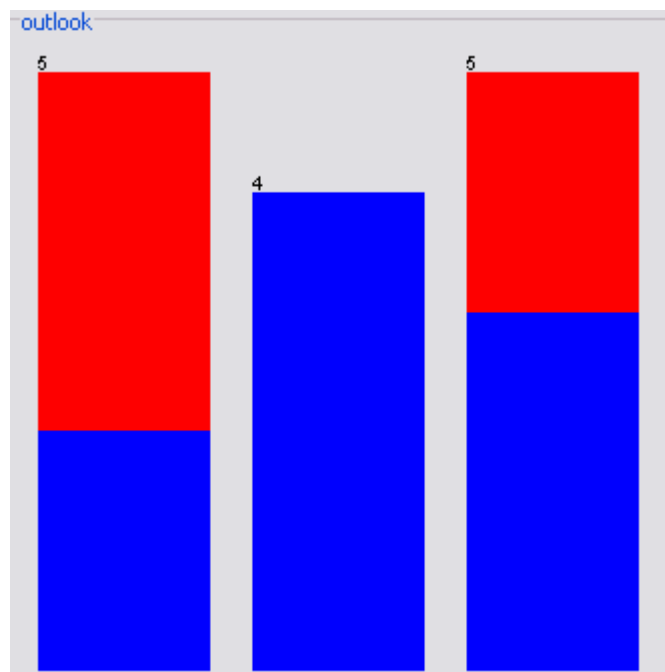


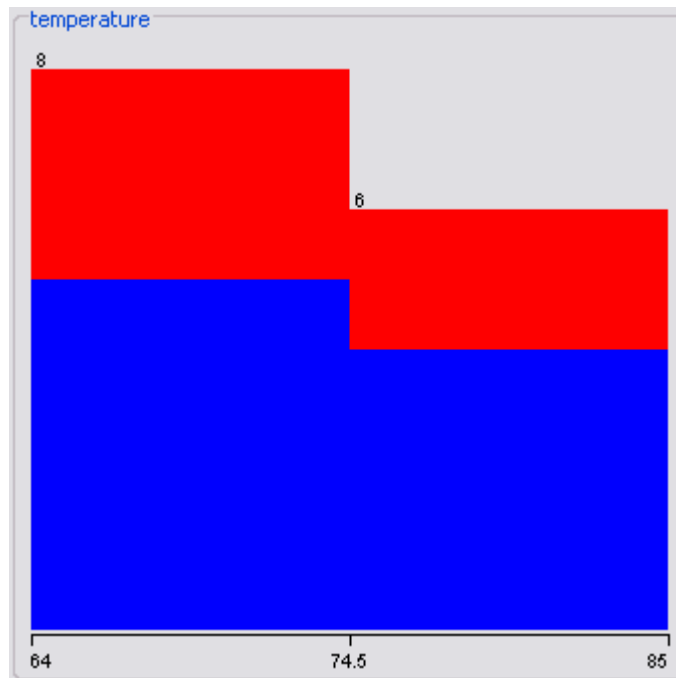
ΕΡΓΑΣΙΑ 02 : DATASET WEATHER ΕΙΡΗΝΗ ΛΥΓΚΩΝΗ

Το dataset weather περιέχει 4 μεταβλητές (outlook, temperature, humidity , windy) και 14 καταχωρήσεις για το καθένα από αυτά. Με βάση αυτές εξετάζεται το γεγονός για το αν θα γίνει το παιχνίδι ή όχι. Η μεταβλητή outlook μπορεί να είναι sunny, overcast, rainy , η μεταβλητή windy μπορεί να είναι true ή false , η μεταβλητή temperature και η μεταβλητή humidity είναι ποσοτικές μεταβλητές (numerical). Όλες αυτές οι μεταβλητές υποδεικνύουν αν θα γίνει το παιχνίδι (play) που μπορεί να πάρει yes ή no. Παρακάτω ακολουθούν διάφορα παράθυρα που προέκυψαν με την βοήθεια του weka:



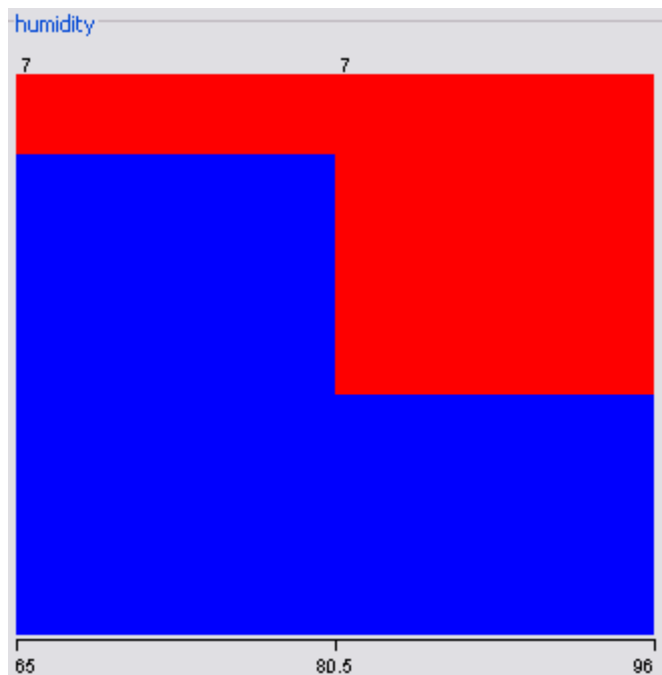
Σχήμα 1

Στο παραπάνω σχήμα παρατηρούμε ότι για το outlook έχουμε 5 sunny (οι 2 είναι yes και οι 3 no), 4 overcast (όλες είναι yes) και 5 rainy (3 είναι yes και οι 2 no) καταχωρήσεις. Το μπλέ χρώμα είναι η καταχώρηση yes και το κόκκινο χρώμα η κατάχώρηση no για το αν θα γίνει το παιχνίδι.



Σχήμα 2

Όσο αφορά την θερμοκρασία (temperature) παρατηρούμε ότι στο διάστημα από 64-74.5 υπάρχουν 8 μετρήσεις-καταχωρήσεις από τις οποίες οι 5 είναι yes (θα γίνει το παιχνίδι) και οι 3 no (δεν θα γίνει το παιχνίδι). Αντίστοιχα στο διάστημα 74.5-85 υπάρχουν 6 μετρήσεις-καταχωρήσεις από τις οποίες οι 4 είναι yes (θα γίνει το παιχνίδι) και οι 2 no (δεν θα γίνει το παιχνίδι). Το μπλέ χρώμα είναι η καταχώρηση yes και το κόκκινο χρώμα η κατά χώρηση no για το αν θα γίνει το παιχνίδι.

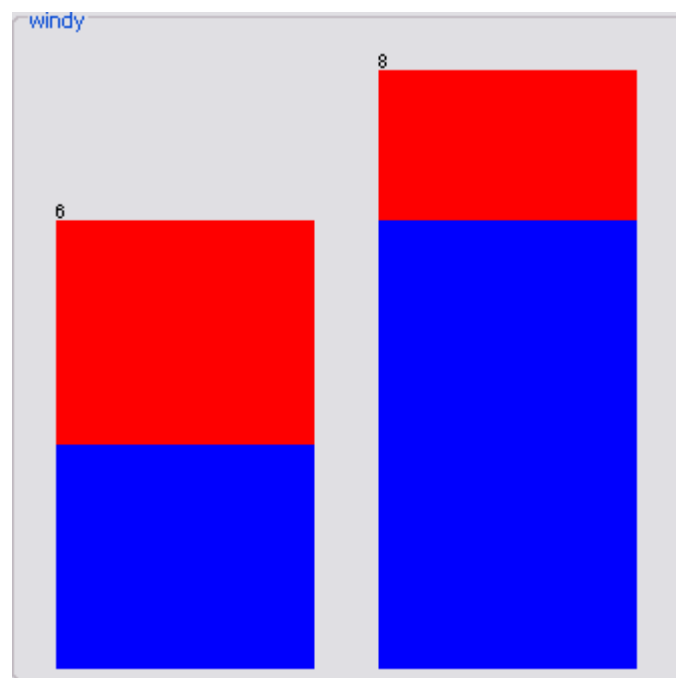


Σχήμα 3

Όσο αφορά την υγρασία (humidity) παρατηρούμε ότι στο διάστημα από 65-80.5 υπάρχουν 7 μετρήσεις-καταχωρήσεις από τις οποίες οι 6 είναι yes (θα γίνει το παιχνίδι) και οι 1 no (δεν θα γίνει το παιχνίδι). Αντίστοιχα στο διάστημα 80.5-96 υπάρχουν 7 μετρήσεις-καταχωρήσεις από τις οποίες οι 4 είναι yes (θα γίνει το παιχνίδι) και οι 3 no (δεν θα γίνει το παιχνίδι). Για το outlook έχουμε 5 sunny (οι 2 είναι yes και οι 3 no), 4 overcast (όλες είναι yes) και 5 runny (3 είναι yes και οι 2 no) καταχωρήσεις. Το μπλέ χρώμα είναι η καταχώρηση yes και το κόκκινο χρώμα η κατά χώρηση no για το αν θα γίνει το παιχνίδι. Ο Πίνακας 1 μας παρουσιάζει επιπλέον το είδος της μεταβλητής (Type:Numeric), αν υπάρχουν ελλειπούσες τιμές (missing) καθώς και την μέγιστη και την ελάχιστη τιμή. Την μέση τιμή και την τυπική απόκλιση των μετρήσεων της συγκεκριμένης μεταβλητής. Όμοια και για την μεταβλητή θερμοκρασίας. Επίσης, το unique είναι τα outliers που έχουμε στην συγκεκριμένη μεταβλητή.

Selected attribute	
Name: humidity	Type: Numeric
Missing: 0 (0%)	Distinct: 10
	Unique: 7 (50%)
Statistic	Value
Minimum	65
Maximum	96
Mean	81.643
StdDev	10.285

Πίνακας 1

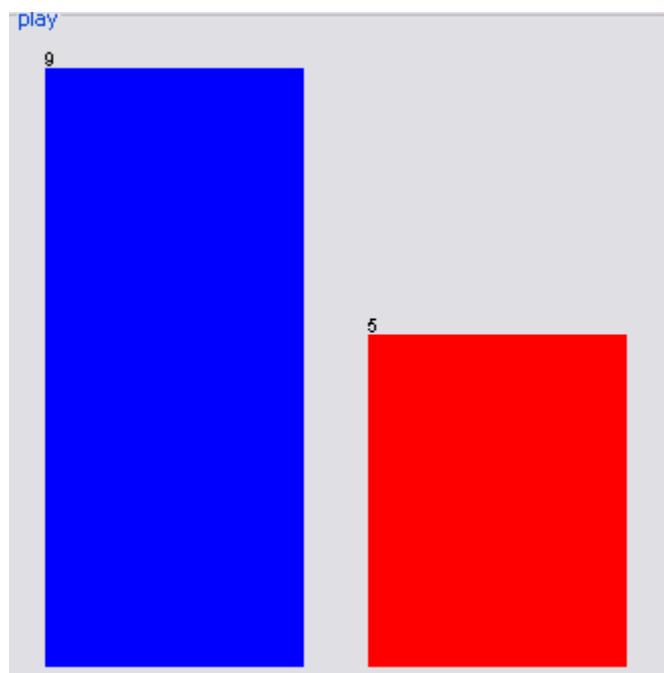


Σχήμα 4

Αντίστοιχο του Σχήματος 1 είναι το Σχήμα 4. παρατηρούμε ότι για το windy έχουμε 8 καταχωρήσεις FALSE (6 yes και 2 no)και 6 TRUE (3 yesκαι 3 no).Το μπλέ χρώμα είναι η καταχώρηση yes και το κόκκινο χρώμα η κατά χώρηση no για το αν θα γίνει το παιχνίδι ή όχι. Ο παρακάτω πίνακας μας δείχνει αν υπάρχουν ελλειπούσες τιμές (missing) στις μετρήσεις τις συγκεκριμένης μεταβλητής windy. Συγκεκριμένα δεν υπάρχουν. Επίσης αναφέρει το τύπο της μεταβλητής (Type: Nominal) καθώς και τον αριθμό των TRUE και FALSE της μεταβλητής. Όμοια και για την μεταβλητή outlook. Εδώ τα outliers δεν υπάρχουν.

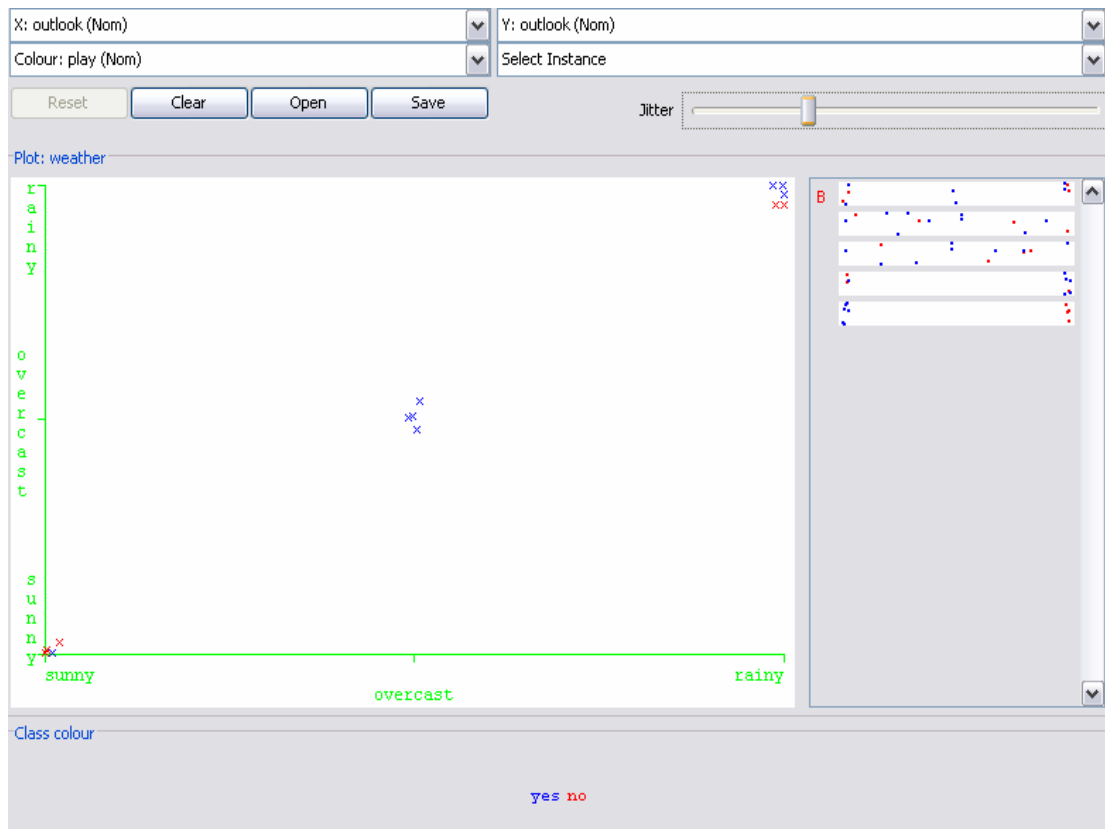
Selected attribute	
Name: windy	Type: Nominal
Missing: 0 (0%)	Distinct: 2
	Unique: 0 (0%)
Label	Count
TRUE	6
FALSE	8

Πίνακας 2



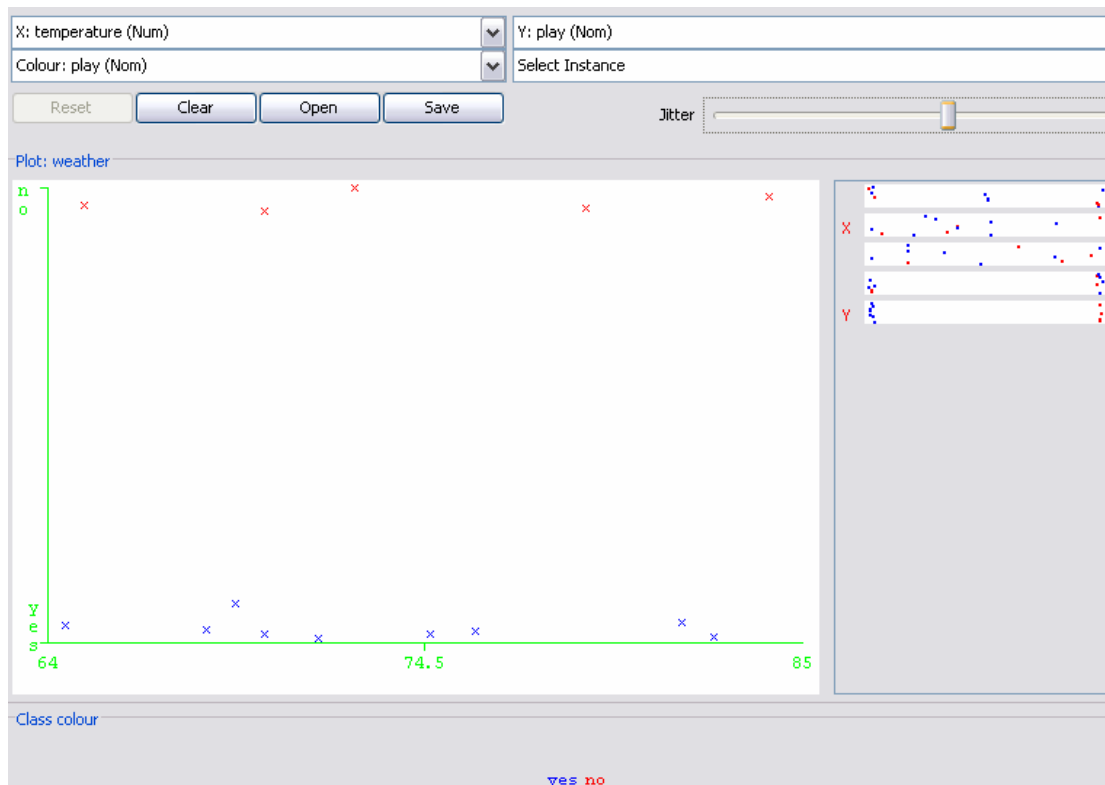
Σχήμα 5

Το παραπάνω σχήμα παρουσιάζει ότι από τις 14 καταχωρήσεις των 4 μεταβλητών που έγιναν στις 9 θα πραγματοποιηθεί το παιχνίδι και στις 5 όχι. Το μπλέ χρώμα είναι η καταχώρηση yes και το κόκκινο χρώμα η κατά χώρηση no για το αν θα γίνει το παιχνίδι ή όχι.



Plot 1

Η παραπάνω γραφική παράσταση μας δείχνει ότι στην περίπτωση της μεταβλητής outlook όπου έχουν γίνει οι εξής καταχωρήσεις: 5 sunny (οι 2 είναι yes και οι 3 no), 4 overcast (όλες είναι yes) και 5 runny (3 είναι yes και οι 2 no) καταχωρήσεις. Το μπλέ χρώμα είναι η καταχώρηση yes και το κόκκινο χρώμα η κατά χώρηση no για το αν θα γίνει το παιχνίδι ή όχι. Όμοιες γραφικές παραστάσεις προκύπτουν και για τις υπόλοιπες μεταβλητές. Παρακάτω ακολουθεί μια γραφική παράσταση της θερμοκρασίας με το παιχνίδι αν θα γίνει ή όχι. Ανάλογα με τη τιμή έχει η θερμοκρασία το παιχνίδι θα γίνει ή όχι.



Plot 2

Με βάση τα δεδομένα μπορεί να προκύψουν κανόνες για το αν θα γίνει το παιχνίδι ή όχι :

1. if outlook=sunny and humidity>83 then play=no
2. if outlook=overcast then play = yes
3. if outlook=rainy, temperature<74.5, humidity>80.6 then play=no
4. if outlook=sunny, temperature<86, humidity<80.5 then play=yes

καθώς και άλλα μπορεί να προκύψουν.