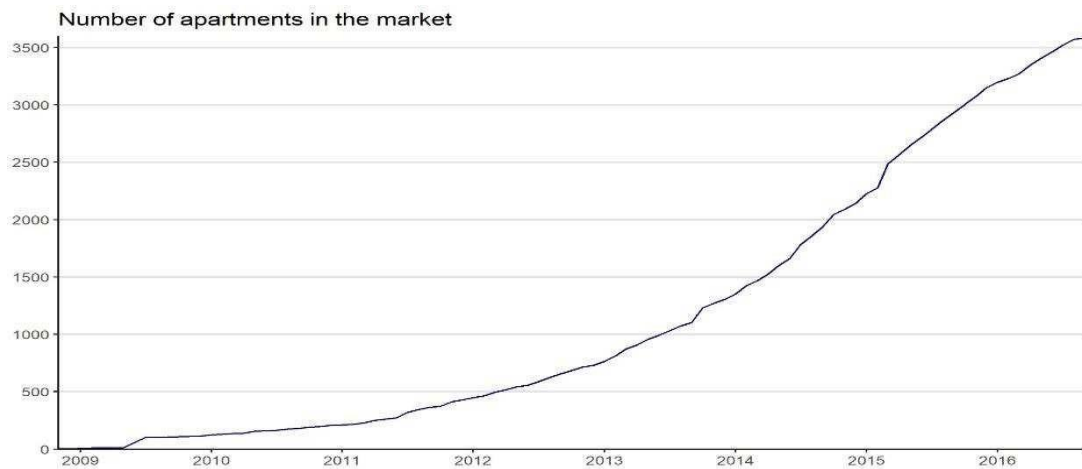
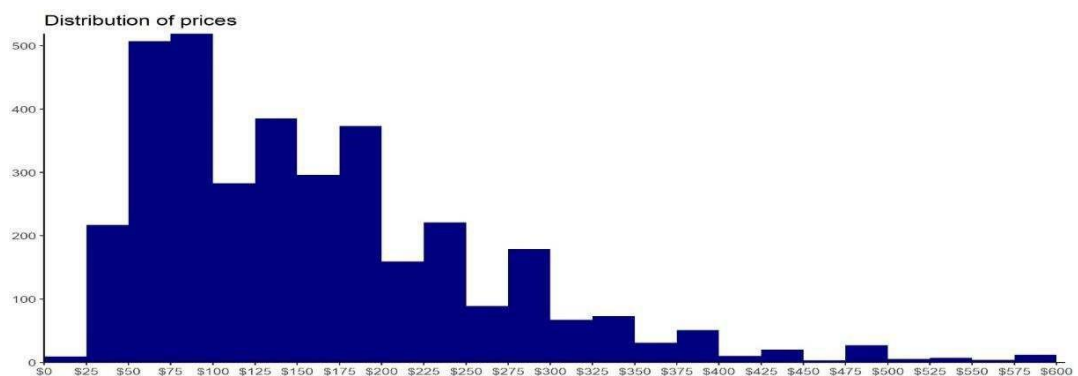


Airbnb Exploratory Data Analysis

סטטיסטיקה תיאורית:



הגרף מראה את מספר הדירות בשוק על פני זמן, רואים פה את המספר המצטבר של הדירות שנכנסו לשוק עד התאריך המסוים, ורואים התמתנות בשיפוע של העלייה מה שאומר שמספר הדירות שנכנסו לשוק ב2016 הוא פחות משאר השנים, נעמיק בזה יותר בעניין של הצמיחה.

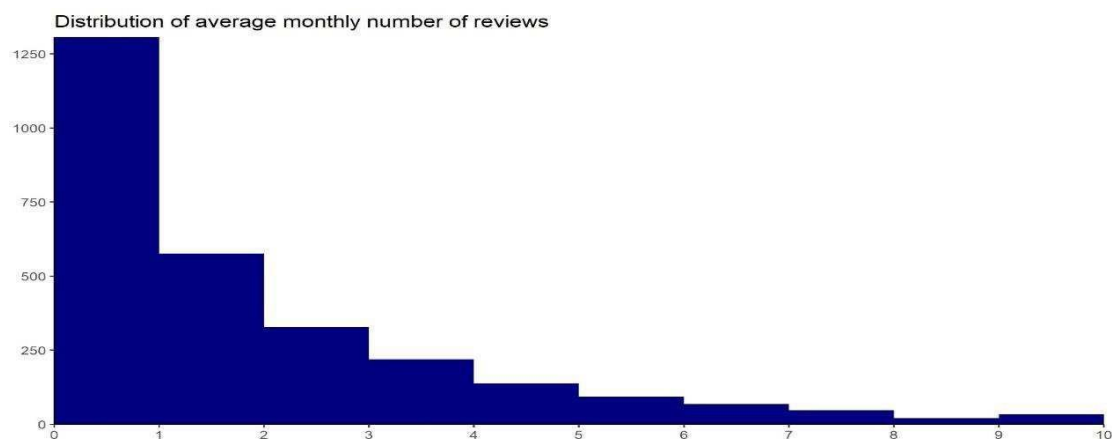


הגרף מראה את התפלגות המחירים¹, רואים התפלגות עם זנב ימני ושרוב הדירות הושכרו במחירים בין 50 ל200 דולר ללילה ויש מעט שהושכרו במחירים מעל \$350. ניתן לראות שבעלי הדירות נוטים לתמחר את הדירות בכפולות של \$50.

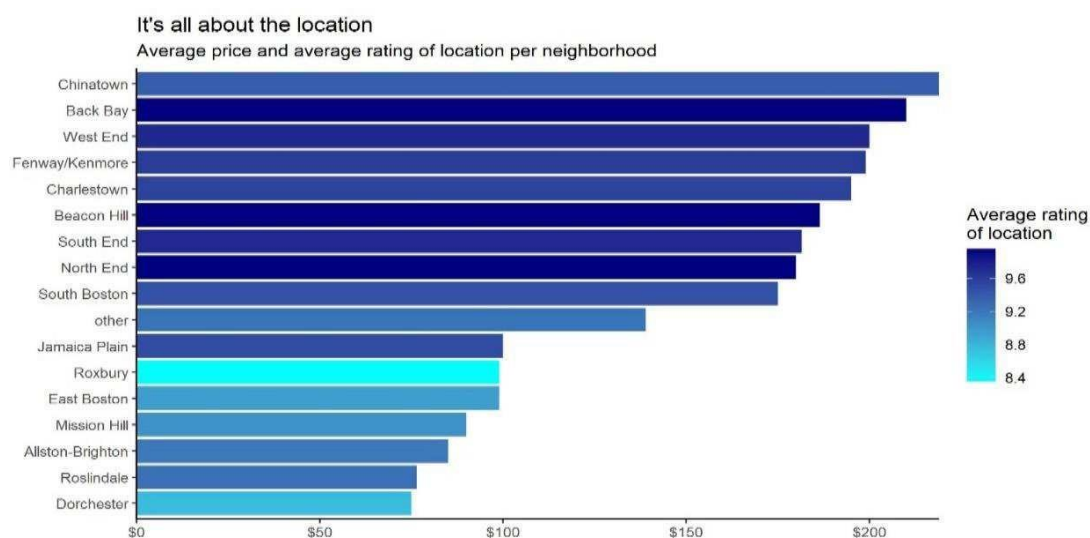
הגרף שמוצג למטה מראה את התפלגות ממוצע מספר הדירוגים בחודש². רואים שרוב הדירות מקבלות דירוג אחד בממוצע בחודש וככל שמספר הדירוגים עולה מספר הדירות יורד ויש מעט דירות שמקבלות 9 דירוגים בחודש בממוצע.

¹ הערכים מעל האחוזון 99 העברנו לערך של האחוזון 99.

² ערכים מעל האחוזון 99 העברנו לערך של האחוזון 99



סטטיסטיקה תיאורית עבור שני משתנים או יותר :



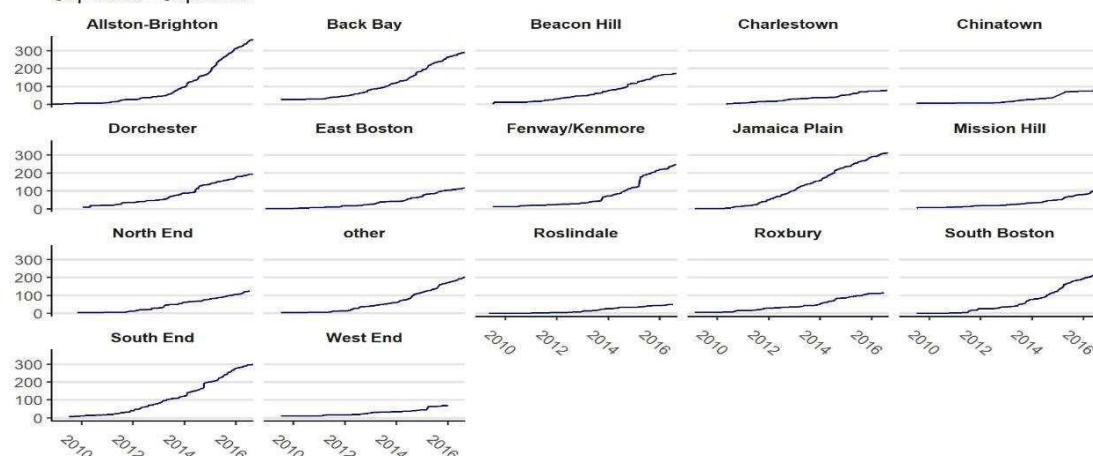
הגרף מראה את המחיר הממוצע ואת הדירוג הממוצע לפי שכונה³ רואים חלוקה של השכונות לשתי קבוצות. שכונות שהמיקום שלהן מדורג גבוה גם המחיר שלהם גבוה לעומת שכונות שהמיקום שלהן מדורג נמוך והדירות בהן זולות. בתוך כל קבוצה הקשר בין המחיר לדירוג המיקום הוא חלש.

הגרף להלן מראה את מספר הדירות לפי שכונה על פני זמן. רואים שלרוב השכונות השנה החלשה ביותר היא שנת 2016 וזה מתואם עם מה שמצאנו בניתוח משתנה `host_since` ולחלק מהשכונות ההתמתנות בשיפוע התחילה במעבר משנת 2014 ל 2015. רואים שיש שונות גדולה בין לא רק גודל השכונות אלא במתי הן גדלו.

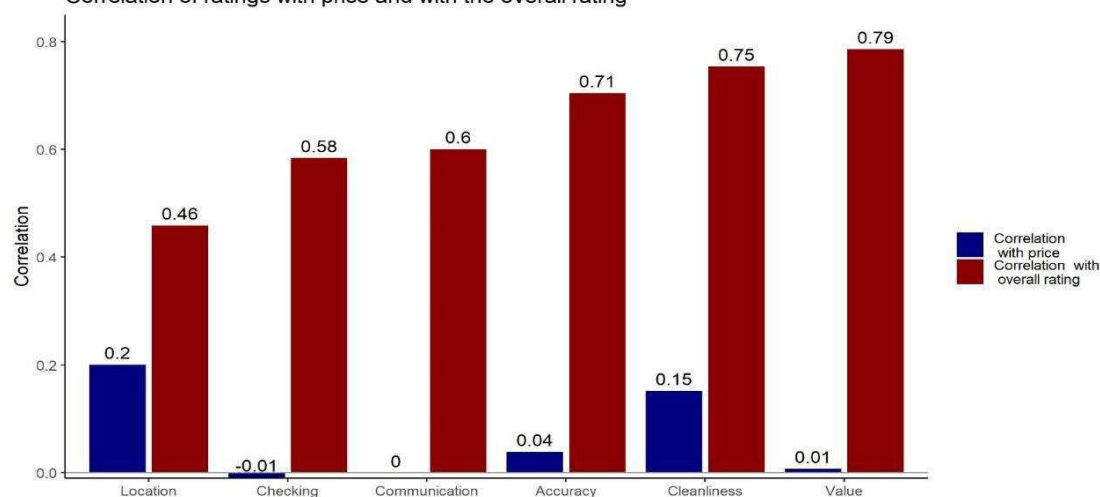
³ שכונות שיש בהן פחות מ-50 נכללו לקטגוריית other.

Number of apartments in each neighborhood

Sep 2009 - Sep 2016



Correlation of ratings with price and with the overall rating



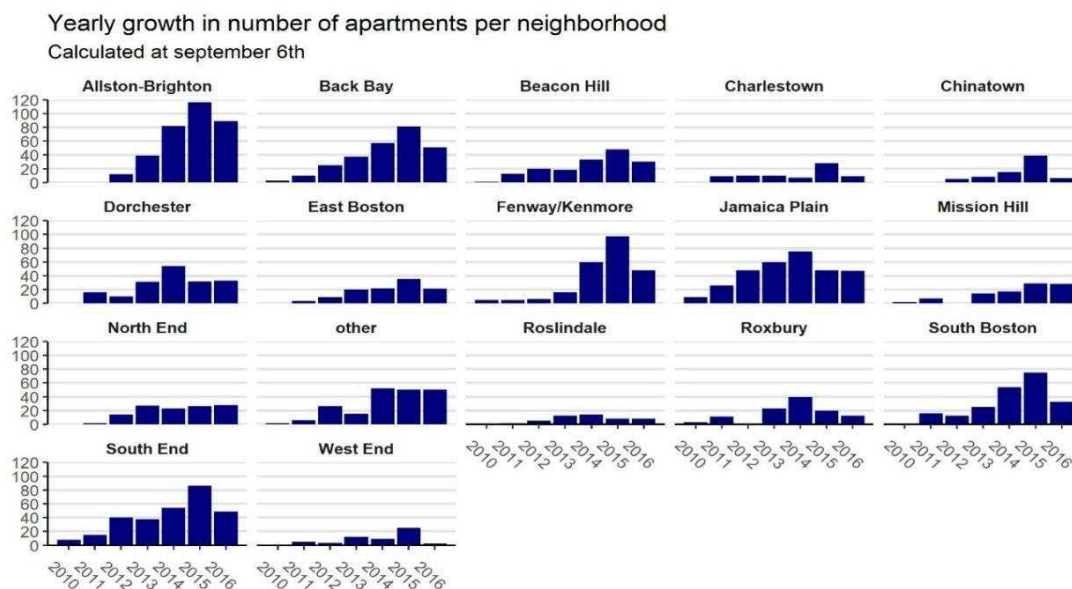
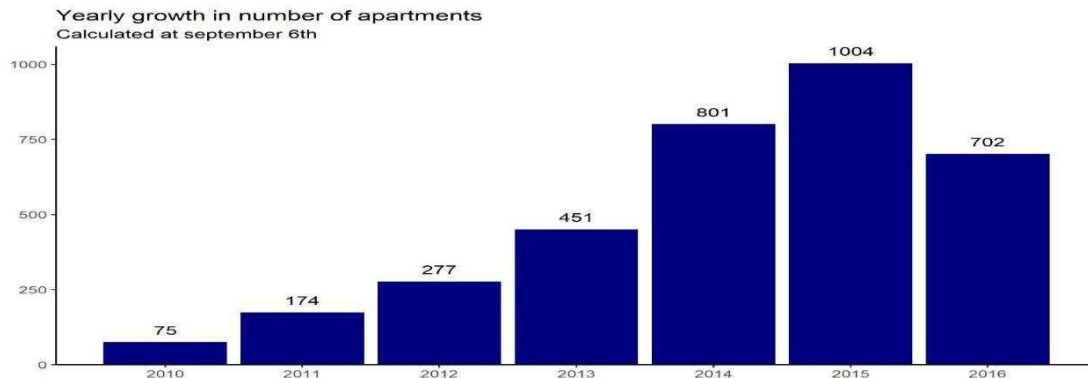
הגרף מראה את הקורלציה בין הדירוגים לבין מחיר והדירוג הכללי. רואים שהדירוג הכי מתואם עם מחיר הוא דירוג המיקום וזה אומר שהפקטור הכי חשוב לאנשים בעת ההחלטה על הדירה זה המיקום. לעומת זאת בקורלציה עם הדירוג הכולל דירוג המיקום הוא הנמוך ביותר ואחרי שזה אומר שאנשים חושבים על מיקום כפרמטר נפרד שמוכנים לשלם עליו ואינו מזוהה ישירות עם הדירה.

מטריקות :

מטריקה לצמיחה (לפי שכונה) : שינוי במספר הדירות החדשות בשוק בכל שנה (לפי שכונה)

המטריקה מאפשרת לנו לראות את התוספת במספר הדירות בכל שנה בשוק כלומר הפוטנציאל של השוק בתפוסה מלאה כמו שרואים בגרפים. הבעיה המרכזית היא שלא יודעים בדיוק כמה אחוז מהדירות הושכר בפועל. בעיה נוספת היא שנראה שהדאטה שאיננו מכיל מידע על דירות שהיו בשוק בעבר אך יצאו ממנו לפני ספטמבר 2016 או לחליפין שחלק מהדירות במאגר יצאו

מהשוק עד תאריך זה אך הדבר אינו בא לידי ביטוי בדאטה⁴. אם היה מידע על השכירות בשנים נוספות היינו רוצים לבחון את השינוי במספר הלילות שהושכרו בין תקופות מקבילות בשנים שונות אך הסתכלות בשנה אחת תיתן בעיקר אפקטים של עונתיות ולא צמיחה.



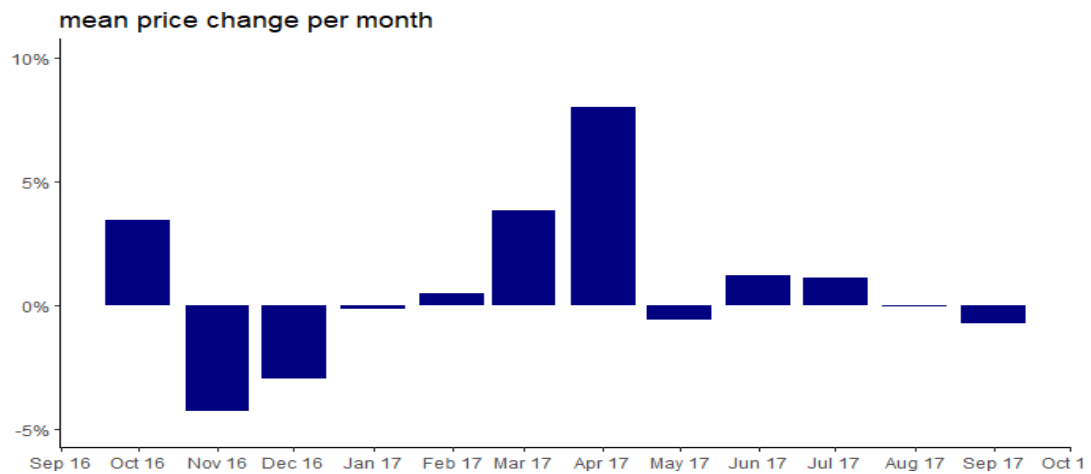
רואים שמספר הדירות בשוק עולה בקצב יורד מתחילת שנת 2016, מספר הדירות החדשות בשוק 2016 הוא פחות ממספר הדירות החדשות שהיו ב-2014 ו-2015 ויכולים לראות את זה גם לפי השכונות שלהרבה מהן שנת 2016 היא שנה חלשה ולחלק מהשכונות ירידת הקצב התחילה ב-2014.

מטריקה למחיר : ממוצע אחוז השינוי החודשי במחיר דירה

המטריקה מאפשרת לנו לראות את ממוצע אחוז השינוי החודשי במחיר הדירות, הבעיה במטריקה הזאת היא שהמחיר הזה הוא מחיר שהוצע ולא דווקא שולם ולכן ייתכן שיש בו הטיה

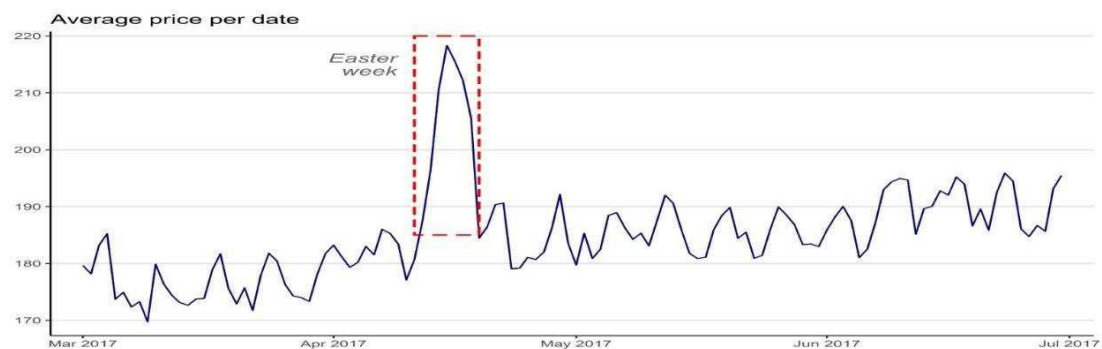
⁴ לכל אחת מהדירות שמופיעות במאגר listings קיימת לפחות רשומה אחת ב Calendar, 688 דירות לא היו זמינות כלל לכל אורך התקופה שנבחתה Calendar. ייתכן שדירות אלה פשטו יצאו מהשוק קודם לכן וייתכן כי דירות אחרות שהיו בשוק בשנים קודמות פשוט אינן נכללות במידע. בכל מקרה, אין לנו דרך להסתכל על הגידול במספר הדירות נטו.

כי בעל הדירה מוריד את המחיר כדי שיוכל להשכיר. בנוסף לכך שההשוואה היא ביחס לחודש קודם ולכן יש פה אפקטים עונתיים ולכן היינו רוצים שהנתונים יכילו נתונים של עוד שנים קודמות כדי להשוות אותן תקופות בכל שנה ולהתגבר על אפקט העונתיות.



ערכים חריגים :

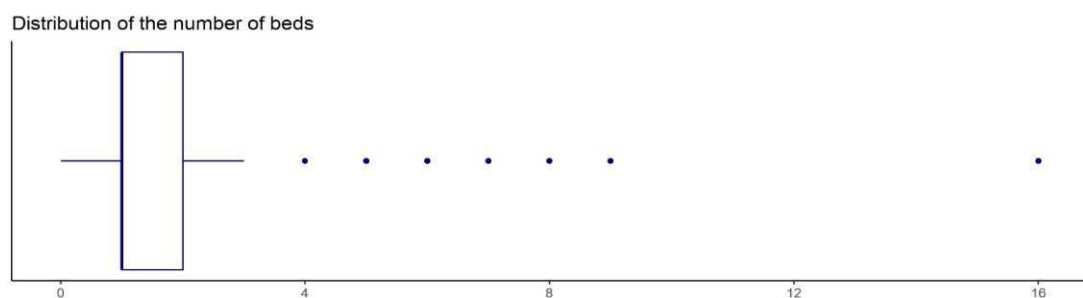
שוות זמנים:



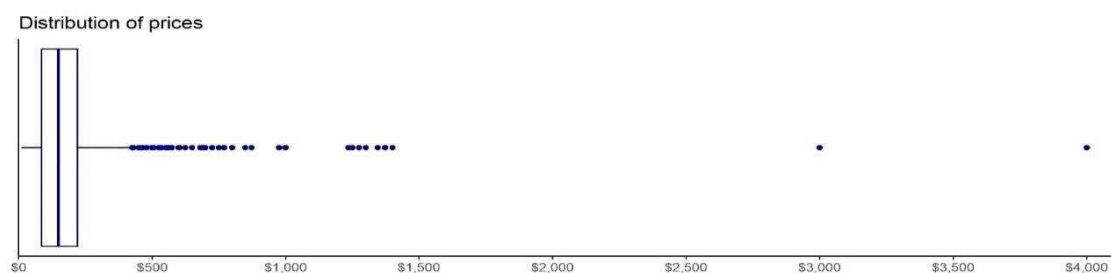
הגרף מראה את ממוצע המחיר על פני זמן. רואים בגרף עלייה חדה במחיר הממוצע לדירות באמצע אפריל, סביר להניח שזה בגלל חג הפסחא ולכן הקפיצה הגיונית ואין בעיה בנתונים ואין צורך לתקן.

מיטות :

רואים בגרף להלן שיש דירה עם 16 מיטות, איתרנו את התצפית וגלינו שהדירה ריקה ואין בה מיטות (מתיאור הדירה), כיוון שהתצפית משפיעה באופן חריג על ההתפלגות היינו מורידים אותה.



מחירים :



כפי שמופיע בגרף, יש שתי דירות שהושכרו במחיר של 3000 ו 4000 דולר ללילה אחת שזה מאוד חריג מהתפלגות המחירים, אחרי איתור הדירות התברר שיש להם הרבה מידע חסר וגם כשהיו זמינות לא הופיע להם מחיר, לכן סביר להניח שזה טעות בהזנת הנתונים והיינו כשמסתכלים על התפלגות המחיר סביר לתקן את הערכים האלו לאחוזון 99 למשל.

מסקנות והמלצה עסקית :

על מנת לבחון את מצב העסק הגדרנו מטריקה חדשה שהיא אחוז הלילות הפנויים שמשקף פוטנציאל גידול ברווחים בהינתן הנכסים הקיימים. רואים שיש אחוזי תפוסה נמוכים ברוב השכונות, בממוצע 49% מהלילות פנויים ובכל השכונות מדובר ביותר מ-30%. ראינו גם שאין הבדל משמעותי. על פני חודש וימי השבוע כלומר ברור שיש בו בעיית רוויה ועודף היצע וכדי לטפל בזה צריך להעלות את אחוזי התפוסה (להפחית את המחיר). לכן האתגר המרכזי של Airbnb בבוסטון הוא להגדיל את אחוז התפוסה ולכן אפשר לתת תמריצים לבעלי דירות שמצליחים להגדיל את מספר הלילות התפוסים.

