

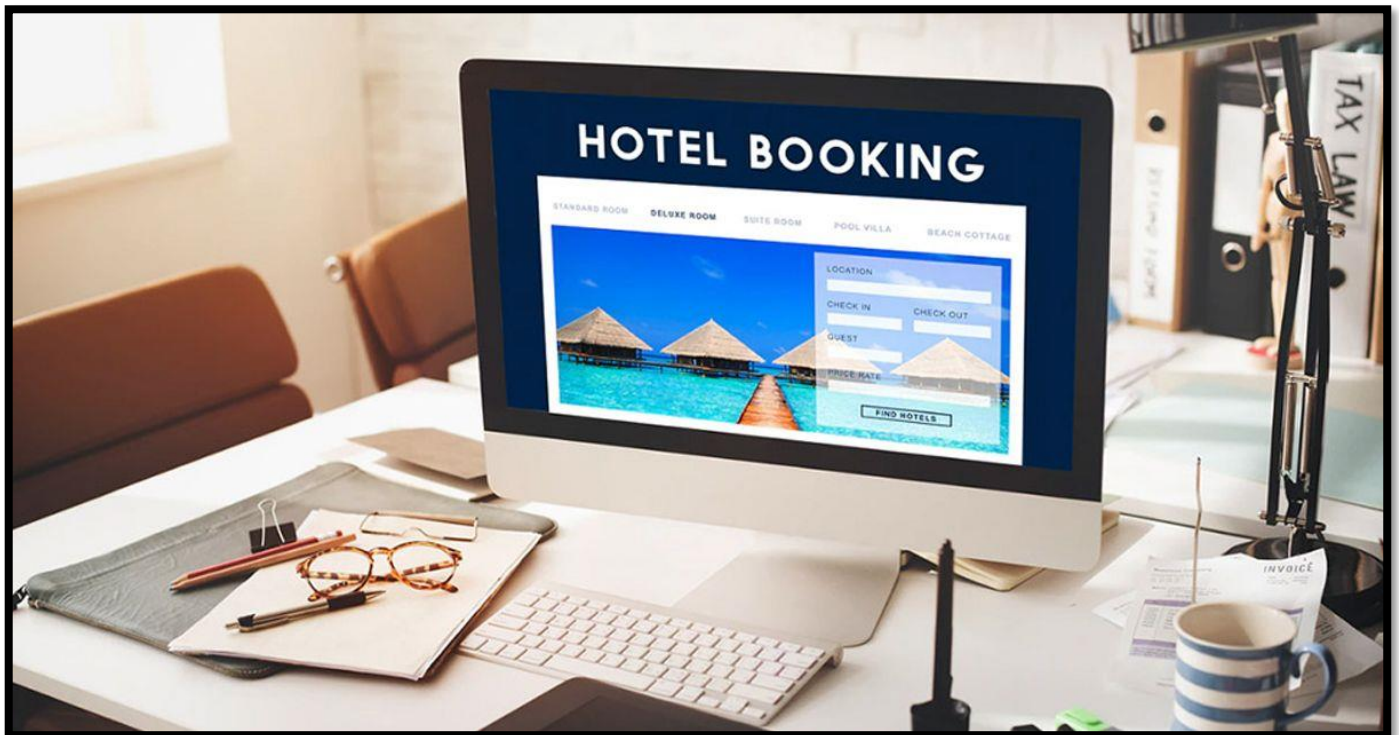
פרויקט בקורס רגרסיה ליניארית

חלק א'

קבוצה 10

שני גואטה – 208499186

רועי עזראי – 206118754



תוכן עניינים

1.	בחירת מאגר נתונים :	3
2.	טבלת משתנים:	3
3.	תיאור המשתנים :	4
4.	תיאור קשרים בין משתנים:	4
5.	ניתוח תיאורי של המשתנים:	6
6.	ניתוח חריגים:	10
7.	פונקציית צפיפות והתפלגות מצטברת:	12
9.	טבלאות שכיחות:	15
	נספחים:	17

1. בחירת מאגר נתונים : מאגר הנתונים שנבחר עבורנו הוא Hotel Booking. מאגר נתונים זה מתאר לנו פרמטרים שונים עבור הזמנת חופשה במלון.

2. טבלת משתנים:

סוג המשתנה	סימון במודל	יחידת מידה	סוג המשתנה	הסבר קצר על המשתנה
מוסבר Y	adr	דולר	רציף	תעריף יומי ממוצע - מחושב על ידי חלוקת סכום כל עסקאות הלינה במספר הכולל של לילות השהייה
מסביר X1	lead_time	מספר שלם	רציף	מספר הימים שחלפו בין תאריך ההזמנה לבין תאריך ההגעה.
מסביר X2	arrival_date_month	טקסטואלי	קטגוריאלי	חודש ההגעה
מסביר X3	arrival_date_day_of_month_	מספר שלם	רציף	יום ההגעה בחודש הרלוונטי (1-31)
מסביר X4	stays_in_week_nights	מספר שלם	רציף	מספר לילות השבוע (שני עד שישי) שהאורח שהה או הזמין לשהות במלון
מסביר X5	adults	מספר שלם	רציף	מספר המבוגרים בחדר
מסביר X6	meal	טקסטואלי	קטגוריאלי	Breakfast & BB – Bed לינה וארוחת בוקר HB - Half Board לינה וחצי פנסיון (ארוחת בוקר וערב) FB - Full Board לינה ופנסיון מלא (ארוחת בוקר צהרים וערב)
מסביר X7	country	טקסטואלי	קטגוריאלי	ארץ מוצא
מסביר X8	reserved_room_type	טקסטואלי	קטגוריאלי	קוד סוג החדר. קוד מוצג במקום ייעוד מטעמי אנונימיות
מסביר X9	total_of_special_requests	מספר שלם	רציף	מספר הבקשות המיוחדות של הלקוח

3. תיאור המשתנים :

המשתנה המוסבר Y -

adr: תעריף יומי ממוצע. מחושב על ידי חלוקת סכום כל עסקאות הלינה במספר הכולל של לילות השהייה.

המשתנים המסבירים-

1.lead_time: מספר הימים שחלפו מתאריך ההזמנה לתאריך ההגעה. משתנה זה יעזור לנו לבחון האם יש קשר בין הזמנה מוקדמת לתעריף היומי הממוצע. בביצוע הזמנה מוקדמת המלון צפוי להיות פנוי יותר ובכך סכום העסקאות לאותה תקופה במלון יהיה קטן יותר.

2.arrival_date_mount: חודש ההגעה. משתנה קטגוריאלי שבעזרתו נרצה לבדוק האם קיים קשר בין חודש ההגעה לבין התעריף היומי הממוצע. יתכן שבחודשי הקיץ, בחודשי החגים/חופשות, ובעונות תיירות הביקוש במלונות יהיה גבוה יותר בהשוואה לחודשים אחרים ובכך עשוי להשפיע על התעריף היומי הממוצע.

3.arrival_date_day_of_month: יום ההגעה בחודש הרלוונטי. בעזרת משתנה זה נרצה לבדוק האם קיים קשר בין היום בחודש בו הוזמן המלון לבין התעריף היומי הממוצע. יתכן שעבור חודש מסוים יהיו ימים בעלי ביקוש גבוה יותר כמו ימים לפני ערבי חגת לפני סופי שבוע או ימים לפני תחריט ספורט וכו'. נצפה בימים או ביקוש גבוה יותר להזמנות במלון ובכך זה עשוי להשפיע על התעריף היומי הממוצע.

4.stays_in_week_nights: מספר לילות השבוע מיום שני עד יום שישי שהאורח שהה או הזמין לשהות במלון. נרצה לבדוק האם קיים קשר בין מספר לילות השבוע (שני עד שישי) לבין התעריף היומי הממוצע. בימי שבת וראשון, שהם ימי סוף שבוע שבהם לרוב האנשים אין עבודה, מספר ההזמנות צפוי להיות גבוה יותר במלון לעומת ימי שני עד שישי שהם ימי חול. סביר להניח שיהיו פחות הזמנות ללילות בימי שני עד שישי, שהם ימי עבודה לרוב האוכלוסייה, ובכך זה עשוי להשפיע על התעריף היומי הממוצע.

5.adults: מספר המבוגרים בחדר. נרצה לבדוק האם קיים קשר בין מספר המבוגרים בחדר לבין התעריף היומי הממוצע. ככל שמספר המבוגרים בחדר גבוה יותר כך מספר החדרים בהזמנה יהיה נמוך יותר כיוון שיותר מבוגרים מתחלקים בפחות חדרים, מכאן שזה עשוי להוריד את המחיר של העסקה ולהשפיע על התעריף היומי הממוצע.

6.meal: סוג הפנסיון (BB, HB, FB). משתנה קטגוריאלי שבעזרתו נרצה לבדוק האם קיים קשר בין סוג הפנסיון לתעריף היומי הממוצע. סביר להניח שהמחיר עבור סוג פנסיון מורחב יותר יהיה יותר יקר מאשר סוג פנסיון בסיסי המספק שירות מוגבל יותר של שירותים. דבר זה עשוי להשפיע על סכום העסקה בהזמנה עבור אותו מספר לילות ולכן יתכן שישפיע על התעריף היומי הממוצע.

7.country: ארץ מוצא. משתנה קטגוריאלי שבעזרתו נרצה לבדוק האם קיים קשר בין המדינה לבין התעריף היומי הממוצע. לכל מדינה יש כלכלה וערך מטבע שונה, יש מדינות יותר מתירות מאחרות או מדינות יותר חמות או קרות. כל אלו ועוד משפיעים על הביקוש, על סכום העסקה, על משך השהיה במלון ועשויים להשפיע על התעריף היומי הממוצע.

8.reserved_room_type: קוד סוג החדר. משתנה קטגוריאלי שבעזרתו נרצה לבדוק האם קיים קשר בין סוג החדר לבין התעריף היומי הממוצע. סביר שיהיה הבדל במחיר בין סוגי חדרים שונים בהתאם לביקוש שלהם ולשירותים/לקיבולת שבהם ומכאן שיש לזה השפעה על סכום העסקה ויתכן שישפיע גם על התעריף היומי הממוצע.

9.total_of_special_requests: מספר הבקשות המיוחדות של הלקוח. נרצה לבדוק האם קיים קשר בין מספר הבקשות המיוחדות של הלקוח לתעריף היומי הממוצע. לקוח יכול לבקש להיות בקומה גבוהה יותר עם נוף לבריכה או באזור עם פחות ילדים, מיטות גדולות יותר או תוספת של מיטה/לול לתינוק לחדר מסוים. בקשות אלו עשויות להיות בתוספת תשלום ולהשפיע על סכום העסקה בהזמנה ומכאן שסביר שמספר הבקשות ישפיע על התעריף היומי הממוצע.

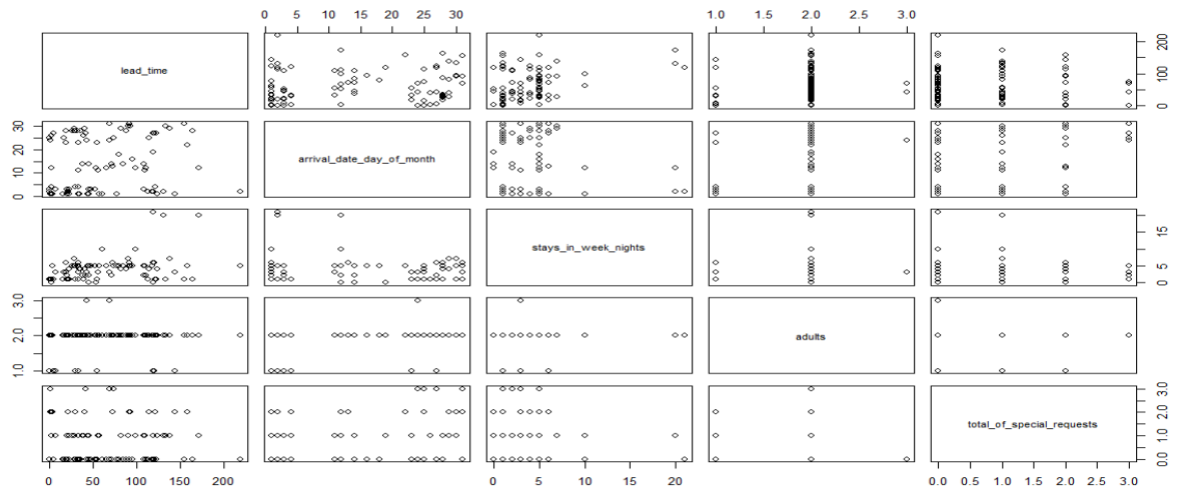
4. תיאור קשרים בין משתנים:

קשר סיבתי: קשר של סיבה ותוצאה בין שני משתנים שאנחנו מניחים שקיים. כאשר מניחים שקיים קשר סיבתי בין משתנה X1 ו X2 נצפה לראות שעבור שינוי חיובי ב X1 יהיה בהתאם שינוי חיובי ב X2 ולהפך. עבור שינוי שלילי ב X1 נצפה לשינוי שלילי ב X2 בהתאם. עבור קשר סיבתי הפוך נצפה שעבור שינוי חיובי ב X1 יהיה שינוי שלילי ב X2.

קשר מדגמי: קשר מדגמי הוא קשר בין 2 משתנים שתלוי במדגם ולא ניתן לחשוב על קשר של סיבה ותוצאה ביניהם. קשר מדגמי נוכל לבחון על ידי בדיקת הקורלציה בין שני המשתנים. במידה וקיימת קורלציה בין המשתנים אך לא ניתן לחשוב על קשר של סיבה ותוצאה, זהו קשר מדגמי.

מקדם המתאם: מסומן באות r . יכול לקבל ערכים בטווח $[-1,1]$. משקף את עוצמת וכיוון הקשר בין 2 משתנים. ככל שערכו מתקרב בערך מוחלט ל1, הוא מתאר קשר חזק יותר בין המשתנים. ככל שערכו מתקרב לאפס הוא מתאר קשר יותר חלש/ אין קשר. בטווח הערכים החיובי שהוא יכול לקבל, הוא מתאר קשר חיובי בין המשתנים. בטווח הערכים השלילי הוא מתאר קשר שלילי בין המשתנים.

Plot על המשתנים המסבירים הרציפים:



ההנחות לגבי קשר סיבתי בין משתנים מסבירים רציפים-

1. קשר בין X_1 ל- X_4 (מספר הימים מתאריך ההזמנה לתאריך ההגעה, מספר לילות בשבוע):

אנחנו משערים שקיים קשר סיבתי בין מספר הימים מתאריך ההזמנה לתאריך ההגעה לבין מספר לילות השבוע. ככל שמתקרבים לזמן ההגעה, למלון יש יותר הזמנות ומכאן שהזמנות במלון יורדת. במועדים הקרובים לזמן ההגעה יתכן שלא תהיה אפשרות לבצע הזמנה למספר רב של לילות וההזמנה תהיה מוגבלת או מנגד, האדם המזמין ירצה להימנע מבעיות אלו ויבצע את ההזמנה מוקדם יותר. נצפה שעבור מספר לילות רב יותר, מספר הימים מתאריך ההזמנה להגעה יהיה גבוה יותר, ולהפך.

2. קשר בין X_1 ל- X_5 (מספר הימים מתאריך ההזמנה לתאריך ההגעה, מספר מבוגרים בחדר):

אנחנו משערים שיש קשר סיבתי הפוך בין מספר הימים מתאריך ההזמנה להגעה לבין מספר המבוגרים בחדר. ככל שמתקרבים לזמן ההזמנה, זמינות המלון יורדת ומספר החדרים הפנויים להזמנה מוגבל כך שיותר מבוגרים יצטרכו להיכנס לחדר מסוים במקום להתחלק ליותר חדרים. נצפה שעבור מספר ימים קטן יותר יהיו יותר מבוגרים בחדר ולהפך.

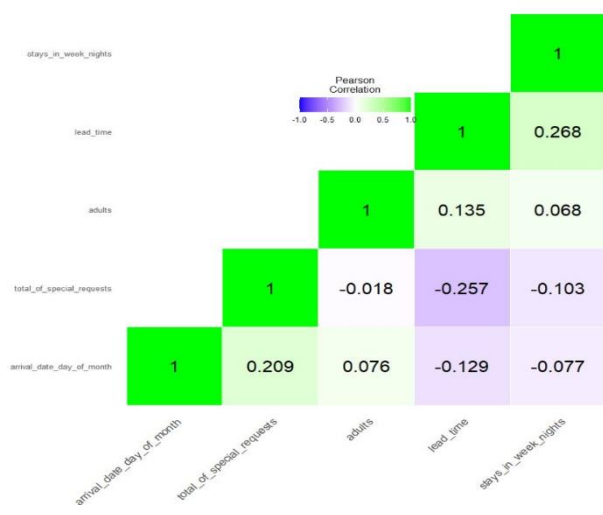
3. קשר בין X_1 ל- X_9 (מספר הימים מתאריך ההזמנה לתאריך ההגעה, מספר בקשות מיוחדות):

אנחנו משערים שיש קשר סיבתי הפוך בין מספר הימים מתאריך ההזמנה להגעה לבין מספר הבקשות המיוחדות של הלקוח. כשמתקרבים לזמן ההגעה, הזמינות במלון יורדת עקב הזמנות אחרות, יש שינויים בסוגי החדרים הזמינים ויתכן שמי שירצה לבצע הזמנה לא ימצא חדר מתאים עבורו אך המלון יוכל לספק לו מענה לבקשות מיוחדות שימלאו את הפער בין מה שהוא צריך למה שהמלון מסוגל לספק. נצפה שככל שמספר הימים יהיה קטן יותר כך מספר הבקשות יהיה גדול יותר ולהפך.

4. קשר בין X_5 ל- X_9 (מספר מבוגרים בחדר, מספר בקשות מיוחדות):

אנחנו משערים שיש קשר סיבתי בין מספר המבוגרים בחדר לבין מספר הבקשות המיוחדות של הלקוח. ככל שמספר המבוגרים גדול יותר בחדר כך הנוחות עלולה להיפגע והלקוח צפוי לבקש יותר בקשות מיוחדות כדי לשפר את שהותו במלון. נצפה שככל שמספר המבוגרים בחדר יגדל כך מספר הבקשות המיוחדות יגדל ולהפך.

גרף קורלציה בין המשתנים המסבירים הרציפים:



1. קשר בין X1 ל- X4 (מספר הימים מתאריך ההזמנה לתאריך ההגעה, מספר לילות בשבוע):
מקדם המתאם בין שני המשתנים הוא 0.268. כפי שהנחנו אכן קיים ביניהם קשר סיבתי, אך הוא חלש מהמצופה. מתקיימת קורלציה חיובית חלשה, כלומר, ככל שמספר הימים מתאריך ההזמנה לתאריך ההגעה יגדל כך מספר הלילות יגדל בהתאם.
2. קשר בין X1 ל- X5 (מספר הימים מתאריך ההזמנה לתאריך ההגעה, מספר מבוגרים בחדר):
מקדם המתאם בין המשתנים הוא 0.135. בשונה ממה שהנחנו מתקיימת קורלציה חיובית חלשה, כלומר, ככל שמספר הימים מתאריך ההזמנה לתאריך ההגעה יגדל כך מספר המבוגרים בחדר יגדל בהתאם. לכן מתקיים קשר מדגמי בין המשתנים עם קורלציה חיובית חלשה.
3. קשר בין X1 ל- X9 (מספר הימים מתאריך ההזמנה לתאריך ההגעה, מספר בקשות מיוחדות):
מקדם המתאם הוא 0.257-. מתקיימת קורלציה שלילית חלשה. בדומה למה ששיערנו אכן מתקיים קשר סיבתי הפוך אך חלש מהמצופה. הקורלציה שלילית חלשה, כלומר, כאשר מספר הימים מתאריך ההזמנה קטן כך מספר הבקשות המיוחדות גדל בהתאם.
4. קשר בין X5 ל- X9 (מספר מבוגרים בחדר, מספר בקשות מיוחדות):
מקדם המתאם הוא 0.018-. בשונה ממה שהנחנו מתקיימת קורלציה שלילית זניחה, כלומר הקשר בין המשתנים זניח. הסבר לכך יכול להיות שיתכן שהחדרים המותאמים ליותר מבוגרים נמצאים באזורים נוחים יותר במלון והם מאובזרים בצידוד טוב יותר מחדרים קטנים המכילים פחות מבוגרים כך שגם כשיש יותר מבוגרים בחדר איכות השהייה לא נפגעת והאורחים לא צריכים לבקש בקשות מיוחדות לשיפור שהותם.
5. קשר בין X3 ל- X9 (יום ההגעה בחודש הרלוונטי, מספר בקשות מיוחדות):
מקדם המתאם הוא 0.209. מתקיימת קורלציה חיובית חלשה. נראה כי אין סיבה לקשר בין יום ההגעה בחודש הרלוונטי לבין מספר הבקשות המיוחדות של הלקוח ומכאן שהקשר הוא מדגמי. לא בהכרח שיום הגעה מאוחר בחודש הרלוונטי יוביל ליותר/לפחות בקשות מיוחדות מהלקוח.

5. ניתוח תיאורי של המשתנים:

ממוצע: הממוצע מחושב מתוך אוסף סופי של מספרים, ומתאר את "מרכז" האוסף מבחינת גודל המספרים.

חציון: החציון מוגדר להיות ערך החוצה את הקבוצה, עם מספר שווה של ערכים מעליו ומתחתיו.

סטיית תקן: סטיית תקן היא מדד סטטיסטי לתיאור הפיזור של נתונים מספריים סביב הממוצע שלהם, התלוי במרחק של הנתונים מן הממוצע שלהם.

תחום בין רבעוני: הוא מדד לפיזור הערכים במדגם שפחות רגיש לתצפיות חריגות. התחום הבין רבעוני הוא תחום בין הרבעונים שבו נמצאים 50% מהתצפיות.

Skewness - (אסימטריה): מדד לסימטריה של ההתפלגות של משתנה מקרי ממשי. כשערכו המוחלט קטן מ-0.5 ניתן לומר בקירוב טוב שההתפלגות סימטרית. עבור ערך גדול מ-0.5 נקבל התפלגות מוטה חיובית בעלת זנב ימני. עבור ערך קטן מ-0.5) נקבל התפלגות מוטה שלילית בעלת זנב שמאלי.

ניתוח של משתנים רציפים:

Y - תעריף יומי ממוצע (משתנה מוסבר) :

נתון סטטיסטי	ערך	הסבר
ממוצע	72.43	ניתן לראות כי הממוצע גבוה מהחציון, ההבדל נובע עקב השונות הגבוהה בהתפלגות התעריפים הממוצעים ליום. כלומר, קיימת כמות לא גבוהה של תצפיות בעלות ערך גבוה מאוד אשר גורמים להעלאת הממוצע, כאשר החציון נקבע לפי כמות התצפיות ולא ערכן ולכן צפוי שנראה שוני ביניהם.
אסימטריה	1.787302	
סטיית תקן	41.76937	
רבעון ראשון	48	
חציון	59.18	
רבעון שלישי	81	

X1 - מספר הימים מתאריך ההזמנה לתאריך ההגעה:

נתון סטטיסטי	ערך	הסבר
ממוצע	84.85	ניתן לראות כי ממוצע הפער מתאריך ההזמנה לתאריך ההגעה גבוה מהחציון. ההבדל נובע עקב השונות הגבוהה בהתפלגות ועקב פיזור גדול הפער בתאריכים. כלומר, קיימים מספר לא גבוה של הזמנות חופשה בהן הפער בין תאריך ההזמנה לתאריך ההגעה גבוה הגורם להעלאת הממוצע, כאשר החציון נקבע לפי כמות התצפיות ולא ערכן ולכן צפוי שנראה שוני ביניהם. ניתן לראות שעבור משתנה זה ישנה אסימטריה חיוביות בינונית המעידה על זנב ימני קטן.
אסימטריה	0.7459269	
סטיית תקן	67.18872	
רבעון ראשון	33	
חציון	69	
רבעון שלישי	119	

X3 - מספר היום בחודש:

נתון סטטיסטי	ערך	הסבר
ממוצע	11.92	ניתן לראות שהממוצע גבוה מהחציון, דבר המעיד על כך כי ישנן תצפיות הרחוקות מהממוצע ומושכות אותו כלפי מעלה. סטיית התקן הגבוהה מעידה על פיזור רחב של ימי ההזמנה בחודש (כמעט שבועיים) מתיישב עם ההנחה הסבירה שברוב החודשים אנשים מזמינים חופשות לאורך כל החודש. ניתן לראות אסימטריה חיוביות ובעלת ערך קטן מ-0.5 המעידה על התפלגות ששואפת להיות סימטרית בקירוב. בנוסף מהטווח הבין רבעוני ניתן להסיק ש 50% מההזמנות שבוצעו בין היום ה 2 ליום ה 25 בחודש, דבר המתיישב עם סטיית התקן של 11 ימים.
אסימטריה	0.4849669	
סטיית תקן	11.52669	
רבעון ראשון	2	
חציון	4	
רבעון שלישי	25	

X4 - מספר לילות השבוע בהזמנה:

נתון סטטיסטי	ערך	הסבר
ממוצע	3.88	ניתן לראות כי הממוצע והחציון כמעט מתלכדים, אסימטריה חיובית וגבוהה מאוד המעידה על זנב ימני גדול הנוצר מערכים בודדים בעלי ערך גבוה. ניתן לראות זאת בפער הגדול בין הערך המקסימלי (21) לערך הרבעון השלישי, אשר מתיישב עם ההנחה הסבירה שרוב האנשים מזמינים חופשה שלא עולה על 5 ימים (אמצ"ש) או שבוע שלם והזמנות של 21 יחשבו למאורע נדיר. ע"פ הטווח הבין רבעוני שהתקבל, 50% מההזמנות בוצעו עבור בין 1 ל 5 לילות, דבר המתיישב טוב עם סטיית התקן שהתקבלה עבור 4 לילות.
אסימטריה	2.753847	
סטיית תקן	3.349531	
רבעון ראשון	1	
חציון	4	
רבעון שלישי	5	

נתון סטטיסטי	ערך	הסבר
ממוצע	1.928	ניתן לראות שערך החציון והממוצע כמעט מתלכדים כשאר החציון גדול יותר מערך הממוצע, מה שמעיד על התפלגות נתונים קרובה להתפלגות סימטרית אך עם זנב שמאלי המתבטא בערך שלילי גדול של המדד, כלומר יש יותר ערכים נמוכים מאשר גבוהים שמושכים את הזנב שמאלה. ע"פ הטווח הבין רבעוני שהתקבל, 50% מההזמנות בוצעו עבור 2 מבוגרים בדיוק, דבר המתיישב עם סטיית התקן הנמוכה שהתקבלה.
אסימטריה	- 1.59935	
סטיית תקן	0.3156151	
רבעון ראשון	2	
חציון	2	
רבעון שלישי	2	

X9 - מספר בקשות מיוחדות:

נתון סטטיסטי	ערך	הסבר
ממוצע	0.6	במדד זה קיבלנו שערך החציון נמוך מערך הממוצע ואסימטריה חיובית המעידה על זנב ימני גדול. מכאן ניתן לראות שברוב ההזמנות לא היו בקשות מיוחדות מלבד מספר הזמנות בודדות בהן מספר הבקשות היה גבוה ותצפיות אלו מושכות את הזנב ימינה. ע"פ הטווח הבין רבעוני שהתקבל, 50% מההזמנות שבוצעו בקשו בקשה מיוחדת אחת בלבד או לא בקשו אף בקשה מיוחדת.
אסימטריה	1.172202	
סטיית תקן	0.8424235	
רבעון ראשון	0	
חציון	0	
רבעון שלישי	1	

ניתוח של משתנים קטגוריאליים:

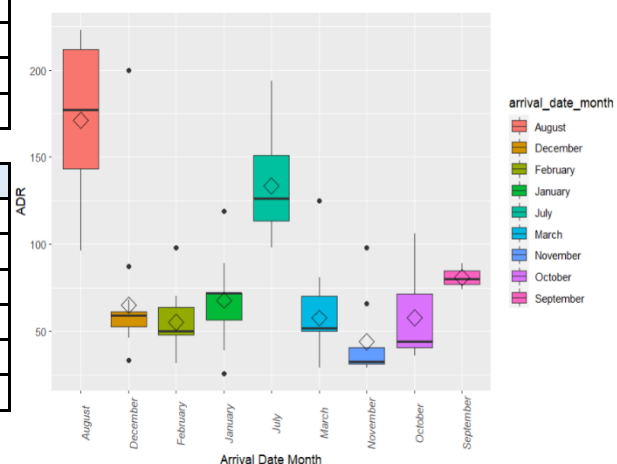
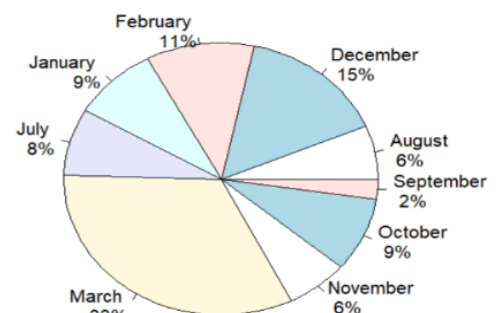
X2 - חודש ההגעה לחופשה במלון:

January	February	March
ממוצע 67.84727	ממוצע 55.185	ממוצע 57.61244
אסימטריה 0.152546	אסימטריה 0.857555	אסימטריה 1.161722
סטיית תקן 25.41069	סטיית תקן 17.139	סטיית תקן 18.70212
רבעון ראשון 56.7	רבעון ראשון 48	רבעון ראשון 50
חציון 72	חציון 50.1	חציון 52
רבעון שלישי 72	רבעון שלישי 64	רבעון שלישי 70

July	August	September
ממוצע 133.419	ממוצע 171.255	ממוצע 81.04667
אסימטריה 0.739018	אסימטריה -0.34695	אסימטריה 0.125512
סטיית תקן 28.29962	סטיית תקן 46.30177	סטיית תקן 7.576393
רבעון ראשון 113.2	רבעון ראשון 143.2	רבעון ראשון 77.04
חציון 126.145	חציון 177.215	חציון 80.08
רבעון שלישי 151	רבעון שלישי 211.8	רבעון שלישי 84.57

October	November	December
ממוצע 57.87091	ממוצע 43.9525	ממוצע 65.13526
אסימטריה 0.681228	אסימטריה 1.242093	אסימטריה 3.108508
סטיית תקן 23.11473	סטיית תקן 25.03062	סטיית תקן 34.34986
רבעון ראשון 40.5	רבעון ראשון 31.36	רבעון ראשון 52.7
חציון 43.93	חציון 32.4	חציון 59.18
רבעון שלישי 71.66	רבעון שלישי 40.8	רבעון שלישי 61.2

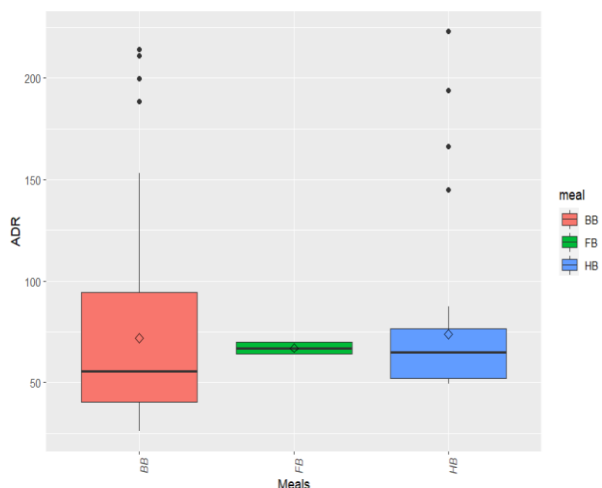
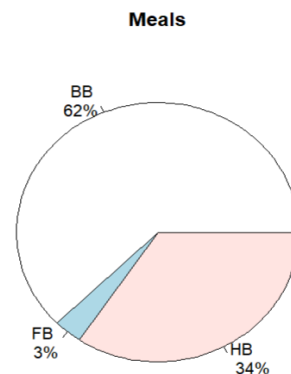
Arrival Date Month



ע"פ תרשים זה ניתן לראות שרוב ההזמנות קורות במהלך תקופת האביב (חודש מרץ) ובתקופת החורף (ינואר עד פברואר) ופחות בעונת הקיץ. ניתן להניח שהתעריף היומי הממוצע המרבי יהיה בתקופות אלו מפני שבחודשים אלו מתבצעות רוב העסקאות במהלך השנה, אך כדי לנתח את המשתנה ה"ל" בצורה ברורה נבחן את הקטגוריות השונות אל מול המשתנה המוסבר. ניתן לראות כי החציון עבור החודשים יולי - אוגוסט גבוה משמעותית משאר החודשים ולכן ניתן להסיק מכך כי אכן למשתנה הקטגורי ה"ל" יש השפעה על המשתנה המוסבר ודבר זה מתחבר לנו עם ההיגיון שבחודשי הקיץ המחירים לחופשות עולות משמעותית ביחס לשאר חודשי השנה.

X6 – סוג פנסיון:

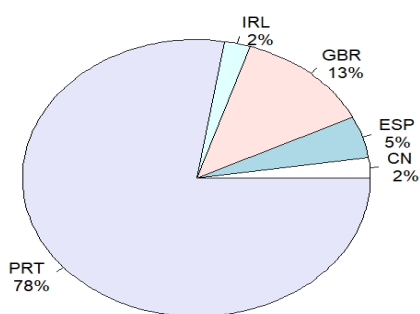
	BB		HB		FB
ממוצע	71.88205	ממוצע	73.9386	ממוצע	67
אסימטריה	1.452294	אסימטריה	2.543399	אסימטריה	0
סטיית תקן	44.94766	סטיית תקן	37.95445	סטיית תקן	3.464102
רבעון ראשון	40.5	רבעון ראשון	52	רבעון ראשון	64
חציון	55.575	חציון	65	חציון	67
רבעון שלישי	94.49	רבעון שלישי	76.5	רבעון שלישי	70



ע"פ תרשים זה ניתן לראות שברוב ההזמנות הלקוחות בחרו בסוג BB, כלומר לינה + ארוחת בוקר. ניתן להניח שיש קשר בין גודל ערך התעריף היומי הממוצע לעסקאות שבוצעו עבור סוג פנסיון BB מאשר סוגי פנסיון אחרים המוצעים ללקוח. נבדוק את הנחה זו ע"י בחינת הקטגוריות השונות אל מול המשתנה המוסבר, ונראה שאין פער גדול בין ערכי החציון של הקטגוריות ולכן לא ניכרת השפעה מדגמית משמעותית.

X7 - יעד החופשה:

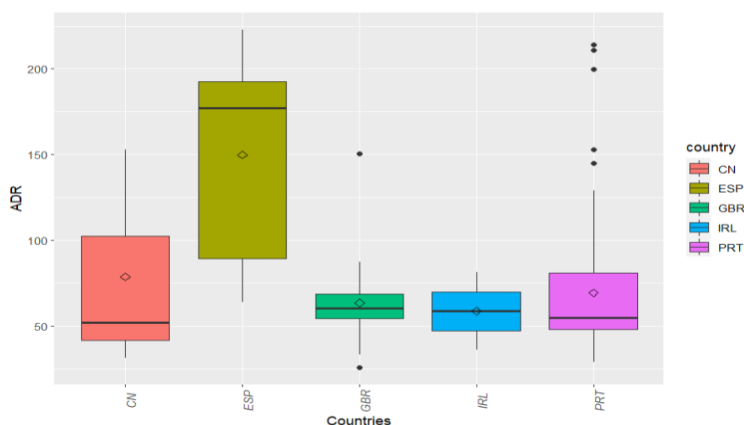
Countries



	ESP
ממוצע	149.905
אסימטריה	-0.37126
סטיית תקן	68.96893
רבעון ראשון	89.53
חציון	177.215
רבעון שלישי	192.57

	CN
ממוצע	78.81667
אסימטריה	0.342152
סטיית תקן	65.06113
רבעון ראשון	41.73
חציון	52
רבעון שלישי	102.5

	PRT
ממוצע	69.33227
אסימטריה	1.857565
סטיית תקן	37.18854
רבעון ראשון	48
חציון	55
רבעון שלישי	81



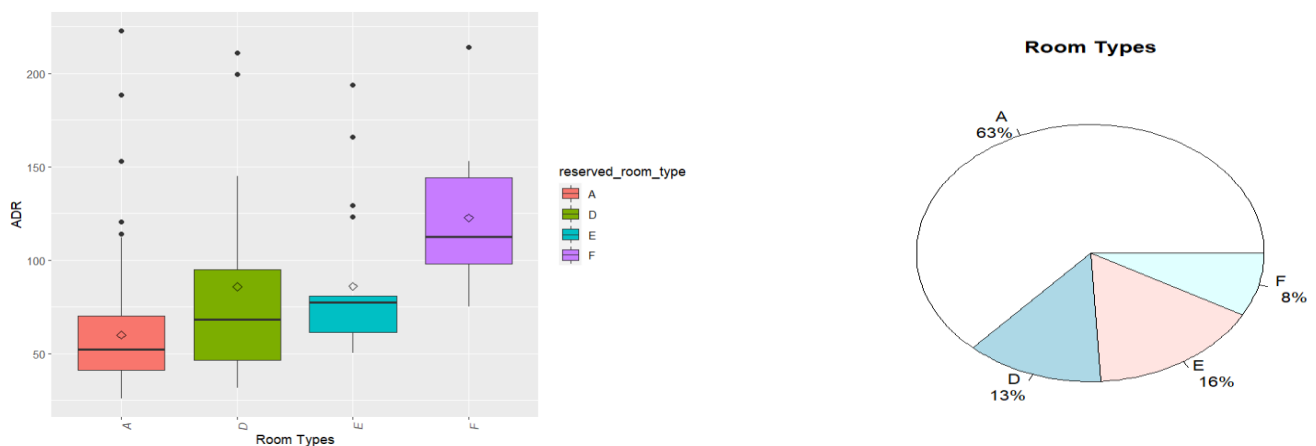
	GBR
ממוצע	63.56313
אסימטריה	1.74284
סטיית תקן	27.57261
רבעון ראשון	54.34
חציון	60.19
רבעון שלישי	68.63

	IRL
ממוצע	58.68
אסימטריה	-0.00088
סטיית תקן	22.67001
רבעון ראשון	47.35
חציון	58.7
רבעון שלישי	70.02

ע"פ תרשים זה ניתן לראות בבירור שרוב ההזמנות מבוצעות עבור חופשות בפורטוגל (PRT). מפני שרוב הזמנות החופשות קורות ביעד זה, ניתן להניח כי התעריף היומי הממוצע יהיה גבוה ביחס לשאר המדינות בשל כמות גדולה יותר של עסקאות ביחס ליעדים אחרים. נבדוק הנחה זו ע"י בחינת הקטגוריות השונות אל מול המשתנה המוסבר.

ע"פ התרשים השמאלי ניתן לראות שיש פערים קטנים בין המדינות בערך החציון, בניגוד לספרד שמציגה ערך חציון גבוה מאוד המעיד על ההשפעה חזקה על התעריף היומי הממוצע בהשוואה לשאר היעדים על אף שרק 5% מההזמנות הן אל יעד זה. מתיישב עם ההיגיון שספרד מדינה יקרה ולכן התעריף היומי הממוצע עשוי להיות גבוה יותר. נוכל להסיק ממצאים אלו שיש השפעה בין קטגוריות המדינה אל המשתנה המוסבר.

X8 – סוג החדר:

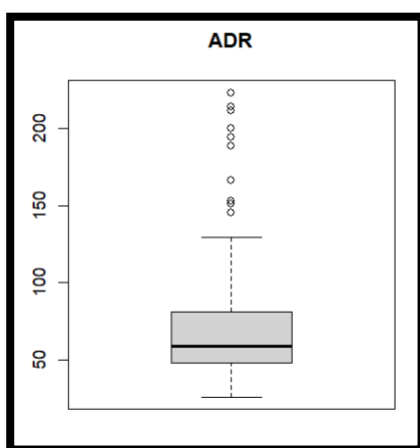


D		A		E		F	
85.72938	ממוצע	59.95367	ממוצע	85.9955	ממוצע	122.624	ממוצע
1.108606	אסימטריה	2.728464	אסימטריה	1.583962	אסימטריה	0.915814	אסימטריה
55.75491	סטיית תקן	32.64584	סטיית תקן	38.05279	סטיית תקן	40.91801	סטיית תקן
46.4	רבעון ראשון	41.25	רבעון ראשון	61.2	רבעון ראשון	98	רבעון ראשון
68.39	חציון	52	חציון	77.19	חציון	112.5	חציון
95.05	רבעון שלישי	69.97	רבעון שלישי	81	רבעון שלישי	144.3	רבעון שלישי

ע"פ תרשים זה, ניתן לראות שחדר מסוג A הוא הנמכר ביותר ומשקלו כ 68% מכלל ההזמנות. כדי לנתח את המשתנה הנ"ל נבחן את הקטגוריות השונות אל מול המשתנה המוסבר - תעריף יומי ממוצע. ניתן לראות כי החציון עבור קטגוריית חדר מסוג F גבוה משמעותית משאר הקטגוריות ולכן ניתן להסיק מכך שלסוגי חדרים שונים יש תמחור שונה ולכן סביר שלמשתנה הקטגוריאלי הנ"ל יש השפעה על המשתנה המוסבר.

6. ניתוח חריגים:

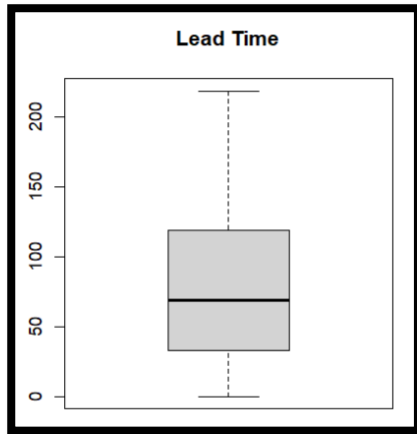
Y - תעריף יומי ממוצע:



במדד זה התקבלו 11 תצפיות חריגות. החריגות שהתקבלו בדומה לניתוח התיאורי בסעיף הקודם הן רחוקות מהחציון ומושכות את הממוצע כלפי מעלה, בנוסף ניתן להסיק שההתפלגות היא אסימטרית עם סטייה חיובית, כלומר זנב ימני. למרות זאת, החלטנו לא להוציא החריגות הללו מהנתונים מכיוון שהן מעידות על תוצאות המשתנה המוסבר, כלומר על תוצאות המחקר ומחיקתן עלולה להשפיע על מהימנות המחקר.

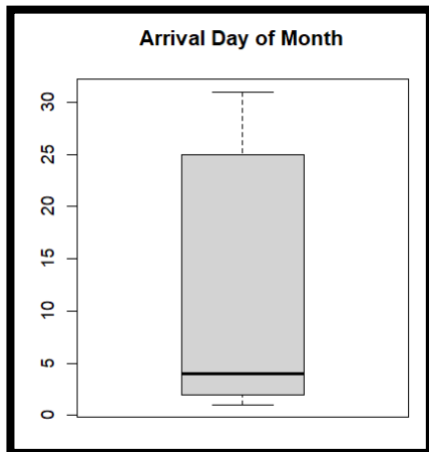
X1 - מספר הימים מתאריך ההזמנה לתאריך ההגעה:

אין תצפיות חריגות עבור המשתנה הנ"ל.



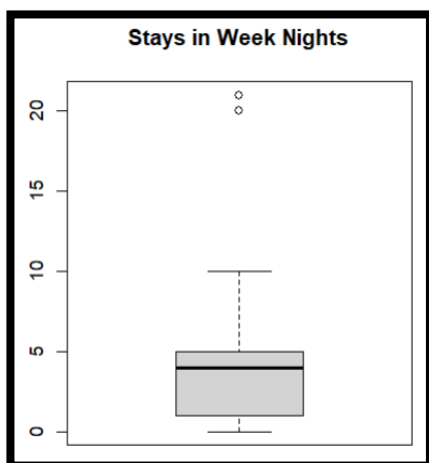
3X - מספר היום בחודש:

אין תצפיות חריגות עבור המשתנה הנ"ל.



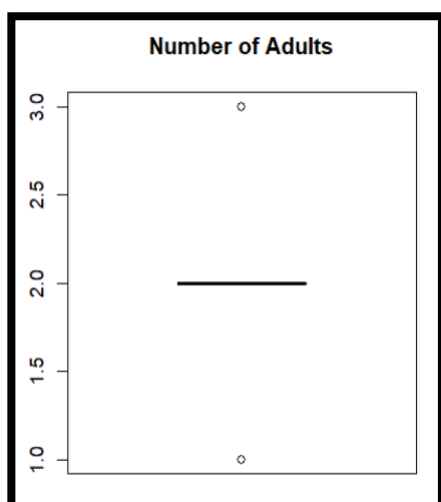
4X - מספר לילות השבוע בהזמנה:

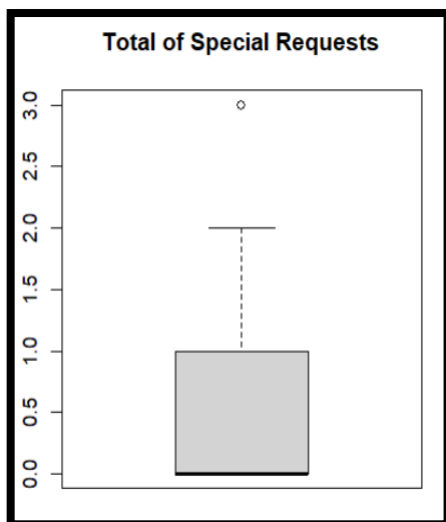
בתרשים זה יש 3 תצפיות חריגות, עבור שני הזמנות חופשה ל 20 לילות ואחת ל 21 לילות, הנובעות מערכים גדולים מאוד ביחס לחציון. החלטנו לנפה את חריגות אלו מכיון שהזמנות כאלו לא שכיחות בתחום החופשות במלון ביעדים אלו. ניתן להניח כי זהו אירוע נקודתי אשר הלקוח מתכנן להישאר לתקופה ארוכה יחסית ביעד החופשה עקב סיבות אישיות מיוחדות\עבודה. # תרשים ה plot המעודכן לאחר ניקוי חריגים נמצא בנספחים.



5X - מספר מבוגרים בחדר:

בתרשים זה יש מספר תצפיות חריגות, 2 תצפיות חריגות עבור הערך 3 ו- 11 תצפיות חריגות עבור הערך 1. חריגות אלו נוצרו מכיוון שערך החציון והממוצע מתלכדים יחד עם האחוזון ה 25 ואחוזון 75, כלומר ש-50% מהערכים במדגם זה נמצאים בקו עם החציון. ניתן לומר על ההתפלגות שהיא אסימטרית שלילית מכיוון שרוב התצפיות הן לכיוון השלילי ומושכות למטה. נבחר שלא להוריד את החריגים במקרה זה מכיון שערכים חריגים אלו אכן מתיישבים עם ההנחה הסבירה שתיתכן הזמנה של בן אדם בודד או קבוצה של שלושה מבוגרים בחדר וההנחה שתמיד יוזמנו חדרים ל 2 מגורים אינה בהכרח נכונה.



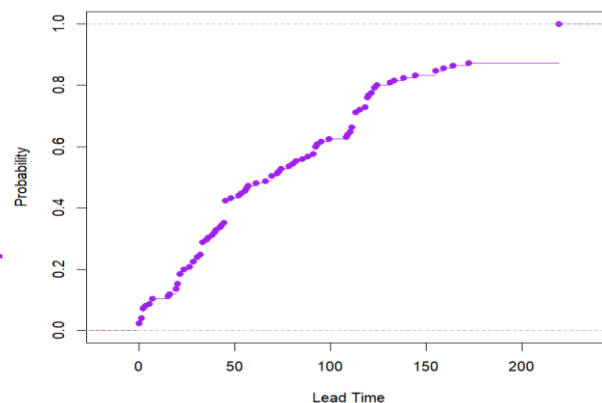
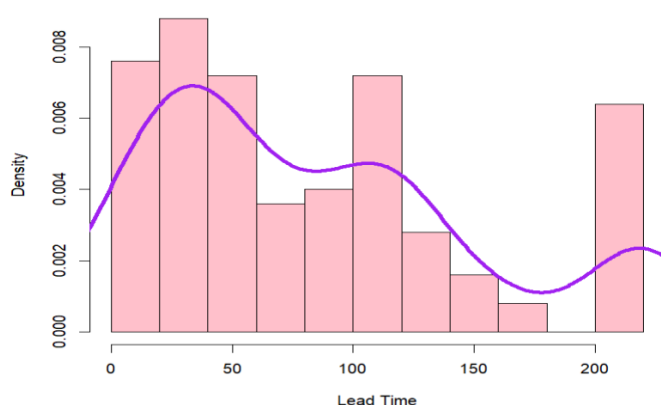


9X - מספר בקשות:

בתרשים זה קיבלנו 4 תצפיות חריגות עם ערך 3, תצפיות אלו גדולות מאוד מהחציון (0 בקשות מיוחדות) ואף מהאחוזון ה-75 של טווח הערכים (בקשה מיוחדת אחת). נבחר להשאיר את חריגים אלו מפני שהערכים אכן מתיישבים עם ההנחה הסבירה שלקוח יבקש יותר מבקשה אחת מיוחדת בהזמנה ולא לפגוע במהימנות המחקר. בקהלים מסוימים ייתכנו יותר בקשות מיוחדות מתוצאות מדגם זה ונרצה להתייחס גם אליהן. (בהנחה שמדובר בקהל ישראלי, הזמנה בעלת 3 בקשות מיוחדות אינה תהיה חריגה בהכרח)

7. פונקציית צפיפות והתפלגות מצטברת:

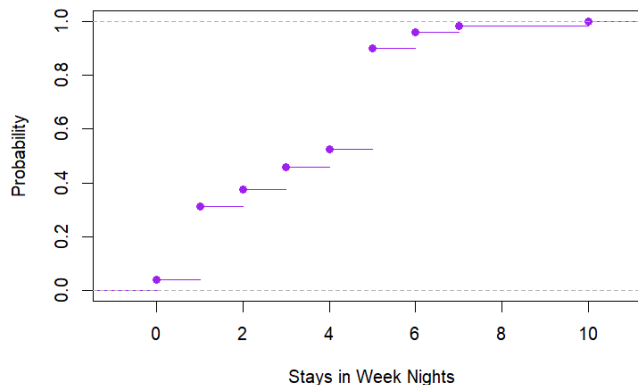
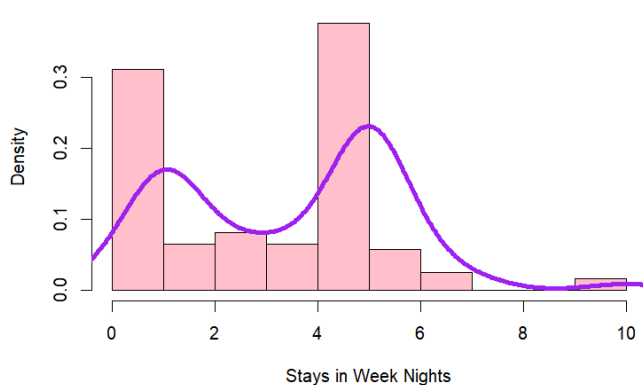
1X - מספר הימים מתאריך ההזמנה לתאריך ההגעה:



פונקציית הצפיפות – ניתן לראות כי ההתפלגות הינה אסימטרית בעלת זנב ימני קטן והשיא שלה נמצא בערך ב-30 ימים כך שמרבית ההזמנות במלון מתבצעות בערך 30 ימים מראש. ניתן לראות שרוב ההזמנות התבצעו בין 0 ל-125 ימים, כלומר אם ניקח הזמנה אקראית רוב הסיכויים שהיא הוזמנה בטווח של 0-125 ימים מראש. ערכים גדולים מ-125 ימים נדירים יותר אך עבור 200 ומעלה ימים יש קפיצה והסיכוי שבדגימה נקבל הזמנה שהוזמנה יותר מ-200 ימים מראש הוא לא מאוד קטן.

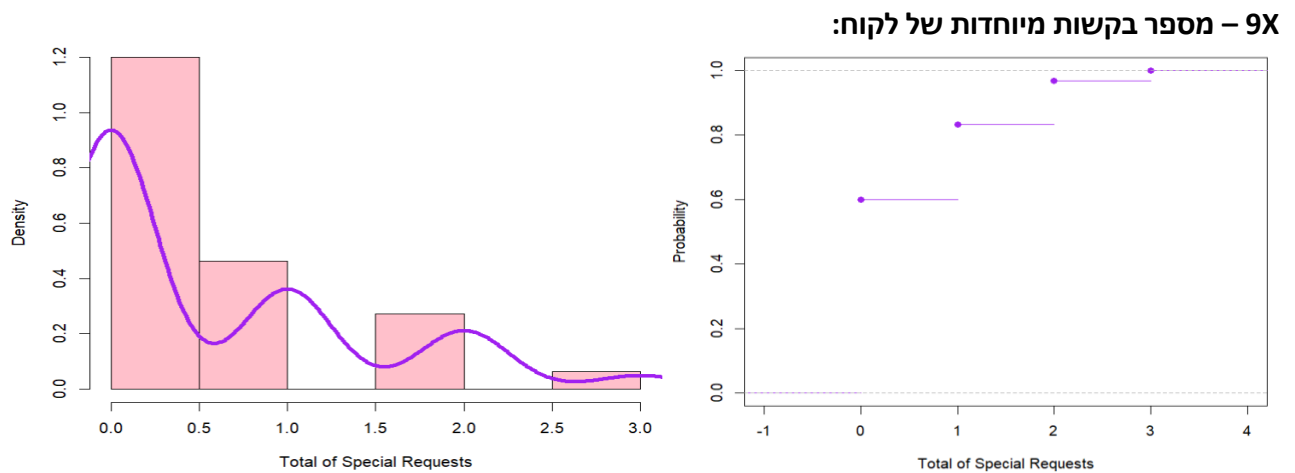
פונקציית ההתפלגות המצטברת - ניתן לראות כי כ-80% מהתצפיות נמצאות בתחום של 0 – 125 ימים מראש כפי שראינו בפונקציית הצפיפות. פחות מ-10% נמצאות בתחום של 125-200 ימים ובערך 15% מהתצפיות בתחום שגדול מ-200 ימים מה שמסקף את הקפיצה שראינו בפונקציית הצפיפות.

4X - מספר לילות השבוע בהזמנה (שני עד שישי):



פונקציית הצפיפות – ניתן לראות כי ההתפלגות הינה אסימטרית בעלת זנב ימני. ריכוז הצפיפות הוא בין הערכים 1 עד 5 לילות השבוע בהזמנה, כלומר אם ניקח הזמנה אקראית רוב הסיכויים שמספר לילות השבוע (שני עד שישי) יהיה בטווח של 1-5 לילות ומצב שבו לקוח הזמין למעלה מ-6 לילות הוא מאורע נדיר. בנוסף ניתן לראות שהשיא נמצא בערך 5 וישנה עליה נוספת נמוכה יותר בערך 1.

פונקציית ההתפלגות המצטברת – ניתן לראות שבערך 30% מזמינים לילה אחד בשבוע, כ-40% מזמינים 5 לילות בשבוע, ורק 10% מזמינים יותר מ-5 לילות. נתונים אלה מחזקים את הנתונים שקיבלנו בפונקציית הצפיפות.

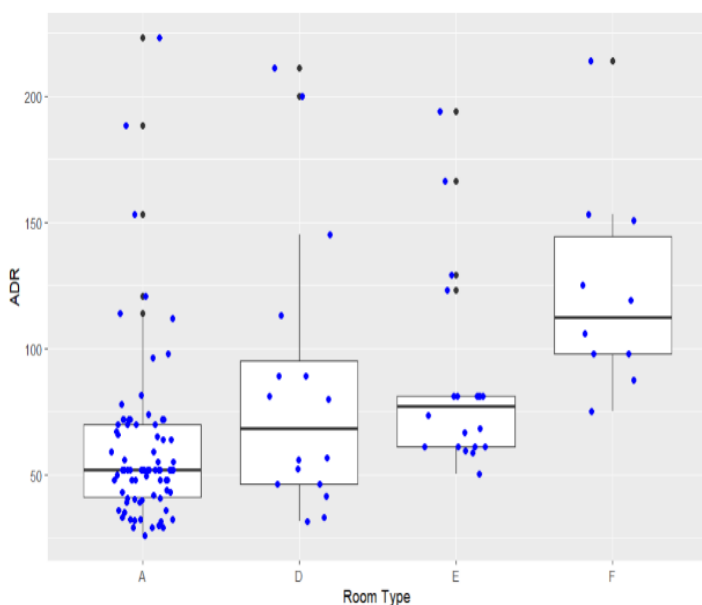


פונקציית הצפיפות – ניתן לראות כי ההתפלגות הינה אסימטרית עם זנב ימני ארוך. השיא שהתקבל הוא 0 בקשות מיוחדות כלומר רוב האורחים במלון לא מבקשים בקשות מיוחדות מעבר להזמנתם. ריכוז הצפיפות הוא בין 0-1 בקשות מיוחדות ומקרים בהם יש יותר מ-2 בקשות מיוחדות הם נדירים יותר.

פונקציית ההתפלגות המצטברת – ניתן לראות כי 60% מהלקוחות לא ביקשו אף בקשה מיוחדת מהמלון כפי שראינו בפונקציית הצפיפות, 22% ביקשו בקשה אחת, 25% בערך ביקשו 2 בקשות מיוחדות ואחוזים בודדים בלבד ביקשו יותר מ-2 בקשות מיוחדות.

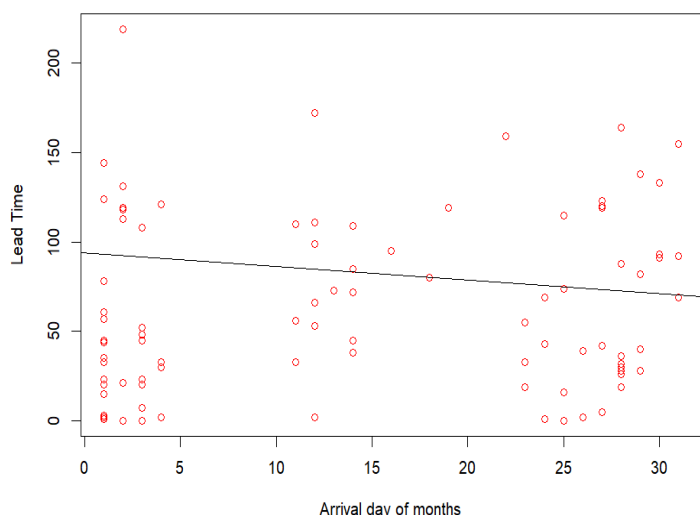
8. ייצוג קשרים בעזרת תרשימים:

1. קשר בין סוג החדר (8X) לתעריף היומי הממוצע (Y):



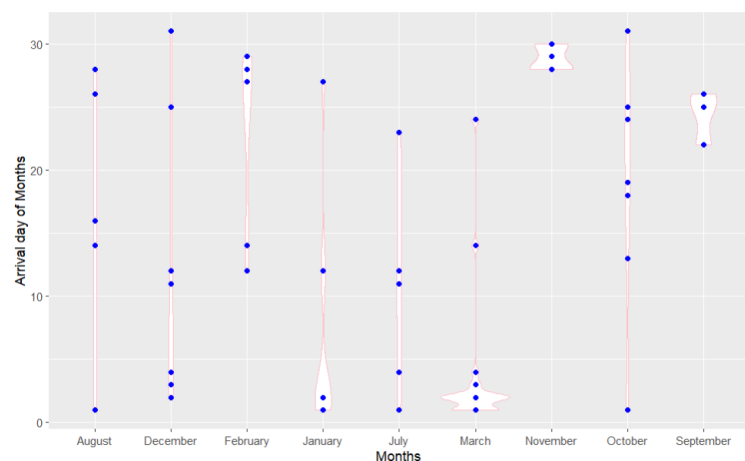
בתרשים זה ציר X מציג את קוד סוגי החדרים, ציר Y מציג את התעריף היומי הממוצע. ניתן לראות בתרשים קשר חזק בין סוג החדר לתעריף היומי הממוצע. עבור חדר מסוג A ממוצע התעריף היומי הממוצע נמוך בהשוואה לסוגי חדרים אחרים ובעל פיזור מצומצם יחסית מכאן ניתן להסיק שברוב ההזמנות שבהם לקוחות הזמינו חדרים מסוג A, התעריף היומי הממוצע היה נמוך. מנגד עבור חדר מסוג F ממוצע התעריף היומי הממוצע היה גבוה בהרבה משאר סוגי החדרים ובעל זנב שמאלי זניח כך שברוב ההזמנות שבהם לקוחות הזמינו חדרים מסוג F התעריף היומי הממוצע גבוה יותר בהשוואה לסוגי חדרים אחרים. חדרים מסוג E ו-D באופן יחסי דומים בנתוני הממוצע התעריף היומי הממוצע שלהם ועבורם הממוצע גבוה יותר מחדרים מסוג A ונמוך מחדרים מסוג F. ממצאים אלה לא מפתיעים אותנו כיוון שסביר שלסוגי חדרים שונים יהיה תעריף שונה והשפעתם על סכום העסקה משפיעה באופן ישיר על חישוב התעריף היומי הממוצע.

2. הקשר בין ימי הגעה בחודש מסוים (X3) לפער בין תאריך הזמנת החופשה לתאריך ההגעה (x1):



בתרשים זה ציר X מציג את יום ההגעה בחודש הרלוונטי וציר Y מציג את מספר הימים מתאריך ההזמנה לתאריך ההגעה. ציפנו לראות קשר בין המשתנים ולמרות זאת התרשים לא מראה קשר מהותי ביניהם. ניתן לראות ריכוזיות הזמנות בתחילת החודש, באמצע ובסוף החודש אך השינוי בציר Y זניח ולא משקף השפעה של המשתנים אחד על השני. בהסתכלות חוזרת ניתן להבין שיתכן שהסיבה נעוצה בחודש הספציפי שאליו מוזמנת החופשה ויש צורך לבצע בדיקה בכל חודש בנפרד כדי לבחון אם קיים קשר. יתכנו חודשים בהם הביקוש לימים מסוימים בחודש גבוה יותר עקב חופשות/חגים/תיירות גבוהה יותר ואנשים יעדיפו לבצע הזמנות הרבה זמן מראש כדי להבטיח את מקומם במלון. מנגד יהיו חודשים שהביקוש יהיה גבוה יותר עבור חופשה בתחילת החודש ולכן אנו רואים בתרשים פיזור לכל אורך החודש.

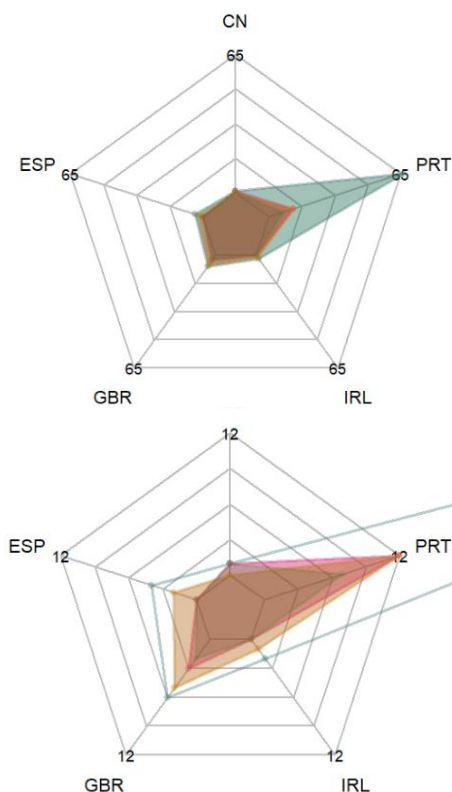
3. הקשר בין חודשי ההגעה (X2) ליום ההגעה באותו החודש (X3):



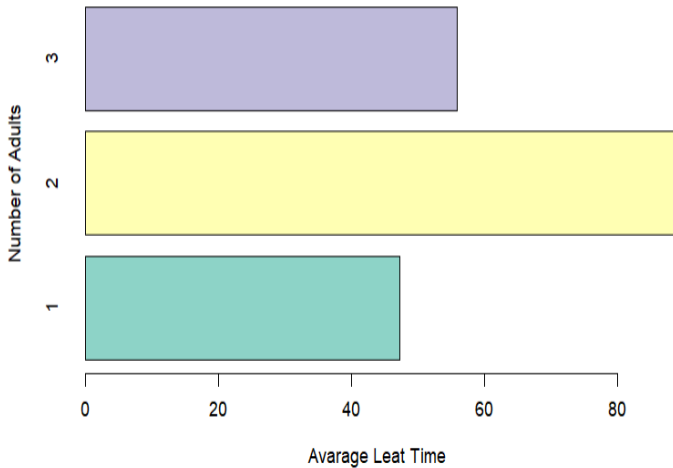
בתרשים זה ציר X מציג חודשי השנה בהם בוצעו הזמנות, ציר Y מציג יום ההגעה לחופשה בחודש. תרשים זה עוזר לנו לבחון את הקשר בין החודש בשנה ליום ההגעה לחופשה באותו החודש. ציפנו לראות קשר מובהק בין שני משתנים אלו אך ניתן לראות שמלבד לחודשים מרץ, ספטמבר ונובמבר לא מתקיים קשר משמעותי. ניתן לראות שעבור החודשים נובמבר וספטמבר יש ריכוזיות גבוהה יותר של הזמנות חופשה לקראת סוף החודש המסמן על תקופה תיירותית יותר ייתכן בשל חופשות/חגים. בחודש מרץ ניכר שרוב ההזמנות מבוצעות לתחילת החודש, דבר זה מסמן בצורה דומה על תקופה תיירותית יותר בחודש. בשאר החודשים ניתן לראות שיש פיזור רחב יותר לאורך כל החודש ללא ריכוז מובהק המראה על קשר משמעותי.

4. קשר בין מדינת החופשה (7X) לבין סוג החדר (8X):

ניתן לראות בכל המדינות כי ישנה עדיפות להזמנת חדר מסוג A על פני שאר סוגי החדרים וסביר להניח שמחרים נמוך בהשוואה לשאר החדרים כך שבכל המדינות זאת האפשרות הכלכלית הפשוטה ביותר שתהיה נוחה לרוב האנשים ללא קשר במדינה עצמה. עבור חדר מסוג F נראה שהביטוי שלו בגרף נבלע בין הגרפים של שאר החדרים ולכן ניתן להסיק שיש ביקוש נמוך מאוד לחדר מסוג זה בכל המדינות, יתכן עקב תמחור גבוה יחסית של החדר/תרבות תיירותית זולה יותר ולא נצפה קשר מובהק לסוג המדינה. ציפנו לראות בתרשים הבדלים בסוגי החדרים בין המדינות, מדינות עם כלכלה חזקה יותר יעדיפו חדרים המתומחרים גבוה יותר ולהפך אך ניתן לראות שבכל המדינות החדר הכי שכיח הוא A והחדר הכי פחות שכיח הוא F, בין חדרים מסוג E ו D ההבדל מזערי ומכאן שהקשר בין המשתנים חלש מאוד/ לא קיים בניגוד להנחה שלנו.



5. קשר מספר המבוגרים בחדר (5X) לבין ממוצע הפער בין תאריך ההזמנה לתאריך הגעה (1X):



בתרשים זה ציר X מציג את ממוצע הימים שחלפו מתאריך ההזמנה לתאריך ההגעה לחופשה וציר Y מציג את מספר המבוגרים בחדר. ניתן לראות שממוצע פער הימים בין התאריכים עבור מבוגר יחיד בחדר הוא הנמוך ביותר, סביר שבן אדם יחיד יהיה יותר ספונטני מאשר מספר אנשים ולכן יבצע הזמנה קרוב יותר למועד החופשה. בחדרים שבהם יש שני מבוגרים אנחנו רואים בגרף שממוצע פער הימים הוא הגבוה ביותר באופן ניכר. ניתן להניח שלרוב מדובר בזוגות אשר נוטים לתכנן את החופשות שלהם זמן רב לפני החופשה. חדרים המכילים שלושה מבוגרים נמצאים באמצע המדד בין מבוגר אחד לשני מבוגרים. ניתן להניח ששלושה מבוגרים יהיו פחות ספונטניים ממבוגר יחיד ויתכננו את חופשתם מוקדם יותר אך יותר ספונטניים מזוגות שעל פי הגרף אנו רואים שהם נוטים לתכנן את חופשתם זמן רב מראש. אנחנו רואים קשר בין המשתנים כאשר מדובר על מבוגר אחד בחדר ועל שני מבוגרים אך עבור שלושה מבוגרים הקשר פחות מובהק.

9. טבלאות שכיחות:

טבלת שכיחות חד ממדית של טווחי התעריף היומי הממוצע:

תחום	0 - 45	45 - 90	90 - 135	135 - 180	180 - 225	סה"כ
שכיחות מוחלטת	27	74	13	5	6	125
שכיחות יחסית	0.22	0.59	0.1	0.04	0.05	1

טבלה זו בוחנת את שכיחות טווחי התעריף היומי הממוצע במדגם במרווחים של 45 דולר. מהנתונים שהתקבלו ניתן לראות שבכמעט 60% מההזמנות התעריף היומי הממוצע היה בטווח של 45-90 דולר, ב-22% מההזמנות התעריף היה נמוך מ-45 ומכאן שב-82% התעריף היומי הממוצע לא עלה על 90 דולר. עבור ערכים גדולים מ-135 דולר, האחוזים אפסיים ומעידים על אסימטריה בפונקציית הצפיפות של המשתנה ועל קיומו של זנב ימני.

טבלת שכיחות חד ממדית של טווחי הלילות בהזמנות:

תחום	0 - 2	2 - 4	4 - 6	6 - 8	8 - 10	סה"כ
שכיחות מוחלטת	41	18	53	3	2	125
שכיחות יחסית	0.34	0.15	0.44	0.03	0.02	1

טבלה זו בוחנת את שכיחות טווחי הלילות השבוע במדגם במרווחים של 2 לילות. מהנתונים שהתקבלו ניתן לראות ש-93% מההזמנות מכילות פחות מ-6 לילות (שני עד שישי). נתונים אלו מתיישבים עם הניתוחים שביצענו על משתנה זה בסעיפים קודמים. האחוזים עבור ערכים של מעל 6 לילות הם אפסיים ומעידים על אסימטריה בפונקציית הצפיפות של המשתנה ועל קיומו של זנב ימני.

טבלת שכיחות דו ממדית של ממוצע התעריף היומי בהשוואה לחודש ההזמנה:

תחום	0 - 25	25 - 50	50 - 75	75 - 100	100 - 125	125 - 150	150 - 175	175 - 200	200 - 225	סה"כ
August	0	0	0	1	1	0	2	1	3	8
December	0	4	13	1	0	0	0	1	0	19
February	0	7	6	1	0	0	0	0	0	14
January	0	3	6	1	1	0	0	0	0	11
July	0	0	0	1	4	2	2	1	0	10
March	0	11	21	8	1	0	0	0	0	41
November	0	6	1	1	0	0	0	0	0	8
October	0	6	3	1	1	0	0	0	0	11
September	0	0	1	2	0	0	0	0	0	3
סה"כ	0	37	51	17	8	2	4	3	3	125

תחום	0 - 25	25 - 50	50 - 75	75 - 100	100 - 125	125 - 150	150 - 175	175 - 200	200 - 225	סה"כ
August	0	0	0	0.008	0.008	0	0.016	0.008	0.024	0.064
December	0	0.032	0.104	0.008	0	0	0	0.008	0	0.152
February	0	0.056	0.048	0.008	0	0	0	0	0	0.112
January	0	0.024	0.048	0.008	0.008	0	0	0	0	0.088
July	0	0	0	0.008	0.032	0.016	0.016	0.008	0	0.08
March	0	0.088	0.168	0.064	0.008	0	0	0	0	0.328
November	0	0.048	0.008	0.008	0	0	0	0	0	0.064
October	0	0.048	0.024	0.008	0.008	0	0	0	0	0.088
September	0	0	0.008	0.016	0	0	0	0	0	0.024
סה"כ	0	0.296	0.408	0.136	0.064	0.016	0.032	0.024	0.024	1

טבלת שכיחות זו בודקת את הקשר בין התעריף היומי הממוצע בטווחים של 25 דולר לבין חודש ההזמנה. מהנתונים עולה כי בחודש יולי טווח התעריף היומי הממוצע הוא בינוני- גבוה ונע בין 75 ל-200 דולר. בחודש אוגוסט רוב ההזמנות נעשות בתעריף יומי ממוצע גבוה מאוד 150-225 דולר. בשאר חודשי השנה התעריף היומי הממוצע לא עולה על 100 דולר ולא יורד מ-25 דולר. מכאן שבחודשי יולי מתחילה עלייה בממוצע התעריף היומי שמתמשכת עד אוגוסט ולאחר מכן בחודש ספטמבר יורד ממוצע התעריף היומי. בחודש מרץ יש הכי הרבה הזמנות והתעריף היומי הממוצע נמוך יחסית 25-100 דולר.

טבלת שכיחויות של ממוצע התעריף היומי, בהשוואה להזמנות במדינות שונות:

תחום	0 - 25	25 - 50	50 - 75	75 - 100	100 - 125	125 - 150	150 - 175	175 - 200	200 - 225	סה"כ
CN	0	1	1	0	0	0	1	0	0	3
ESP	0	0	2	0	0	0	1	2	1	6
GBR	0	4	10	1	0	0	1	0	0	16
IRL	0	1	1	1	0	0	0	0	0	3
PRT	0	31	37	15	8	2	1	1	2	97
סה"כ	0	37	51	17	8	2	4	3	3	125

טבלת שכיחות זו מראה את הקשר בין המדינה בה מתבצעת החופשה (המשתנה המסביר) לבין התעריף היומי הממוצע

תחום	0 - 25	25 - 50	50 - 75	75 - 100	100 - 125	125 - 150	150 - 175	175 - 200	200 - 225	סה"כ
CN	0	0.008	0.008	0	0	0	0.008	0	0	0.024
ESP	0	0	0.016	0	0	0	0.008	0.016	0.008	0.048
GBR	0	0.032	0.08	0.008	0	0	0.008	0	0	0.128
IRL	0	0.008	0.01	0.008	0	0	0	0	0	0.026
PRT	0	0.248	0.296	0.12	0.064	0.016	0.008	0.008	0.016	0.776
סה"כ	0	0.296	0.41	0.136	0.064	0.016	0.032	0.024	0.024	1

(המשתנה המוסבר) בטווחים של 25 דולר. מהנתונים עולה כי במדינת פורטוגל מתבצעות 77.6% מההזמנות ובעיקר בתעריף יומי ממוצע נמוך (25 – 100) מנגד בספרד מתבצעות פחות מ-5% מההזמנות ולכן טווח התעריף היומי הממוצע גבוה מאוד (150 – 225). עבור שאר המדינות השכיחות לתעריפים נמוכים גדולה יותר מאשר השכיחות לתעריפים גבוהים ונע בטווחים הנמוכים מ-100 דולר.

סעיף 4:



סעיף 5:

```
> summary(dataset$adr)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 25.92  48.00   59.18   72.43  81.00  223.00
> sd(dataset$adr, na.rm = TRUE)
[1] 41.76937
> skewness(dataset$adr, na.rm=TRUE)
[1] 1.787302
```

```
> summary(dataset$lead_time)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   33.00   69.00   84.85 119.00  219.00
> sd(dataset$lead_time, na.rm = TRUE)
[1] 67.18872
> skewness(dataset$lead_time, na.rm=TRUE)
[1] 0.7459269
```

```
> summary(dataset$arrival_date_day_of_month)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.00   2.00   4.00   11.92  25.00   31.00
> sd(dataset$arrival_date_day_of_month, na.rm = TRUE)
[1] 11.52669
> skewness(dataset$arrival_date_day_of_month, na.rm=TRUE)
[1] 0.4849669
```

```
> summary(dataset$stays_in_week_nights)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   1.00   4.00   3.88   5.00   21.00
> sd(dataset$stays_in_week_nights, na.rm = TRUE)
[1] 3.349531
> skewness(dataset$stays_in_week_nights, na.rm=TRUE)
[1] 2.753847
```

```
> summary(dataset$adults)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.000   2.000   1.928  2.000   3.000
> sd(dataset$adults,na.rm = TRUE)
[1] 0.3156151
> skewness(dataset$adults,na.rm=TRUE)
[1] -1.59935
```

```
> summary(dataset$total_of_special_requests)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0      0.0      0.0      0.6      1.0      3.0
> sd(dataset$total_of_special_requests,na.rm = TRUE)
[1] 0.8424235
> skewness(dataset$total_of_special_requests,na.rm=TRUE)
[1] 1.172202
```

פלטים למשתנים קטגוריאליים:

2X – חודשים:

```
> skewness(JanuaryTable$adr)
[1] 0.1525458
> summary(JanuaryTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults
Min. : 2.00 Length:11 Min. : 1.000 Min. : 1.000 Min. : 1.000
1st Qu.: 59.50 Class :character 1st Qu.: 2.000 1st Qu.: 1.000 1st Qu.: 2.000
Median :113.00 Mode :character Median : 2.000 Median : 1.000 Median : 2.000
Mean : 88.73 Mean : 5.909 Mean : 3.636 Mean : 1.909
3rd Qu.:113.00 3rd Qu.: 7.000 3rd Qu.: 1.500 3rd Qu.: 2.000
Max. :172.00 Max. :27.000 Max. :20.000 Max. :2.000

  meal country reserved_room_type total_of_special_requests adr
Length:11 Length:11 Length:11 Min. :0.0000 Min. : 25.92
Class :character Class :character Class :character 1st Qu.:0.0000 1st Qu.: 56.70
Mode :character Mode :character Mode :character Median :0.0000 Median : 72.00
Mean :0.2727 Mean : 67.85
3rd Qu.:0.5000 3rd Qu.: 72.00
Max. :1.0000 Max. :119.00

> skewness(FebruaryTable$adr)
[1] 0.8575547
> summary(FebruaryTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal
Min. : 19.00 Length:14 Min. :12.00 Min. :1.000 Min. :1.000 Length:14
1st Qu.: 40.50 Class :character 1st Qu.:17.25 1st Qu.:2.000 1st Qu.:2.000 Class :character
Median :110.00 Mode :character Median :27.00 Median :3.000 Median :2.000 Mode :character
Mean : 84.79 Mean :23.57 Mean :3.786 Mean :1.857
3rd Qu.:119.75 3rd Qu.:28.00 3rd Qu.:6.000 3rd Qu.:2.000
Max. :138.00 Max. :29.00 Max. :7.000 Max. :2.000

  country reserved_room_type total_of_special_requests adr
Length:14 Length:14 Min. :0.0000 Min. :31.59
Class :character Class :character 1st Qu.:0.0000 1st Qu.:48.00
Mode :character Mode :character Median :0.0000 Median :50.10
Mean :0.5714 Mean :55.19
3rd Qu.:1.0000 3rd Qu.:64.00
Max. :3.0000 Max. :98.00

> skewness(MarchTable$adr)
[1] 1.161722
> summary(MarchTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal
Min. : 0.0 Length:41 Min. : 1.000 Min. : 0.000 Min. :1.000 Length:41
1st Qu.: 45.0 Class :character 1st Qu.: 1.000 1st Qu.: 3.000 1st Qu.:2.000 Class :character
Median : 69.0 Mode :character Median : 2.000 Median : 5.000 Median :2.000 Mode :character
Mean :114.6 Mean : 3.512 Mean : 4.683 Mean :1.878
3rd Qu.:219.0 3rd Qu.: 2.000 3rd Qu.: 5.000 3rd Qu.:2.000
Max. :219.0 Max. :24.000 Max. :20.000 Max. :3.000

  country reserved_room_type total_of_special_requests adr
Length:41 Length:41 Min. :0.0000 Min. : 29.00
Class :character Class :character 1st Qu.:0.0000 1st Qu.: 50.00
Mode :character Mode :character Median :0.0000 Median : 52.00
Mean :0.2439 Mean : 57.61
3rd Qu.:0.0000 3rd Qu.: 70.00
Max. :2.0000 Max. :125.00

> skewness(JulyTable$adr)
[1] 0.7390178
> summary(JulyTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 2.0 Length:10 Min. : 1.00 Min. :0.0 Min. :1.0 Length:10 Length:10
1st Qu.: 25.5 Class :character 1st Qu.: 5.75 1st Qu.:1.0 1st Qu.:2.0 Class :character Class :character
Median : 33.0 Mode :character Median :11.50 Median :3.0 Median :2.0 Mode :character Mode :character
Mean : 39.6 Mean :12.10 Mean :2.6 Mean :1.9
3rd Qu.: 48.5 3rd Qu.:20.25 3rd Qu.:4.0 3rd Qu.:2.0
Max. :110.0 Max. :23.00 Max. :5.0 Max. :2.0

  reserved_room_type total_of_special_requests adr
Length:10 Length:10 Min. : 98.0
Class :character 1st Qu.:0.00 1st Qu.:113.2
Mode :character Median :0.00 Median :126.1
Mean :0.40 Mean :133.4
3rd Qu.:0.75 3rd Qu.:151.0
Max. :2.00 Max. :194.0

> skewness(AugustTable$adr)
[1] -0.3469545
> summary(AugustTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 1.00 Length:8 Min. : 1.00 Min. :1.00 Min. :2 Length:8 Length:8
1st Qu.: 2.75 Class :character 1st Qu.: 1.00 1st Qu.:1.75 1st Qu.:2 Class :character Class :character
Median : 29.00 Mode :character Median :15.00 Median :4.50 Median :2 Mode :character Mode :character
Mean : 40.12 Mean :14.38 Mean :3.50 Mean :2
3rd Qu.: 52.25 3rd Qu.:26.50 3rd Qu.:5.00 3rd Qu.:2
Max. :124.00 Max. :28.00 Max. :5.00 Max. :2

  reserved_room_type total_of_special_requests adr
Length:8 Length:8 Min. : 96.3
Class :character 1st Qu.:0.000 1st Qu.:143.2
Mode :character Median :1.000 Median :177.2
Mean :0.875 Mean :171.3
3rd Qu.:1.250 3rd Qu.:211.8
Max. :2.000 Max. :223.0
```

```

> skewness(SeptemberTable$adr)
[1] 0.1255123
> summary(SeptemberTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 39.00 Length:3 Min. :22.00 Min. :1.000 Min. :2 Length:3 Length:3
1st Qu.: 56.50 Class :character 1st Qu.:23.50 1st Qu.:3.000 1st Qu.:2 Class :character Class :character
Median : 74.00 Mode :character Median :25.00 Median :5.000 Median :2 Class :character Mode :character
Mean : 90.67 Mean :24.33 Mean :3.667 Mean :2 Class :character Mode :character
3rd Qu.:116.50 3rd Qu.:25.50 3rd Qu.:5.000 3rd Qu.:2
Max. :159.00 Max. :26.00 Max. :5.000 Max. :2

reserved_room_type total_of_special_requests adr
Length:3 Min. :1.0 Min. :74.00
Class :character 1st Qu.:1.5 1st Qu.:77.04
Mode :character Median :2.0 Median :80.08
Mean :2.0 Mean :81.05
3rd Qu.:2.5 3rd Qu.:84.57
Max. :3.0 Max. :89.06

> skewness(OctoberTable$adr)
[1] 0.6812277
> summary(OctoberTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 1.00 Length:11 Min. : 1.00 Min. :0.000 Min. :2 Length:11 Length:11
1st Qu.: 58.50 Class :character 1st Qu.:15.50 1st Qu.:1.000 1st Qu.:2 Class :character Class :character
Median : 80.00 Mode :character Median :19.00 Median :3.000 Median :2 Class :character Mode :character
Mean : 86.82 Mean :18.82 Mean :2.818 Mean :2 Class :character Mode :character
3rd Qu.:119.00 3rd Qu.:25.00 3rd Qu.:5.000 3rd Qu.:2
Max. :155.00 Max. :31.00 Max. :6.000 Max. :2

reserved_room_type total_of_special_requests adr
Length:11 Min. :0.0000 Min. : 36.00
Class :character 1st Qu.:0.0000 1st Qu.: 40.50
Mode :character Median :1.0000 Median : 43.93
Mean :0.9091 Mean : 57.87
3rd Qu.:1.5000 3rd Qu.: 71.66
Max. :3.0000 Max. :106.00

> skewness(NovemberTable$adr)
[1] 1.242093
> summary(NovemberTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 26.00 Length:8 Min. :28.00 Min. :1.000 Min. :2 Length:8 Length:8
1st Qu.: 70.50 Class :character 1st Qu.:28.00 1st Qu.:3.250 1st Qu.:2 Class :character Class :character
Median : 89.50 Mode :character Median :28.50 Median :4.000 Median :2 Class :character Mode :character
Mean : 89.12 Mean :28.88 Mean :3.875 Mean :2 Class :character Mode :character
3rd Qu.:103.00 3rd Qu.:30.00 3rd Qu.:5.000 3rd Qu.:2
Max. :164.00 Max. :30.00 Max. :7.000 Max. :2

reserved_room_type total_of_special_requests adr
Length:8 Min. :0.00 Min. :29.16
Class :character 1st Qu.:0.00 1st Qu.:31.36
Mode :character Median :1.00 Median :32.40
Mean :0.75 Mean :43.95
3rd Qu.:1.00 3rd Qu.:40.80
Max. :2.00 Max. :98.00

> skewness(DecemberTable$adr)
[1] 3.108508
> summary(DecemberTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 0.00 Length:19 Min. : 2.00 Min. : 1.000 Min. :2 Length:19 Length:19
1st Qu.: 21.00 Class :character 1st Qu.: 2.50 1st Qu.: 1.000 1st Qu.:2 Class :character Class :character
Median : 52.00 Mode :character Median : 3.00 Median : 3.000 Median :2 Class :character Mode :character
Mean : 57.21 Mean :10.74 Mean : 3.842 Mean :2 Class :character Mode :character
3rd Qu.: 92.00 3rd Qu.:18.50 3rd Qu.: 5.000 3rd Qu.:2
Max. :121.00 Max. :31.00 Max. :21.000 Max. :2

reserved_room_type total_of_special_requests adr
Length:19 Min. :0.000 Min. : 33.30
Class :character 1st Qu.:0.000 1st Qu.: 52.70
Mode :character Median :1.000 Median : 59.18
Mean :1.105 Mean : 65.14
3rd Qu.:2.000 3rd Qu.: 61.20
Max. :3.000 Max. :199.73

```

ארוחות – 6X

```

> skewness(BBTable$adr)
[1] 1.452294
> summary(BBTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 0.00 Length:78 Min. : 1.00 Min. : 0.000 Min. :1.00 Length:78 Length:78
1st Qu.: 21.00 Class :character 1st Qu.: 3.00 1st Qu.: 1.000 1st Qu.:2.00 Class :character Class :character
Median : 45.00 Mode :character Median :14.00 Median : 3.000 Median :2.00 Class :character Mode :character
Mean : 59.97 Mean :15.13 Mean : 3.538 Mean :1.91
3rd Qu.: 94.50 3rd Qu.:27.00 3rd Qu.: 5.000 3rd Qu.:2.00
Max. :172.00 Max. :31.00 Max. :20.000 Max. :3.00

reserved_room_type total_of_special_requests adr
Length:78 Min. :0.0000 Min. : 25.92
Class :character 1st Qu.:0.0000 1st Qu.: 40.50
Mode :character Median :1.0000 Median : 55.58
Mean :0.8333 Mean : 71.88
3rd Qu.:1.0000 3rd Qu.: 94.49
Max. :3.0000 Max. :214.00

> skewness(HBTable$adr)
[1] 2.543399
> summary(HBTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 28 Length:43 Min. : 1.000 Min. : 1.000 Min. :1.000 Length:43 Length:43
1st Qu.: 45 Class :character 1st Qu.: 1.500 1st Qu.: 5.000 1st Qu.:2.000 Class :character Class :character
Median :113 Mode :character Median : 2.000 Median : 5.000 Median :2.000 Class :character Mode :character
Mean :127 Mean :5.395 Mean :4.721 Mean :1.953
3rd Qu.:219 3rd Qu.: 2.000 3rd Qu.: 5.000 3rd Qu.:2.000
Max. :219 Max. :31.000 Max. :21.000 Max. :2.000

reserved_room_type total_of_special_requests adr
Length:43 Min. :0.0000 Min. : 49.50
Class :character 1st Qu.:0.0000 1st Qu.: 52.00
Mode :character Median :0.0000 Median : 65.00
Mean :0.2326 Mean : 73.94
3rd Qu.:0.0000 3rd Qu.: 76.50
Max. :2.0000 Max. :223.00

> skewness(FBTable$adr)
[1] 0
> summary(FBTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. :111 Length:4 Min. :12.0 Min. :1.0 Min. :2 Length:4 Length:4
1st Qu.:111 Class :character 1st Qu.:12.0 1st Qu.:1.0 1st Qu.:2 Class :character Class :character
Median :117 Mode :character Median :19.5 Median :1.5 Median :2 Class :character Mode :character
Mean :117 Mean :19.5 Mean :1.5 Mean :2
3rd Qu.:123 3rd Qu.:27.0 3rd Qu.:2.0 3rd Qu.:2
Max. :123 Max. :27.0 Max. :2.0 Max. :2

reserved_room_type total_of_special_requests adr
Length:4 Min. :0 Min. :64
Class :character 1st Qu.:0 1st Qu.:64
Mode :character Median :0 Median :67
Mean :0 Mean :67
3rd Qu.:0 3rd Qu.:70
Max. :0 Max. :70

```

```

> skewness(CNTable$adr)
[1] 0.342152
> summary(CNTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min.   : 53.0   Length:3           Min.   : 2.000           Min.   : 0.000   Min.   : 2   Length:3   Length:3
1st Qu.: 92.0   Class :character       1st Qu.: 2.000           1st Qu.: 2.500   1st Qu.: 2   Class :character Class :character
Median :131.0   Mode  :character       Median : 2.000           Median : 5.000   Median : 2   Mode  :character Mode  :character
Mean   :134.3           Mean : 5.333           Mean : 8.333     Mean : 2
3rd Qu.:175.0           3rd Qu.: 7.000           3rd Qu.:12.500   3rd Qu.: 2
Max.   :219.0           Max.   :12.000           Max.   :20.000   Max.   : 2
reserved_room_type total_of_special_requests adr
Length:3           Min.   :0.0000   Min.   : 31.45
Class :character   1st Qu.:0.0000   1st Qu.: 41.73
Mode  :character   Median :0.0000   Median : 52.00
                    Mean :0.3333   Mean : 78.82
                    3rd Qu.:0.5000   3rd Qu.:102.50
                    Max.   :1.0000   Max.   :153.00

> skewness(ESPTTable$adr)
[1] -0.3712624
> summary(ESPTTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min.   : 3.00   Length:6           Min.   : 1.00           Min.   :1.000   Min.   : 2   Length:6   Length:6
1st Qu.: 29.25   Class :character       1st Qu.:16.25           1st Qu.:1.250   1st Qu.: 2   Class :character Class :character
Median : 35.50   Mode  :character       Median :25.00           Median :2.500   Median : 2   Mode  :character Mode  :character
Mean   : 58.00           Mean :20.00           Mean :2.833     Mean : 2
3rd Qu.:101.75           3rd Qu.:27.00           3rd Qu.:4.500   3rd Qu.: 2
Max.   :123.00           Max.   :28.00           Max.   :5.000   Max.   : 2
reserved_room_type total_of_special_requests adr
Length:6           Min.   :0.0000   Min.   : 64.00
Class :character   1st Qu.:0.0000   1st Qu.: 89.53
Mode  :character   Median :0.5000   Median :177.22
                    Mean :0.6667   Mean :149.91
                    3rd Qu.:1.0000   3rd Qu.:192.57
                    Max.   :2.0000   Max.   :223.00

> skewness(GBRTTable$adr)
[1] 1.74284
> summary(GBRTTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min.   : 21.00   Length:16           Min.   : 1.000           Min.   : 1.000   Min.   : 2   Length:16   Length:16
1st Qu.: 40.50   Class :character       1st Qu.: 2.000           1st Qu.: 4.250   1st Qu.: 2   Class :character Class :character
Median : 59.00   Mode  :character       Median : 3.000           Median : 5.000   Median : 2   Mode  :character Mode  :character
Mean   : 67.62           Mean : 9.312           Mean : 6.062     Mean : 2
3rd Qu.: 92.00           3rd Qu.:12.250           3rd Qu.: 5.000   3rd Qu.: 2
Max.   :172.00           Max.   :31.000           Max.   :21.000   Max.   : 2
reserved_room_type total_of_special_requests adr
Length:16           Min.   :0.0000   Min.   : 25.92
Class :character   1st Qu.:0.0000   1st Qu.: 54.34
Mode  :character   Median :1.0000   Median : 60.19
                    Mean :0.9375   Mean : 63.56
                    3rd Qu.:2.0000   3rd Qu.: 68.63
                    Max.   :2.0000   Max.   :150.71

> skewness(IRLTable$adr)
[1] -0.000822223
> summary(IRLTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min.   : 56.00   Length:3           Min.   :11.0           Min.   : 3.0   Min.   : 2   Length:3   Length:3
1st Qu.: 68.00   Class :character       1st Qu.:14.5           1st Qu.:3.5   1st Qu.: 2   Class :character Class :character
Median : 80.00   Mode  :character       Median :18.0           Median :4.0   Median : 2   Mode  :character Mode  :character
Mean   : 83.67           Mean :18.0           Mean :4.0         Mean : 2
3rd Qu.: 97.50           3rd Qu.:21.5           3rd Qu.:4.5       3rd Qu.: 2
Max.   :115.00           Max.   :25.0           Max.   :5.0       Max.   : 2
reserved_room_type total_of_special_requests adr
Length:3           Min.   :0.0       Min.   :36.00
Class :character   1st Qu.:0.5       1st Qu.:47.35
Mode  :character   Median :1.0       Median :58.70
                    Mean :1.0       Mean :58.68
                    3rd Qu.:1.5       3rd Qu.:70.02
                    Max.   :2.0       Max.   :81.34

> skewness(PRTTable$adr)
[1] 1.857565
> summary(PRTTable)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min.   : 0.00   Length:97           Min.   : 1.00           Min.   : 0.000   Min.   :1.000   Length:97   Length:97
1st Qu.: 32.00   Class :character       1st Qu.: 2.00           1st Qu.: 1.000   1st Qu.:2.000   Class :character Class :character
Median : 72.00   Mode  :character       Median : 4.00           Median : 4.000   Median :2.000   Mode  :character Mode  :character
Mean   : 87.86           Mean :11.87           Mean : 3.443     Mean :1.907
3rd Qu.:120.00           3rd Qu.:25.00           3rd Qu.: 5.000   3rd Qu.:2.000
Max.   :219.00           Max.   :31.00           Max.   :10.000   Max.   :3.000
reserved_room_type total_of_special_requests adr
Length:97           Min.   :0.0000   Min.   : 29.00
Class :character   1st Qu.:0.0000   1st Qu.: 48.00
Mode  :character   Median :0.0000   Median : 55.00
                    Mean :0.5361   Mean : 69.33
                    3rd Qu.:1.0000   3rd Qu.: 81.00
                    Max.   :3.0000   Max.   :214.00

```

```

> skewness(A_Table$adr)
[1] 2.728464
> summary(A_Table)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 0.0 Length:79 Min. : 1.00 Min. : 0.000 Min. : 1.000 Length:79 Length:79
1st Qu.: 39.0 Class :character 1st Qu.: 2.00 1st Qu.: 1.000 1st Qu.: 2.000 Class :character Class :character
Median :111.0 Mode :character Median : 4.00 Median : 5.000 Median : 2.000 Mode :character Mode :character
Mean :105.3 Mean :12.89 Mean : 3.747 Mean : 1.861
3rd Qu.:155.0 3rd Qu.:27.00 3rd Qu.: 5.000 3rd Qu.: 2.000
Max. :219.0 Max. :31.00 Max. :20.000 Max. :2.000

reserved_room_type total_of_special_requests adr
Length:79 Min. :0.0000 Min. : 25.92
Class :character 1st Qu.:0.0000 1st Qu.: 41.25
Mode :character Median :0.0000 Median : 52.00
Mean :0.5443 Mean : 59.95
3rd Qu.:1.0000 3rd Qu.: 69.97
Max. :3.0000 Max. :223.00

> skewness(D_Table$adr)
[1] 1.108606
> summary(D_Table)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 1.00 Length:16 Min. : 1.00 Min. : 1.000 Min. : 2.000 Length:16 Length:16
1st Qu.: 30.50 Class :character 1st Qu.: 2.75 1st Qu.: 1.000 1st Qu.: 2.000 Class :character Class :character
Median : 44.00 Mode :character Median : 3.50 Median : 2.500 Median : 2.000 Mode :character Mode :character
Mean : 52.31 Mean :10.31 Mean : 4.062 Mean : 2.062
3rd Qu.: 70.25 3rd Qu.:16.50 3rd Qu.: 5.000 3rd Qu.: 2.000
Max. :131.00 Max. :31.00 Max. :20.000 Max. :3.000

reserved_room_type total_of_special_requests adr
Length:16 Min. :0 Min. : 31.45
Class :character 1st Qu.:0 1st Qu.: 46.40
Mode :character Median :1 Median : 68.39
Mean :1 Mean : 85.73
3rd Qu.:1 3rd Qu.: 95.05
Max. :3 Max. :211.00

> skewness(E_Table$adr)
[1] 1.583962
> summary(E_Table)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 21.00 Length:20 Min. : 1.0 Min. : 1.00 Min. : 2 Length:20 Length:20
1st Qu.: 26.75 Class :character 1st Qu.: 1.0 1st Qu.: 2.50 1st Qu.: 2 Class :character Class :character
Median : 45.00 Mode :character Median : 2.0 Median : 5.00 Median : 2 Mode :character Mode :character
Mean : 50.85 Mean : 7.6 Mean : 4.65 Mean : 2
3rd Qu.: 60.25 3rd Qu.:11.5 3rd Qu.: 5.00 3rd Qu.: 2
Max. :121.00 Max. :31.0 Max. :21.00 Max. :2

reserved_room_type total_of_special_requests adr
Length:20 Min. :0.0 Min. : 50.40
Class :character 1st Qu.:0.0 1st Qu.: 61.20
Mode :character Median :0.0 Median : 77.19
Mean :0.7 Mean : 86.00
3rd Qu.:2.0 3rd Qu.: 81.00
Max. :2.0 Max. :194.00

> skewness(F_Table$adr)
[1] 0.9158144
> summary(F_Table)
  lead_time arrival_date_month arrival_date_day_of_month stays_in_week_nights adults meal country
Min. : 2.0 Length:10 Min. : 1.00 Min. : 0.0 Min. : 2.0 Length:10 Length:10
1st Qu.:21.5 Class :character 1st Qu.: 6.00 1st Qu.: 2.0 1st Qu.: 2.0 Class :character Class :character
Median :41.0 Mode :character Median :15.00 Median : 3.0 Median : 2.0 Mode :character Mode :character
Mean :43.5 Mean :15.50 Mean : 3.1 Mean : 2.1
3rd Qu.:65.0 3rd Qu.:24.75 3rd Qu.: 5.0 3rd Qu.: 2.0
Max. :95.0 Max. :28.00 Max. :5.0 Max. :3.0

reserved_room_type total_of_special_requests adr
Length:10 Min. :0.0 Min. : 75.0
Class :character 1st Qu.:0.0 1st Qu.: 98.0
Mode :character Median :0.0 Median :112.5
Mean :0.2 Mean :122.6
3rd Qu.:0.0 3rd Qu.:144.3
Max. :1.0 Max. :214.0

```

סעיף 6:

```

# Box Plots for Exception analysis.

# Y
bpADR<-boxplot(dataset$adr, main='ADR')
bpADR$out
length(bpADR$out)

# X1 - None
bpLT<-boxplot(dataset$lead_time, main='Lead Time')

# X3 - None
bpADDM<-boxplot(dataset$arrival_date_day_of_month, main='Arrival Day of Month')

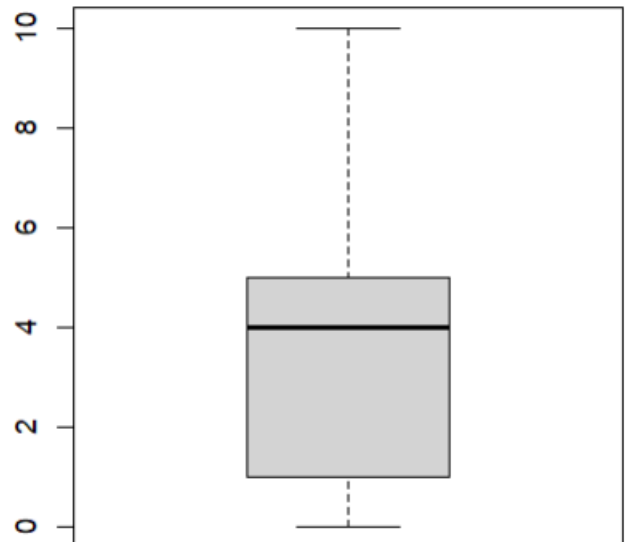
# X4
bpAWN<-boxplot(dataset$stays_in_week_nights, main='Stays in Week Nights')
bpAWN$out
length(bpAWN$out)
# X4 new
bpAWN1<-subset(dataset,dataset$stays_in_week_nights<20)
bpAWN1new<-boxplot(bpAWN1$stays_in_week_nights, main='Stays in Week Nights - new ')

# X5
bpAD<-boxplot(dataset$adults, main='Number of Adults')
bpAD$out
length(bpAD$out)

# X9
bpTSR<-boxplot(dataset$total_of_special_requests, main='Total of Special Requests')
bpTSR$out
length(bpTSR$out)

```

Stays in Week Nights - new



סעיף 7:

```
# X1 - Lead Time
hist(dataset$lead_time,prob=TRUE, main='',ylab = 'Density',xlab = 'Lead Time', col = "pink")
lines(density(dataset$lead_time),col="purple",lwd=4)
plot.ecdf(dataset$lead_time,main = "",col = "purple",ylab = 'Probability', xlab = 'Lead Time')

# X4 - Stays in Week Nights - without irregulars (bpAWNX)
hist(bpAWNX$stays_in_week_nights,prob=TRUE, main='',ylab = 'Density',xlab = 'Stays in Week Nights', col = "pink")
lines(density(bpAWNX$stays_in_week_nights),col="purple",lwd=4)
plot.ecdf(bpAWNX$stays_in_week_nights,main = "",col = "purple",ylab = 'Probability', xlab = 'Stays in Week Nights')

# X9 - Total of Special Requests
hist(dataset$total_of_special_requests,prob=TRUE, main='',ylab = 'Density',xlab = 'Total of Special Requests', col = "pink")
lines(density(dataset$total_of_special_requests),col="purple",lwd=4)
plot.ecdf(dataset$total_of_special_requests,main = "",col = "purple",ylab = 'Probability', xlab = 'Total of Special Requests')
```

סעיף 8:

```
# ADR and Room Type - Y & X8 :
Y_X5_gg <- ggplot(dataset, aes(x = reserved_room_type, y = adr))+
  xlab("Room Type")+ylab("ADR")+
  geom_boxplot()
Y_X5_gg+geom_jitter(shape=16, position=position_jitter(0.2), col = "blue")

# Lead Time and Arrival day of months - X1 & X4 :
plot( x = dataset$arrival_date_day_of_month, y =dataset$lead_time, col='red', pch=1,
  xlab = "Arrival day of months", ylab = "Lead Time")+
  abline(lm(dataset$lead_time ~ dataset$arrival_date_day_of_month , col = "red"))

# Months and Arrival day of months - X2 & X4 :
ggplot(dataset, aes(x = arrival_date_month , y = arrival_date_day_of_month))+
  geom_violin( col = "pink")+
  geom_point( col = "blue")+xlab("Months")+ylab("Arrival day of Months")

# Adults groups and AVG(Lead Time) - X5 & AVG(X1) :
A<-sqldf('SELECT adults, AVG(lead_time) FROM dataset GROUP BY adults')
data <- data.frame(
  name=A$adults,
  value=A$`AVG(lead_time)`[,3])
# Specific color for each bar? Use a well known palette
library(RColorBrewer)
coul <- brewer.pal(5, "Set3")
barplot(height=data$value, names=data$name, col=coul,hORIZ=T,xlab="Avarage Leat Time",
  ylab="Number of Adults", )
```



```

# Countries and Room Types :
library(fmsb)
T_A<-sqldf('SELECT country, Count(reserved_room_type) FROM dataset WHERE reserved_room_type = "A" GROUP BY country')
T_D<-sqldf('SELECT country, Count(reserved_room_type) FROM dataset WHERE reserved_room_type = "D" GROUP BY country')
T_E<-sqldf('SELECT country, Count(reserved_room_type) FROM dataset WHERE reserved_room_type = "E" GROUP BY country')
T_F<-sqldf('SELECT country, Count(reserved_room_type) FROM dataset WHERE reserved_room_type = "F" GROUP BY country')

M<-matrix(nrow=5,ncol=4)
M<-cbind(c(1,4,6,2,66),c(1,0,3,0,12),c(0,2,5,1,12),c(1,0,2,0,7))
M.t<-t(M)
data <- as.data.frame(M.t)
colnames(data) <- unique(dataset$country)
rownames(data) <- unique(dataset$reserved_room_type)
# To use the fmsb package, I have to add 2 lines to the dataframe: the max and min of each variable to show on the plot!
data <- rbind(rep(12,1) , rep(0,5) , data)
# Color vector
colors_border=c(rgb(0.3,0.5,0.5,0.4), rgb(0.9,0.2,0.5,0.4) , rgb(0.8,0.5,0.1,0.4) ,rgb(0.3,0.5,0.1,0.2))
colors_in=c( rgb(0.1,0.4,0.3,0.4), rgb(0.8,0.1,0.4,0.4) , rgb(0.7,0.4,0.1,0.4) ,rgb(0.2,0.4,0.1,0.2))

# plot with default options:
radarchart( data , axistype=2 , pfc=colors_in ,pcol=colors_border,
# custom polygon
plwd=3 , plty=1,
# custom the grid
cglcol="grey", cglty=1, axislabcol="black", caxislabels=seq(0,20,5), cglwd=2,
# custom labels
vlicex=1.2 )
# Add a legend
legend(x=1.3, y=1.2, legend = rownames(data[-c(1,2),]), bty = "n", pch=16 , col=colors_in , text.col = "grey", cex=1.3, pt.cex=3)

```

סעיף 9:

```

> # Y
> cbind(Freq=table(cut(dataset$adr, breaks=seq(0,225,45))),
+       relative=round(prop.table(table(cut(dataset$adr, breaks = seq(0,225,45))))),2))
      Freq relative
(0,45]    27    0.22
(45,90]    74    0.59
(90,135]   13    0.10
(135,180]    5    0.04
(180,225]    6    0.05
>
> # X4 - Stays in Week Nights - without irregulars (bpAWN)
> cbind(Freq=table(cut(bpAWN$stays_in_week_nights, breaks=seq(0,10,2))),
+       relative=round(prop.table(table(cut(bpAWN$stays_in_week_nights, breaks = seq(0,10,2))))),2))
      Freq relative
(0,2]     41    0.35
(2,4]     18    0.15
(4,6]     53    0.45
(6,8]      3    0.03
(8,10]     2    0.02
>
> # Y and X7 - Countries
> cbind(Freq=table(dataset$country, cut(dataset$adr, breaks = seq(0,225,25))))
      (0,25] (25,50] (50,75] (75,100] (100,125] (125,150] (150,175] (175,200] (200,225]
CN         0         1         1         0         0         0         1         0         0
ESP        0         0         2         0         0         0         1         2         1
GBR        0         4        10         1         0         0         1         0         0
IRL        0         1         1         1         0         0         0         0         0
PRT        0        31        37        15         8         2         1         1         2
> cbind(relative=round(prop.table(table(dataset$country, cut(dataset$adr, breaks = seq(0,225,25))))),3))
      (0,25] (25,50] (50,75] (75,100] (100,125] (125,150] (150,175] (175,200] (200,225]
CN         0 0.008 0.008 0.000 0.000 0.000 0.008 0.000 0.000
ESP        0 0.000 0.016 0.000 0.000 0.000 0.008 0.016 0.008
GBR        0 0.032 0.080 0.008 0.000 0.000 0.008 0.000 0.000
IRL        0 0.008 0.008 0.008 0.000 0.000 0.000 0.000 0.000
PRT        0 0.248 0.296 0.120 0.064 0.016 0.008 0.008 0.016
> # Y and X1 - Months
> cbind(Freq=table(dataset$arrival_date_month, cut(dataset$adr, breaks = seq(0,225,25))))
      (0,25] (25,50] (50,75] (75,100] (100,125] (125,150] (150,175] (175,200] (200,225]
August      0         0         0         1         1         0         2         1         3
December    0         4        13         1         0         0         0         1         0
February    0         7         6         1         0         0         0         0         0
January     0         3         6         1         1         0         0         0         0
July        0         0         0         1         4         2         2         1         0
March       0        11        21         8         1         0         0         0         0
November    0         6         1         1         0         0         0         0         0
October     0         6         3         1         1         0         0         0         0
September   0         0         1         2         0         0         0         0         0
> cbind(relative=round(prop.table(table(dataset$arrival_date_month, cut(dataset$adr, breaks = seq(0,225,25))))),3))
      (0,25] (25,50] (50,75] (75,100] (100,125] (125,150] (150,175] (175,200] (200,225]
August      0 0.000 0.000 0.008 0.008 0.000 0.016 0.008 0.024
December    0 0.032 0.104 0.008 0.000 0.000 0.000 0.008 0.000
February    0 0.056 0.048 0.008 0.000 0.000 0.000 0.000 0.000
January     0 0.024 0.048 0.008 0.008 0.000 0.000 0.000 0.000
July        0 0.000 0.000 0.008 0.032 0.016 0.016 0.008 0.000
March       0 0.088 0.168 0.064 0.008 0.000 0.000 0.000 0.000
November    0 0.048 0.008 0.008 0.000 0.000 0.000 0.000 0.000
October     0 0.048 0.024 0.008 0.008 0.000 0.000 0.000 0.000
September   0 0.000 0.008 0.016 0.000 0.000 0.000 0.000 0.000

```