

## **BETT IYVONNE CHEPKOECH**

### **Data Science Assignment**

#### **Heart disease prediction:**

##### **Description:**

This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4.

Attribute Information:

1. Age - age in years
2. Sex - (1 = male; 0 = female)
3. Cp - chest pain type
4. Trestbps - resting blood pressure (in mm Hg on admission to the hospital)
5. Chol - serum cholestoral in mg/dl
6. Fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. Restecg- resting electrocardiographic results
8. Thalach- maximum heart rate achieved
9. Exang - exercise induced angina (1 = yes; 0 = no)
10. Oldpeak - ST depression induced by exercise relative to rest
11. Slope - the slope of the peak exercise ST segment
12. Ca- number of major vessels (0-3) colored by flourosopy
13. Thal- 3 = normal; 6 = fixed defect; 7 = reversable defect
14. Target - 1 or 0

##### **Inspiration**

Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0).

See if you can find any other trends in heart data to predict certain cardiovascular events or find any clear indications of heart health.

```
#bar chart
```

```
df.groupby('age')['trestbps'].mean().plot(kind='bar', color='green')  
plt.title('Age vs blood pressure')  
plt.xlabel('age')  
plt.ylabel('blood pressure')  
plt.show()
```

```
#age vs blood sugar
```

```
df.groupby('age')['fbs'].mean().plot(kind='bar', color='green')  
plt.title('Age vs blood sugar')  
plt.xlabel('age')
```

```
plt.ylabel('blood sugar')
plt.show()
```

```
df.groupby('age')['chol'].mean().plot(kind='bar', color='green')
plt.title('Age vs cholestral')
plt.xlabel('age')
plt.ylabel('cholestral')
plt.show()
```

```
figure, ax = plt.subplots()
ax.hist(df['age'], color = 'blue')
ax.set_xlabel('age')
ax.set_ylabel('freq.')
ax.set_title('Age')
plt.show()
```

```
df.groupby('age')['thalach'].mean().plot(kind='bar', color='green')
plt.title('Age vs heart rate')
plt.xlabel('age')
plt.ylabel('heart rate')
plt.show()
```