**Data Overview:**

**EDA:**

**Need to do:**

1. Describe
2. Renaming columns
3. Visualizations
4. Tell types of each column
5. Structure, Granulairty, Scope, temporality, faithfulness
6. No joins
7. Visualize to examine individuals/pairs
8. Quantitaive individuals (hist, density curves, boxplots)
9. Quantitative pairs( scatters, hex)
10. Hist(mean, mode, skewness, tails, outliers) , overlaying
11. Box(quartiles)

**Blog Post Format:**

1. Inspiration
2. Goals( research qs)
3. Dataset (column explanation, column types, granularity, structure, shape
4. EDA (overlapping historgrams with kdes, other visualizations)
5. Modeling,
6. conculsion

**Finalize Research Questions:**

**Objective:**

* Our Objective for this project is to conduct an Epidemiological study of risk factors to predict mortality by congestive heart failure based on specific risk factors: age, high blood pressure, diabetes, serum creatinine, creatinine phosphokinase, ejection fraction, and serum sodium.

**Research Questions:**

1. To investigate the prevalence of various risk factors in heart patients (cross tabs)
2. How strongly are various risk factors associated with the heart disease.
3. To predict the mortality of the patients based on risk factors.

**Dataset:**

​