

כריית נתונים ב-R // תרגיל בית מספר 4

מטרת התרגיל: חיזוי הביקוש לתחבורה ציבורית

חיזוי ביקוש לתחבורה ציבורית חיוני למטרות מרובות כגון תכנון כוח אדם, ובקרת מלאי. בשנת 2006, מערכת התחבורה הציבורית בסנטיאגו דה צ'ילה עברה תהליכי שיפור משמעותיים (www.transantiago.cl). במסגרת זו התקיימה באוקטובר 2006 תחרות בתחום המודיעין העסקי, שהתמקדה בחיזוי הביקוש לתחבורה הציבורית. להלן פרטים מתוך אתר התחרות, שאינו זמין יותר, בתרגום לעברית:

תיאור הבעיה:

חברת התחבורה הציבורית צופה עליה משמעותית בביקוש לשירותים שלה ומתכננת לרכוש אוטובוסים חדשים ולהרחיב את המסופים. השקעות אלו מחייבות תחזית מהימנה של ביקוש עתידי, אשר צריכה להתבסס על ביקוש היסטורי המאוחסן במחסן הנתונים של החברה. המידע ההסטורי מכיל את מספר הנוסעים המגיעים את הטרמינל עבור במקטעים של 15 דקות בין 6:30 ל-22:00. כיועץ חיזוי התבקש לחזות את מספר הנוסעים המגיעים למסוף.

נתונים זמינים:

חלק מהמידע ההיסטורי מאוחסן בקובץ `bicup2006.csv`. קובץ זה מכיל מידע היסטורי עם ביקוש ידוע לתקופה של 3 שבועות, מופרדים לפרקים של 15 דקות, ותאריכים ושעות עבור תקופה חיזוי של 3 ימים.

שאלות:

- קרא את הנתונים לתוכנת R, בפורמט של time series.
 - טיפ: בפונקציית `ts` הזן עבור פרמטרים `start` ו-`end` את מספרי הימים בחודש ואת מיספר המקטע בו היום מתחיל, ובפרמטר `freq` הכנס את מספר המקטעים ביום.
 - לדוגמא, התאריך 2/3/2005 בשעה 15:07, יוזן כצמד המספרים `c(2, 4)` – היום השני, המקטע הרביעי.
- (5 נק') צור גרף של המידע ההיסטורי (בלבד). הקפד על שמות צירים אינפורמטיביים. **העתק את הגרף לקובץ הפלט.**
- (2 נק') **בקובץ הפלט: פרט את רכיבי הסדרה העיתית, לפי הגרף בסעיף 2.** (בדומה לשקף 5 מהכיתה)

4. (3 נק') התאם מודל רגרסיה המכיל את רכיבי ה- level, trend בלבד. השתמש ברגרסיה

שיצרת על מנת לחזות את הביקוש בשלושת הימים החסרים. דווח לפלט את ביצועי

המודל לפי מדד RMSE.

5. (3 נק' בונוס) התאם מודל רגרסיה לנתונים והשתמש בה על מנת לחזות את הביקוש

בשלושת הימים החסרים. דווח לפלט את ביצועי המודל לפי מדד RMSE.

אופן הגשה:

✓ הגשה דרך אתר למידה

✓ יש להגיש קובץ R מתועד,

✓ קובץ פלט בפורמט word,

✓ למעוניינים: קובץ פתרון לשאלת בונוס.