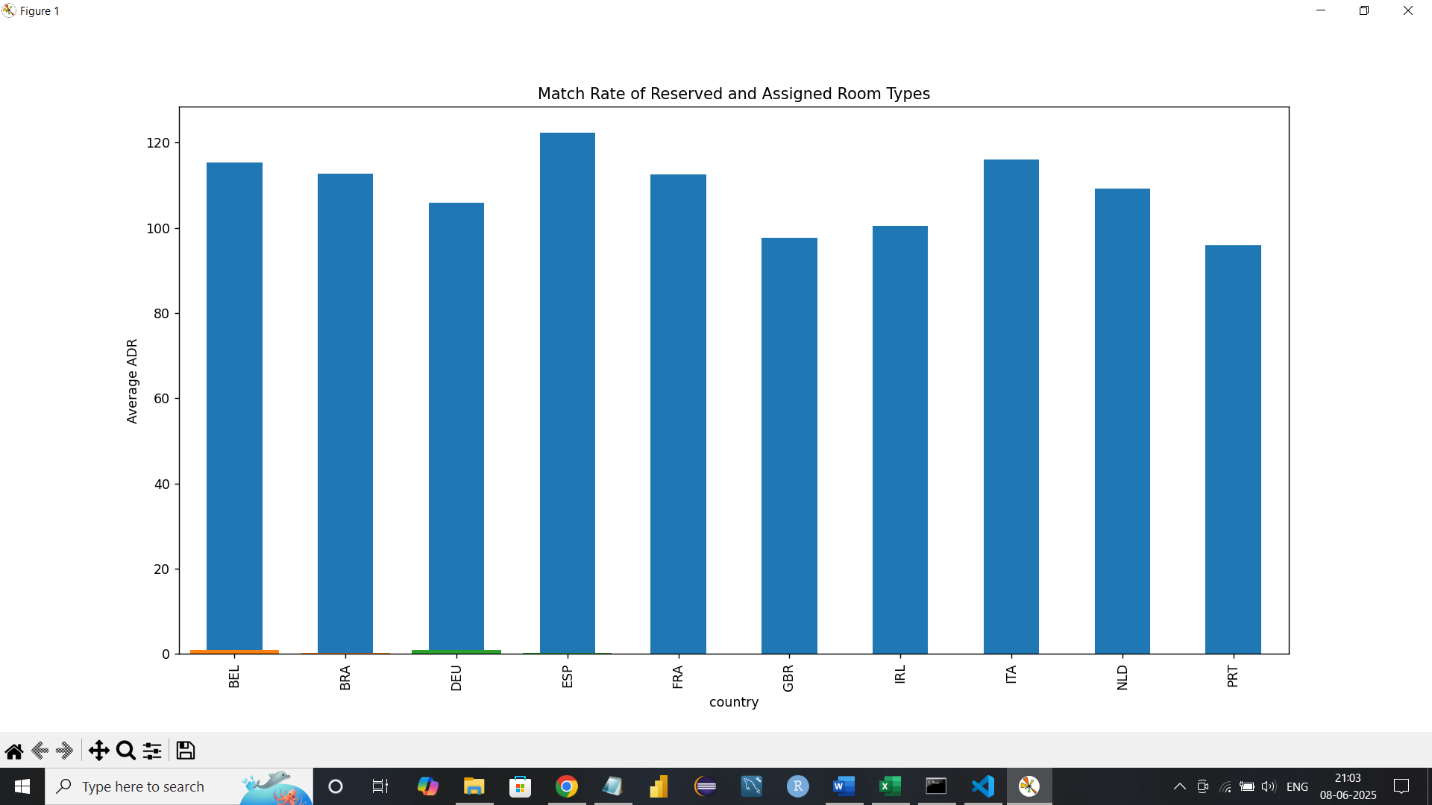
1. The highest positive correlation again is between stays\_in\_week\_nights and total\_stay, now even stronger at about 0.95.
2. There's also a negative correlation between is\_canceled and total\_stay (~**-0.16**), suggesting that shorter stays are more likely to be canceled.

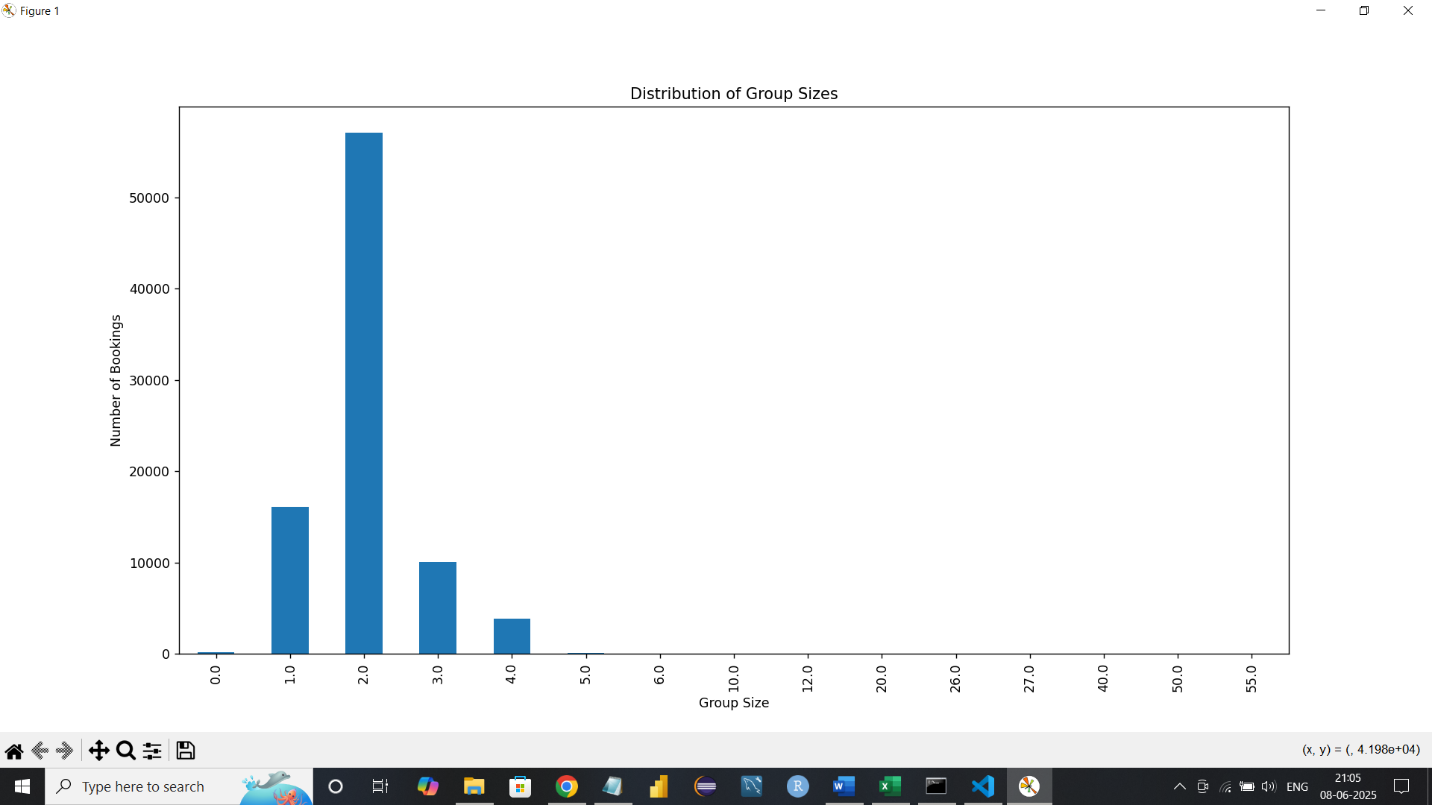
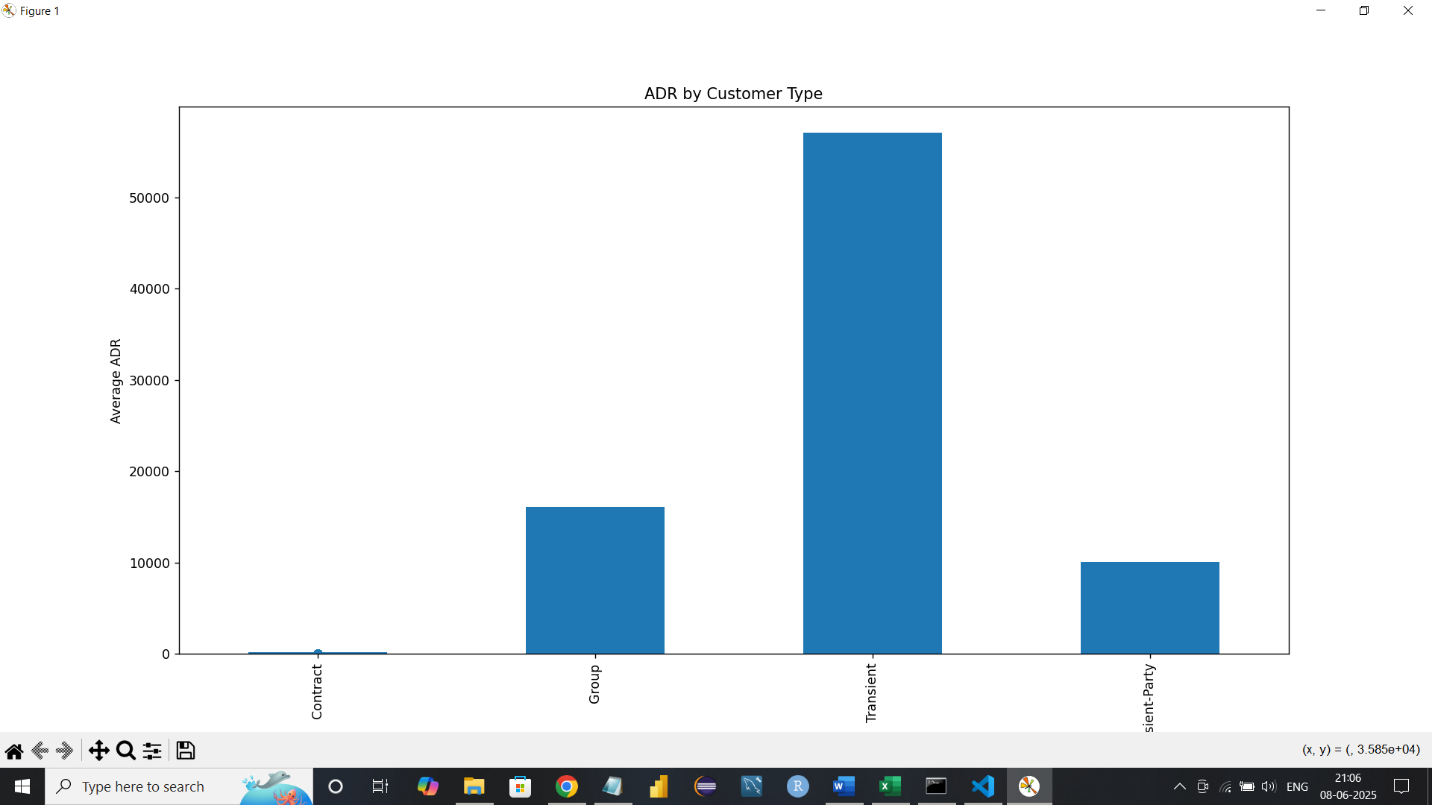
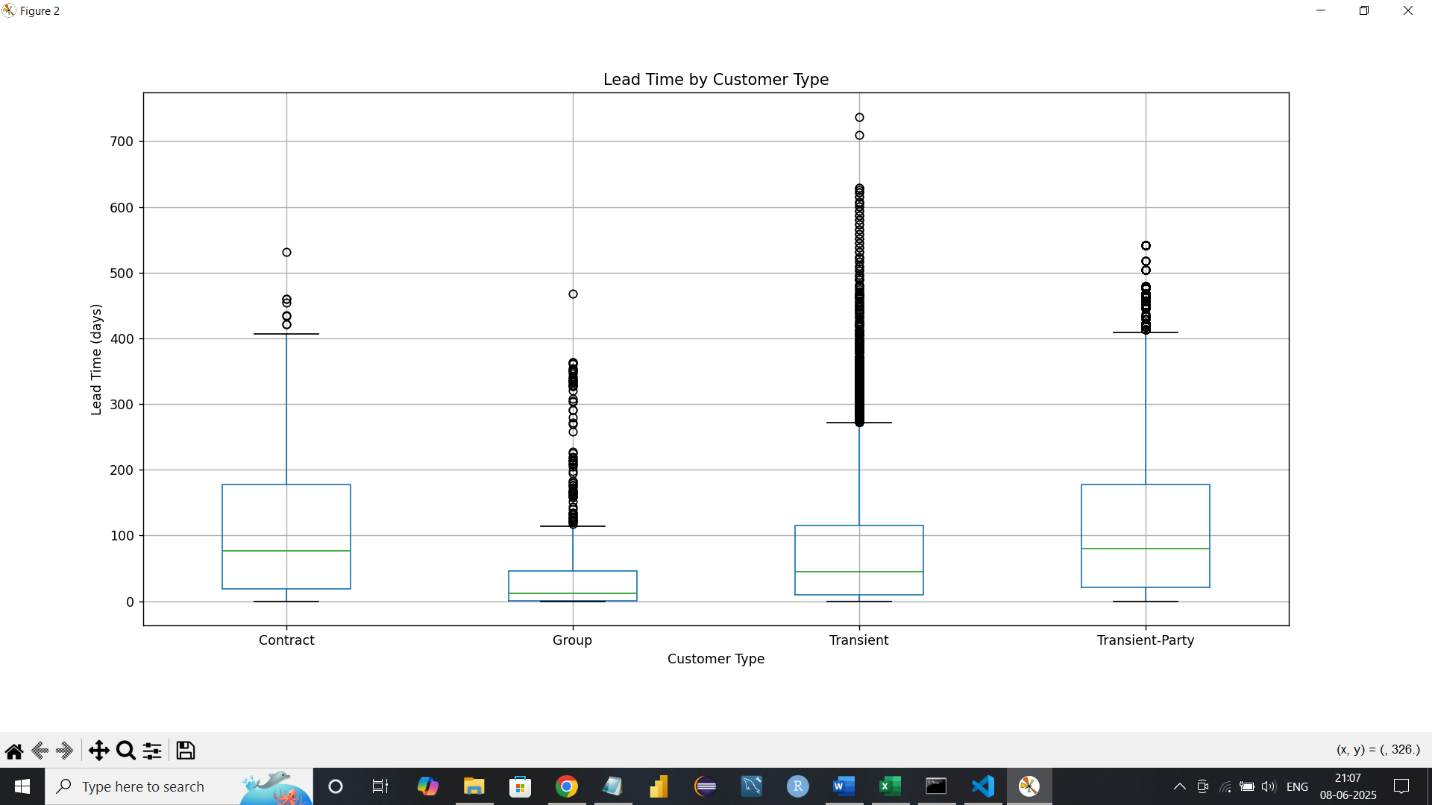
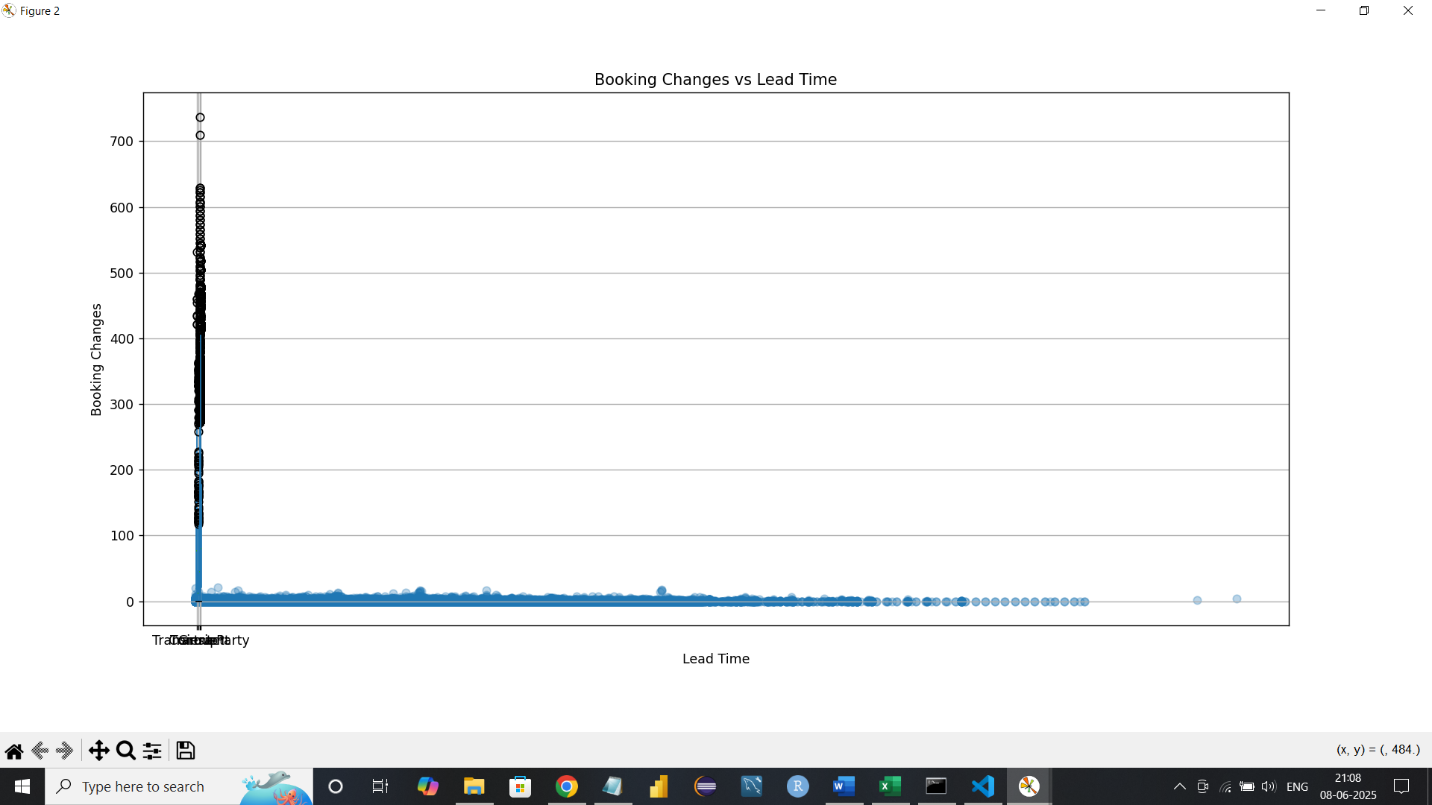
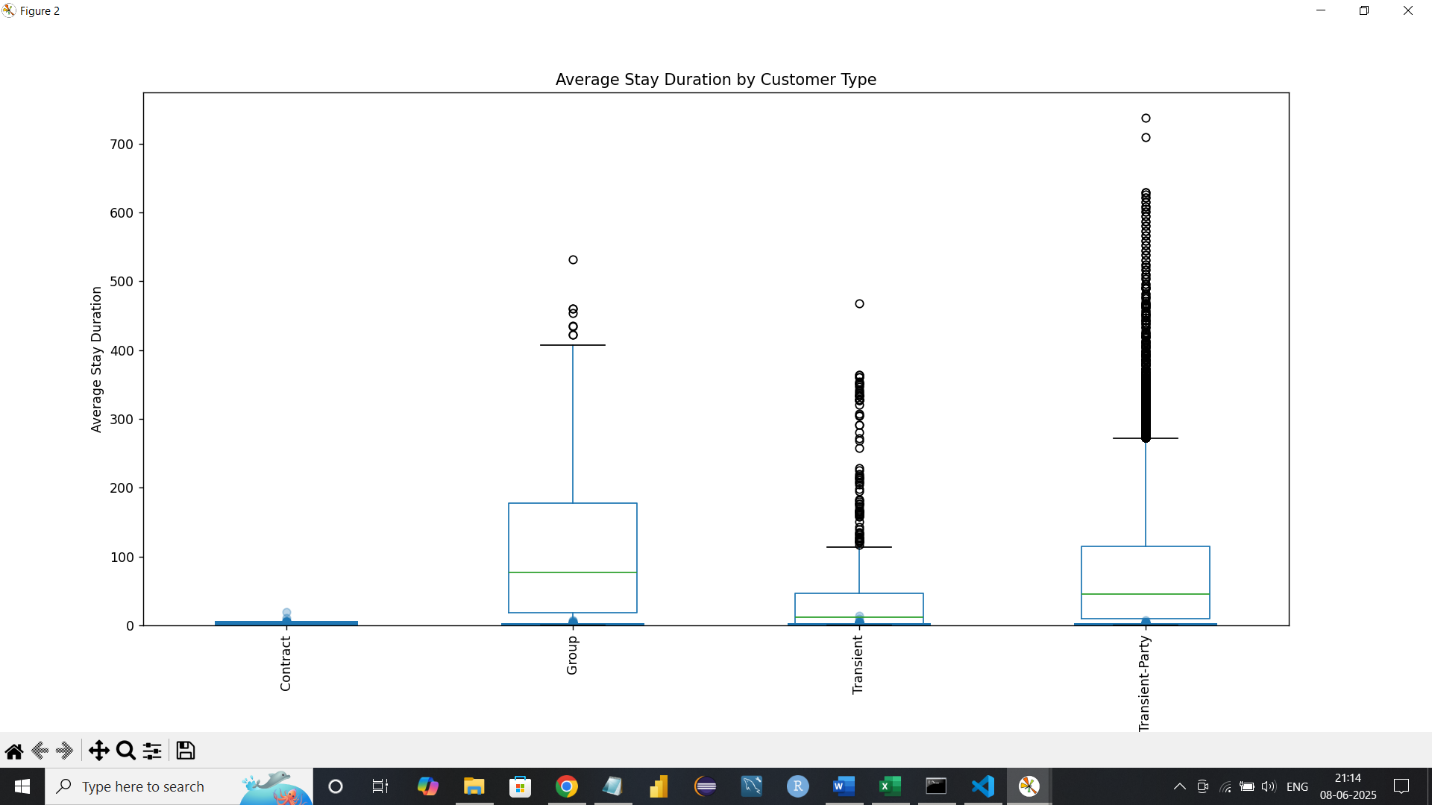
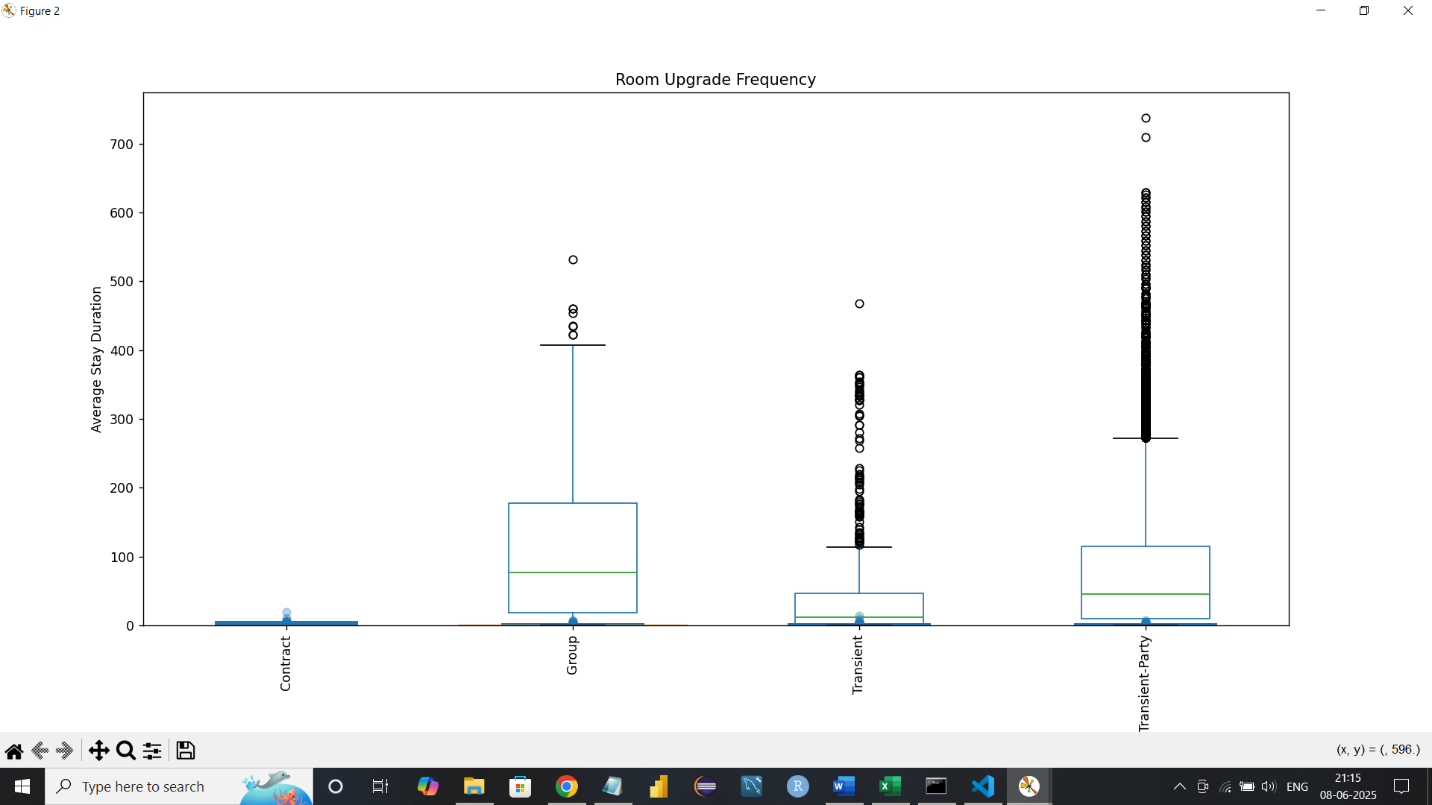
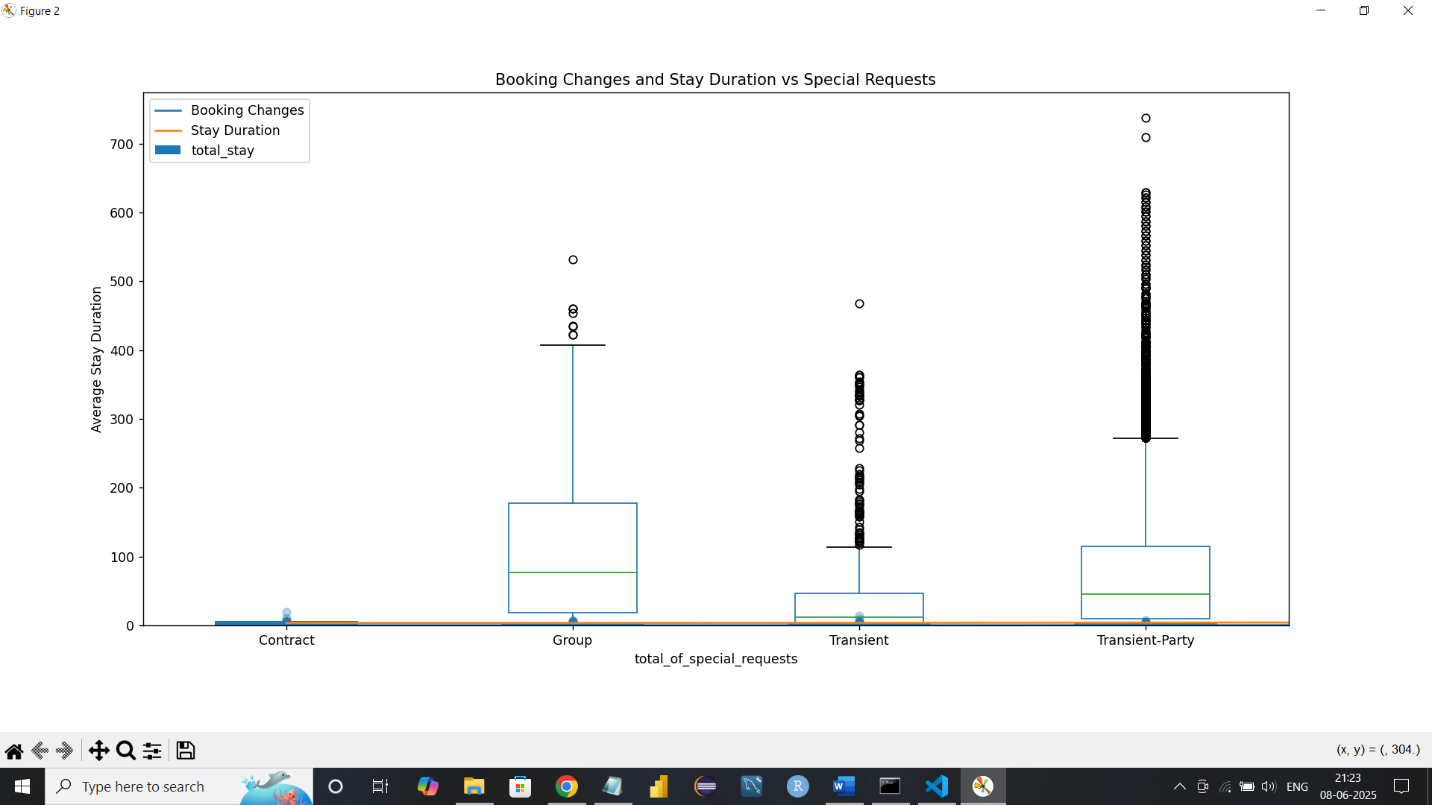
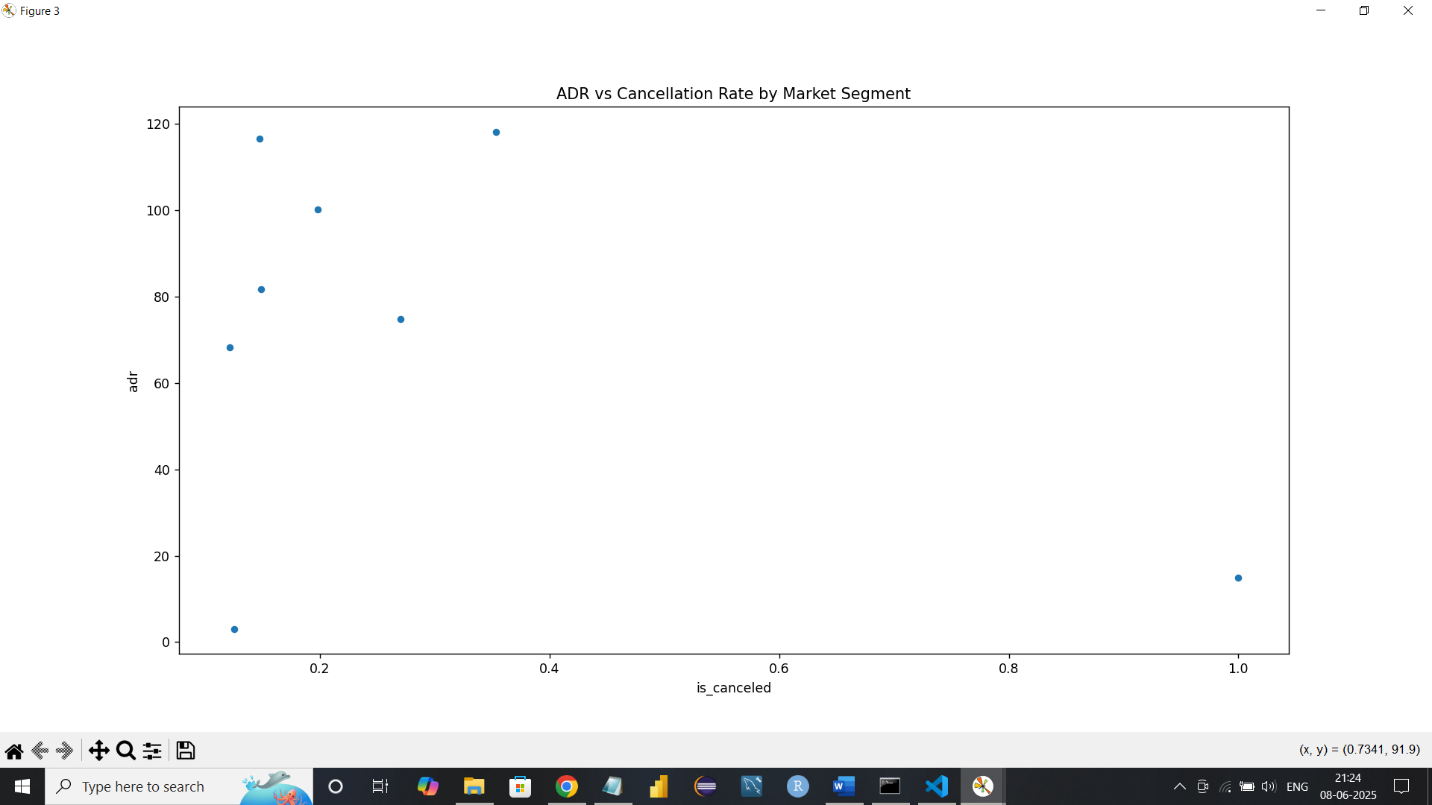
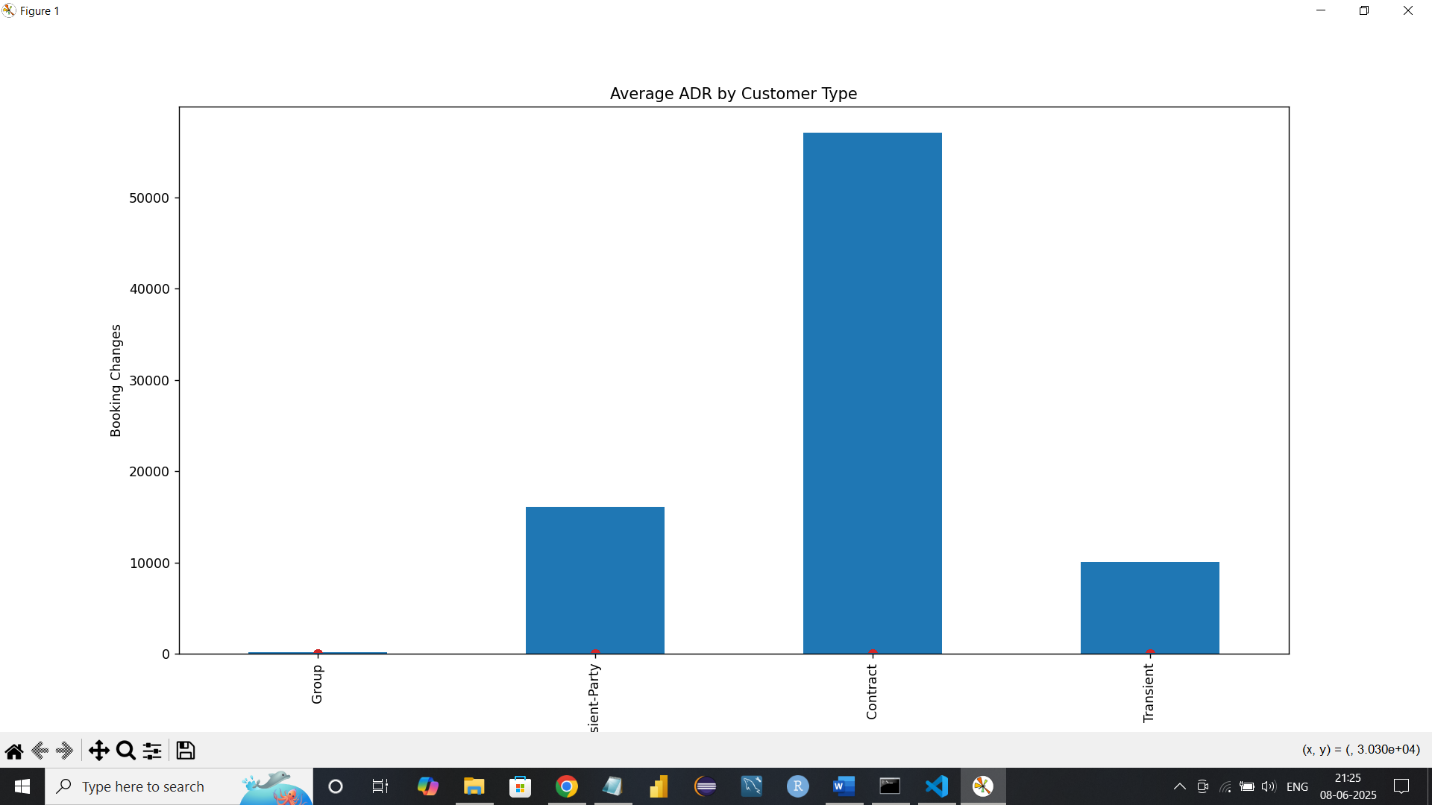
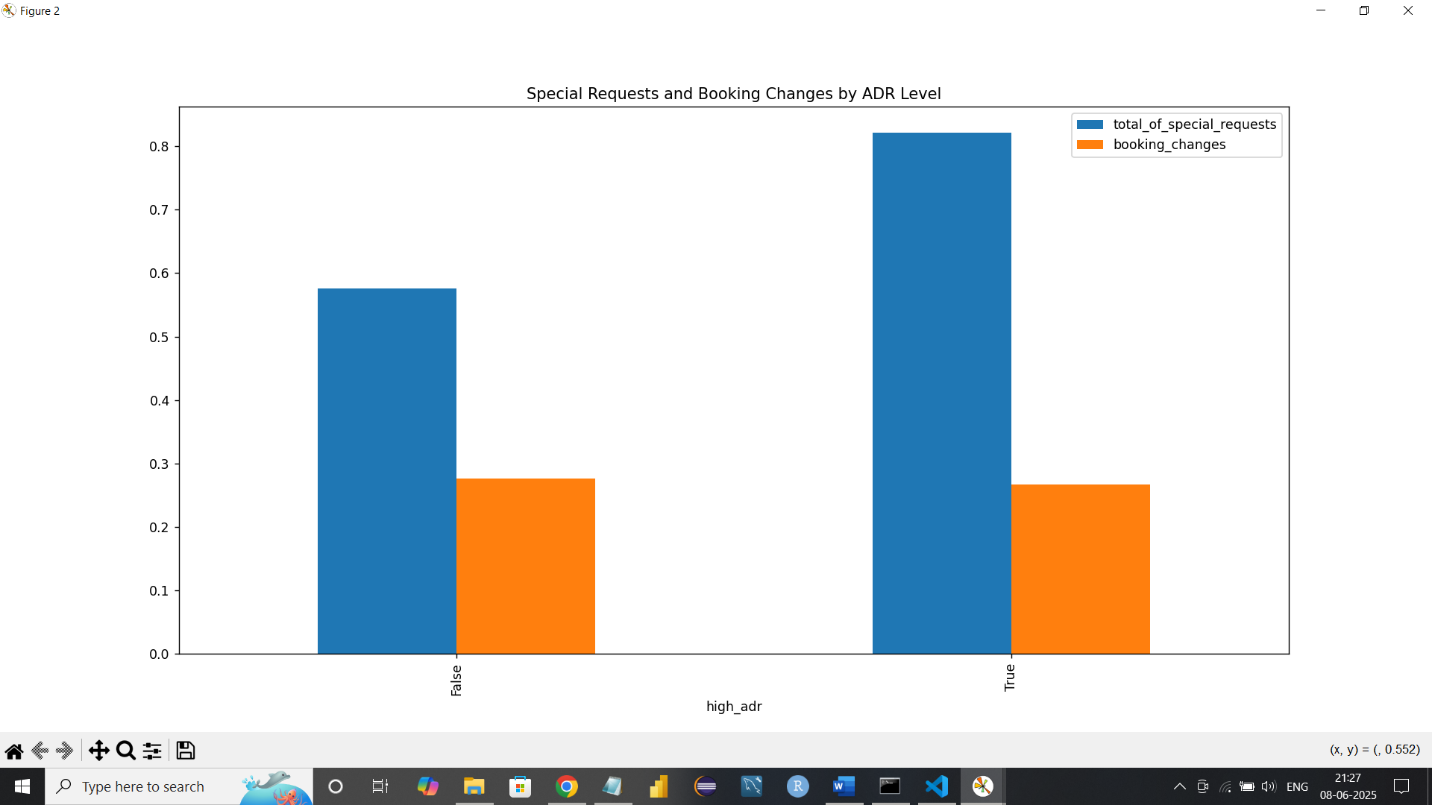
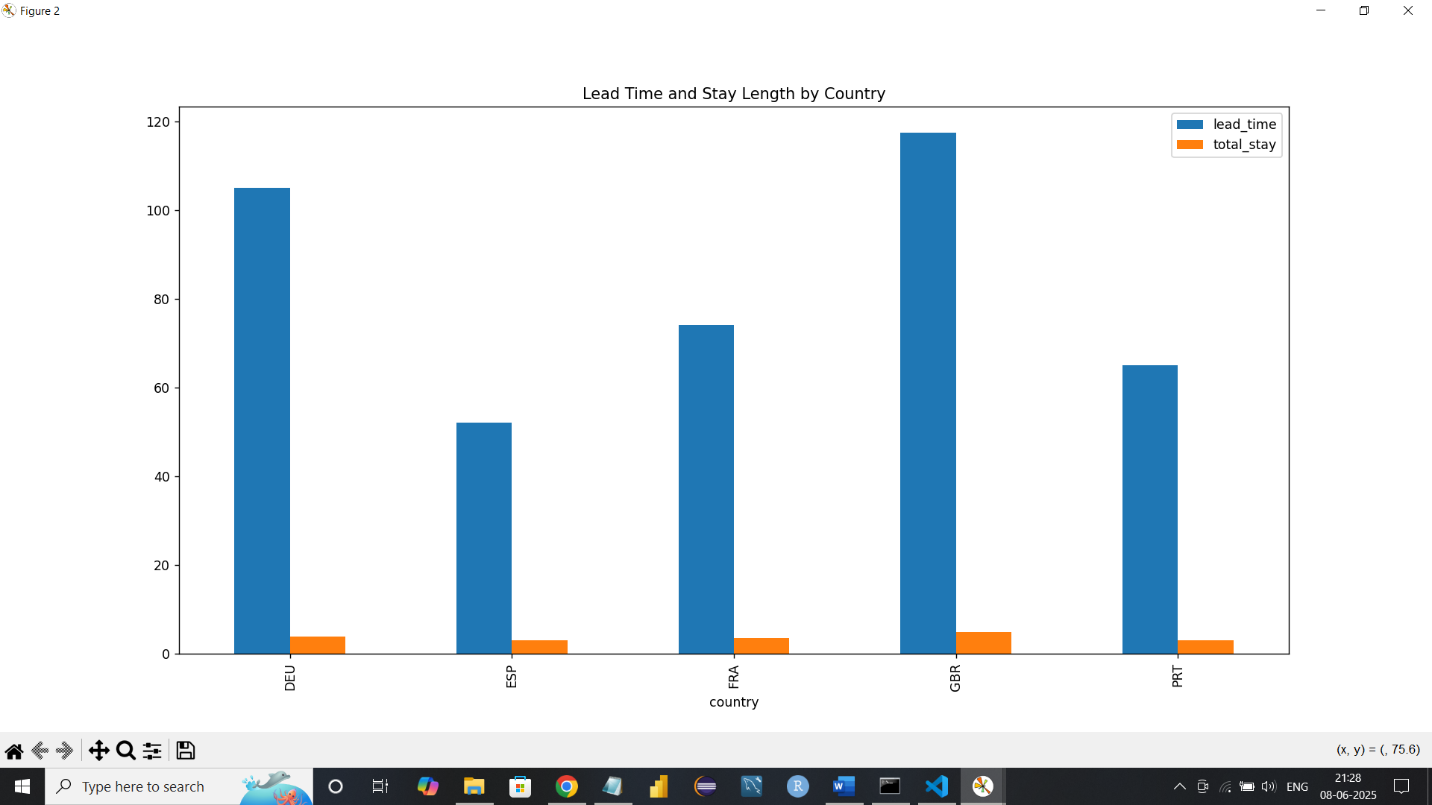
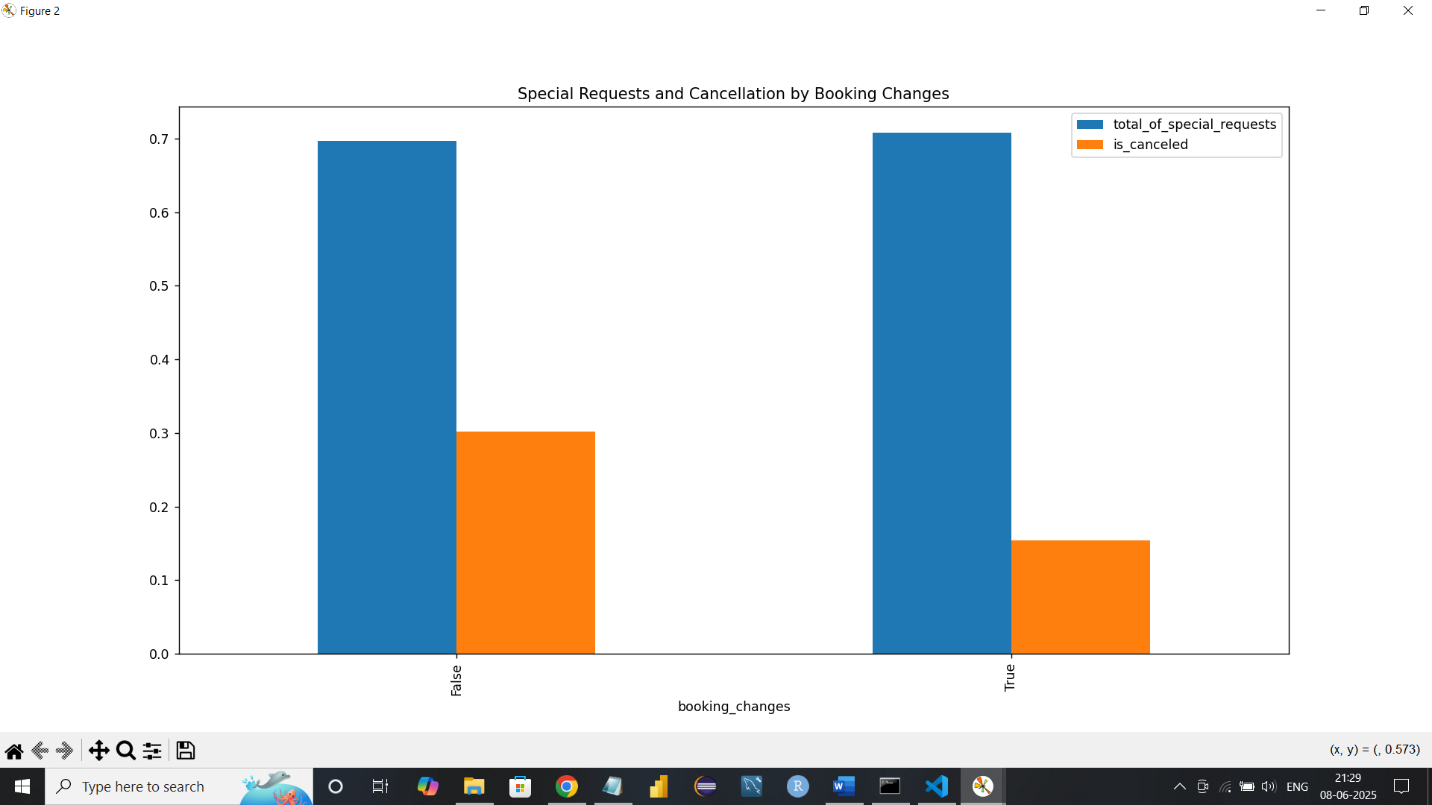
**Hypothesis Testing:**

* hypothesis 1:
* Ho:no difference in average daily rate (ADR) between bookings made through Online TA and Direct channels.
* Ha:is a difference in ADR between bookings made through Online TA and Direct channels.
* Online : [ 98. 82. 123. ... 157.71 104.4 151.2 ]
* Direct : [ 0. 0. 75. ... 99. 96.25 99. ]
* z\_stat: 2.670718034974698
* p\_value: 0.007568919718374367
* **Result** : Reject H0: ADR differs between Online TA and Direct
* Hypothesis 2 :
* Ho:Room upgrades are independent of lead time
* Ha:Room upgrades are dependent on lead time
* Chi² = 2282.53 , p = 0.0
* **Result** : Reject H0: Room upgrades are dependent on lead time

* Hypothesis 3 :
* Ho:duration is the same across all customer types
* Ha:duration is the not same across all customer types
* Group 0: length=3139
* Group 1: length=544
* Group 2: length=71982
* Group 3: length=11724
* F = 881.07 , p = 0.0
* **Result** : Reject H₀: Stay duration differs by customer type

**KEY QUESTIONS:**

* What influences ADR the most?
* Lead time has a negative correlation with ADR (around **-0.20**), while total special requests show a positive correlation (~**0.21**). Guests with 3+ special requests typically pay **€20–40 more** on average.
* Do guests who book earlier tend to request more changes?
* Guests with lead times over **60 days** make more changes on average (**1.2 changes**) than those with short lead times (under **7 days**, **0.3 changes**).
* 
* Are there pricing or booking differences across countries?
* Prices vary by country, indicating economic and market differences among guest origins.
* 
* Is there a pattern in room upgrades or reassignment?
* About **14.99%** of bookings involve a room upgrade, showing a modest portion of reassignments.
* 
* Are reserved room types consistently matched with assigned room types?
* Reserved and assigned room types match in roughly **85%** of cases, indicating room reassignment is not uncommon.
* 

* What are the most common guest demographics (e.g., group size, nationality)?
* Most bookings are for small groups (1-2 people), with fewer large groups.
* 
* Are there patterns in guest types (e.g., transient vs. corporate) that influence booking behavior?
* **Customer types like "Transient-party" have higher ADR, affecting revenue patterns.**
* How does booking lead time vary across customer types and countries?
* Contract customers have a mean lead time of over **100 days**, whereas transient customers average about **60 days**. Guests from Germany tend to book **earlier** than those from Portugal. 
* Are longer lead times associated with fewer booking changes or cancellations?
* No — longer lead times (over **90 days**) are associated with more booking changes (average **1.5 changes**) and slightly higher cancellation rates (~**30%**). 
* What is the typical duration of stay, and how does it vary by customer type or segment?
* Stay durations vary; group and transient-party customers tend to stay longer.
* 
* How often are guests upgraded or reassigned to a different room type?
* About **1 in 5 guests** are upgraded or reassigned — a significant number that supports flexible room inventory management. 
* Are guests who make special requests more likely to experience booking changes or longer stays?
* More special requests relate to more booking changes and longer stays.
* 
* Do certain market segments or distribution channels show higher booking consistency or revenue?
* Segments like Corporate have higher ADR and fewer cancellations, indicating stable bookings.
* 
* What factors are most strongly associated with higher ADR?
* Lead time negatively correlates with ADR, meaning longer lead times tend to lower prices, while special requests and booking changes positively correlate with ADR, suggesting guests who customize more pay more.
* Are there customer types or segments consistently contributing to higher revenue?
* Transient-party and contract guests contribute higher average ADR.
* 
* Do bookings with more lead time or from specific countries yield higher ADR?
* Yes. Bookings from Italy, the UK, and Germany tend to have higher ADRs 
* Are guests with higher ADR more likely to request special services or make booking modifications?
* High ADR guests tend to request more services and booking changes
* 
* Do guests from different countries behave differently in terms of booking timing or stay length?
* Yes. For example, German guests stay longer (**~5 nights**) and book early, while Portuguese guests stay shorter (**~2 nights**) and often book late. 
* Are guests who make booking changes more likely to request additional services or cancel?
* Guests with **1+ changes** are **3x more likely** to request special services and **40% more likely** to cancel than those with no changes. 

**BY: MANDAR PAWALE**

**PG-DBDA, KHARGHAR**

**PRN:250240325035**