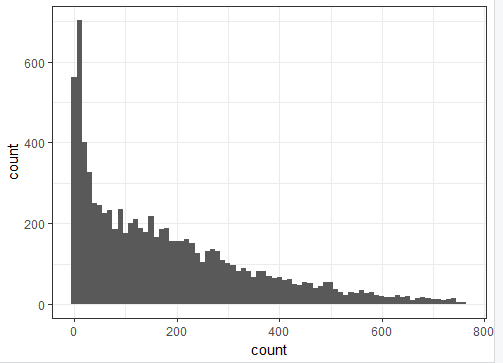
**כלכלה – BIG DATA:**

1. שאלה 1:
   1. ניקוי והגמשת data:
      * לקחנו את כל ה-data עד ה-quantile ה-99.
      * בדקנו שאין ערכי N/A ב-DataSet.
      * יצרנו featureים חדשים של שעה (hour), יום בשבוע (weekday), זמן (time), תאריך (date), חודש (month) ושנה (year) המבוססים על ה-datetime.
      * הפכנו משתנים לקטגוריאלים (ממספר לשם)
   2. מסקנות כלליות:

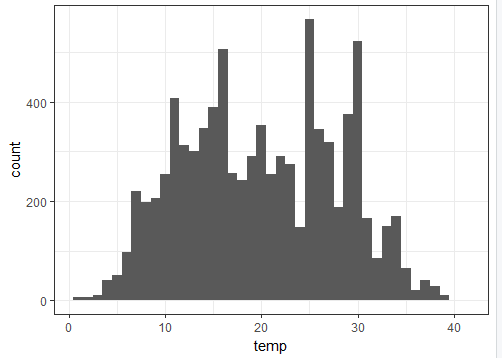
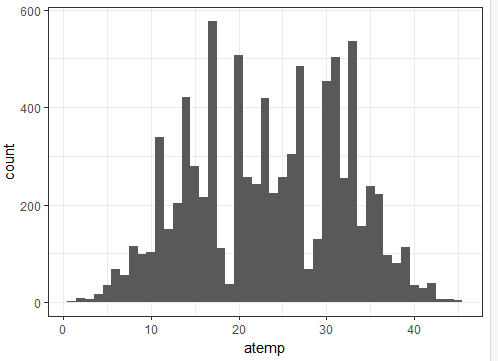
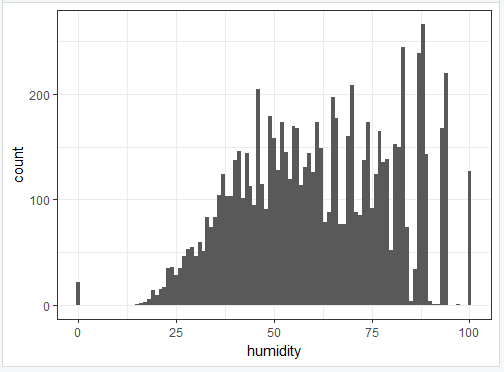
כמות ההשכרות של האופניים נעה בטווח רחב של בין השכרה אחת ביום לבין כ-760 כאשר הממוצע הוא 183 ליום:

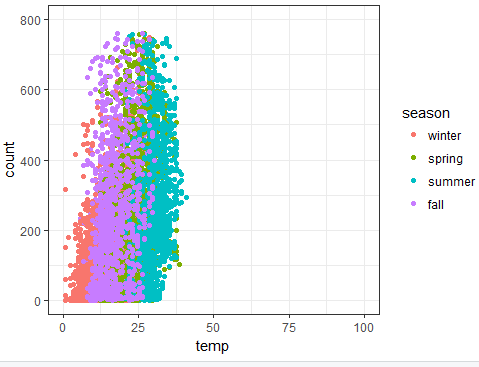


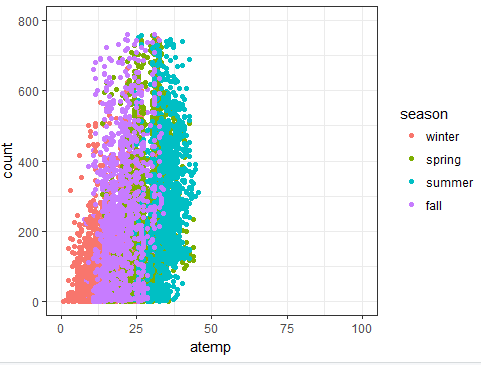
ניתן לראות שבאופן כללי feature ה-count מתנהג כמו רשת חברתית, זאת לפי ההיסטוגרמה שלו:

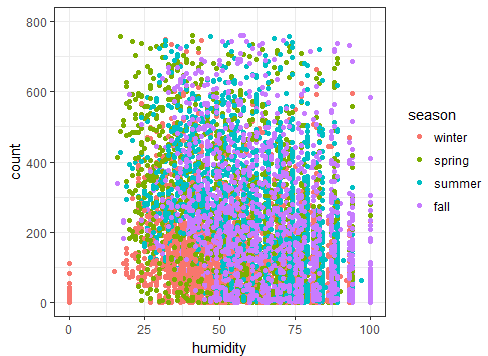


מבחינת טמפורטורה, לחות וטמפטורה מורגשת (temp, humidity, atemp) ניכר שאין הם משפיעים רבות על כמות אירועי ההשכרות מלבד בטמפרטורות הקיצוניות. כמו-כן לא ניתן ללמוד מכך רבות אלא רק לצפות במגמה כי מדובר בכמות אירועי ההשכרות במצבים הללו ולא כמה השכרות בוצעו בפועל במהלך תצפית (אירוע השכרה):

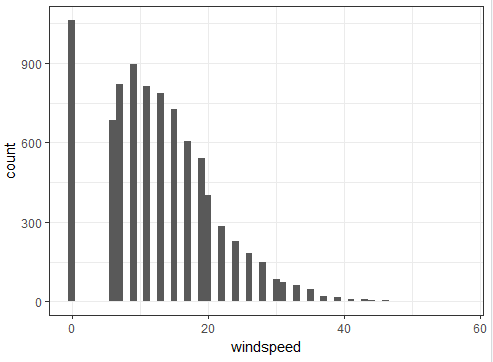


ניתן לראות שלמרות כי היינו חושבים שהלחות והמטפורות ישפיעו מאוד, הדבר אינו תואם למציאות:

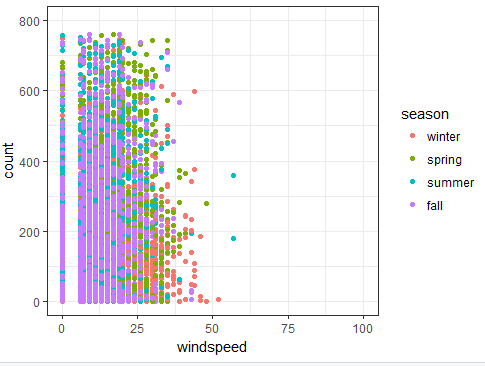




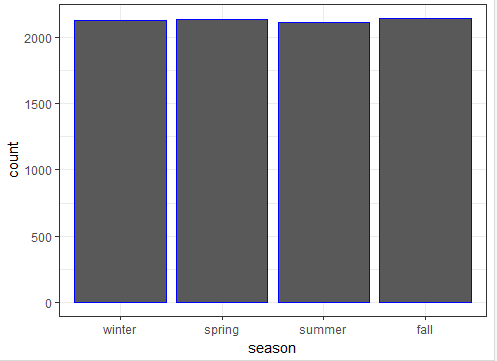
לעומת זאת, כן ניתן לראות במגמה כזאת מבחינת מהירות הרוח (windspeed) והיא שככל שהרוח חלשה ישנן יותר השכרות:



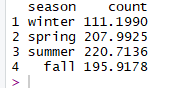
ניתן לראות גם מבחינת מהירות הרוח ביחס לעונות – ככה שהרוח מתחזקת ישנם פחות ופחות אירועי השכרות.

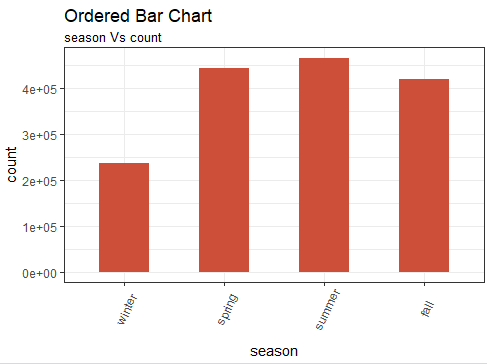


מבחינת משתנים קטגוריאלים, נראה כי מבחינת אירועי השכרות, לעונה אין השפעה:

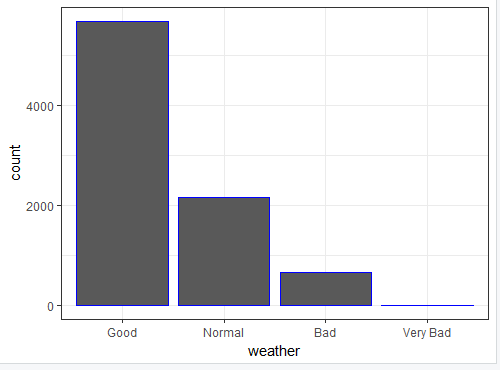


אך כאשר בודקים את הממוצע פר עונה ניתן לראות כי יש העדפה ברורה לקיץ, לאחר מכן לסתיו ולבסוף הכי פחות בחורף.

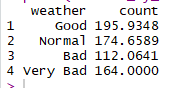


נראה זאת באמצעות bar plot על הכמות:

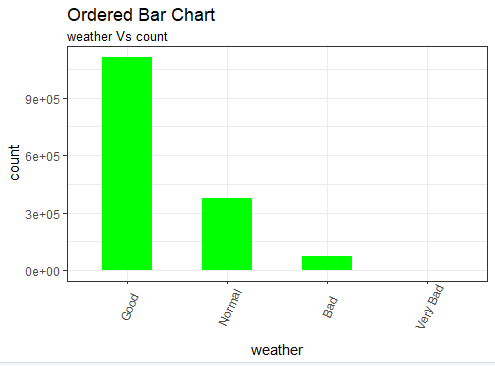
אולם מבחינת מזג אוויר (**weather**) נראה כי באופן מגמתי יש יותר אירועי השכרות כאשר מזג האוויר **נחשב טוב**:



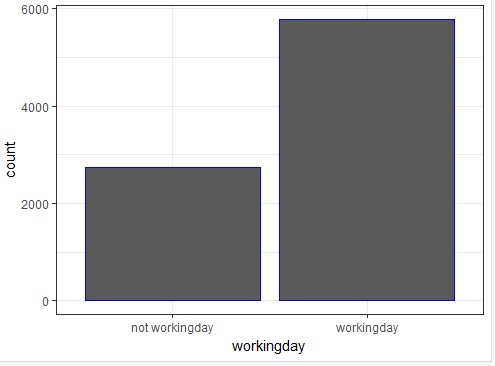
ניתן לראות כי בממוצע אכן הדבר מתקיים, מלבד כך שבמזג אוויר רע יש פחות ממזג אוויר טוב:



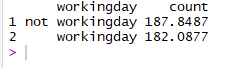
מבחינת כמויות בסה"כ לפי מזע האוויר – אכן הדבר זהה ל-avg. נראה זאת באמצעות bar-plot:



כאשר מסתכלים על יום עבודה לעומת יום חופשה (סופ"ש או חג) מקבלים כי ישנם יותר אירועי השכרה ב- **workingday**:



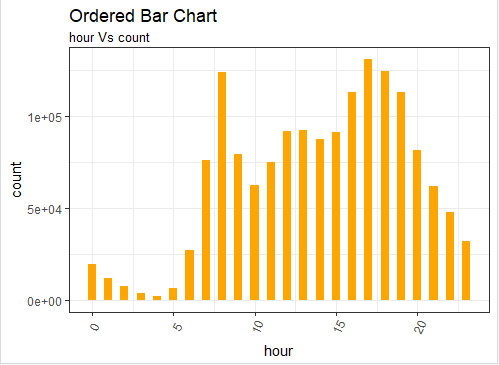
אולם ישנם יותר ימי עבודה מימי חופשה. לכן, עפ"י ממוצע, ניתן לראות כי מספר אירועי ההשכרה **דומה**:



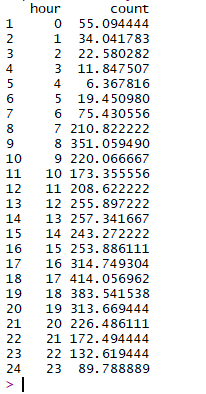
**כעת, נדון על סטטיסטיקות של זמנים בהתאם לכמויות. נחלק זאת ל-3 משתנים עיקריים שבהם נתון, חודש (month), יום בשבוע (weekday) ושעה (hour).**

**שעה:**

נראה כי השעות בהן ישנן הכי הרבה השכרות אופניים הם 08:00 בבוקר, 17:00 ו-18:00:

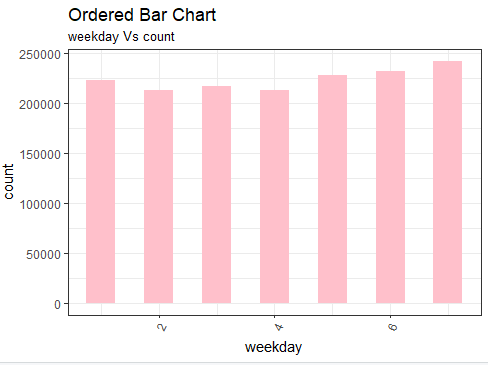


הדבר נתמך גם באופן **ממוצע**:

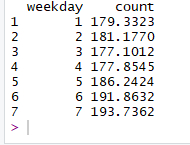


**יום בשבוע:**

נראה כי באופן יחסי, בסופ"ש ובראשון ישנן יותר השכרות אופניים בסה"כ:

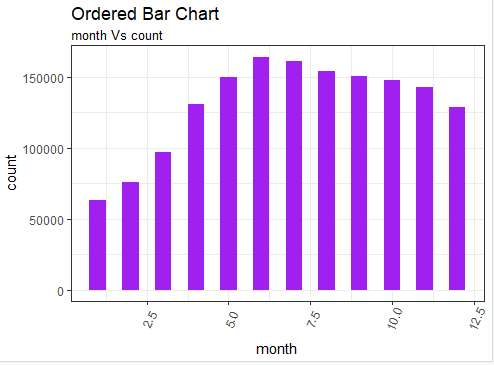


אולם מבחינת **ממוצע**, הדבר הינו קצת **שונה** ובעיקר בשישי-שבת יש כ-15 יותר השכרות:

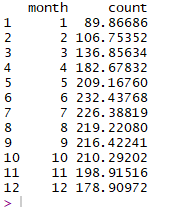


**חודש:**

מבחינת חודשים נראה כי חודשי הקיץ הינם החודשים בהם יש הכי הרבה השכרות:



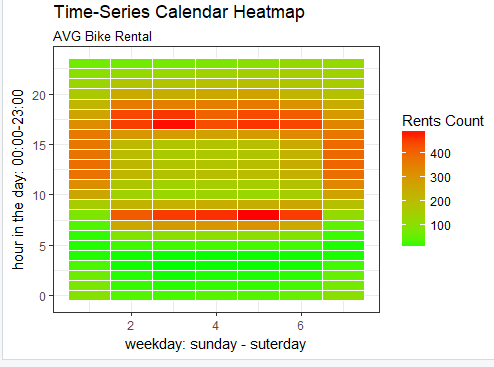
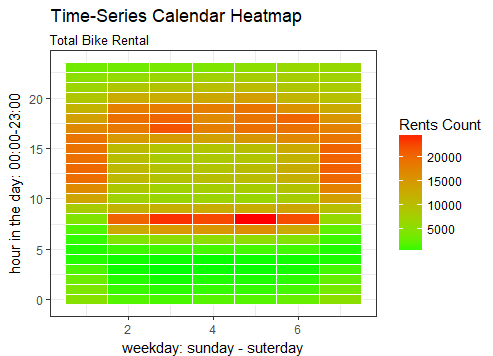
הדבר נתמך גם מבחינת ממוצע בחודשים הללו לאירועי השכרה:



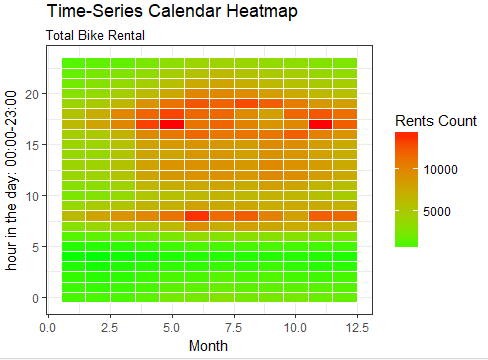
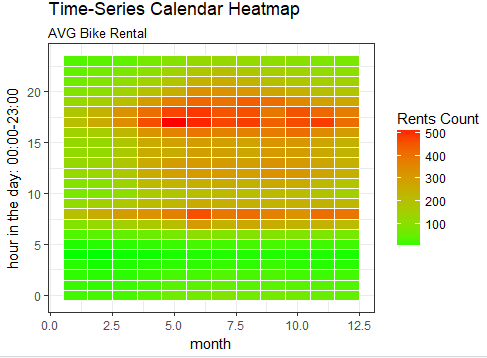
**כעת נראה מפות חום הממחישות את הנתונים הללו.**

לכל זוג נתונים נחקר (יחד עם נתון ה-count בתור נתון שלישי), נראה מבחינת **ממוצע** (ימין) ומבחינת **כמות** השכרות כוללת (שמאל):

**יום VS שעה VS כמות:**

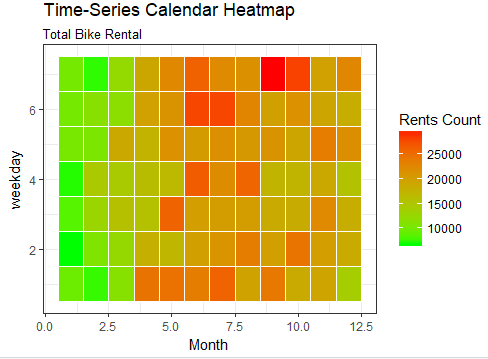
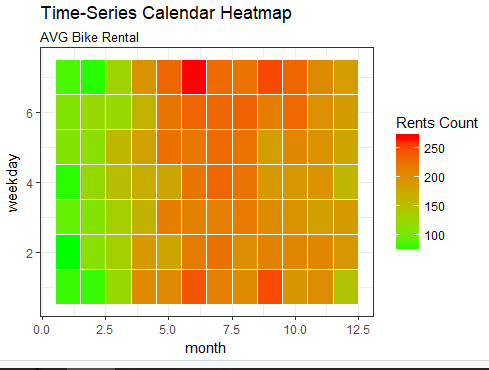


ניתן לראות בצורה נורא יפה, כי בסופ"ש (שבת וראשון) רב ההשכרות מתרחשות יותר בזמנים של בוקר מאוחר-צהרים ואילו בימות השבוע, רב ההשכרות מבוצעות בבקרים ואחה"צ.

**שעה VS חודש VS כמות:**

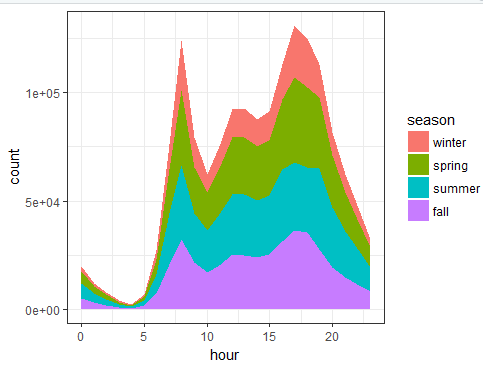
ניתן לראות כי אכן חודשי הקיץ הינם חודשים מובילים כצפוי ממקודם.

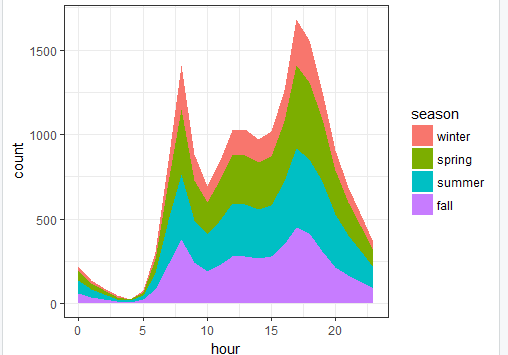
**יום VS חודש VS כמות:**



ניתן לראות כי הזמנים המשמעותיים ביותר להשכרות הינם בקיץ ובסופ"שים, כמקודם.

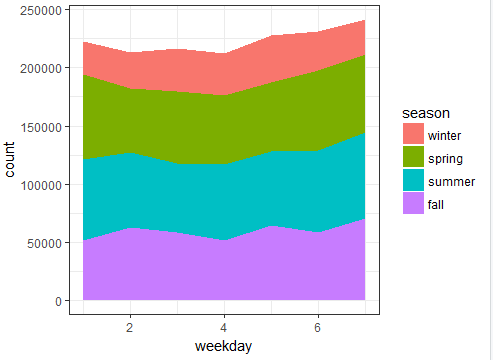
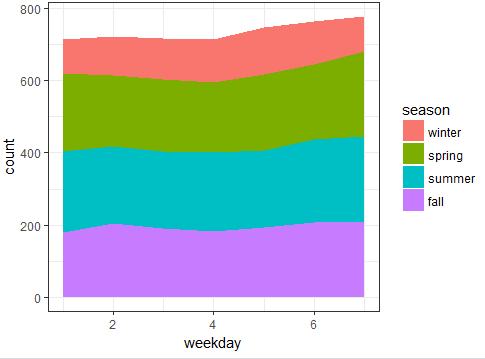
דרך נוספת להראות זאתהינה בגרף שטח **מבחינת העונות ביחס לשעות:**





ניתן לראות **שאכן בקיץ ובאביב** ההשכרות הן הרבות ביותר במיוחד בשעות הנ"ל.

באופן דומה מבחינת **ימות השבוע** ולא שעות:

* להראות מבחינת חודשים ועונות השנה אינו מוסיף מידע משמעותי.