

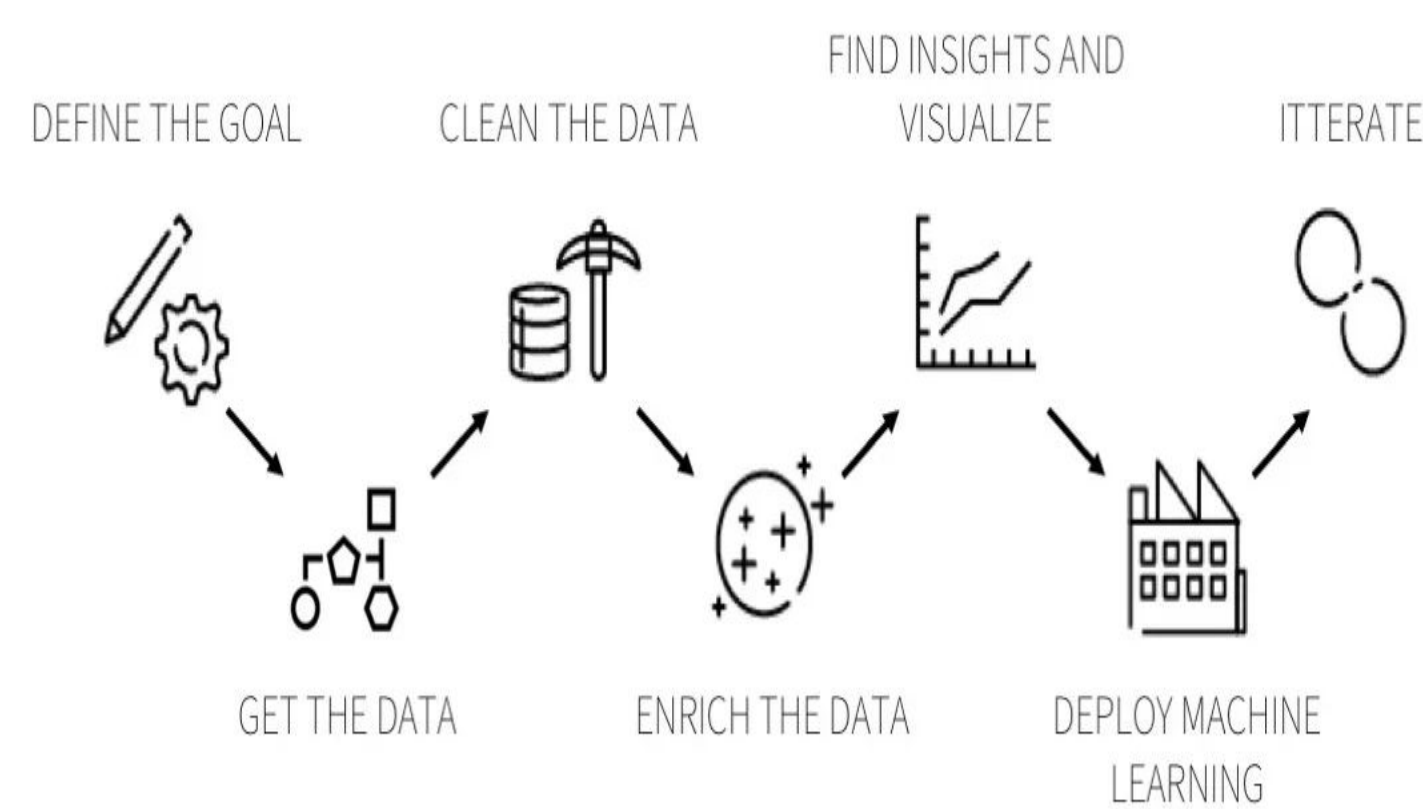
EXPLORATORY DATA ANALYSIS

INTRODUCTION

Data Analytics has a key role in improving once business as it is used to gather hidden insights, generate reports, perform market analysis, and improve business requirements. Data is extracted from various sources and is cleaned and categorized to analyze various behavioral patterns. The techniques and the tools used vary according to the organization or individual. understand once field of Business Administration and have the capability to perform Exploratory Data Analysis, to gather the required information, results in a good outcome and strong standard of career in Data Analytics.

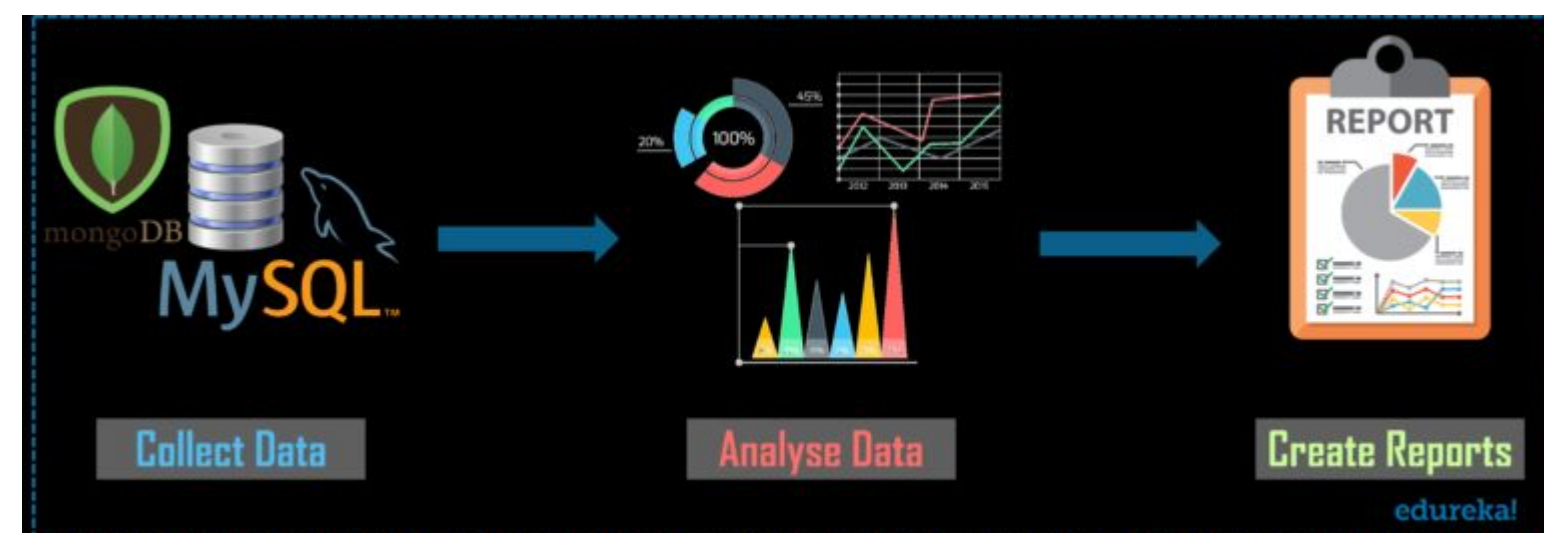
Once we have identified the patterns and derived the necessary insights from the given data. A project of this scale can easily be done with Python, and for the packages, we can use pandas, NumPy, seaborn, and matplotlib.

BLOCK DIAGRAM



- 1.Data Requirement Gathering(from different source in Google)
2. Data Collection(heart failure analysis)
3. Data Cleaning(removed the duplicates and repetition)
4. Data Analysis(understanding the data)
5. Data Interpretation
6. Data visualization(output)

METHODOLGY



The data collected is in form of csv (comma separated values) data set – predicting mortality caused by Heart Failure (heart_failure_clinical_records_dataset.csv) Heart failure is a common event caused by CVDs and this dataset contains 13 features that can be used to predict mortality by heart failure.

Analyze data : different graphs , maps and chart are obtained from the data set using python coding by imports libraries and working on it.

Create result/reports - From the obtained output graphs ,based on different parameters insights from these causing more mortality or less risking of life/prediction od chances of death due to different parameters are analyzed

WORKING PRINCIPLE

It is a associated to different fields of coding, creating and analysing .Thus taking a dataset that predicts the number of heart failure mortality .The main principle involved in this is Data cleaning by removing the duplicates and repetitions ,Data Analysis understanding the data by working on it to get different output , Data Interpretation and Data visualization are finally used to get insights and good presentation views.Using python and its imported libraries in any open source available they are easily analysed

- Getting the dataset
- Removing the duplicates
- Analysing getting hands on code
- Output graphs are obtained
- A finial report or insights from those graphs are reported

COMPONENTS USED

The main components of Data Science are given below:

1. **Statistics:** Statistics is a way to collect and analyze the numerical data in a large amount and finding meaningful insights from it.
2. **Domain Expertise:** Domain expertise means specialized knowledge or skills of a particular area.
3. **Data engineering:** involves acquiring, storing, retrieving, and transforming the data.
4. **Visualization:** Data visualization is meant by representing data in a visual context so that people can easily understand the significance of data.

5. **Advanced computing:** Advanced computing involves designing, writing, debugging, and maintaining the source code of computer programs.

This project completely uses only software components, the following are the software components used

- A programming language – python is used in this project for data analyzing.
- A open source platform – working on the finest and fastest supportive open source for getting graphical outputs
Jupyter lab – An extensible environment for interactive and reproducible computing.
- A data set – to analysis and visualize data
- Python coding
- Finally a system to work on the code and online open source

RESULTS

The python code is run over jupyter lab and different plots , graphs , charts are obtained . The main purpose of selecting this project was on note to Analyze data using Python , Importing Datasets , Data frame manipulation , Explore and clean the given data , and to get basic understandable insights. As we know getting useful insights from a dataset determines a good data analyzer. And being a good data analyzer is not a easy task.

So justifying our objectives of this project we have understood the importing and the different types of libraries that can be used in order to get a well desired graph. We have been working on the code and understood the functioning used to get the output.

From the dataset heart_failure_clinical_records_dataset.csv , we have got some analyzations of the data on how or on what factor or by how much percentage death factor depends upon with respect to the different parameters specified in the dataset

Some insights based on age which leads to heart failure 1):Age wise 40 to 80 the spread is High , less than 40 age and higher than 80 age people are very low with heart failure.

Based on Analysis in Age on Survival Status :2)Survival is high on 40 to70.The spread of Not survival is going through all ages.

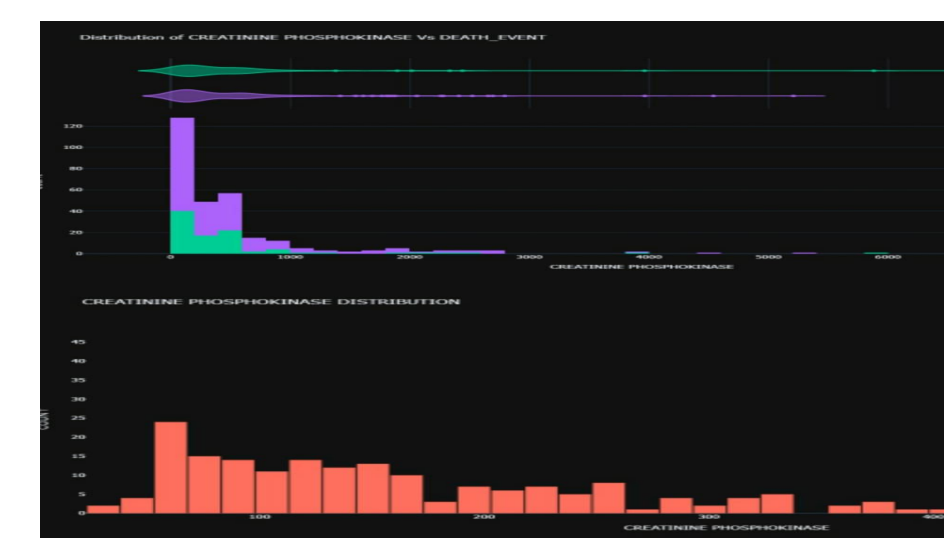
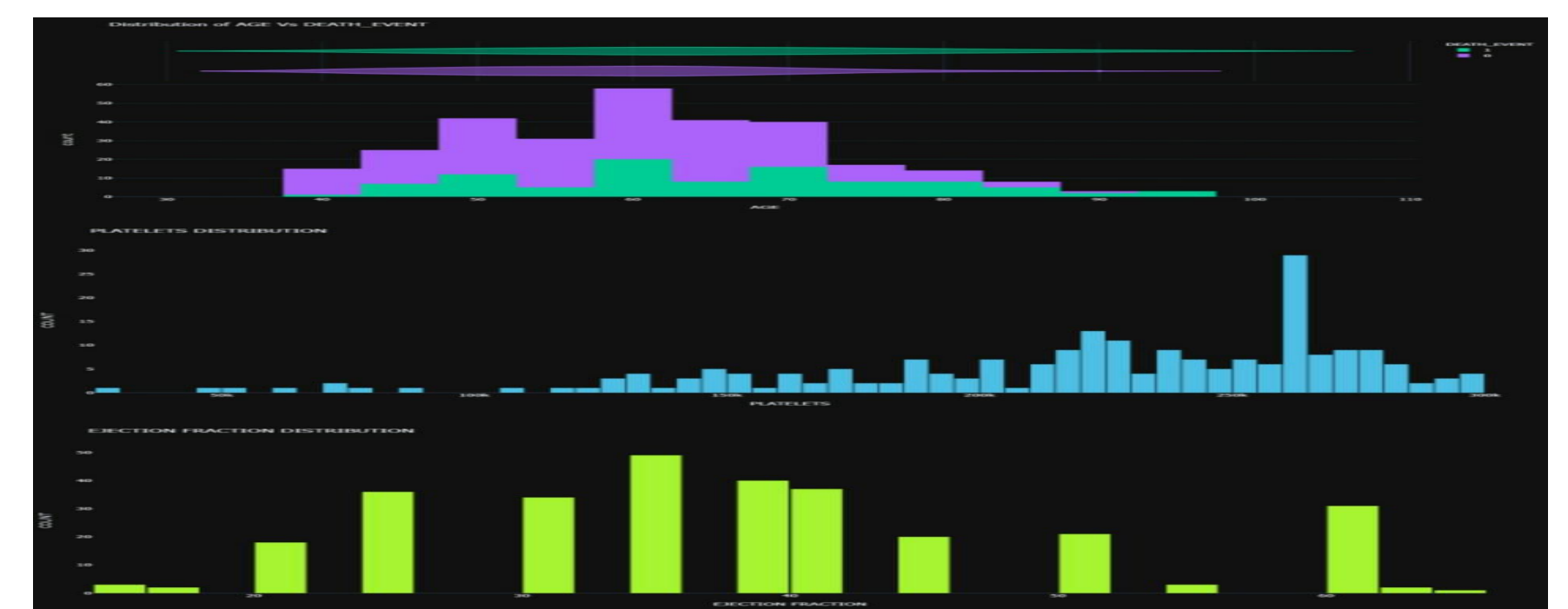
3)The Survival is high for not smoking person 55 to 65, while for smoking person it is between 50 to 60

Death event for smoking person is high than not smoking person.

Other insights are been shared in ppt

The above diagram shows how Data insights a visualization are obtained in general .

SNAPSHOTS OF YOUR PROJECT



RESULTS & CONCLUSIONS

Thus on understanding different modules and views of the dataset and on working over this project , we have got some basic insights .The conclusion of the project leads us to the understanding python code, jupyter lab, and methods of manipulation and understanding ways of getting insights from a data .Keeping these in mind we would be further working on huge dataset helping us analyze much better and making us go innovative with different way of understanding data.

References:Source: Dataset from Davide Chicco, Giuseppe Jurman: Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction
<https://www.kaggle.com/>
<https://www.freecodecamp.org/>
<https://towardsdatascience.com/exploratory-data-analysis-in-python-c9a77dfa39ce>
Udmeay , coursera