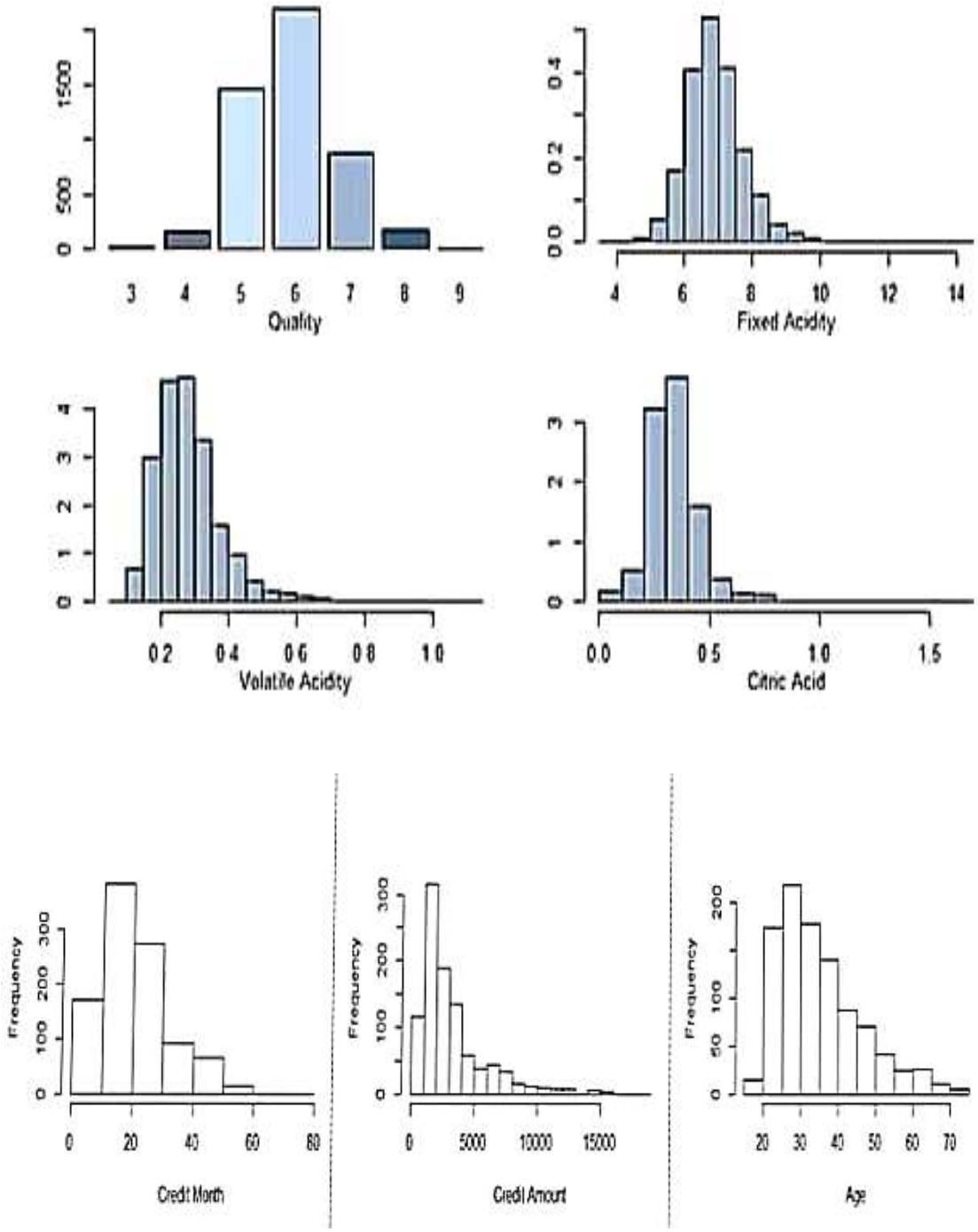
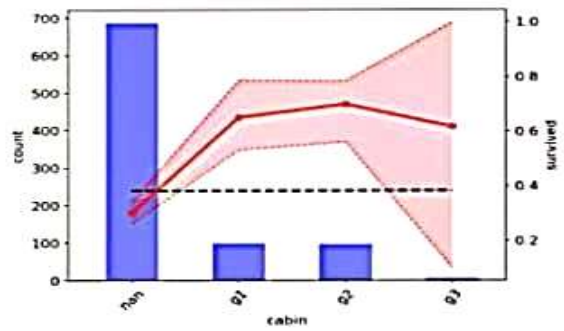
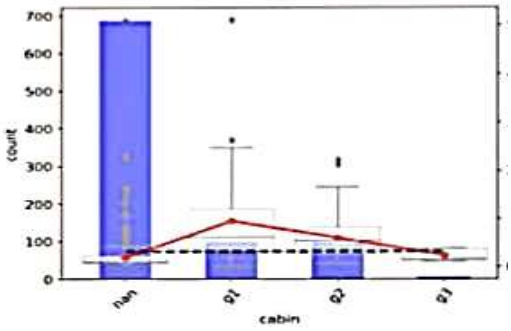
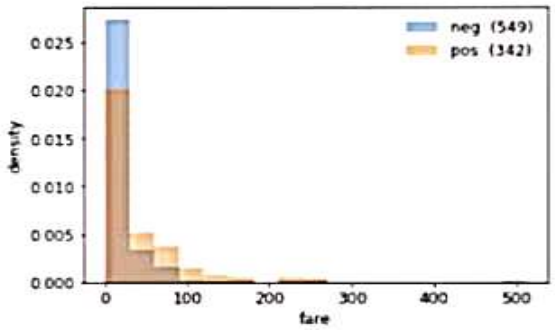
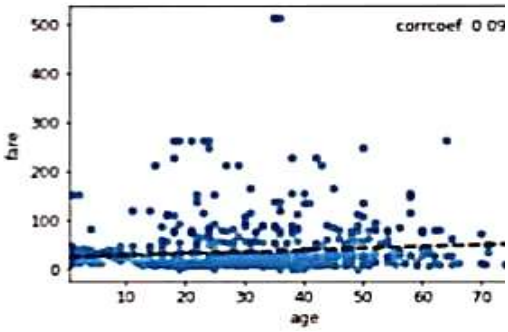
EXPLORATION OF DATA

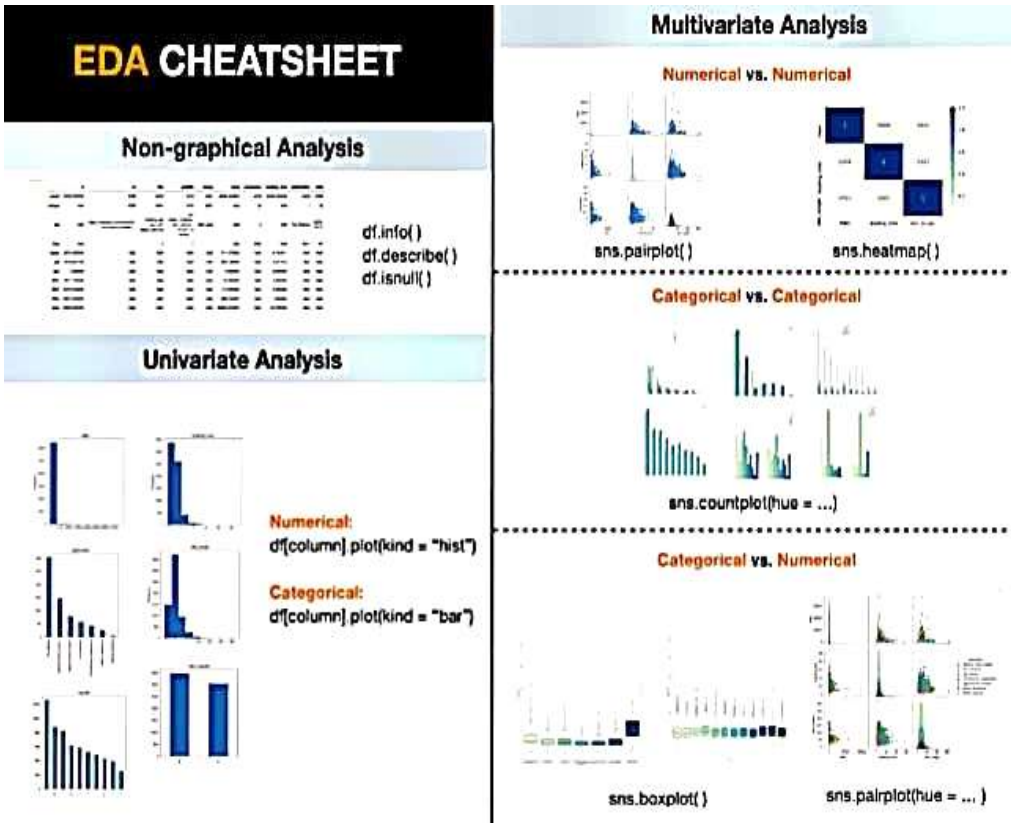


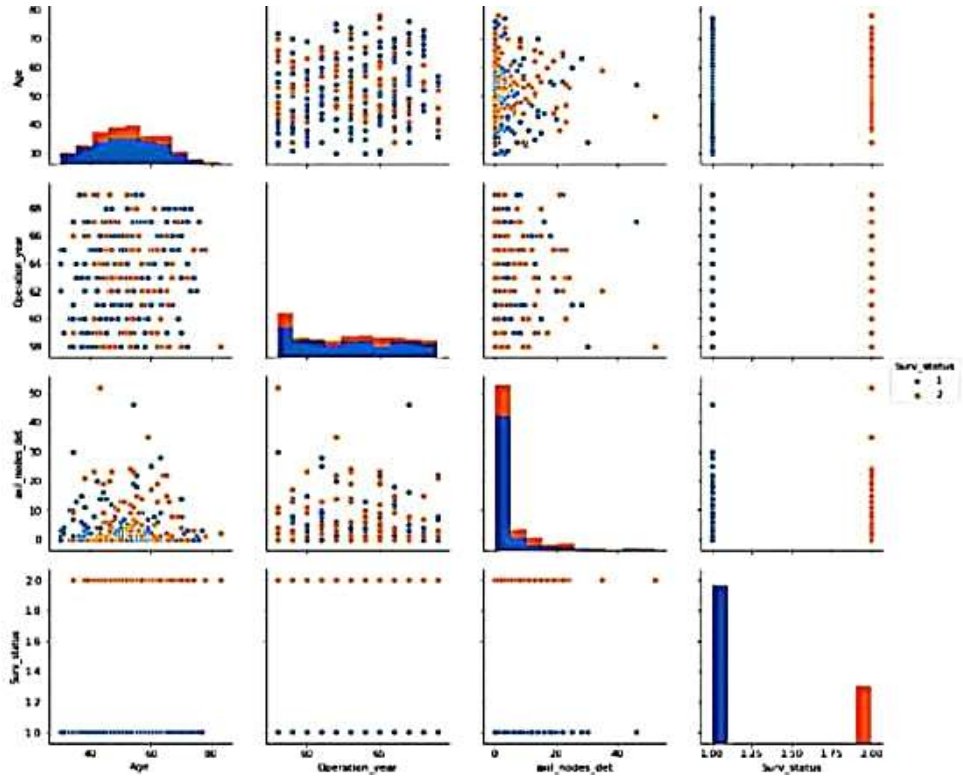

count
survived
cabin
nan
g1
g2
g3

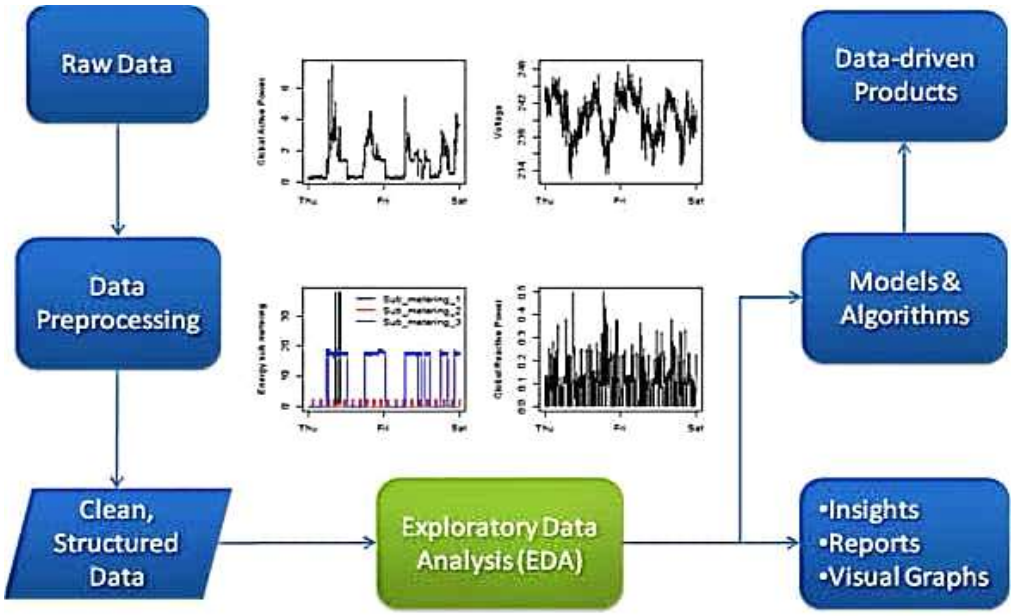

count
cabin
nan
g1
g2
g3

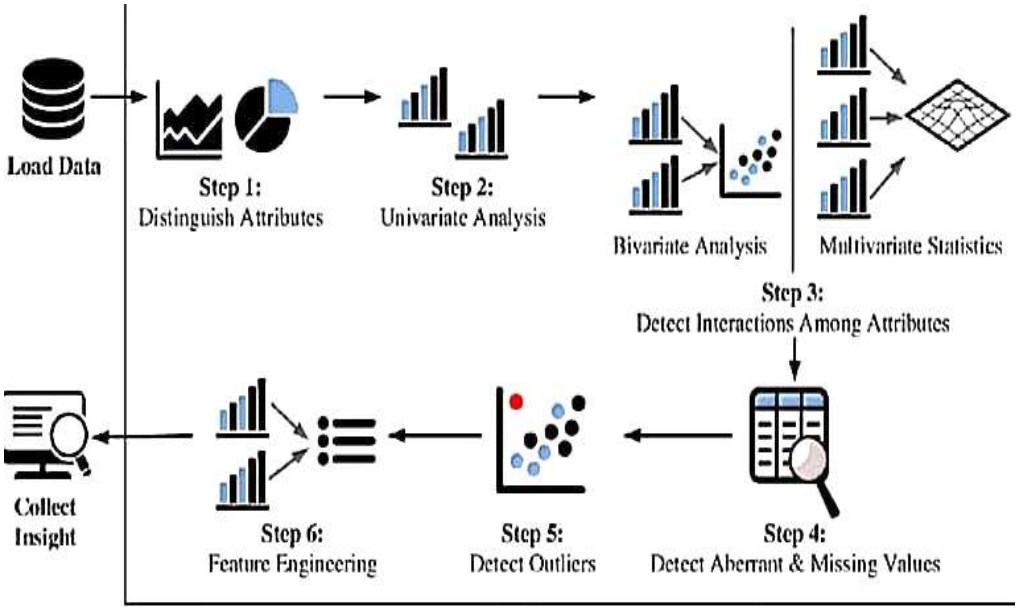

neg (549)
pos (342)
density
fare

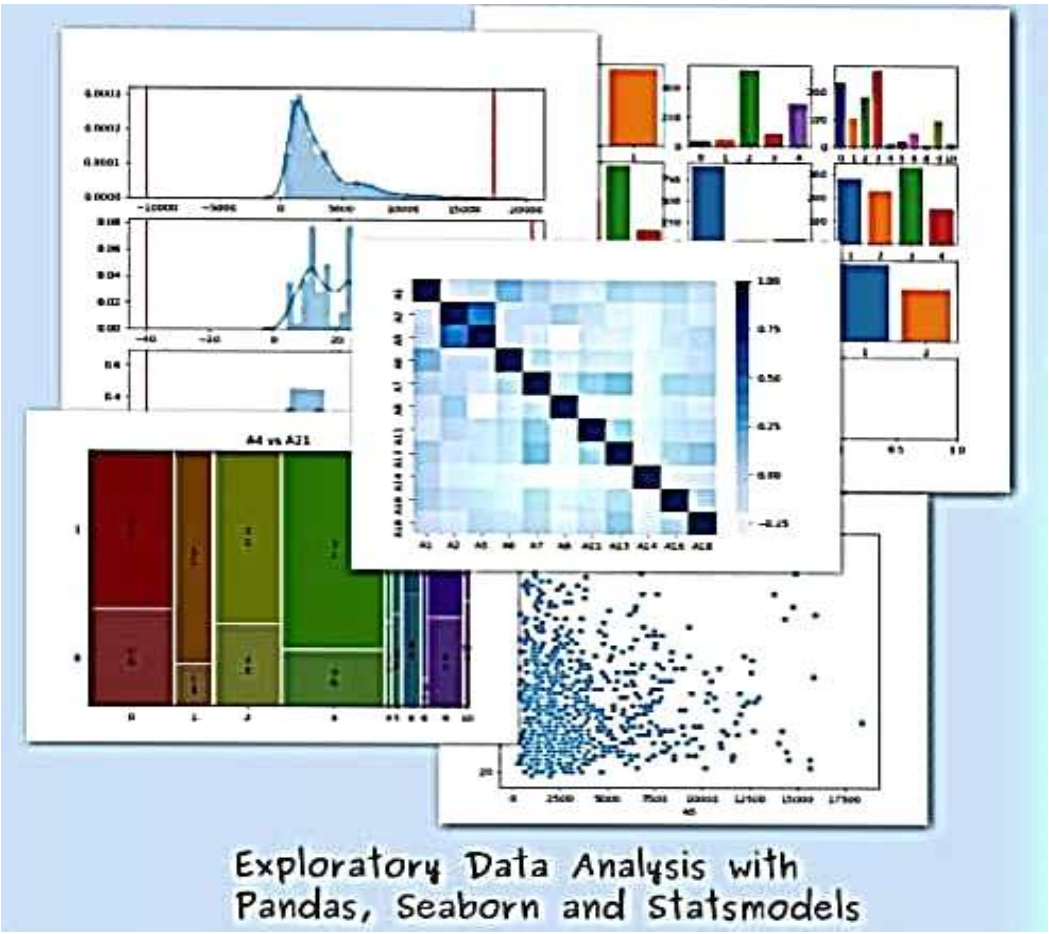

corrcoef 0.09
fare
age


EDA CHEATSHEET
Non-graphical Analysis
df.info()
df.describe()
df.isnull()
Univariate Analysis
Numerical:
df[column].plot(kind = "hist")
Categorical:
df[column].plot(kind = "bar")
Multivariate Analysis
Numerical vs. Numerical
sns.pairplot()
sns.heatmap()
Categorical vs. Categorical
sns.countplot(hue = ...)
Categorical vs. Numerical
sns.boxplot()
sns.pairplot(hue = ...)


Age
Operation_year
axil_nodes_det
Surv_status


Raw Data
Data Preprocessing
Clean, Structured Data
Exploratory Data Analysis (EDA)
Models & Algorithms
Data-driven Products
•Insights
•Reports
•Visual Graphs


Load Data
Step 1:
Distinguish Attributes
Step 2:
Univariate Analysis
Bivariate Analysis
Multivariate Statistics
Step 3:
Detect Interactions Among Attributes
Step 4:
Detect Aberrant & Missing Values
Step 5:
Detect Outliers
Step 6:
Feature Engineering
Collect
Insight


A4 vs A21
Exploratory Data Analysis with
Pandas, Seaborn and Statsmodels