

## תרגיל בית 2

איל טורץ 207906884

טל דולברג 205914609

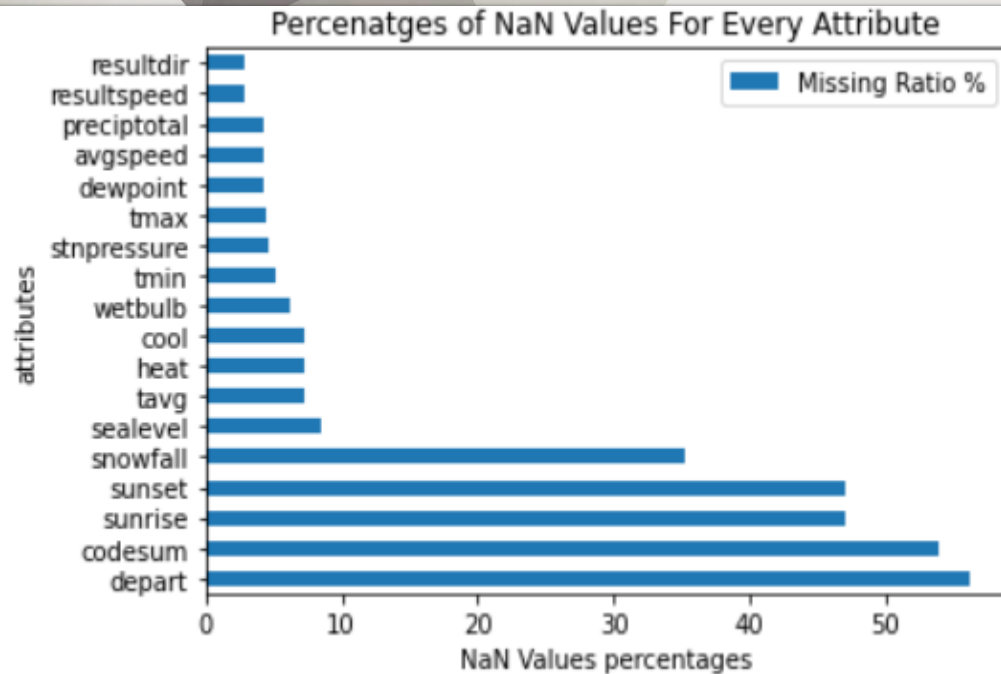
# השלמת מידע חסר

◇ השלמת מידע חסר הוא חלק

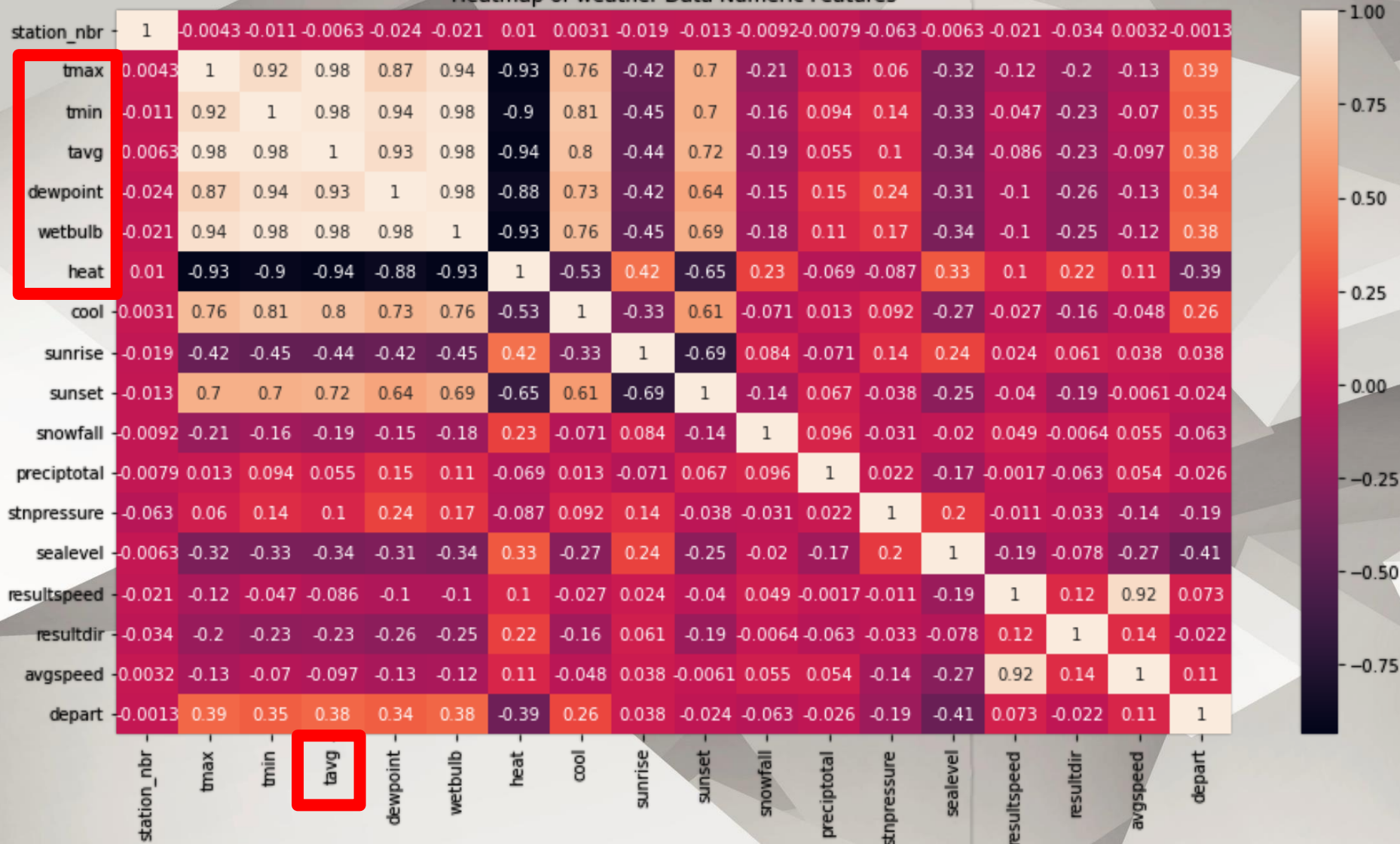
קריטי בעיבוד המידע

◇ הוויזואליזציה הבאה עזרה לדעת

באילו תכונות להתמקד



Heatmap of weather Data Numeric Features



# Imputation

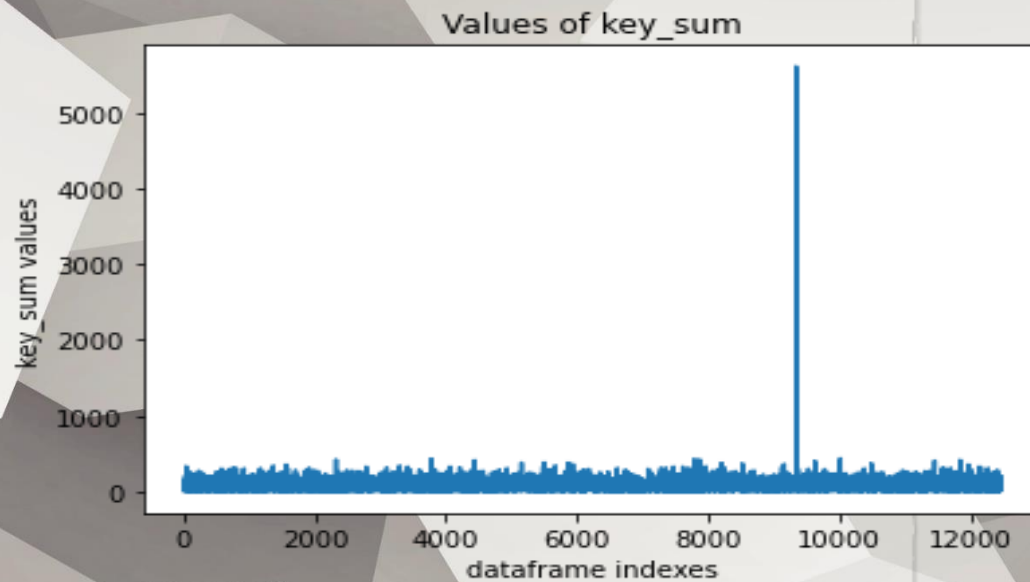
זמן הריצה הוא דבר  
קריטי ותכונות בעלות  
קורלציה גבוהה  
תורמות מידע דומה  
למודל ולכן ניתן  
להשמיט את חלקן.

בחרנו להשמיט את:  
tmax, tmin,  
dewpoint, wetbulb,  
heat

# Imputation

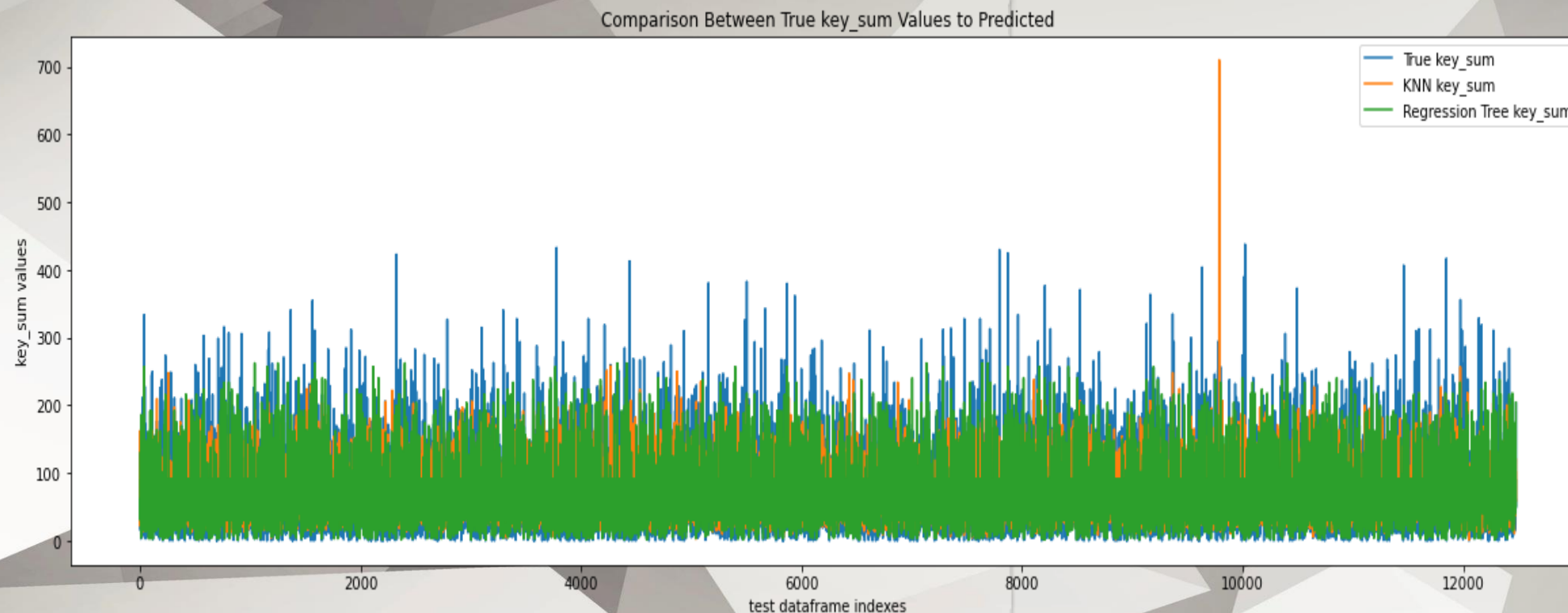
◇ במהלך ההשוואה בין ביצועי המודלים התגלתה תצפית חריגה בנתוני `test`.

◇ החלטנו להוריד את תצפית זו, כיוון שהיא לא משקפת את הרוב המוחלט של הנתונים ולכן תפגע בהערכה נכונה של דיוק המודלים.





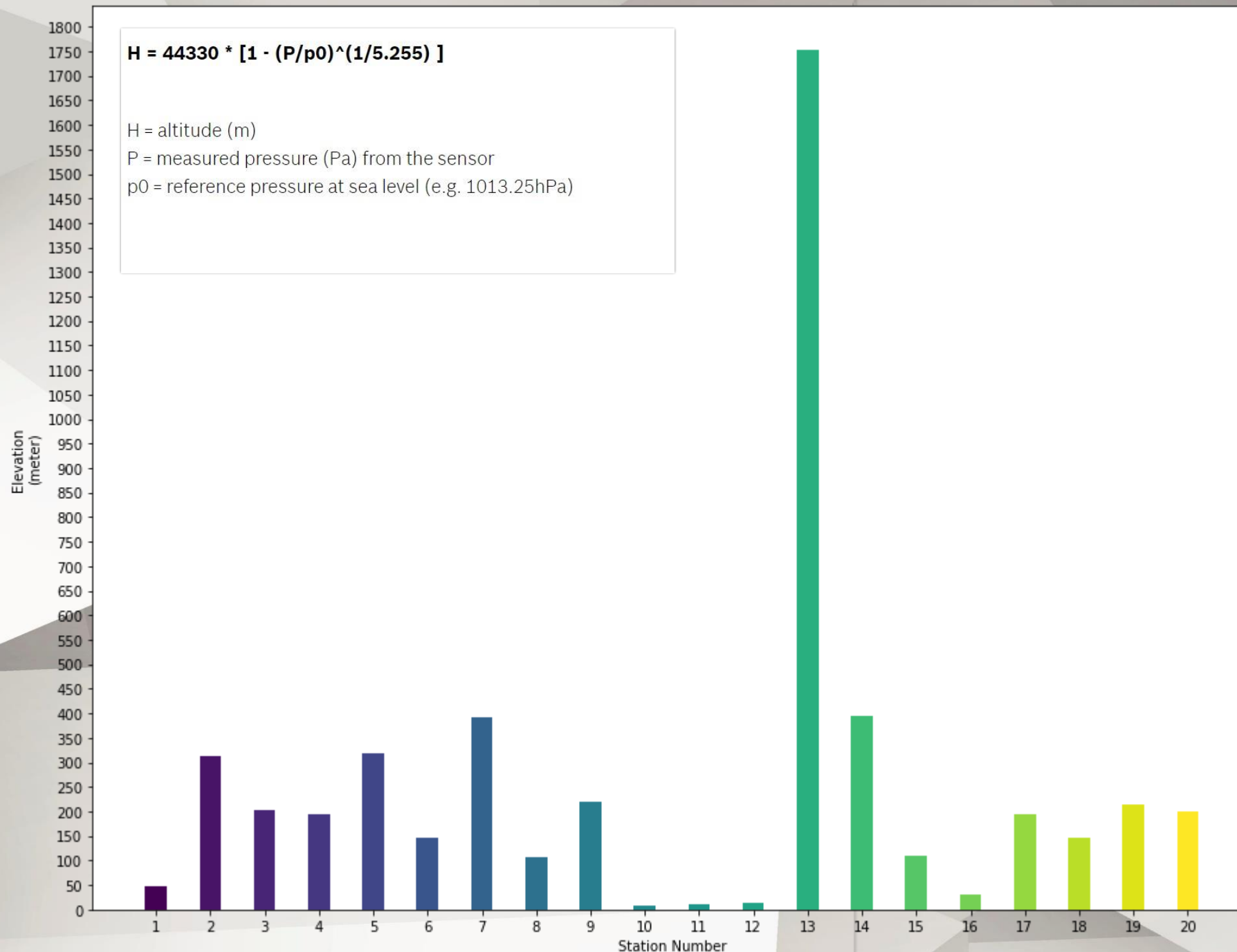
# הערכת מודלים



◇ על מנת להשוות בין המודלים השתמשנו בחלק C במידת דיוק וזמן ריצה.

◇ ניתן לראות את ההשוואה בין ערכי key\_sum האמיתיים בסט המבחן לבין הערכים שהמודלים העריכו לגבי כל תצפית.

	MSE	run time(seconds)
KNN Regressor	2045.002	170.504590
Decision Tree Regressor	1629.083	99.390431



# הערכת גובה

◇ על מנת להעריך  
את הגובה של כל  
תחנת מזג אוויר  
השתמשנו  
במשוואה  
מטאורולוגית  
אשר משתמשת  
בלחץ האוויר  
בגובה פני הים  
ולחץ האוויר  
הנמדד בכל  
תחנה