**פרויקט בקורס רגרסיה ליניארית**

**חלק א'**



**תוכן עניינים**

[1. בחירת מאגר נתונים : 3](#_Toc101876709)

[2. טבלת משתנים: 3](#_Toc101876710)

[3. תיאור המשתנים : 4](#_Toc101876711)

[4. תיאור קשרים בין משתנים: 4](#_Toc101876712)

[5. ניתוח תיאורי של המשתנים: 6](#_Toc101876713)

[6. ניתוח חריגים: 10](#_Toc101876714)

[7. פונקציית צפיפות והתפלגות מצטברת: 12](#_Toc101876715)

[9. טבלאות שכיחות: 15](#_Toc101876716)

[נספחים: 17](#_Toc101876717)

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 | BB – Bed & Breakfast  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 תחרויות ספורט וכו'. נצפה בימים או ביקוש גבוהה 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 מתוירות 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 בX2 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.

Chart, waterfall chart

Description automatically generated**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: Calendar

Description automatically generated**  
**ההנחות לגבי קשר סיבתי בין משתנים מסבירים רציפים-**   
**1. קשר בין X1 ל- X4 (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 שהזמינות 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 X1 ל- X5 (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 X1 ל- X9 (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 X5 ל- X9 (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.

1. **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.

1. **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.

1. **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 בציוד 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.

1. **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 החריגות. התחום הבין רבעוני הוא תחום בין הרבעונים שבו נמצאים 50% מהתצפיות.

**Skewness - (אסימטריה):** מדד לסימטריה של ההתפלגות של משתנה מקרי ממשי. כשערכו המוחלט קטן מ0.5 ניתן לומר בקירוב טוב שההתפלגות סימטרית. עבור ערך גדול מ0.5 נקבל התפלגות מוטה חיובית בעלת זנב ימני. עבור ערך קטן מ(0.5-) נקבל התפלגות מוטה שלילית בעלת זנב שמאלי**.**

**ניתוח של משתנים רציפים:**

|  |  |  |
| --- | --- | --- |
| 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.  ניתן לראות שע"פ מדד האסימטריה שהתקבל הינו 1.787302. ערך זה גבוה במיוחד בהתאמה לפער בין הממוצע לחציון. מערך מדד האסימטריה ניתן להסביר שמדובר בהתפלגות מוטה חיובית בעלת זנב ימני. ניתן לראות שערך המקסימום, אשר מייצג את התעריף היומי הממוצע הגבוה ביותר במדגם, גבוה מאוד ביחס לערך הרבעון השלישי. בנוסף מהטווח הבין רבעוני ניתן להסיק ש 50% התעריפים היומיים בן בין 48 ~ 81 דולר. |
| אסימטריה | 1.787302 |
| סטיית תקן | 41.76937 |
| רבעון ראשון | 48 |
| חציון | 59.18 |
| רבעון שלישי | 81 |

**Y - תעריף יומי ממוצע (משתנה מוסבר) :**

|  |  |  |
| --- | --- | --- |
| 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 |

**X1 - מספר הימים מתאריך ההזמנה לתאריך ההגעה:**

**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 מ0.5 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 |

|  |  |  |
| --- | --- | --- |
| 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 |

**X4 - מספר לילות השבוע בהזמנה:**

|  |  |  |
| --- | --- | --- |
| 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 |

**X5 -מספר מבוגרים בחדר:**

|  |  |  |
| --- | --- | --- |
| 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 |

**X9 - מספר בקשות מיוחדות:**

**ניתוח של משתנים קטגוריאליים:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | January |  |  | February |  |  | March |
| 67.84727 | ממוצע |  | 55.185 | ממוצע |  | 57.61244 | ממוצע |
| 0.152546 | אסימטריה |  | 0.857555 | אסימטריה |  | 1.161722 | אסימטריה |
| 25.41069 | סטיית תקן |  | 17.139 | סטיית תקן |  | 18.70212 | סטיית תקן |
| 56.7 | רבעון ראשון |  | 48 | רבעון ראשון |  | 50 | רבעון ראשון |
| 72 | חציון |  | 50.1 | חציון |  | 52 | חציון |
| 72 | רבעון שלישי |  | 64 | רבעון שלישי |  | 70 | רבעון שלישי |
|  |  |  |  |  |  |  |  |
|  | July |  |  | August |  |  | September |
| 133.419 | ממוצע |  | 171.255 | ממוצע |  | 81.04667 | ממוצע |
| 0.739018 | אסימטריה |  | -0.34695 | אסימטריה |  | 0.125512 | אסימטריה |
| 28.29962 | סטיית תקן |  | 46.30177 | סטיית תקן |  | 7.576393 | סטיית תקן |
| 113.2 | רבעון ראשון |  | 143.2 | רבעון ראשון |  | 77.04 | רבעון ראשון |
| 126.145 | חציון |  | 177.215 | חציון |  | 80.08 | חציון |
| 151 | רבעון שלישי |  | 211.8 | רבעון שלישי |  | 84.57 | רבעון שלישי |
|  |  |  |  |  |  |  |  |
|  | October |  |  | November |  |  | December |
| 57.87091 | ממוצע |  | 43.9525 | ממוצע |  | 65.13526 | ממוצע |
| 0.681228 | אסימטריה |  | 1.242093 | אסימטריה |  | 3.108508 | אסימטריה |
| 23.11473 | סטיית תקן |  | 25.03062 | סטיית תקן |  | 34.34986 | סטיית תקן |
| 40.5 | רבעון ראשון |  | 31.36 | רבעון ראשון |  | 52.7 | רבעון ראשון |
| 43.93 | חציון |  | 32.4 | חציון |  | 59.18 | חציון |
| 71.66 | רבעון שלישי |  | 40.8 | רבעון שלישי |  | 61.2 | רבעון שלישי |

Chart, pie chart

Description automatically generated**X2 - חודש ההגעה לחופשה במלון:**

Chart, box and whisker chart

Description automatically generated

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 הנ"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 הקטגוריאלי הנ"ל יש השפעה על המשתנה המוסבר ודבר זה מתחבר לנו עם ההיגיון שבחודשי הקיץ המחירים לחופשות עולות משמעותית ביחס לשאר חודשי השנה.

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

Chart, pie chart

Description automatically generated**X6 – סוג פנסיון:**

Chart, box and whisker chart

Description automatically generated

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.

Chart

Description automatically generated**X7 - יעד החופשה:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | ESP |  |  | CN |  |  | PRT |
| 149.905 | ממוצע |  | 78.81667 | ממוצע |  | 69.33227 | ממוצע |
| -0.37126 | אסימטריה |  | 0.342152 | אסימטריה |  | 1.857565 | אסימטריה |
| 68.96893 | סטיית תקן |  | 65.06113 | סטיית תקן |  | 37.18854 | סטיית תקן |
| 89.53 | רבעון ראשון |  | 41.73 | רבעון ראשון |  | 48 | רבעון ראשון |
| 177.215 | חציון |  | 52 | חציון |  | 55 | חציון |
| 192.57 | רבעון שלישי |  | 102.5 | רבעון שלישי |  | 81 | רבעון שלישי |
|  |  |  |  |  |  |  |  |
|  |  |  |  | GBR |  |  | IRL |
|  |  |  | 63.56313 | ממוצע |  | 58.68 | ממוצע |
|  |  |  | 1.74284 | אסימטריה |  | -0.00088 | אסימטריה |
|  |  |  | 27.57261 | סטיית תקן |  | 22.67001 | סטיית תקן |
|  |  |  | 54.34 | רבעון ראשון |  | 47.35 | רבעון ראשון |
| Chart, box and whisker chart  Description automatically generated |  |  | 60.19 | חציון |  | 58.7 | חציון |
|  |  |  | 68.63 | רבעון שלישי |  | 70.02 | רבעון שלישי |

ע"פ תרשים זה ניתן לראות בבירור שרוב ההזמנות מבוצעות עבור חופשות בפורטוגל (PRT).  
מפני שרוב הזמנת 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 יחסית ולכן התעריף היומי הממוצע עשוי להיות גבוה יותר. נוכל להסיק ממצאים אלו שיש השפעה בין קטגוריות המדינה אל המשתנה המוסבר.

Chart, box and whisker chart

Description automatically generatedChart, pie chart

Description automatically generated **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:

Chart, box and whisker chart

Description automatically generated

**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.

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

**X5 - 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 ש50% 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 מכוון שערכים חריגים אלו אכן מתיישבים עם ההנחה הסבירה שתיתכן הזמנה של בן אדם בודד או קבוצה של שלושה מבוגרים בחדר וההנחה שתמיד יוזמנו חדרים ל2 מגורים אינה בהכרח נכונה.

**X4 - מספר לילות השבוע בהזמנה:**בתרשים זה יש 3 תצפיות חריגות, עבור שני הזמנות חופשה ל 20 לילות ואחת ל21 לילות, הנובעות מערכים גדולים מאוד ביחס לחציון.  
החלטנו לנפה את חריגות אלו מכוון 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.

**X3 - מספר היום בחודש:**  
אין תצפיות חריגות עבור המשתנה הנ"ל.

**X1 - מספר הימים מתאריך ההזמנה לתאריך ההגעה:**אין תצפיות חריגות עבור המשתנה הנ"ל.

Chart, box and whisker chart

Description automatically generated

**X9 - 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:

**X1 - מספר הימים מתאריך ההזמנה לתאריך ההגעה:**

Chart, scatter chart

Description automatically generatedChart, histogram

Description automatically generated  
**פונקציית הצפיפות** – ניתן לראות כי ההתפלגות הינה אסימטרית בעלת זנב ימני קטן והשיא שלה נמצא בערך ב30 ימים כך שמרבית ההזמנות במלון מתבצעות בערך 30 ימים מראש. ניתן לראות שרוב ההזמנות התבצעו בין 0 ל125 ימים, כלומר אם ניקח הזמנה אקראית רוב הסיכויים שהיא הוזמנה בטווח של 0-125 ימים מראש. ערכים גדולים מ125 ימים נדירים יותר אך עבור 200 ומעלה ימים יש קפיצה והסיכוי שבדגימה נקבל הזמנה שהוזמנה יותר מ200 ימים מראש הוא לא מאוד קטן.

**פונקציית ההתפלגות המצטברת** - ניתן לראות כי כ80% מהתצפיות נמצאות בתחום של 0 – 125 ימים מראש כפי שראינו בפונקציית הצפיפות. פחות מ10% נמצאות בתחום של 125-200 ימים ובערך 15% מהתצפיות בתחום שגדול מ200 ימים מה שמשקף את הקפיצה שראינו בפונקציית הצפיפות.

**X4 - מספר לילות השבוע בהזמנה (שני עד שישי):**

**Chart

Description automatically generatedChart, histogram

Description automatically generated**

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 מ6 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 מ5 לילות. נתונים אלה מחזקים את הנתונים שקיבלנו בפונקציית הצפיפות.**

**X9 – מספר בקשות מיוחדות של לקוח:**

Graphical user interface, chart, box and whisker chart

Description automatically generatedChart, histogram

Description automatically generated

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 מ2 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 מ2 בקשות מיוחדות.

8. ייצוג קשרים בעזרת תרשימים:

**1. קשר בין סוג החדר (X8) 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 D וE 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.

**A screenshot of a computer

Description automatically generated with medium confidence**

**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 ולתאריך ההגעה. ציפינו 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.

**Chart, scatter chart

Description automatically generated**

**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.  
ציפינו 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 החושים ניתן לראות שיש פיזור רחב יותר לאורך כל החודש ללא ריכוז מובהק המראה על קשר משמעותי.

**Chart, scatter chart

Description automatically generated**

Chart, radar chart

Description automatically generated

Chart, radar chart

Description automatically generated

**4. קשר בין מדינת החופשה (X7) לבין סוג החדר (X8):**

ניתן לראות בכל המדינות כי ישנה עדיפות להזמנת חדר מסוג A על פני שאר סוגי החדרים וסביר להניח שמחירם 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. ציפינו 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 D ו E ההבדל מזערי ומכאן שהקשר בין המשתנים חלש מאוד/ לא קיים בניגוד להנחה שלנו.

**5. קשר מספר המבוגרים בחדר (X5) לבין ממוצע הפער בין תאריך ההזמנה לתאריך הגעה (X1):**

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.

**Chart, bar chart

Description automatically generated**

# 9. טבלאות שכיחות:

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

**טבלת שכיחות חד ממדית של טווחי התעריף היומי הממוצע:**

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, ב22% מההזמנות התעריף היה נמוך מ45 ומכאן שב82% 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.

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

**טבלת שכיחות חד ממדית של טווחי הלילות בהזמנות:**

טבלה זו בוחנת את שכיחות טווחי לילות השבוע במדגם במרווחים של 2 לילות. מהנתונים שהתקבלו ניתן לראות ש 93% מההזמנות מכילות פחות מ6 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:**

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

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

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 ל200 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 מ25 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.

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

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

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

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

# נספחים:

סעיף 4:

A picture containing calendar

Description automatically generated

סעיף 5 :

Text

Description automatically generatedText

Description automatically generated Text

Description automatically generatedText

Description automatically generated Text

Description automatically generated Text

Description automatically generated

פלטים למשתנים קטגוריאליים:

X2 – חודשים :

Text

Description automatically generated

Graphical user interface, text

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

Text

Description automatically generated

A screen shot of a computer

Description automatically generated with low confidence

Text

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

X6 – ארוחות:

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

Text

Description automatically generated

X7 – מדינות:  
A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screen shot of a computer

Description automatically generated with low confidence

Graphical user interface

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with medium confidence

X8 – סוגי חדר:

A screen shot of a computer

Description automatically generated with low confidence  
Graphical user interface

Description automatically generated with medium confidence Text

Description automatically generated with medium confidence

A picture containing graphical user interface

Description automatically generated

סעיף 6:

Text

Description automatically generated

Chart, box and whisker chart

Description automatically generated

סעיף 7:

A screenshot of a computer

Description automatically generated with medium confidence

סעיף 8:

Text

Description automatically generated

Text

Description automatically generated

סעיף 9:

Text

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence